

CT4 – P C – 13

Combined Materials Pack

ActEd Study Materials: 2013 Examinations

Subject CT4

Contents

Study Guide for the 2013 exams

Course Notes

Question and Answer Bank

Series X Assignments*

***Note:** The Series X Assignment Solutions should also be supplied with this pack unless you chose not to receive them with your study material.

If you think that any pages are missing from this pack, please contact ActEd's admin team by email at ActEd@bpp.com.

How to use the Combined Materials Pack

Guidance on how and when to use the Combined Materials Pack is set out in the *Study Guide for the 2013 exams*.

Important: Copyright Agreement

This study material is copyright and is sold for the exclusive use of the purchaser. You may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of it. You must take care of your material to ensure that it is not used or copied by anybody else. By opening this pack you agree to these conditions.

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

2013 Study Guide

Subject CT4

Introduction



This Study Guide contains all the information that you will need before starting to study Subject CT4 for the 2013 exams. **Please read this Study Guide carefully before reading the Course Notes**, even if you have studied for some actuarial exams before.

When studying for the UK actuarial exams, you will need a copy of the **Formulae and Tables for Examinations of the Faculty of Actuaries and the Institute of Actuaries, 2nd Edition (2002)**. These are often referred to as simply the yellow **Tables** and are available separately from the Publications shop of the Actuarial Profession. You will also need a ‘permitted’ scientific calculator from the list published by the Profession. Please check the list carefully since it is reviewed each year. You will find the list of permitted calculators and a link to the Publications shop in the profession’s website at www.actuaries.org.uk.

Contents:

Section 1	The Subject CT4 course structure	Page 2
Section 2	ActEd study support	Page 3
Section 3	Core Reading, the Syllabus and the Profession	Page 10
Section 4	Study skills	Page 14
Section 5	Frequently asked questions	Page 14
Section 6	Syllabus	Page 19
Section 7	Assignment deadlines	Page 26

1 The Subject CT4 course structure

There are four parts to the Subject CT4 course. This should help you plan your progress across the study session. The parts cover related topics and have broadly equal lengths. The parts are broken down into chapters.

The following table shows how the parts, the chapters and the syllabus items relate to each other. The end columns show how the chapters relate to the days of the regular tutorials.

Part	Chapter	Title	No of pages	Syllabus objectives	2 full days	3 full days
1	1	Principles of actuarial modelling	25	(i)	1	1
	2	Stochastic processes	35	(ii)		
	3	Markov chains	66	(iii)		
	4	The two-state Markov model	37	(v)8, (vii)1-2		
2	5	Time-homogeneous Markov jump processes	65	(iv)1-4, (vii)2-3	2	2
	6	Time-inhomogeneous Markov jump processes	60	(iv)1, 3-8		
3	7	Survival models	42	(v)1-7	2	2
	8	Estimating the lifetime distribution function	60	(vi)1-4		
	9	Proportional hazards models	49	(vi)5		
4	10	The Binomial and Poisson models	42	(vii)4, (viii)	3	3
	11	Exposed to risk	30	(ix)		
	12	Graduation and statistical tests	72	(x)1-3		
	13	Methods of graduation	44	(x)4-7		

2 **ActEd study support**

This section lists the study support available from ActEd for Subject CT4.

Course Notes

The Course Notes will help you develop the basic knowledge and understanding of principles needed to pass the exam. They incorporate the complete Core Reading and include full explanation of all the syllabus objectives, with worked examples and short questions to test your understanding.

Each chapter includes the relevant syllabus objectives, a chapter summary and, where appropriate, a page of important formulae or definitions.

Question and Answer Bank

The Question and Answer Bank provides a comprehensive bank of questions (including some past exam questions) with full solutions and comments.

The Question and Answer Bank is divided into five parts. The first four parts include a range of short and long questions to test your understanding of the corresponding part of the Course Notes. Part five consists of 100 marks of exam-style questions.

Assignments

The four Series X Assignments (X1 to X4) cover the material in Parts 1 to 4 respectively. Assignments X1 and X2 are 80-mark tests and should take you two and a half hours to complete. Assignments X3 and X4 are 100-mark tests and should take you three hours to complete. The actual Subject CT4 examination will have a total of 100 marks.

Combined Materials Pack (CMP)

The Combined Materials Pack (CMP) comprises the Course Notes, the Question and Answer Bank and the Series X Assignments.

Mock Exam

A 100-mark mock exam paper (Mock Exam A) is available for students as a realistic test of their exam preparation. The mock is based on Mock Exam A from last year but it has been updated to reflect any changes to the Syllabus and Core Reading.

Additional Mock Pack (AMP)

The Additional Mock Pack (AMP) consists of two further 100-mark mock exam papers – Mock Exam B and Mock Exam C. This is ideal for students who are retaking and have already sat Mock Exam A, or for those who just want some extra question practice. If you are retaking this subject you should note that the mock exams in the AMP use many of the same questions from Mock Exam B for the 2011 exams and the Y Assignments for the 2011 exams.

ActEd Solutions with Exam Technique (ASET)

The ActEd Solutions with Exam Technique (ASET) contains ActEd's solutions to the previous four years' exam papers, *ie* eight papers, plus comment and explanation. In particular it will highlight how questions might have been analysed and interpreted so as to produce a good solution with a wide range of relevant points. This will be valuable in approaching questions in subsequent examinations.

A "Mini-ASET" will also be available in the summer session covering the April Exam only.

CMP Upgrade

The CMP Upgrade lists all significant changes to the Core Reading and ActEd material so that you can manually amend last year's study material to make it suitable for study for this year. The Upgrade includes replacement pages and additional pages where appropriate. If a large proportion of the material has changed significantly, making it inappropriate to include all changes, the upgrade will still explain what has changed and if necessary recommend that students purchase a replacement CMP or Course Notes at a significantly reduced price. The CMP Upgrade can be downloaded free of charge from our website at www.ActEd.co.uk.

Revision Notes

ActEd's Revision Notes have been designed with input from students to help you revise efficiently. They are suitable for first-time sitters who have worked through the ActEd Course Notes or for retakers (who should find them much more useful and challenging than simply reading through the course again). The Revision Notes are a set of nine A5 spiral-bound booklets – perfect for revising on the train or tube to work. Each booklet covers one main theme of the course and includes Core Reading (with a set of integrated short questions to develop your bookwork knowledge), relevant past exam questions (with concise solutions) from the last ten years, detailed analysis of key past exam questions and other useful revision aids.

Flashcards

Flashcards are a set of A6-sized cards that cover the key points of the subject that most students want to commit to memory. Each flashcard has questions on one side and the answers on the reverse. We recommend that you use the cards actively and test yourself as you go.

Flashcards may be used to complement your other study and revision materials. They are not a substitute for question practice but they should help you learn the essential material required.

Marking

We are happy to mark your attempts at any of the currently available assignments, Mock Exam A or the mock exams included within AMP. Marking is not included with the products themselves and you need to order it separately. You can submit your scripts by email, fax or post.

Series Marking and Mock Exam marking

Series Marking (for the Series X Assignments) and Mock Exam Marking (for Mock Exam A) apply to a specified subject, session and student. If you purchase Series Marking or Mock Exam Marking, you will **not** be able to defer the marking to a future exam sitting or transfer it to a different subject or student.

If you order marking at the same time as you order the assignments or mock exam, you can choose whether or not to receive a copy of the solutions in advance. If you choose not to receive the solutions in advance, we will send the solutions to you when we return your marked script (or following the deadline date if you don't submit).

If you are having your attempts at the assignments marked by ActEd, you should submit your scripts regularly throughout the session, in accordance with the schedule of recommended dates set out in the summary at the end of this document. This will help you to pace your study throughout the session and leave an adequate amount of time for revision and question practice.

Any script submitted after the relevant final deadline date will not be marked. It is your responsibility to ensure that scripts are posted in good time.



Important information

The recommended submission dates are realistic targets for the majority of students. Your scripts will be returned more quickly if you submit them well before the final deadline dates.

Marking Vouchers

Marking Vouchers give the holder the right to submit a script for marking at any time, irrespective of the individual assignment deadlines, study session, subject or person.

Marking Vouchers can be used for any assignment, Mock Exam A, or the mock exams contained within the AMP. Please note that attempts at the AMP can **only** be marked using Marking Vouchers.

Marking Vouchers are valid for four years from the date of purchase and can be refunded at any time up to the expiry date.



Important information

Although you may submit your script with a Marking Voucher at any time, you will need to adhere to the explicit Marking Voucher deadline dates to ensure that your script is returned before the date of the exam. The deadline dates are given at the end of this study guide.

If you live outside the UK you must ensure that your last script reaches the ActEd office earlier than this to allow the extra time needed to return your marked script.

Tutorials

ActEd tutorials are specifically designed to develop the knowledge that you will acquire from the course material into the higher level understanding that is needed to pass the exam. We expect you to have read the relevant part of the Course Notes before attending the tutorial so that the group can spend time on exam questions and discussion to develop understanding rather than basic bookwork.

ActEd run a range of different tutorials at various locations. Full details are set out in ActEd's *Tuition Bulletin*, which is sent regularly to all students based in the UK, Eire and South Africa and is also available from the ActEd website at www.ActEd.co.uk.

Taught Courses

We offer 4-day Taught Courses in some subjects. Each day will cover one part of the course. You will not be expected to have read the relevant part of the notes before attending as the tutor will introduce the material, key concepts and principles that you will need to master for the exam. The course will not replace the need to read the notes, but it will enable you to work through the material at a far quicker pace following the tutorial than would otherwise be the case.

Although the courses will involve more tutor-led sessions, simple examples, exercises and questions will be used to develop your understanding. Therefore, some active student participation will still be required. If you attend a taught course, you may also like to consider a Block Tutorial or a Revision Day closer to the exams to provide some guided exam-style question practice and additional support.

See the Tuition Bulletin for further details.

Regular and Block Tutorials

You can choose **one** of the following types of tutorial:

- **Regular Tutorials** (usually two or three days) spread over the session.
- **A Block Tutorial** (two or three days) held 2 to 8 weeks before the exam.

The Regular Tutorials provide an even progression through the course. Block Tutorials cover the whole course.

Revision Days

Revision Days are intensive one-day tutorials in the final run-up to the exam. They are particularly suitable for first-time sitters who attended Regular Tutorials and would like to spend a day close to the exam focusing on further question practice or retakers who have already attended ActEd tutorials. Revision Days give you the opportunity to practise interpreting and answering past exam questions and to raise any outstanding queries with an ActEd tutor. These courses are most suitable if you have previously attended Regular Tutorials or a Block Tutorial in the same subject.

Details of how to apply for ActEd's tutorials are set out in our *Tuition Bulletin*, which is sent regularly to all students based in the UK, Eire and South Africa and is also available from the ActEd website at www.ActEd.co.uk.

Online Classroom

The Online Classroom is an exciting new approach to studying for the actuarial exams. It acts as either a valuable add-on to a face-to-face tutorial or a great alternative to a tutorial, particularly if you're not based in the UK or near a tutorial venue. At the heart of the Online Classroom in each subject is a comprehensive, easily-searched collection of over 100 tutorial units. These are a mix of:

- teaching units, helping you to really get to grips with the course material, and
- guided questions, enabling you to learn the most efficient ways to answer questions and avoid common exam pitfalls.

The best way to discover the Online Classroom is to see it in action. You can watch a sample of the Online Classroom tutorial units on the ActEd website at www.ActEd.co.uk.

Queries and feedback

From time to time you may come across something in the study material that is unclear to you. The easiest way to solve such problems is often through discussion with friends, colleagues and peers – they will probably have had similar experiences whilst studying. If there's no-one at work to talk to then use ActEd's discussion forum at www.ActEd.co.uk/forums (or use the link from our home page at www.ActEd.co.uk).

Our online forum is dedicated to actuarial students so that you can get help from fellow students on any aspect of your studies from technical issues to study advice. You could also use it to get ideas for revision or for further reading around the subject that you are studying. ActEd Tutors will visit the site from time to time to ensure that you are not being led astray and we also post other frequently asked questions from students on the forum as they arise.

If you are still stuck, then you can send queries by email to **CT4@bpp.com** or by fax to 01235 550085 (but we recommend that you try the forum first). We will endeavour to contact you as soon as possible after receiving your query but you should be aware that it may take some time to reply to queries, particularly when tutors are away from the office running tutorials. At the busiest teaching times of year, it may take us more than a week to get back to you.

If you have many queries on the course material, you should raise them at a tutorial or book a personal tuition session with an ActEd Tutor. Information about personal tuition is set out in our current brochure. Please email **ActEd@bpp.com** for more details.

If you find an error in the course, please check the corrections page of our website (www.ActEd.co.uk/Html/paper_corrections.htm) to see if the correction has already been dealt with. Otherwise please send details via email to **CT4@bpp.com** or send a fax to **01235 550085**.

Each year ActEd Tutors work hard to improve the quality of the study material and to ensure that the courses are as clear as possible and free from errors. We are always happy to receive feedback from students, particularly details concerning any errors, contradictions or unclear statements in the courses. If you have any comments on this course please email them to **CT4@bpp.com** or fax them to **01235 550085**.

The ActEd Tutors also work with the profession to suggest developments and improvements to the Syllabus and Core Reading. If you have any comments or concerns about the Syllabus or Core Reading, these can be passed on via ActEd. Alternatively, you can address them directly to the Profession's Examination Team at Napier House, 4 Worcester Street, Oxford, OX1 2AW or by email to examinations@actuaries.org.uk.

3 Core Reading, the Syllabus and the Profession

Core Reading

The Syllabus for Subject CT4 has been written by the profession to state the requirements of the examiners. The relevant individual Syllabus Objectives are included at the start of each course chapter and a complete copy of the Syllabus is included in Section 6 of the Study Guide. We recommend that you use the Syllabus as an important part of your study. The Syllabus is supplemented by Core Reading, which has also been written by the profession. The purpose of Core Reading is to give the examiners, tutors and students a clear, shared understanding of the depth and breadth of treatment required by the Syllabus. In examinations students are expected to demonstrate their understanding of the concepts in Core Reading. Examiners have the Core Reading available when setting papers.

Core Reading deals with each Syllabus objective. Core Reading covers what is needed to pass the exam but the tuition material that has been written by ActEd enhances it by giving examples and further explanation of key points. The Subject CT4 Course Notes include the Core Reading in full, integrated throughout the course. Here is an excerpt from some ActEd Course Notes to show you how to identify Core Reading and the ActEd material. **Core Reading is shown in this bold font.**

Note that in the example given above, the index *will* fall if the actual share price goes below the theoretical ex-rights share price. Again, this is consistent with what would happen to an underlying portfolio.

After allowing for chain-linking, **the formula for the investment index becomes:**

$$I(t) = \frac{\sum_i N_{i,t} P_{i,t}}{B(t)}$$

where $N_{i,t}$ is the number of shares issued for the i th constituent at time t ;
 $B(t)$ is the base value, or divisor, at time t .

This is
ActEd
text

This is
Core
Reading

Core Reading accreditation

The Institute and Faculty of Actuaries would like to thank the numerous people who have helped in the development of this material and in the previous versions of Core Reading.

The following paper has been used as the basis for several Units:

Macdonald A S, *An Actuarial Survey of Statistical Models for Decrement and Transition Data*, British Actuarial Journal 2 (1996), by permission of the Editor of the British Actuarial Journal.

Changes to the Syllabus and Core Reading

The Syllabus and Core Reading are updated as at 31 May each year. The exams in April and September 2013 will be based on the Syllabus and Core Reading as at 31 May 2012.

We recommend that you always use the up-to-date Core Reading to prepare for the exams.

The Profession's Copyright

All of the course material is copyright. The copyright belongs to Institute and Faculty Education Ltd, a subsidiary of the Institute and Faculty of Actuaries. The material is sold to you for your own exclusive use. You may not hire out, lend, give, sell, transmit electronically, store electronically or photocopy any part of it. You must take care of your material to ensure it is not used or copied by anyone at any time.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Past exam papers

You can download some past exam papers and Examiners' Reports from the profession's website at www.actuaries.org.uk.

Further reading

The exam will be based on the relevant Syllabus and Core Reading and the ActEd course material will be the main source of tuition for students.

However, some students may find it useful to obtain a different viewpoint on a particular topic covered in Subject CT4. The following list of further reading for Subject CT4 has been prepared by the Institute and Faculty. This list is not exhaustive and other useful material may be available.

Basic stochastic processes; A course through exercises. – *Brzezniak, Zdzislaw; Zastawniak, Tomasz.* – Springer, 1998. – x, 225 pages. – ISBN: 3 540 76175 6.

Available from the Publications Unit.

Introduction to actuarial modeling. – *Hickman, James C.* North American Actuarial Journal (1997) 1(3) 1-5.

URL: http://www.soa.org/bookstore/naaj_archive.html

Modeling, analysis, design, and control of stochastic systems. – *Kulkarni, Vidyadhar G.* – Springer, 1999. – xiv, 374 pages. – ISBN: 0 387 98725 8.

Probability and random processes. – *Grimmett, Geoffrey; Stirzaker, David.* – 3rd ed. – Oxford University Press, 2001. – xii, 596 pages. – ISBN: 0 19 857222 0.

Actuarial mathematics. - *Bowers, Newton L; Gerber, Hans U; Hickman, James C; Jones, Donald A; Nesbitt, Cecil J.* - 2nd ed. - Society of Actuaries, 1997. - xxvi, 753 pages. - ISBN: 0 938959 46 8.

Actuarial models for disability insurance. – *Haberman, Steven; Pitacco, Ermanno* .– Chapman & Hall, 1999. – xviii, 280 pages. – ISBN: 0 8493 0389 3.

Available from the Publications Unit.

Analysing survival data from clinical trials and observational studies. – *Marubini, Ettore; Valsecchi, Maria Grazia.* – John Wiley, 1995. – xvi, 414 pages. – ISBN: 0 471 93987 0.

Life contingencies. – *Neill, Alistair.* – Heinemann, 1977. – vii, 452 pages. – ISBN: 0 434 91440 1.

(This book is no longer in print, but has been used as a textbook under earlier education strategies. You should find it relatively easy to borrow a copy from a colleague. Alternatively you can borrow it from the libraries.)

Life insurance mathematics. – *Gerber, Hans U.* – 3rd ed. – Springer. Swiss Association of Actuaries, 1997. – 217 pages. – ISBN: 3 540 62242 X.
Available from the Publications Unit.

Mortality studies. – *Scott, William F.* – Department of Mathematical Sciences, University of Aberdeen, 2000. – 147 pages.
Available from the Publications Unit.

Survival models and data analysis. – *Elandt-Johnson, Regina C; Johnson, Norman L.* – Classics Library ed. – John Wiley & Sons, 1999. – xvi, 457 pages. – ISBN: 0 471 34992 5.
Available from the Publications Unit.

Calculators

Please refer to the profession's website for the latest advice on which calculators are permitted in the exams.

4 **Study skills**

The CT Subject exams

The Core Reading and exam papers for these subjects tend to be very technical. The exams themselves have many calculation and manipulation questions. The emphasis in the exam will therefore be on understanding the mathematical techniques and applying them to various, frequently unfamiliar, situations. It is important to have a feel for what the numerical answer should be by having a deep understanding of the material and by doing reasonableness checks.

Subjects CT2 and CT7 are more “wordy” than the other subjects, including an “essay-style” question in Subject CT7.

Since there will be a high level of mathematics required in the courses it is important that your mathematical skills are extremely good. If you are a little rusty you may wish to consider buying the Foundation ActEd Course (FAC) available from ActEd. This covers all of the mathematical techniques that are required for the CT Subjects, some of which are beyond A-Level (or Higher) standard. It is a reference document to which you can refer when you need help on a particular topic.

You will have sat many exams before and will have mastered the exam and revision techniques that suit you. However it is important to note that due to the high volume of work involved in the CT Subjects it is not possible to leave all your revision to the last minute. Students who prepare well in advance have a better chance of passing their exams on the first sitting.

Unprepared students find that they are under time pressure in the exam. Therefore it is important to find ways of maximising your score in the shortest possible time. Part of your preparation should be to practise a large number of exam-style questions under timed exam conditions as soon as possible. This will:

- help you to develop the necessary understanding of the techniques required
- highlight which are the key topics that crop up regularly in many different contexts and questions
- help you to practise the specific skills that you will need to pass the exam.

There are many sources of exam-style questions. You can use past exam papers, the Question and Answer Bank (which includes many past exam questions), assignments, mock exams, the Revision Notes and ASET.

Overall study plan

We suggest that you develop a realistic study plan, building in time for relaxation and allowing some time for contingencies. Be aware of busy times at work, when you may not be able to take as much study leave as you would like. Once you have set your plan, be determined to stick to it. You don't have to be too prescriptive at this stage about what precisely you do on each study day. The main thing is to be clear that you will cover all the important activities in an appropriate manner and leave plenty of time for revision and question practice.

Aim to manage your study so as to allow plenty of time for the concepts you meet in this course to “bed down” in your mind. Most successful students will probably aim to complete the course at least a month before the exam, thereby leaving a sufficient amount of time for revision. By finishing the course as quickly as possible, you will have a much clearer view of the big picture. It will also allow you to structure your revision so that you can concentrate on the important and difficult areas of the course.

A sample CT subject study plan is available on our website at:

www.ActEd.co.uk/Html/help_and_advice_study_plans.htm

It includes details of useful dates, including assignment deadlines and tutorial finalisation dates.

Study sessions

Only do activities that will increase your chance of passing. Try to avoid including activities for the sake of it and don't spend time reviewing material that you already understand. You will only improve your chances of passing the exam by getting on top of the material that you currently find difficult.

Ideally, each study session should have a specific purpose and be based on a specific task, *eg “Finish reading Chapter 3 and attempt Questions 1.4, 1.7 and 1.12 from the Question and Answer Bank”*, as opposed to a specific amount of time, *eg “Three hours studying the material in Chapter 3”*.

Try to study somewhere quiet and free from distractions (*eg a library or a desk at home dedicated to study*). Find out when you operate at your peak, and endeavour to study at those times of the day. This might be between 8am and 10am or could be in the evening. Take short breaks during your study to remain focused – it's definitely time for a short break if you find that your brain is tired and that your concentration has started to drift from the information in front of you.

Order of study

We suggest that you work through each of the chapters in turn. To get the maximum benefit from each chapter you should proceed in the following order:

1. Read the Syllabus Objectives. These are set out in the box on Page 1 of each chapter.
2. Read the Chapter Summary at the end of each chapter. This will give you a useful overview of the material that you are about to study and help you to appreciate the context of the ideas that you meet.
3. Study the Course Notes in detail, annotating them and possibly making your own notes. Try the self-assessment questions as you come to them. Our suggested solutions are at the end of each chapter. As you study, pay particular attention to the listing of the Syllabus Objectives and to the Core Reading.
4. Read the Chapter Summary again carefully. If there are any ideas that you can't remember covering in the Course Notes, read the relevant section of the notes again to refresh your memory.

You may like to attempt some questions from the Question and Answer Bank when you have completed a part of the course. It's a good idea to annotate the questions with details of when you attempted each one. This makes it easier to ensure that you try all of the questions as part of your revision without repeating any that you got right first time.

Once you've read the relevant part of the notes and tried a selection of questions from the Question and Answer Bank (and attended a tutorial, if appropriate) you should attempt the corresponding assignment. If you submit your assignment for marking, spend some time looking through it carefully when it is returned. It can seem a bit depressing to analyse the errors you made, but you will increase your chances of passing the exam by learning from your mistakes. The markers will try their best to provide practical comments to help you to improve.

It's a fact that people are more likely to remember something if they review it from time to time. So, do look over the chapters you have studied so far from time to time. It is useful to re-read the Chapter Summaries or to try the self-assessment questions again a few days after reading the chapter itself.

To be really prepared for the exam, you should not only know and understand the Core Reading but also be aware of what the examiners will expect. Your revision programme should include plenty of question practice so that you are aware of the typical style, content and marking structure of exam questions. You should attempt as many questions as you can from the Question and Answer Bank and past exam papers.

Active study

Here are some techniques that may help you to study actively.

1. Don't believe everything you read! Good students tend to question everything that they read. They will ask "why, how, what for, when?" when confronted with a new concept, and they will apply their own judgement. This contrasts with those who unquestioningly believe what they are told, learn it thoroughly, and reproduce it (unquestioningly?) in response to exam questions.
2. Another useful technique as you read the Course Notes is to think of possible questions that the examiners could ask. This will help you to understand the examiners' point of view and should mean that there are fewer nasty surprises in the exam room! Use the Syllabus to help you make up questions.
3. Annotate your notes with your own ideas and questions. This will make you study more actively and will help when you come to review and revise the material. Do not simply copy out the notes without thinking about the issues.
4. Attempt the questions in the notes as you work through the course. Write down your answer before you check against the solution.
5. Attempt other questions and assignments on a similar basis, *ie* write down your answer before looking at the solution provided. Attempting the assignments under exam conditions has some particular benefits:
 - It forces you to think and act in a way that is similar to how you will behave in the exam.
 - When you have your assignments marked it is *much* more useful if the marker's comments can show you how to improve your performance under exam conditions than your performance when you have access to the notes and are under no time pressure.
 - The knowledge that you are going to do an assignment under exam conditions and then submit it (however good or bad) for marking can act as a powerful incentive to make you study each part as well as possible.
 - It is also quicker than trying to write perfect answers.
6. Sit a mock exam four to six weeks before the real exam to identify your weaknesses and work to improve them. You could use a mock exam written by ActEd or a past exam paper.

5 Frequently asked questions

Q: What knowledge of earlier subjects should I have?

A: The Course Notes are written on the assumption that students have studied Subject CT3.

Q: What level of mathematics is required?

A: The level of maths you need for this course is broadly A-level standard. However, there may be some symbols (*eg* the gamma function) that are not usually included on A-level syllabuses. You will find the course (and the exam!) much easier if you feel comfortable with the mathematical techniques used in the course and you feel confident in applying them yourself. If you feel that you need to brush up on your mathematical skills before starting the course, you may find it useful to study the Foundation ActEd Course (FAC) or read an appropriate textbook. The full Syllabus for FAC, a sample of the Course Notes and an Initial Assessment to test your mathematical skills can be found on our website at www.ActEd.co.uk.

Q: What calculators am I allowed to use in the exam?

A: Please refer to the Profession's website for the latest advice.

6 Syllabus

The full Syllabus for Subject CT4 is given here. The numbers to the right of each objective are the chapter numbers in which the objective is covered in the ActEd course.

Aim

The aim of the Models subject is to provide a grounding in stochastic processes and survival models and their application.

Links to other subjects

Subject CT1 — Financial Mathematics: provides an introduction to stochastic interest rates.

Subject CT3 — Probability and Mathematical Statistics: introduces the concepts of statistical distributions and modelling.

Subject CT5 — Contingencies develops the application of Markov chains.

Subject CT8 — Financial Economics: develops the concepts introduced here further.

Subject CA1 — Actuarial Risk Management and the Specialist Technical subjects use the models and principles introduced in this subject.

Objectives

On completion of the subject the trainee actuary will be able to:

- (i) Describe the principles of actuarial modelling. (Chapter 1)
1. Describe why and how models are used.
 2. Explain the benefits and limitations of modelling.
 3. Explain the difference between a stochastic and a deterministic model, and identify the advantages/disadvantages of each.
 4. Describe, in general terms, how to decide whether a model is suitable for any particular application.

5. Explain the difference between the short-run and long-run properties of a model, and how this may be relevant in deciding whether a model is suitable for any particular application.
 6. Describe, in general terms, how to analyse the potential output from a model, and explain why this is relevant to the choice of model.
 7. Describe the process of sensitivity testing of assumptions and explain why this forms an important part of the modelling process.
 8. Explain the factors that must be considered when communicating the results following the application of a model.
- (ii) Describe the general principles of stochastic processes, and their classification into different types. (Chapter 2)
1. Define in general terms a stochastic process and in particular a counting process.
 2. Classify a stochastic process according to whether it:
 - operates in continuous or discrete time
 - has a continuous or a discrete state space
 - is a mixed typeand give examples of each type of process.
 3. Describe possible applications of mixed processes.
 4. Explain what is meant by the Markov property in the context of a stochastic process and in terms of filtrations.
- (iii) Define and apply a Markov chain. (Chapter 3)
1. State the essential features of a Markov chain model.
 2. State the Chapman-Kolmogorov equations that represent a Markov chain.
 3. Calculate the stationary distribution for a Markov chain in simple cases.
 4. Describe a system of frequency based experience rating in terms of a Markov chain and describe other simple applications.

5. Describe a time-inhomogeneous Markov chain model and describe simple applications.
 6. Demonstrate how Markov chains can be used as a tool for modelling and how they can be simulated.
- (iv) Define and apply a Markov process. (Chapters 5 and 6)
1. State the essential features of a Markov process model.
 2. Define a Poisson process, derive the distribution of the number of events in a given time interval, derive the distribution of inter-event times, and apply these results.
 3. Derive the Kolmogorov equations for a Markov process with time independent and time/age dependent transition intensities.
 4. Solve the Kolmogorov equations in simple cases.
 5. Describe simple survival models, sickness models and marriage models in terms of Markov processes and describe other simple applications.
 6. State the Kolmogorov equations for a model where the transition intensities depend not only on age/time, but also on the duration of stay in one or more states.
 7. Describe sickness and marriage models in terms of duration dependent Markov processes and describe other simple applications.
 8. Demonstrate how Markov jump processes can be used as a tool for modelling and how they can be simulated.
- (v) Explain the concept of survival models.
1. Describe the model of lifetime or failure time from age x as a random variable. (Chapter 7)
 2. State the consistency condition between the random variable representing lifetimes from different ages. (Chapter 7)
 3. Define the distribution and density functions of the random future lifetime, the survival function, the force of mortality or hazard rate, and derive relationships between them. (Chapter 7)

4. Define the actuarial symbols $_t p_x$ and $_t q_x$ and derive integral formulae for them. (Chapter 7)
 5. State the Gompertz and Makeham laws of mortality. (Chapter 7)
 6. Define the curtate future lifetime from age x and state its probability function. (Chapter 7)
 7. Define the expected value and variance of the complete and curtate future lifetimes and derive expressions for them. Define the symbols e_x and \mathring{e}_x and derive an approximate relation between them. (Chapter 7)
 8. Describe the two-state model of a single decrement and compare its assumptions with those of the random lifetime model. (Chapter 4)
- (vi) Describe estimation procedures for lifetime distributions.
1. Describe the various ways in which lifetime data might be censored. (Chapter 8)
 2. Describe the estimation of the empirical survival function in the absence of censoring, and what problems are introduced by censoring. (Chapter 8)
 3. Describe the Kaplan-Meier (or product limit) estimate of the survival function in the presence of censoring, compute it from typical data and estimate its variance. (Chapter 8)
 4. Describe the Nelson-Aalen estimate of the cumulative hazard rate in the presence of censoring, compute it from typical data and estimate its variance. (Chapter 8)
 5. Describe the Cox model for proportional hazards, derive the partial likelihood estimate in the absence of ties, and state its asymptotic distribution. (Chapter 9)
- (vii) Derive maximum likelihood estimators for the transition intensities in models of transfers between states with piecewise constant transition intensities.
1. Describe an observational plan in respect of a finite number of individuals observed during a finite period of time, and define the resulting statistics, including the waiting times. (Chapter 8)

2. Derive the likelihood function for constant transition intensities in a Markov model of transfers between states given the statistics in 1. (Chapter 5)
 3. Derive maximum likelihood estimators for the transition intensities in 2. and state their asymptotic joint distribution. (Chapter 5)
 4. Describe the Poisson approximation to the estimator in 3. in the case of a single decrement and its advantages and disadvantages. (Chapter 10)
- (viii) Describe the Binomial model of mortality, derive a maximum likelihood estimator for the probability of death and compare the Binomial model with the multiple state models. (Chapter 10)
1. Describe the Binomial model of the mortality of a group of identical individuals subject to no other decrements between two given ages.
 2. Derive the maximum likelihood estimator for the rate of mortality in the Binomial model and its mean and variance.
 3. Describe the advantages and disadvantages of the multiple state model and the Binomial model, including consistency, efficiency, simplicity of the estimators and their distributions, application to practical observational plans and generality.
- (ix) Describe how to estimate transition intensities depending on age, exactly or using the census approximation. (Chapter 11)
1. Explain the importance of dividing the data into homogeneous classes, including subdivision by age and sex.
 2. Describe the principle of correspondence and explain its fundamental importance in the estimation procedure.
 3. Specify the data needed for the exact calculation of a central exposed to risk (waiting time) depending on age and sex.
 4. Calculate a central exposed to risk given the data in 3.
 5. Explain how to obtain estimates of transition probabilities, including in the single decrement model the actuarial estimate based on the simple adjustment to the central exposed to risk.

6. Explain the assumptions underlying the census approximation of waiting times.
7. Explain the concept of rate interval.
8. Develop census formulae given age at birthday where the age may be classified as next, last, or nearest relative to the birthday as appropriate.

The deaths and census data may use different definitions of age.

9. Specify the age to which estimates of transition intensities or probabilities in 8 apply.
- (x) Describe how to test crude estimates for consistency with a standard table or a set of graduated estimates, and describe the process of graduation.
1. Describe the following statistical tests of crude estimates, for comparison with a standard table: (Chapter 12)
 - chi-square test
 - standardised deviations test
 - sign test
 - cumulative deviation test
 - grouping of signs test
 - serial correlations test
 For each test describe:
 - the formulation of the hypothesis
 - the test statistic
 - the distribution of the test statistic using approximations where appropriate
 - the application of the test statistic
 2. Describe the reasons for graduating crude estimates of transition intensities or probabilities, and state the desirable properties of a set of graduated estimates. (Chapter 12)
 3. Describe a test for smoothness of a set of graduated estimates. (Chapter 12)

4. Describe the process of graduation by the following methods, and state the advantages and disadvantages of each: (Chapter 13)
 - parametric formula
 - standard table
 - graphical

(The student will not be required to carry out a graduation.)
5. Describe how the tests in 1. should be amended to compare crude and graduated sets of estimates. (Chapter 13)
6. Describe how the tests in 1. should be amended to allow for the presence of duplicate policies. (Chapter 13)
7. Carry out a comparison of a set of crude estimates and a standard table, or of a set of crude estimates and a set of graduated estimates. (Chapter 13)

7 Assignment Deadlines – CT Subjects

For the session leading to the April 2013 exams – CT Subjects

Marking vouchers

Subjects	Assignments	Mocks
CT1, CT2, CT5, CT8	20 March 2013	26 March 2013
CT3, CT4, CT6, CT7	26 March 2013	3 April 2013

Series X Assignments

Subjects	Assignment	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	X1	14 November 2012	16 January 2013
CT3, CT4, CT6, CT7		21 November 2012	23 January 2013
CT1, CT2, CT5, CT8	X2	28 November 2012	6 February 2013
CT3, CT4, CT6, CT7		5 December 2012	13 February 2013
CT1, CT2, CT5, CT8	X3	23 January 2013	27 February 2013
CT3, CT4, CT6, CT7		30 January 2013	6 March 2013
CT1, CT2, CT5, CT8	X4	13 February 2013	13 March 2013
CT3, CT4, CT6, CT7		20 February 2013	20 March 2013

Mock Exams

Subjects	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	20 March 2013	26 March 2013
CT3, CT4, CT6, CT7	26 March 2013	3 April 2013

We encourage you to work to the recommended submission dates where possible. Please remember that the turnaround of your script is likely to be quicker if you submit it well before the final deadline date.

For the session leading to the September/October 2013 exams – CT Subjects

Marking vouchers

Subjects	Assignments	Mocks
CT1, CT2, CT5, CT8	28 August 2013	4 September 2013
CT3, CT4, C6, CT7	4 September 2013	11 September 2013

Series X Assignments

Subjects	Assignment	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	X1	12 June 2013	3 July 2013
CT3, CT4, C6, CT7		19 June 2013	10 July 2013
CT1, CT2, CT5, CT8	X2	3 July 2013	24 July 2013
CT3, CT4, C6, CT7		10 July 2013	31 July 2013
CT1, CT2, CT5, CT8	X3	24 July 2013	7 August 2013
CT3, CT4, C6, CT7		31 July 2013	14 August 2013
CT1, CT2, CT5, CT8	X4	7 August 2013	21 August 2013
CT3, CT4, C6, CT7		14 August 2013	28 August 2013

Mock Exams

Subjects	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	21 August 2013	4 September 2013
CT3, CT4, C6, CT7	28 August 2013	11 September 2013

We encourage you to work to the recommended submission dates where possible. Please remember that the turnaround of your script is likely to be quicker if you submit it well before the final deadline date.

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

CT4 Index

Actuarial estimate (see Binomial model)

Binomial model

Actuarial estimate	Ch10	p10-13
Baldacci assumption.....	Ch10	p9
Basic model	Ch10	p3-6
Comparison with multi-state and Poisson models...	Ch10	p22-27
Generalised model	Ch10	p7-13
Initial exposed to risk	Ch10	p11-12
Strengths and weaknesses.....	Ch10	p14-15

Censoring

Informative and non-informative censoring	Ch8	p7-8
Interval censoring	Ch8	p6
Left censoring	Ch8	p5-6
Random censoring	Ch8	p6-7
Right censoring.....	Ch8	p6
Type I censoring	Ch8	p8
Type II censoring.....	Ch8	p9

Chapman-Kolmogorov equations.....	Ch3	p5-8
	Ch5	p14
	Ch6	p3

Comparison of the binomial, Poisson and multi-state models

Ch10 p22-28

Compound Poisson process (see Stochastic processes)

Cox regression model (see Proportional hazards models)

Differential equations

Backward equations (time-homogeneous case).....	Ch5	p25-26
Backward equations (time-inhomogeneous case) ...	Ch6	p10-12
Forward equations (time-homogeneous case)	Ch5	p20-24
Forward equations (time-inhomogeneous case)	Ch6	p4-8
Techniques for solving differential equations	Ch4	p21-24

Exponential hazard (see Future lifetime)

Exposed to risk

Census approximation.....	Ch11	p13-16
Definitions of age.....	Ch11	p17-21
Exact calculation of the central exposed to risk	Ch11	p10-12
Exact calculation of the initial exposed to risk	Ch10	p12
Principle of correspondence.....	Ch11	p9

Future lifetime

Complete expectation of life.....	Ch7	p15-16, 18-19
Curtate expectation of life.....	Ch7	p16-19
Distribution function of T_x	Ch7	p4
Exponential hazard	Ch7	p24
	Ch8	p29-35
Force of mortality	Ch7	p6-7
Gompertz' and Makeham's laws	Ch7	p26-29
Initial and central rates of mortality.....	Ch7	p12-14
Integral formulae.....	Ch7	p21-23
Probabilities of death and survival.....	Ch7	p5, p7-8
Probability density function of T_x	Ch7	p9
Survival function of T_x	Ch7	p3
Variance of future lifetime random variables	Ch7	p19
Weibull hazard.....	Ch7	p25
	Ch8	p28

Graduation

Chi-squared test	Ch12	p22-25, p28-31
	Ch13	p23-24
Comparing a graduation with another experience ...	Ch12	p5-6
Comparing a graduation with a standard table	Ch12	p6-7
Comparison of the methods of graduation.....	Ch13	p17-21
Cumulative deviations test.....	Ch12	p41-43
Desirable features of a graduation	Ch12	p11-13
Duplicate policies	Ch13	p25-29
Graphical method.....	Ch13	p13-16
Grouping of signs test.....	Ch12	p44-48
Parametric formula method	Ch13	p3-8
Reasons for graduation	Ch12	p9-10
Serial correlation test	Ch12	p49-53
Signs test	Ch12	p37-40
Smoothness	Ch12	p11-12, p16-18
Standard table method	Ch13	p9-12
Standardised deviations test.....	Ch12	p32-36

Hessian matrix	Ch5	p44
Homogeneity	Ch11	p6-8
Indicator random variables	Ch4	p8
Kaplan-Meier model		
Assumptions and notation	Ch8	p10-14
Calculating the Kaplan-Meier estimate of $S(t)$	Ch8	p16-21
Discrete hazard function	Ch8	p15
Variance of the distribution function estimator	Ch8	p22
Life table functions.....	Ch7	p10-11
Likelihood functions.....	Ch4	p24-26
Limiting age.....	Ch7	p3
Markov chains		
Chapman-Kolmogorov equations	Ch3	p5-9
Estimating transition probabilities	Ch3	p39-40
Irreducibility	Ch3	p32-33
Long-term distribution.....	Ch3	p27-38
Model of accident proneness	Ch3	p24-25
No claims discount models.....	Ch3	p14-19
Periodicity.....	Ch3	p34-35
Simple random walks	Ch3	p20-23
Simulation.....	Ch3	p41-42
Stationary distributions	Ch3	p27
Testing the fit of a model.....	Ch3	p40
Time-homogeneous	Ch3	p9-11
Time-inhomogeneous	Ch3	p13
Transition matrix	Ch3	p3
Transition probabilities	Ch3	p5, 9

Modelling

Analysing the output of a model.....	Ch1	p16
Benefits of modelling.....	Ch1	p7
Communicating the results of a model	Ch1	p18
Deterministic models	Ch1	p9-12
Key steps in a modelling process.....	Ch1	p5-6
Limitations of modelling	Ch1	p7-8
Sensitivity testing.....	Ch1	p17
Simulation.....	Ch1	p11
Short and long-term properties of a model	Ch1	p15
Stochastic models	Ch1	p9-12
Suitability of a model.....	Ch1	p14

Monte Carlo simulation	Ch1	p11
------------------------------	-----	-----

Nelson-Aalen model

Calculating estimates	Ch8	p25
Integrated hazard function	Ch8	p23-24
Relationship with Kaplan-Meier.....	Ch8	p26-27
Variance of the integrated hazard estimator	Ch8	p25

Parametric models	Ch8	p28-34
-------------------------	-----	--------

Poisson model.....	Ch10	p16-21
--------------------	------	--------

Poisson process (see Stochastic processes)

Proportional hazards models

Breslow's approximation.....	Ch9	p21-22
Building models.....	Ch9	p28
Covariates	Ch9	p2
Definition	Ch9	p3
Estimating the regression parameters	Ch9	p16-22
Cox model.....	Ch9	p11
Hypothesis tests	Ch9	p26-27
Parametric models for the hazard function	Ch9	p4-10
Partial likelihood.....	Ch9	p16-25

Semi-parametric models	Ch9	p12
------------------------------	-----	-----

Solving first-order differential equations	Ch4	p22-25
--	-----	--------

State spaces.....	Ch1	p13
	Ch2	p3-6

Stochastic processes

Compound Poisson process	Ch2	p22
Counting process	Ch2	p8
Definition.....	Ch2	p3
Increments	Ch2	p15-16
Filtration	Ch2	p18
Markov property	Ch2	p16-17
Mixed process.....	Ch2	p7
Poisson process.....	Ch2	p21
	Ch5	p6-13, 27
Random walk.....	Ch2	p9-10, 20-21
	Ch3	p19-22
Sample paths.....	Ch2	p9, 11
Stationarity.....	Ch2	p12-13
Weak stationarity	Ch2	p13-14
White noise	Ch2	p10, 20

Time-homogeneous Markov jump processes

Chapman-Kolmogorov equations.....	Ch5	p14
Expected time to reach a given state	Ch5	p35-37
Generator matrix.....	Ch5	p17
Health, sickness, death model.....	Ch5	p18-19
Holding times and occupancy probabilities.....	Ch5	p28-34
Jump chains	Ch5	p38-39
Kolmogorov's backward differential equations	Ch5	p25-26
Kolmogorov's forward differential equations	Ch5	p20-24
ML estimators of the transition rates	Ch5	p42-48
Poisson process.....	Ch5	p6-13, 27
Notation	Ch5	p4-5
Transition matrix	Ch5	p14-15
Transition rates	Ch5	p15-17

Time-inhomogeneous Markov jump processes

Chapman-Kolmogorov equations	Ch6	p3
Duration dependence	Ch6	p28-31
Health, sickness, death model.....	Ch6	p24-31
Integrated form of the backward equation.....	Ch6	p17-18
Integrated form of the forward equation.....	Ch6	p19-22
Kolmogorov's backward differential equations	Ch6	p10-12
Kolmogorov's forward differential equations	Ch6	p4-8
Marriage model.....	Ch6	p23-24
Modelling and simulation	Ch6	p32-38
Occupancy probabilities	Ch6	p9
Residual holding times.....	Ch6	p15-16
Transition rates	Ch6	p3

Time sets.....	Ch1	p13
	Ch2	p3-7

Two-state Markov model

Assumptions.....	Ch4	p2-3
Cramér-Rao lower bound.....	Ch4	p20
Future lifetime random variable	Ch4	p8
Joint density function.....	Ch4	p10-11
Maximum likelihood estimator of μ	Ch4	p13-20
Survival probabilities.....	Ch4	p5
Waiting time	Ch4	p9

Weibull model (see Future lifetime)

Chapter 1

Principles of actuarial modelling



Syllabus objectives

- (i) *Describe the principles of actuarial modelling.*
1. *Describe why and how models are used.*
 2. *Explain the benefits and limitations of modelling.*
 3. *Explain the relative suitability of deterministic and stochastic models.*
 4. *Describe, in general terms, how to decide whether a model is suitable for any particular application.*
 5. *Explain the difference between the short-run and long-run properties of a model, and how this may be relevant in deciding whether a model is suitable for any particular application.*
 6. *Describe, in general terms, how to analyse the potential output from a model, and explain why this is relevant to the choice of model.*
 7. *Describe the process of sensitivity testing of assumptions and explain why this forms an important part of the modelling process.*
 8. *Explain the factors that must be considered when communicating the results following the application of a model.*

0 ***Introduction***

This introductory chapter gives useful background information about modelling, in particular within an actuarial context. The following general principles are covered:

- why we want to model
- how to model
- testing the suitability of the model
- analysing the output
- communicating the results.

Most exam questions are numerical and hence based on the models discussed in later chapters. Once you have studied these examples you may find it useful to reread the general principles outlined here to appreciate the broad picture. You could be asked to comment on a particular model or modelling procedure for which this kind of background would be appropriate.

1 Models

1.1 Why models are used

A model is an imitation of a real world system or process. Models of many activities can be developed, for example, the economy of a country, the workings of the human heart, the future cashflows of the broker distribution channel of a life insurance company.

The expression “broker distribution channel” refers to the fact that in life insurance it is common for payments to go through a broker who acts as an intermediary between the policyholder and the insurance company. This is also true of general insurance.

Suppose we wished to “predict” the effect that a real world change would have on these three models. In some cases it might be too risky, or too expensive or too slow, to try a proposed change in the real world even on a sample basis. Trying out the change first without the benefit of a model could have serious consequences. The economy might go into recession costing a government the next election, the patient might die and the life office could suffer a surge in new business but at highly unprofitable premium rates.

Parameters

A model enables the possible consequences to be investigated. The effect of changing certain input parameters can be studied before a decision is made to implement the plans in the real world.

Suppose, for example, we want to buy shares in some company. In order to calculate what price we should pay we would like to have some idea of the future price of these shares, so we should develop a model relating the price of the shares to various factors. In our model we *might* assume (among other things) a fixed value for future interest rates, say 6%. This value is an input parameter. It is fixed within the model, but we can vary it to obtain different models. Using the figure 6% will give us one value for the price we should pay, but we cannot be certain that 6% was the correct input. What if interest rates rise to 10%, or fall to 3% and so on? By varying the parameters in the model we can answer questions such as these.

Note that if we model interest rates by a random variable, *eg* as a lognormal distribution with parameters μ and σ , then although the interest rate itself is not fixed, its distribution is. The input parameters are then μ and σ .

To build a model of a system or process, a set of mathematical or logical assumptions about how it works needs to be developed. The complexity of a model is determined by the complexity of the relationships between the various model parameters. For example, in modelling a life office, consideration must be given to issues such as regulations, taxation and cancellation terms. Future events affecting investment returns, inflation, new business, lapses, mortality and expenses also affect these relationships.

In insurance terms, a cancellation occurs when a policyholder terminates their insurance cover, *eg* when you change your car insurance to a different insurer mid-year. A lapse occurs when you don't pay a premium that is due, *eg* when you change to a new insurer when your policy with the old one is up for renewal.

Data

In order to produce the model and determine suitable parameters, data need to be considered and judgements need to be made as to the relevance of the observed data to the future environment. Such data may result from past observations, from current observations (such as the rate of inflation) or from expectations of future changes.

An example of a future change in an insurance context would be if the government decided to increase the rate of insurance tax.

Where observed data are considered to be suitable for producing the parameters for a chosen model, statistical methods can be used to fit the data.

For example, if we were to model the daily changes in the price of shares in some company by a normal distribution, then the parameters would be the mean and the variance. We could estimate these by looking at past data and calculating the sample mean and sample variance. These would be sensible estimates.

Objectives

Before finalising the choice of model and parameters, it is important to consider the objectives for creation and use of the model. For example, in many cases there may not be a desire to create the most accurate model, but instead to create a model that will not underestimate costs or other risks that may be involved.

Moreover, we need to decide what the “best” model is, within the context of its intended use. For example, we might wish to model the number of claims made to an insurance company for various amounts. To simplify matters, suppose there are only two possible probability distributions that we are choosing between. If most claims are for small amounts and one distribution models these small amounts better than the other, then this might seem a natural choice. However, if this same distribution models the large claims relatively poorly, then although there are not many of these large claims, the fact that they are for large amounts is significant. The decisions resulting from a poor model of the higher claims may be more costly to the company than those based on a model that is slightly inaccurate on a lot of small claims.

1.2 **How models are used**

While in reality a modelling process does not follow a rigid pattern of prescribed steps, it is helpful in introducing the topic to imagine a set of key steps. In practice, actuaries who build and use models move back and forth between these key steps continuously to improve the model.

The key steps in a modelling process can be described as follows:

- 1. Develop a well-defined set of objectives that need to be met by the modelling process.**

Continuing our last example on modelling the size of insurance claims, in addition to the basic purpose of being able to predict the number of claims of different sizes, one objective that might be demanded is to give as accurate a prediction as possible for 95% of the claims.

- 2. Plan the modelling process and how the model will be validated.**

The validation of the model will involve a series of diagnostic tests to ensure that the model meets the objectives we want.

- 3. Collect and analyse the necessary data for the model.**

- 4. Define the parameters for the model and consider appropriate parameter values.**

- 5. Define the model initially by capturing the essence of the real world system. Refining the level of detail in the model can come at a later stage.**

- 6. Involve experts on the real world system you are trying to imitate so as to get feedback on the validity of the conceptual model.**

7. Decide on whether a simulation package or a general purpose language is appropriate for the implementation of the model. Choose a statistically reliable random number generator that will perform adequately in the context of the complexity of the model.

Models based on a deterministic approach would not need this.

8. Write the computer program for the model.

After this stage we can run the model.

9. Debug the program to make sure it performs the intended operations in the model definition.

10. Test the reasonableness of the output from the model.

11. Review and carefully consider the appropriateness of the model in the light of small changes in input parameters.

Suppose, for example, that small changes in the input parameters lead to large changes in the output. If these parameters cannot be known with great accuracy then we cannot be sure that our predictions will be valid. However, we could still use the model to come up with a range of possible outputs by assuming a range of values for the input parameters. This is covered in Section 8 on sensitivity testing.

12. Analyse the output from the model.

13. Ensure that any relevant professional guidance has been complied with. For example the Board for Actuarial Standards has issued Technical Actuarial Standards on data, modelling and reporting: TAS D, TAS M and TAS R. (A knowledge of the detail of these TASs is not required for CT4.)

14. Communicate and document the results and the model.

In the Core Applications (CA) series subjects you will meet the concept of the “Actuarial Control Cycle.” This refers not only to analysing problems and developing solutions in a logical way, similar to that above, but also to the possibility of improving the model by repeating Steps 3 to 12 at a later date in the light of data obtained since the original predictions were made.

2 ***Modelling – the benefits and limitations***

2.1 ***Advantages of models***

In actuarial work, one of the most important benefits of modelling is that systems with long time frames – such as the operation of a pension fund – can be studied in compressed time.

Other benefits include:

- Complex systems with stochastic elements such as the operation of a life insurance company cannot be properly described by a mathematical or logical model that is capable of producing results that are easy to interpret. Simulation modelling is a way of studying the operation of a life insurance company.

So, where randomness is an essential part of the process being modelled, this randomness must be included in the model in order to get meaningful results. We'll be explaining what "stochastic" means in the next section.

- Different future strategies or possible actions can be compared to see which best suits the requirements or constraints of a user.
- In a model of a complex system we can usually get control over the experimental conditions so that we can reduce the variance of the results output from the model without upsetting their mean values.

2.2 ***Disadvantages***

However, models are not the simple solution to all actuarial problems – they have drawbacks that must be understood when interpreting the output from a model and communicating the results to clients.

The drawbacks include:

- Model development requires a considerable investment of time, and expertise. The financial costs of development can be quite large given the need to check the validity of the model's assumptions, the computer code, the reasonableness of results and the way in which results can be interpreted in plain language by the target audience.

In an actuarial context, the target audience could be, for example:

- a life office policyholder, who wants an idea of how much (s)he will receive from his/her endowment policy when it matures
 - a pension fund client who needs to know how much to pay into the company's pension fund next year
 - the finance director of a general insurance company who wants an estimate of the end of year profit figures.
- **In a stochastic model, for any given set of inputs each run gives only estimates of a model's outputs. So to study the outputs for any given set of inputs, several independent runs of the model are needed.**

A stochastic model allows for randomness and each computer “run” would give the figures for one possible outcome.

- **As a general rule, models are more useful for comparing the results of input variations than for optimising outputs.**

In other words it's easier to use a model to predict what *might* happen than it is to determine what inputs would be required to obtain a particular outcome.

- **Models can look impressive when run on a computer so that there is a danger that one gets lulled into a false sense of confidence. If a model has not passed the tests of validity and verification its impressive output is a poor substitute for its ability to imitate its corresponding real world system.**
- **Models rely heavily on the data input. If the data quality is poor or lacks credibility then the output from the model is likely to be flawed.**

In actuarial terminology, “credibility” refers to the extent to which data can be relied on.

- **It is important that the users of the model understand the model and the uses to which it can be safely put. There is a danger of using a model as a "black box" from which it is assumed that all results are valid without considering the appropriateness of using that model for the particular data input and the output expected.**
- **It is not possible to include all future events in a model. For example a change in legislation could invalidate the results of a model, but may be impossible to predict when the model is constructed.**
- **It may be difficult to interpret some of the outputs of the model. They may only be valid in relative rather than absolute terms, as when, for example, comparing the level of risk of the outputs associated with different inputs.**

“Risk” refers to the level of uncertainty associated with an outcome.

3 **Stochastic and deterministic models**

If it is desired to represent reality as accurately as possible the model needs to imitate the random nature of the variables. A *stochastic* model is one that recognises the random nature of the input components. A model that does not contain any random component is *deterministic* in nature.

In a deterministic model, the output is determined once the set of fixed inputs and the relationships between them have been defined. By contrast, in a stochastic model the output is random in nature – like the inputs, which are random variables. The output is only a snapshot or an estimate of the characteristics of the model for a given set of inputs. Several independent runs are required for each set of inputs so that statistical theory can be used to help in the study of the implications of the set of inputs.

A deterministic model is really just a special (simplified) case of a stochastic model.

The example on the next page illustrates the difference between the two approaches.



Example

An investor has bought shares worth £5,000 and wants to estimate how much they will be worth in a year's time. Give both a deterministic and stochastic model based on an expected growth rate of 7%.

Solution

Deterministic model

The outcome from a deterministic model would be the prediction that the value in a year's time would be:

$$5000 \times 1.07 = \text{£}5350$$

Stochastic model

A stochastic model would have to allow for randomness in the growth rate. For example, it might be decided (based on past performance of the company and prospects for the company, the investment sector and the economy in general) that the probabilities of different growth rates for the shares are:

$$\text{growth} = \begin{cases} 20\% & \text{with probability 0.1} \\ 10\% & \text{with probability 0.6} \\ 0\% & \text{with probability 0.2} \\ -10\% & \text{with probability 0.1} \end{cases}$$

Note that these probabilities add up to 1 and that the expected growth rate is still 7% for this model since:

$$0.1 \times 20\% + 0.6 \times 10\% + 0.2 \times 0\% + 0.1 \times (-10\%) = 7\%$$

The outcome from this model, if it was run 100 times, would be a list of 100 predicted values, which might look like this:

£5500, £5000, £5000, £6000, £5000, ... , £4500, £6000

This situation is very similar to the stochastic interest rate models in Subject CT1.

Whether one wishes to use a deterministic or a stochastic model depends on whether one is interested in the results of a single “scenario” or in the distribution of results of possible “scenarios”. A deterministic model will give one the results of the relevant calculations for a single scenario; a stochastic model gives distributions of the relevant results for a distribution of scenarios. If the stochastic model is investigated by using “Monte Carlo” simulation, then this provides a collection of a suitably large number of different deterministic models, each of which is considered equally likely.

“Monte Carlo” simulation is where a computer is set up to run a stochastic model a number of times, eg 10,000 times, using pseudo-random numbers generated by the computer. The results look like the list of numbers in our stochastic model example. The numbers generated by the computer are pseudo-random rather than truly random because they are generated using a prescribed method.



Question 1.1

Give examples of when stochastic models might be more useful than the deterministic ones.

The results for a deterministic model can often be obtained by direct calculation, but sometimes it is necessary to use numerical approximations, either to integrate functions or to solve differential equations.

If a stochastic model is sufficiently tractable, it may be possible to derive the results one wishes by analytical methods. If this can be done it is often preferable to, and also often much quicker than, Monte Carlo simulation. One gets precise results and can analyse the effect of changes in the assumptions more readily. Monte Carlo simulation is covered in Subject CT6.

We met situations like this when we studied stochastic interest rates in Subject CT1. With certain assumed interest rate distributions it was possible to determine explicitly the distribution of, for example, the accumulated value after 5 years.

Many practical problems, however, are too complicated for analytical methods to be used easily, and Monte Carlo simulation is an extremely powerful method for solving complicated problems. But if even part of a model can be treated analytically, it may provide a check on any simulation method used. It may be possible to use a deterministic method to calculate the expected values, or possibly the median values, for a complicated problem, where the distributions around these central values are estimated by simulation.

One also needs to recognise that a simulation method generally provides “what if?” answers; what the results are on the basis of the assumptions that have been made. It is much harder to use simulation to provide the optimum solution; in other words to find the set of assumptions that maximises or minimises some desired result.

A further limitation is that the precision of a simulated result depends on the number of simulations that are carried out. This is covered in more detail in Subject CT6.

4 ***Discrete and continuous state spaces and time sets***

The state of a model is the set of variables that describe the system at a particular point in time taking into account the goals of the study. It is possible to represent any future scenarios as states, as will be developed in the later chapters.

Discrete states are where the variables exhibit step function changes in time, for example, from a state of alive to dead, or an increase in the number of policies for an insurer. Continuous states are where the variables change continuously with respect to time: for example, real time changes in values of investments.

The decision to use a discrete or continuous state model for a particular system is driven by the objectives of the study, rather than whether or not the system itself is of a discrete or continuous nature. A model may also consider time in a discrete or a continuous way. This may reflect the fact that outputs from the model are only required at discrete points in time, or may be to satisfy the objectives of the modelling.

One cannot in practice use Monte Carlo simulation for a continuous time problem; one has to discretise the time step.

For example, a model of notification of insurance claims might look at how many claims are recorded each day. The precise time of day would not be needed.

This can be done with whatever precision one likes, but the higher the precision the longer the time taken to process any particular model. This may or may not be a limitation in practice. And it should be remembered that some results for continuous time, continuous space models can not be obtained by discrete simulation at all.

Processes with continuous state spaces will be discussed in Subjects CT6, CT8 and ST6.

5 **Suitability of a model**

In assessing the suitability of a model for a particular exercise it is important to consider the following:

- The objectives of the modelling exercise.
- The validity of the model for the purpose to which it is to be put.
- The validity of the data to be used.
- The possible errors associated with the model or parameters used not being a perfect representation of the real world situation being modelled.
- The impact of correlations between the random variables that “drive” the model.
- The extent of correlations between the various results produced from the model.
- The current relevance of models written and used in the past.
- The credibility of the data input.
- The credibility of the results output.
- The dangers of spurious accuracy.
- The ease with which the model and its results can be communicated.

The important actuarial/investment concept of “matching” of assets and liabilities relies on the fact that the value of the matched assets and liabilities will tend to move together, *ie* they are positively correlated. In models of such situations, it would be essential to incorporate this correlation.

An example of “spurious accuracy” would be if someone said to you: “The value of our company pension fund’s assets is £46,279,312.86. The company printout says so.” This is spurious accuracy because the market value of the investment will change by the minute and this level of accuracy cannot be justified. A more appropriate figure to give you would have been £46.3m say.

Question 1.2



Suppose the investor in the example on Page 10 has a portfolio consisting of shares in 10 different companies. Each of her 10 shareholdings is currently worth £5,000. She simulates the value of her portfolio in one year’s time by running the stochastic model in that example one hundred times, and pooling the results. She then uses the results to estimate the probability that she will have the £50,000 she requires. Why might her estimated probability not be valid?

6 ***Short-term and long-term properties of a model***

The stability of the relationships incorporated in the model may not be realistic in the longer term. For example exponential growth can appear linear if surveyed over a short period of time. If changes can be predicted they can be incorporated in the model, but often it must be accepted that longer-term models are suspect.

Models are by definition, simplified versions of the real world. They may therefore ignore "higher order" relationships which are of little importance in the short term, but which may accumulate in the longer term.

There is an analogy here with series. For small values of time t , the exponential function e^t appears linear, since:

$$e^t = 1 + t + o(t^2)$$

But if t is greater, we need to include the higher order terms:

$$e^t = 1 + t + \frac{1}{2}t^2 + \dots$$

7 ***Analysing the output of a model***

Statistical sampling techniques are needed to analyse the output of a model, as a simulation is just a computer-aided statistical sampling project. The actuary must exercise great care and judgement at this stage of the modelling process as the observations in the process are correlated with each other and the distributions of the successive observations change over time. The useless and fatally attractive temptation of assuming that the observations are independent and identically distributed is to be avoided.

If there is a real world system against which results can be compared, a “Turing” test should be used. In a Turing test, experts on the real world system are asked to compare several sets of real world and model data without being told which are which. If these experts can differentiate between the real world and model data, their techniques for doing so could be used to improve the model.

This is an extension of the original meaning of a Turing test named after the British mathematician and early computer pioneer, Alan Turing. This refers to one objective of artificial intelligence researchers, which is to write a computer program that cannot be distinguished from a real person by a user asking questions over a computer link.

8 Sensitivity testing

It is important to test the reasonableness of the output from the model against the real world if possible. An examination of the sensitivity of the outputs to small changes in the inputs or their statistical distributions should be carried out. The appropriateness of the model should then be reviewed, particularly if small changes in inputs or their statistical distributions give rise to large changes in the outputs. In this way, the key inputs and relationships to which particular attention should be given in designing and using the model can be determined.

Note that if small changes in the inputs give rise to large changes in the outputs then our initial choices are more crucial. How confident are we in our choices of input? If the resulting changes in output are small then our initial choices are less important in this respect.

Sensitivity testing also refers to the approach of using a deterministic model with slight changes in one or more of the assumptions to see the range of possible outcomes that might occur. For example, personal pension plans provided in the UK are required to give illustrations to policyholders showing how much pension they would get if growth rates were 5%, 7% and 9% *pa* and interest rates used for converting the fund into a pension were 4%, 6% and 8% *pa*. This allows the policyholder to gauge the extent to which the resulting pension will be affected by changes in the growth rate.

The model should be tested by designing appropriate simulation experiments. Through this process the model can be refined.

An approach that has been traditionally used by actuaries in the fields of insurance, pensions and investment is to carry out a set of deterministic calculations using different actuarial bases, *ie* under different sets of assumptions. By varying the assumptions, the actuary could use the model to arrive at figures that were “best estimates” (the most likely, or median, result) or “optimistic” (if things worked out favourably) or “cautious” (if things worked out badly).

9 **Communication of the results**

The final step in the modelling process is the communication and documentation of the results and the model itself to others. The communication must be such that it takes account of the knowledge of the target audience and their viewpoint. A key issue here is to make sure that the client accepts the model as being valid and a useful tool in decision making. It is important to ensure that any limitations on the use and validity of the model are fully appreciated.

The following example illustrates one possible limitation of a model that would need to be explained to a client.



Example

An actuarial student working at a pension consultancy is asked by a client how expensive it would be for the client's company pension scheme to offer more generous benefits to members who left the company before reaching retirement age.

As the company had in the past had a very stable workforce with very little turnover of staff, the model we were using at the time assumed that everyone would stay with the company until retirement age. (If somebody did leave, this would simply lead to a small "profit" or "loss" at that time.) So the client's question could not be answered using the existing model.

The solution, however, was easy. We simply needed to extend the model to allow for an appropriate rate of withdrawal of staff before retirement age, and then compare the costs of the scheme with the existing withdrawal benefit with the costs for the "improved" one.

10 Specimen exam questions

At the end of each of the chapters in this course we have included an exam-style question for you to attempt. The first one of these is coming up at the end of this chapter.

We suggest two different ways that you might wish to use these questions to help you progress through the course:

- (1) You could attempt the questions as soon as you reach them in your studies. You may find them quite difficult on the first attempt, and we would expect you to refer back to the notes in order to answer them. However, by tackling them as you go through the course, you will get to know more quickly the level you need to be aiming for in order to pass the Subject CT4 exam. But you should not be worried if your answers appear far from perfect on these first attempts.
- (2) Alternatively, you could miss them out until you get to the end of each part. At this point you should be aiming to tackle a good sample of questions from the Question and Answer Bank prior to attempting the relevant assignment. Immediately before the assignment, you could go back and attempt all the exam-style questions from the relevant part, which should help your preparation for tackling the assignment.

Whichever of these you follow, you are likely to benefit from a fresh second attempt at these questions as part of your revision. On these second attempts you should be looking to do the questions under exam conditions, and strictly within the time available – remember that 1 mark = 1.8 minutes of exam time! We suggest that you don't try all the questions in one sitting, but tackle them one at a time once you have fully revised the topics involved.

Of course, you may wish to use the questions in other ways. These are just suggestions.

11 Exam-style question

Here is the first of the specimen exam questions that we have included in this course.



Question 1.3

Subject CT4, April 2005, Question A2

You have been commissioned to develop a model to project the assets and liabilities of an insurer after one year. This has been requested following a change in the regulatory capital requirement. Sufficient capital must now be held such that there is less than a 0.5% chance of liabilities exceeding assets after one year.

The company does not have any existing stochastic models, but estimates have been made in the planning process of “worst case” scenarios.

Set out the steps you would take in the development of the model.

[7]



Chapter 1 Summary

A model is an imitation of a real world system or process.

It enables possible consequences to be investigated without carrying out the actions themselves.

There are benefits to modelling, such as the possibility of looking at long-term phenomena in an accelerated time frame, but there are also limitations that must be considered such as the time and expertise required to develop and run a model.

The following 14 key steps can be considered in the construction and use of a model:

1. Develop a well-defined set of objectives that need to be met by the modelling process.
2. Plan the modelling process and how the model will be validated.
3. Collect and analyse the necessary data for the model.
4. Define the parameters for the model and consider appropriate parameter values.
5. Define the model initially by capturing the essence of the real-world system. Refining the level of detail in the model can come at a later stage.
6. Involve the experts on the real-world system you are trying to model in order to get feedback on the validity of the conceptual model.
7. Decide on whether a simulation package or general-purpose language is appropriate for the implementation of the model. If necessary, choose a statistically reliable random number generator that will perform adequately in the context and complexity of the model.
8. Write the computer program for the model.
9. Debug the program to make sure it performs the intended operations in the model definition.
10. Test the reasonableness of the output from the model.
11. Review and carefully consider the appropriateness of the model in the light of small changes to the input parameters.
12. Analyse the output from the model.
13. Ensure that any relevant professional guidance has been complied with. For example the Board for Actuarial Standards has issued Technical Actuarial Standards on data, modelling and reporting: TAS D, TAS M and TAS R.
14. Communicate and document the results and the model.

To build a model the suitability of that model to the objectives should be borne in mind. Some mathematical or logical assumptions about the workings of the real world system must be made. Input parameters need to be chosen, possibly by statistical analysis of past data. Sensitivity analysis of the dependence of the output on these parameters can be carried out.

A model may be stochastic or deterministic. For stochastic models it is better to use direct calculation if possible, but in complex situations it may be necessary to use Monte Carlo simulation on a computer.

Models may also be constructed in discrete or continuous time and with discrete or continuous state spaces. Monte Carlo simulations have to be run in discrete time.

The output needs to be analysed. This can be done using the idea of a Turing test – can an expert tell the difference between a set of simulated outputs and actual occurrences?

The results need to be communicated to other people. The key question when framing the level of communication is “to whom?”

Chapter 1 Solutions

Solution 1.1

The investor in the previous example might be using a similar model, with different assumptions for all the investments in her portfolio. The stochastic model would allow her to estimate how likely it was that the entire portfolio was worth at least £50,000 (say) in a year's time. She may need to have £50,000 available for a specific purpose.

Stochastic models can also be used to estimate shortfall probabilities and confidence intervals.

This wouldn't be possible using a deterministic model.

Solution 1.2

In practice, the share prices are likely to be positively correlated, *ie* there is a tendency for the prices of different shares to move together because of general factors affecting the whole market, *eg* changes in interest rates. A stochastic model such as the one described does not allow for this. It would assume that the changes in the share price were independent and therefore tended to "average out". As a result it would underestimate the probability that all the shares might fall together.

This positive correlation is the reason why investment strategists always recommend diversification, ie investing in a range of different assets whose behaviour is as far as possible independent.

In addition, the model she uses may not be representative of the real-life system, the parameters may be wrong, and the results will be subject to random sampling error.

Solution 1.3

Comment

This is a tough question because it asks you to respond to a particular situation rather than just to repeat 'bookwork' ideas from the chapter.

A bit of informal 'brainstorming' before writing an answer should ensure that you don't miss any points and that you highlight the important ones.

We are asked to ‘set out the steps in the development of the model’. This means a sequential list of things that need to be done. We do not need to say anything about how the model output would be used.

So what are these steps? Think carefully about how you have built models in the past. Words that come to mind are:

- Specification, ie writing down the equations, listing the statistical assumptions.
- Estimation, ie collecting the data, calculating estimates of the model parameters.
- Testing or verification, ie making sure that the model ‘fits the data’ and ‘produces realistic answers’.
- Using the model to get some results, ie how much capital is needed?
- Sensitivity analysis to see how robust our results are to changes in the model parameters.

The criterion we are being asked to investigation is stochastic. The probability relates to the difference of two quantities, the value of assets and liabilities respectively in a year’s time.

We will need to model the value of the assets and the value of the liabilities separately. Things will be simpler if we concentrate on current assets and liabilities, and ignore (or mention in passing) new assets and liabilities arising during the year.

Similar factors will affect the value of both the assets and the liabilities. This means the models should be run simultaneously and the value of the assets minus the value of the liabilities should be calculated at the end of each run.

The results of lots of runs (how many?) can be summarised in a frequency distribution of net liabilities. The key summary measure is the frequency of positive net liabilities. We are expected to ensure that this frequency is less than 0.5%.

‘The company does not have any existing stochastic models’ – so we are going to have to say something about the choice of models. However we are told – ‘estimates have been made of worse case scenarios’. This information will help to restrict the choice of models.

The value of the current assets (capital) is under our control so we can change this to make sure that the 0.5% requirement is met. Other parameters, eg future long-term interest rates, future inflation rates are not under our control. So we will need to test how sensitive our answers are to changes in these parameter values.

Example solution

Steps

1. Set out the modelling objectives. For example, to predict the value of assets and the value of liabilities in one year's time.
2. Set out the validation process. For example, run the model using past data and ensure that the objectives are met.
3. Collect and analyse data the necessary data for the model, *eg* information about the current yield curve for interest rates.
4. Define the parameters and assign appropriate parameter values, *eg* for the interest rate to use.
5. Define the model, *eg* that future long-term interest rates will follow a lognormal distribution.
6. Consult experts with experience of the real world system on the appropriateness of the specified model.
7. Decide on how the simulation will be carried out, *eg* how many runs to do.
8. Write the necessary computer programs for the model simulation.
9. Test the computer programs to ensure they operate correctly.
10. Carry out simulations of assets and liabilities in one year's time and check that the simulated distributions appear reasonable. This checking might involve the use of past data or consulting experts.
11. Test the sensitivity of the simulation results to the choice of the parameter values estimated in step 3. Use the 'worse case scenarios' to choose a suitable range of parameter values. Determine how sensitive the capital needed is to this variation in the parameter values.
12. Analyse the results.
13. Check that all relevant professional guidance has been complied with.
14. Document the simulation results, and decide how the key results should be communicated to those who are required to make decisions about the capital needed.

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

ActEd Tutorials

Have you booked yours yet?



"Tutor was excellent, able to answer all questions, very vibrant and engaging. Tutorials were very enjoyable."



Order at www.ActEd.co.uk/estore

Visit www.ActEd.co.uk for more details or to see what's on offer..

If you are unable or unwilling to travel to one of the advertised courses but would like to attend a course in a location near you, submit a request for a tutorial on the estore at www.ActEd.co.uk/estore by clicking on "Tutorials" and then "Request a Tutorial (non-scheduled)". You can state your chosen subject, course type and your preferred location.



www.ActEd.co.uk/forums

Chapter 2

Stochastic processes



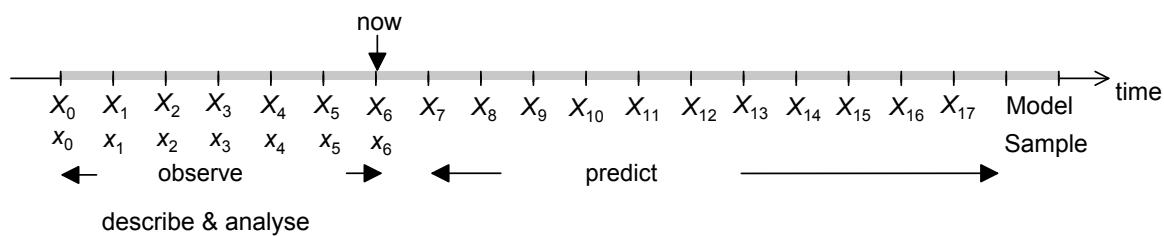
Syllabus objectives

- (ii) *Describe the general principles of stochastic processes, and their classification into different types.*
1. *Define in general terms a stochastic process and in particular a counting process.*
 2. *Classify a stochastic process according to whether it:*
 - (a) *operates in continuous or discrete time*
 - (b) *has a continuous or a discrete state space*
 - (c) *is a mixed type**and give examples of each type of process.*
 3. *Describe possible applications of mixed processes.*
 4. *Explain what is meant by the Markov property in the context of a stochastic process and in terms of filtrations.*

0 *Introduction*

In this chapter we define the concept of a *stochastic process*. A stochastic process is a family or set of ordered random variables. The order is indicated by indexing each random variable in the family by a subscript. Usually the ordering is a result of the random variables being observed over time, so X_t is a random variable that models the value of the stochastic process at time t . The random variables in the set can be dependent on one another reflecting the nature of the process being modelled.

As in all statistical modelling we will collect a sample of data from the process being modelled. So for example, X_t might model the price of a stock at time t , and we have observations of the stock price for the last 6 trading days. We can use these data to describe the process and to analyse the nature of its past behaviour over time. The data may also be used to estimate the parameters of our stochastic process model. We could then use the estimated stochastic process model to predict the future behaviour of the stock price. It is the dependence between the random variables in the set that allows us to make predictions by extrapolating past patterns into the future.



We start with some definitions and then give examples of different types of processes. The final parts of the first section discuss some of the properties that a stochastic process may possess, *ie stationarity, independent increments and the Markov property*.

In Section 3 we look at several examples of stochastic processes, and use these to illustrate the definitions and properties we have given in the previous section. Some of these examples will be discussed in more detail in later chapters.

1 Types of stochastic processes

We begin with the definition of a stochastic process.



Definition

A **stochastic process** is a model for a time-dependent random phenomenon. So, just as a single random variable describes a static random phenomenon, a stochastic process is a collection of random variables X_t , one for each time t in some set J .

The process is denoted $\{X_t : t \in J\}$.

The set of values that the random variables X_t are capable of taking is called the **state space of the process**, S .

The values in the set J are called the *time domain* or *time set* of the process. The terminology is one that you will have seen before. If $y = f(x)$, the values taken by x are the domain of the function and the values taken by y are the range of the function.

The set of random variables may be dependent. So in order to describe their statistical properties we will need to know their joint distribution. If the random variables in the set were independent (as is the case for example with statistical models of the sampling process) it would be sufficient to know the (marginal) distribution of each random variable in the set.

The random variables in the set need not be identically distributed. However, we will discover that processes that have identically distributed random variables are particularly important in the study of stochastic processes.

The state space of the stochastic process will include all of the values that can be taken by any of the random variables in the set. However, for particular random variables in the set, some of these values may have a zero probability of occurring.

For example, we might model the closing value of the FTSE100 index by a stochastic process $\{X_t\}$. The random variable X_t models the value at the end of day t .



Question 2.1

Are the state space and the time domain for this process discrete or continuous?

A possible model might say that the value of X_t depends on the values at the end of the two previous trading days X_{t-1} and X_{t-2} . If this dependence does not change with t , then we could use the model to predict future values of X_t . Note that these predictions may not be exact even if the model is good.

The first choice that one faces when selecting a stochastic process to model a real life situation is that of the nature (discrete or continuous) of the time set J and of the state space S .

While the process being modelled could have its value recorded continuously or at very frequent discrete times, we may choose to model only the values at discrete or less frequent discrete time points. This may be because we are only able to record measurements at these times. This could be because of physical limitations on the measurement process or because the measurement process is very expensive and we cannot afford more frequent measurements. Aside from these considerations we may be content just to model the process at these time points because, for example, predictions using this frequency will be perfectly adequate for our needs. No purpose is served by using a more elaborate model than is necessary.

There is no requirement that the labels used in the set J should be actual calendar times, merely that they should put the random variables in order.

In statistical modelling it is common to approximate (the state space of) discrete random variables by continuous random variables when the number of discrete values becomes large enough. So, for example, we often approximate a discrete binomial random variable by a continuous normal random variable when the binomial random variable has more than 20 or 30 discrete values. We know that the continuous model is not an exact representation of what is being modelled but it is adequate for our purposes and is easier to use.

1.1 Discrete state space with discrete time changes

Here is an example of a process with a discrete state space and discrete time changes.

A motor insurance company reviews the status of its customers yearly. Three levels of discount are possible (0, 25%, 40%) depending on the accident record of the driver. In this case the appropriate state space is $S = \{0, 25, 40\}$ and the time set is $J = \{0, 1, 2, \dots\}$ where each interval represents a year. Problems involving no claims discount systems (such as this) are studied in Chapter 3.

Note that the time set often starts at 0 (whether continuous or discrete). The time 0 is taken to be the start, so that after one unit of time (day, minute *etc*) we have $t = 1$.

In principle the company could record the discount status of each policyholder on a continuous basis, but discount levels are usually only changed on the annual renewal date of the policy. So it makes sense just to model and record these values. A model with more frequent recording will be more complicated (and expensive) yet is unlikely to be more useful in managing a portfolio of motor policies.

The time domain is discrete. The state space contains three discrete values.

1.2 ***Discrete state space with continuous time changes***

A life insurance company classifies its policyholders as Healthy, Sick or Dead. Hence the state space $S = \{H, S, D\}$. As for the time set, it is natural to take $J = [0, \infty)$ as illness or death can occur at any time. On the other hand, it may be sufficient to count time in units of days, thus using $J = \{0, 1, 2, \dots\}$. This model is studied in some detail in Chapters 5 and 6.

The description suggests that the time domain could be continuous or discrete with very frequent recording. In using the model it will be important that we are able to answer questions like ‘What is the probability that a life that is healthy at time s is sick at time t ?’ This suggests that a model with a continuous time domain will be more useful.

Here the state space contains three discrete values.

1.3 ***Continuous state space***

Claims of unpredictable amounts reach an insurance company at unpredictable times; the company needs to forecast the *cumulative claims* over $[0, n]$ in order to assess the risk that it might not be able to meet its liabilities. It is standard practice to use $[0, \infty)$ both for S and J in this problem. However, other choices are possible: claims come in units of a penny and do not really form a continuum. Similarly the intra-day arrival time of a claim, that is the time at which it arrives on a particular day, is of little significance, so that $\{0, 1, 2, \dots\}$ is a possible choice for J and/or S .

The choice of a discrete or continuous time domain would be influenced by the availability of data, *eg* are cumulative claims figures recorded on a daily basis or are figures only available at the end of each quarter; and by the purpose of the modelling *eg* will predictions of the cumulative claims at the end of each quarter be sufficient or are predictions needed more frequently.

Claim amounts may be recorded in pence or to the nearest pound and so in principle the state space is discrete, but it contains a very large number of non-negative values and so a continuous approximation will be perfectly adequate.

An important class of models having a continuous state space and a discrete time domain is *time series*. Many economic and financial stochastic processes fall into this class, *eg* daily prices at the close of trading for a company's shares.

1.4 **Displaying observed data**

When we take observations on a process we obtain a sample of each of the random variables in the set making up the stochastic process.

In displaying the data we want to convey information about three features. These are:

- the size of the values, *ie* to give an idea of the means of the random variables in the set
- the volatility of the values, *ie* to give an idea of the variances of the random variables in the set
- the relationships between the values of the random variables, *ie* to give an idea of the covariances between the random variables in the set.

We do this by plotting X_t against t . Even when the time domain is discrete we can join up the plotted points with straight lines. There are no observations on these lines, but they can help to show the "shape" of the time series. You will have seen plots like this in newspapers, *eg* showing the price of a stock each day or the interest rate at the end of each quarter. Plots like this can be used even when the state space of the process is discrete.



Question 2.2

You are thinking of moving to live in Edinburgh. As part of your research into what it's like to live in Edinburgh you have collected the following sets of data:

- the maximum daily temperature each day since 1 January 2006
 - whether or not it rained for each day since 1 January 2006
 - the number of cyclists injured in road accidents since 1 January 2006.
- (i) For each data set choose an appropriate state space and a time domain. In each case state whether the state space and the time domain are discrete or continuous, and give the units of measurement, *eg* dollars, weeks.
- (ii) Imagine that you have data for each process. Draw sketches to display these samples of data.

1.5 Processes of mixed type

Just because a stochastic process operates in continuous time does not mean that it cannot also change value at predetermined discrete instants; such processes are said to be of mixed type. As an example, consider a pension scheme in which members have the option to retire on any birthday between ages 60 and 65. The number of people electing to take retirement at each year of age between 60 and 65 cannot be predicted exactly, nor can the time and number of deaths among active members. Hence the number of contributors to the pension scheme can be modelled as a stochastic process of mixed type with state space $S = \{1, 2, 3, \dots\}$ and time set or domain $J = [0, \infty)$. Decrement amounts will occur at fixed dates due to retirement as well as at random dates due to death.

So this process is a combination (mixture) of two processes:

- a stochastic process modelling the number of deaths which has a discrete state space and a continuous time domain
- a stochastic process modelling the number of early retirements which has a discrete state space and a discrete time domain.

We model the total number of decrements and observe the initial number of members less the total number of decrements in $(0, t)$. This is a mixed process.



Question 2.3

You run a business that sells and provides service for a range of expensive sports cars. Each car sells for between £40,000 and £50,000 (cash only) and you sell about 10 to 20 each year. The “life blood” of the business is the regular servicing and maintenance of the cars you have sold previously.

Describe the characteristics of a stochastic process that might be a suitable model for the balance on your company’s bank account.

As a rule, one can say that continuous time and continuous state space stochastic processes, although conceptually more difficult than discrete ones, are also ultimately more flexible (in the same way as it is easier to calculate an integral than to sum an infinite series).

It is important to be able to conceptualise the nature of the state space of any process which is to be analysed, and to establish whether it is most usefully modelled using a discrete, a continuous, or a mixed time domain. Usually the choice of state space will be clear from the nature of the process being studied (as, for example, with the Healthy-Sick-Dead model), but whether a continuous or discrete time set is used will often depend on the specific aspects of the process that are of interest, and upon practical issues like the time points for which data are available.

1.6 Counting processes

A counting process is a stochastic process, X , in discrete or continuous time, whose state space S is the collection of natural numbers $\{0, 1, 2, \dots\}$, with the property that $X(t)$ is a non-decreasing function of t .

2 Defining a stochastic process

2.1 Sample paths

Having selected a time set and a state space, it remains to define the process $\{X_t : t \in J\}$ itself. This amounts to specifying the joint distribution of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ for all t_1, t_2, \dots, t_n in J and all integers n . This appears to be a formidable task; in practice this is almost invariably done indirectly, through some simple intermediary process (see Section 2.3 and Section 3).

Consider a simple model of the price of a stock measured in pence. Each trading day $t = 0, 1, 2, \dots$ the price increases by 1 pence or decreases by 1 pence with probabilities p and $1 - p$ respectively. The changes each day are independent. Let the price at time t be denoted X_t and assume $X_0 = 100$, so that the initial price (time 0 in our model) is £1.

This completely determines a stochastic process, even though we haven't *explicitly* given all the joint distributions. This particular one is called a *simple random walk* with a barrier at 0. The time set (measured in days) is $J = \{0, 1, 2, \dots\}$. The state space (measured in pence) is the set of non-negative integers $\{0, 1, 2, \dots\}$.

In fact, what we've done is to specify the process in terms of its *increments* at each time. These changes $Z_t = X_t - X_{t-1}$ form another stochastic process – the intermediary process, which was referred to above.

The stochastic process of these increments is:

$$Z_t = \begin{cases} +1 & \text{with probability } p \\ -1 & \text{with probability } 1 - p \end{cases}$$

Since we have assumed that these random variables are independent, $\{Z_t\}_{t \in J}$ is a set of independent and identically distributed (IID) random variables. Note, however, that the X_t 's themselves are neither independent nor identically distributed.



Question 2.4

Explain why the random variables X_t defined above are neither independent nor identically distributed.



White noise

White noise is a stochastic process that consists of a set of independent and identically distributed random variables. The random variables can be either discrete or continuous and the time set can be either discrete or continuous.

Returning to our example: a stochastic process – the simple random walk – $\{X_t\}_{t \in J}$ has been defined in terms of its increments $\{Z_t\}_{t \in J}$ such that:

$$X_t = X_{t-1} + Z_t$$

where Z_t is a given white noise process.

Knowing the probabilities p and $1-p$ allows us to calculate the joint probability distributions of X_{t_1}, \dots, X_{t_n} for all n and all t_1, \dots, t_n . So we have a complete specification of X_t .



Example

Calculate $P(X_2 = 102, X_5 = 103 | X_0 = 100)$ for the random walk discussed above.

Solution

We can do this by considering all the different ways of starting from 100 at time 0, arriving at 102 at time 2, and finishing at 103 at time 5. In order for this to occur, the price must increase on the first two days, which happens with probability p^2 . Independently it must then increase on another two days and decrease on one day, not necessarily in that order. The decrease in price can occur at times 3, 4 or 5, giving three different possibilities. Each of these has probability $p^2(1-p)$. So:

$$\begin{aligned} P(X_2 = 102, X_5 = 103 | X_0 = 100) &= p^2 \times 3p^2(1-p) \\ &= 3p^4(1-p) \end{aligned}$$

Other joint probabilities can be calculated in the same way, so all the joint distributions can be determined and the stochastic process X_t is completely specified.

**Question 2.5**

For the random walk described above:

(i) calculate:

(a) $P(X_2 = 100, X_4 = 103 | X_0 = 100)$

(b) $P(X_2 = 100, X_4 = 102 | X_0 = 100).$

(ii) write down the joint distribution of X_2, X_4 given $X_0 = 100$.

**Question 2.6**

A student declares that to specify a stochastic process $\{X_t\}$ you need to define the probability distribution of X_t for all times t in the time set J . Comment on this assertion.

**Sample paths**

A joint realisation of the random variables X_t for all t in J is called a sample path of the process; this is a function from J to S .

Think of a random event as an experiment with outcomes of varying probability. The outcome of a given experiment would be a realisation of that random variable. A sample path is then just the sequence of outcomes of a particular set of experiments.

For example, suppose we toss a coin at times $\{0,1,2,3,4\}$. The outcome of each toss (*ie* experiment) will be a head H or tail T , so that the expression $HHTHT$ denotes an example sample path. The set of all sample paths will be represented by the set of all sequences of H and T of length five.

For the random walk described above the term “sample path” is very suitable since the path is the route taken by the walk.

To say that a sample path is a function from J to S means that with each time, *ie* with each member of J , we associate the outcome of the experiment carried out at that time, which is a member of S .

The properties of the sample paths of the process must match those observed in real life (at least in a statistical sense). If this is the case, the model is regarded as successful and can be used for prediction purposes. It is essential that at least the broad features of the real life problem be reproduced by the model; the most important of these are discussed in the next subsections.

Suppose we toss a *biased* coin 1,000 times and on each toss the coin only has a one in four chance of landing tails say (a very biased coin!). A naive model of the series of tosses with equal probabilities for heads and tails would lead to sample paths that tended to have similar numbers of heads and tails. The real life experiments, however, would differ substantially. This discrepancy between observed sample paths and predicted paths would highlight the weakness of the model.

2.2 Stationarity

Stationarity is defined as follows.



Strict stationarity

A stochastic process is said to be **stationary**, or **strictly stationary**, if the joint distributions of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ and $X_{k+t_1}, X_{k+t_2}, \dots, X_{k+t_n}$ are identical for all t_1, t_2, \dots, t_n and $k + t_1, k + t_2, \dots, k + t_n$ in J and all integers n . This means that the statistical properties of the process remain unchanged as time elapses.

Here “statistical properties” refers to probabilities, expected values, variances, and so on. A stationary process will be statistically “the same” over the time period 5 to 10 and the time period from 120 to 125, for example.

In particular, the distribution of X_t will be identical to the distribution of X_{t+k} for all t and k such that t and $t+k$ are in J . Moreover, this in turn implies that expectations $E(X_t)$ and variances $\text{var}(X_t)$ must be constant over time. The failure of any one of these conditions to hold could be used to show a process was not stationary. Showing that they all hold may be difficult though.



Question 2.7

Consider a random walk as described above, with probabilities p and $1-p$ of moving one step to the right or left respectively. Assume $X_0 = 0$.

What is the probability that $X_{10} = 10$? What is the probability that $X_2 = 10$?

Is the random walk stationary?

Recall the example of Section 1.2 with the three states Healthy, Sick and Dead. One would certainly not use a strictly stationary process in this situation, as the probability of being alive in 10 years' time should depend on the age of the individual and hence will vary over time.

Strict stationarity is a stringent requirement which may be difficult to test fully in real life. For this reason another condition, known as *weak stationarity* is also in use. This requires that the mean of the process $m(t) = E(X_t)$ is constant and that the covariance of the process:

$$\text{cov}(X_s, X_t) = E[(X_s - m(s))(X_t - m(t))]$$

depends only on the time difference $t - s$.

The time difference $t - s$ is referred to as the *lag*. Recall also that the covariance can be written:

$$\text{cov}(X_s, X_t) = E(X_s X_t) - E(X_s)E(X_t)$$

If a process is strictly stationary then it will also be weakly stationary. A weakly stationary process is not necessarily strictly stationary.

Weak stationarity considers only the first two moments of the joint distribution of the set of random variables X_t .



Weak stationarity

A process will be weakly stationary if:

- $E(X_t)$ is constant for all t , and
- $\text{cov}(X_t, X_{t+k})$ depends only on the lag, k .

To be weakly stationary a process must pass both these tests. If it fails either of the tests then it is not weakly stationary. So when checking for stationarity, start with the easiest condition (the mean), then check the covariances.

To carry out these checks you will need to use the properties of the expectation and variance functions that you learned in Subject CT3. Remember that:

$$\text{var}(X_t) = \text{cov}(X_t, X_t)$$

ie the variance is equal to the covariance at lag 0. So $\text{var}(X_t)$ will be constant for a weakly stationary process. You will also need to be familiar with the following properties of the covariance function:

If W, X and Y are random variables and a, b and c are constants, then the following hold with *no assumptions* required:

- $\text{cov}(Y, X) = \text{cov}(X, Y)$
- $\text{cov}(X, c) = 0$
- $\text{cov}(aX, bY) = ab \text{cov}(X, Y)$
- $\text{cov}(X + Y, W) = \text{cov}(X, W) + \text{cov}(Y, W)$.



Question 2.8

Show that:

$$\text{cov}(aX + bY, cW + d) = ac \text{cov}(W, X) + cb \text{cov}(W, Y)$$

by using the properties of the covariance function.

2.3 Increments



Increments

An increment of a process is the amount by which its value changes over a period of time, eg $X_{t+u} - X_t$ (where $u > 0$).

The *increments* of a process often have simpler properties than the process itself.

Example

Let S_t denote the price of one share of a specific stock. It might be considered reasonable to assume that the distribution of the return over a period of duration u , $\frac{S_{t+u}}{S_t}$, depends on u but not on t . Accordingly the log-price process

$X_t = \log S_t$ would have stationary increments:

$$X_{t+u} - X_t = \log \frac{S_{t+u}}{S_t}$$

even though X_t itself is unlikely to be stationary.

Independent increments

A process X_t is said to have *independent increments* if for all t and every $u > 0$ the increment $X_{t+u} - X_t$ is independent of all the past of the process $\{X_s : 0 \leq s \leq t\}$.

In the last example it is a form of the efficient market hypothesis to assume that $X_t = \log S_t$ has independent increments.

The example of Section 1.3 can also be modelled by a process with stationary independent increments. Many processes are defined through their increments: see Section 3.

We have already seen that a random walk may be defined through its increments. The increment with $u=1$ is the process $\nabla X_{t+1} = X_{t+1} - X_t = Z_{t+1}$ we discussed before. These are independent of the past of the process X_t . In fact for any $u > 0$ the increment $X_{t+u} - X_t$ is independent of the past values of the process. (Here u has to be a positive integer.)

2.4 The Markov property

A major simplification occurs if the future development of a process can be predicted from its present state alone, without any reference to its past history.

Suppose that we are at time s and the value of the process at time s is x . (In symbols we have $X_s = x$).

Stated precisely the Markov property reads:

$$P[X_t \in A | X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_n} = x_n, X_s = x] = P[X_t \in A | X_s = x]$$

for all times $s_1 < s_2 < \dots < s_n < s < t$, all states x_1, x_2, \dots, x_n and x in S and all subsets A of S . This is called the *Markov property*.

The necessity to work with subsets $A \subseteq S$ (rather than just having $X_t = a \in S$) is to cover the continuous state space cases. For these the probability that X_t takes on a particular value is zero. We therefore need to work with probabilities of X_t lying in some interval of S , or more generally in some subset. For discrete state spaces the Markov property has the following simplification.



Markov property for a stochastic process with a discrete state space

A stochastic process with a discrete state space has the Markov property if:

$$P[X_t = a | X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_n} = x_n, X_s = x] = P[X_t = a | X_s = x]$$

for all times $s_1 < s_2 < \dots < s_n < s < t$ and all states a, x_1, \dots, x_n of S .

It can be argued that the example of Section 1.2 can be modelled by a Markov process: if there is full recovery from the sick state to the healthy state, past sickness history should have no effect on future health prospects.

**Important result**

A process with independent increments has the Markov property.

Proof

$$\begin{aligned}
 P[X_t \in A | X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_n} = x_n, X_s = x] \\
 &= P[X_t - X_s + x \in A | X_{s_1} = x_1, X_{s_2} = x_2, \dots, X_{s_n} = x_n, X_s = x] \\
 &= P[X_t - X_s + x \in A | X_s = x] \\
 &= P[X_t \in A | X_s = x]
 \end{aligned}$$

**Question 2.9**

Explain the three steps in this proof.

If you need to check whether or not a stochastic process is Markov, then:

- first check if it has independent increments, if yes then it is Markov.
- if it does not have independent increments, then it may still be Markov if it satisfies the Markov definition.
- sometimes it is very difficult to check a process using the Markov definition, so you may need to resort to some general reasoning arguments to try and demonstrate that the Markov definition is satisfied.

**Question 2.10**

Prove that a random walk has the Markov property.

**Question 2.11**

Consider a discrete time process on the integers defined as follows: $X_t = X_{t-1} + I_t$ where I_t are random variables taking the value $+1$ or -1 with probabilities $p_t = e^{-|X_{t-1}|}$ and $q_t = 1 - e^{-|X_{t-1}|}$ respectively. Is this process Markov? Does it have independent increments?

2.5 Filtrations

Using the preliminaries in this section we can now show by a series of examples how to define a stochastic process in terms of filtrations.

The following structures underlie any stochastic process X_t :

- a sample space Ω : each outcome ω in Ω determines a sample path $X_t(\omega)$
- a set of events F : this is a collection of events, by which is meant subsets of Ω , to which a probability can be attached
- for each time t , a smaller collection of events $F_t \subset F$: this is the set of those events whose truth or otherwise are known at time t . In other words an event A is in F_t if it depends only on X_s , $0 \leq s \leq t$.

As t increases, so does F_t : $F_t \subset F_u$, $t \leq u$. Taken collectively, the family $(F_t)_{t \geq 0}$ is known as the (natural) filtration associated with the stochastic process X_t , $t \geq 0$; it describes the information gained by observing the process or the internal history of X_t up to time t .

This is the key thing that you need to know about the (natural) filtration – that F_t gives you the history of the process up to time t . If we had a process with a discrete time set $t = 0, 1, 2, \dots$, then we could write the history of the process up to time n as follows:

$$X_n = x_n, X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_1 = x_1, X_0 = x_0$$

Alternatively, we could denote this set of events by F_n .

However, if we had a process with a continuous time set, we couldn't list the complete set of events up to time n , even if we used the “...” notation. (This is because there is an uncountable number of points in the set $[0, n]$.) So in this case the complete history of the process up to time n has to be represented in terms of the filtration.

The process X_t can be said to have the Markov property if:

$$P[X_t \leq x | F_s] = P[X_t \leq x | X_s]$$

for all $t \geq s \geq 0$.

When a Markov process has a discrete state space and a discrete time set it is called a *Markov chain*; Markov chains are studied in Chapter 3. When the state space is discrete but the time set is continuous, one uses the term *Markov jump process*; Markov jump processes are studied in Chapters 4, 5 and 6.

Using the preliminaries in this section we can now show by a series of examples how to define a stochastic process.

3 Examples

In each of the examples in this section we try to appreciate which of the following properties hold:

- stationarity in the weak sense
- independent increments
- the Markov property.

3.1 White noise

Consider a discrete-time stochastic process consisting of a sequence of independent random variables X_1, \dots, X_n, \dots .

The Markov property holds in a trivial way.

Even though the process does not have independent increments, the Markov definition is satisfied.

The process is stationary if and only if all the random variables X_n have the same distribution. Such sequences of independent identically distributed (IID for short) random variables are sometimes described as a discrete-time *white noise*; their main use is as a starting point to construct more elaborate processes below.

3.2 General random walk

Start with a sequence of IID random variables Y_1, \dots, Y_j, \dots and define the process

$$X_n = \sum_{j=1}^n Y_j \text{ with initial condition } X_0 = 0. \text{ This is a process with stationary}$$

independent increments, and thus a discrete-time Markov process. It is known as a general random walk. The process is not even weakly stationary, as its mean and variance are both proportional to n .

The log of the closing value of the FTSE100 index could be modelled by a general random walk, the value one day being the value on the previous day plus some random adjustment

In the special case where the steps Y_j of the walk take only the values +1 and -1, the process is known as a *simple random walk*.

In addition, if:

$$Y_j = \begin{cases} +1 & \text{with probability 0.5} \\ -1 & \text{with probability 0.5} \end{cases}$$

the process is known as a *simple symmetric random walk*.

3.3 Poisson process

A Poisson process with rate λ is a continuous-time integer-valued process N_t , $t \geq 0$ with the following properties:

- (i) $N_0 = 0$
- (ii) N_t has independent increments
- (iii) N_t has Poisson distributed stationary increments:

$$P[N_t - N_s = n] = \frac{[\lambda(t-s)]^n e^{-\lambda(t-s)}}{n!}, \quad s < t, \quad n = 0, 1, \dots$$

This is a Markov jump process with state space $S = \{0, 1, 2, \dots\}$. It is not stationary: as in the case for the random walk, both the mean and variance increase linearly with time.

The process counts the number of events (that are occurring at a rate λ per unit time) that occur between time s and time t . It is sometimes called a counting process.

This process is of fundamental importance when counting the cumulative number of occurrences of some event over $[0, t]$, irrespective of the nature of the event (car accident, claim to insurance company, arrival of customer at a service point). A detailed study of this process and its extensions is one of the subjects of Chapter 5.

Since the increments have a Poisson distribution they can only take the values 0, 1, 2, ... It follows that the process must be increasing, that is $N_t \geq N_s$ for all s and t such that $t \geq s$. The expectations must therefore be increasing, and the process is not weakly stationary.

In fact, we shall see in Chapter 5 that the process can only increase by one step at a time, making it a natural counting process. It is very often used to model the number of insurance claims made by time t . The rate parameter λ is the expected number of claims arriving per unit time.

We will also see that there are other equivalent definitions of a Poisson process.

3.4 Compound Poisson process

Compound Poisson processes and ruin theory, as defined below, are covered in more detail in Subjects CT6 and Subject ST3.

Start with a Poisson process $N_t, t \geq 0$ and a sequence of IID random variables $Y_j, j \geq 1$ (all defined on the same probability space). A compound Poisson process is defined by:

$$X_t = \sum_{j=1}^{N_t} Y_j, t \geq 0 \quad (2.1)$$

When $N_t = 0$ we define $X_t = 0$. The Y_j may be discrete or continuous random variables. For example, N_t could be the number of storms up to time t , and Y_j could be the number of claims arising from the j th storm (discrete), or the cost of claims from the j th storm (continuous).

This process has independent increments and thus the Markov property holds. It serves as a model for the cumulative claim amount reaching an insurance company during $[0, t]$: N_t is the total number of claims over the period and Y_j is the amount of the j -th claim.

The basic problem of classical risk theory consists of estimating the probability of ruin:

$$\psi(u) = P[u + ct - X_t < 0 \text{ for some } t > 0]$$

for a given initial capital u , premium rate c , X_t defined as in (2.1), and some fixed distribution of the claim sizes.

Note here that if we receive income from premiums at a rate of c then by time t we will have received an amount ct . Also, X_t models the cumulative amount of claims incurred by the company. Starting with an initial surplus u we will therefore have a surplus of $u + ct - X_t$ at time t . The probability of ruin is therefore just the probability that at some point in the future we will be ruined, ie the probability that the surplus is less than 0.

4 Exam-style questions

Here are two short questions on the topics in this chapter.

**Question 2.12****Subject 103, September 2003, Question 2, part (i)**

Calculate the covariance between the values $X(t), X(t+s)$ taken by a Poisson process $X(t)$ with constant rate λ at the two times t and $t+s$, where $s > 0$. [2]

**Question 2.13****Subject CT4, April 2005, Question A1**

- (i) Define each of the following examples of a stochastic process:
 - (a) a symmetric simple random walk
 - (b) a compound Poisson process. [2]
- (ii) For each of the processes in (i), classify it as a stochastic process according to its state space and the time that it operates on. [2]

[Total 4]

This page has been left blank so that you can keep the chapter summaries together for revision purposes.



Chapter 2 Summary

Stochastic processes

A *stochastic process* is a model for a time-dependent random phenomenon, a collection of random variables $\{X_t : t \in J\}$, one for each time t in the *time set* J . The time set may be discrete or continuous. The set of values that the X_t are capable of taking is called the *state space* S . The state space may also be discrete or continuous.

Defining such a process amounts to specifying the joint distribution of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ for all t_1, t_2, \dots, t_n in the time set J and all integers n .

Sample paths

A joint realisation of the random variables X_t for all t in J is called a sample path of the process; this is a function from J to S .

Stationarity

If the statistical properties of a process do not vary over time, the process is *stationary*. This makes the modelling process much easier.

Mathematically, stationarity requires the joint distribution of any set of values $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ to be the same as the joint distribution of $\{X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}\}$, ie when all times are shifted across by k . This is the *strict* definition.

In practice, it is only necessary to have *weak* stationarity. This requires only the first two moments not to vary over time, ie $E(X_t)$ and $\text{var}(X_t)$ are constant, and $\text{cov}(X_{t_1}, X_{t_2})$ depends only on the lag $t_2 - t_1$.

Independent Increments

An *increment* of a stochastic process (that has a numerical state space) is just the change in the value between two times, ie $X_{t_2} - X_{t_1}$. If this is independent of the past values of the process up to and including time t_1 then the process is said to have *independent increments*.

Filtration

We often need to look at expectations of the future value of a process, conditional on the known past history. For example, for a discrete-time process, we might be interested in $P[X_n | X_1, X_2, \dots, X_{n-1}]$.

For a continuous-time process this presents a theoretical difficulty, since it is impossible to list the values at all past times. The *filtration* notation F is used here, and we write, for example, $P[X_t | F_s]$ to represent the probability distribution at a future time t , conditional on the values up to the earlier time s .

The filtration notation can (and should) be used for both discrete-time and continuous-time processes.

Markov property

If the probabilities for the future values of a process are dependent only on the latest available value, the process has the Markov property.

Mathematically, for a process with time set $\{1, 2, 3, \dots\}$ and a discrete state space:

$$\begin{aligned} P(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_1 = x_1) \\ = P(X_n = x_n | X_{n-1} = x_{n-1}) \end{aligned}$$

For a continuous-time process with a discrete state space, you need to express this in the form:

$$P(X_n = x_n | F_s) = P(X_n = x_n | X_s)$$

For a continuous-time process with a continuous state space, you need to express this in the form:

$$P(X_n \in A | F_s) = P(X_n \in A | X_s)$$

White noise

White noise is a stochastic process that consists of a set of independent and identically distributed random variables. The random variables can be either discrete or continuous and the time set can be either discrete or continuous. White noise processes are stationary and have the Markov property.

Chapter 2 Solutions

Solution 2.1

The state space is continuous and the time domain is discrete.

Solution 2.2

Maximum daily temperature

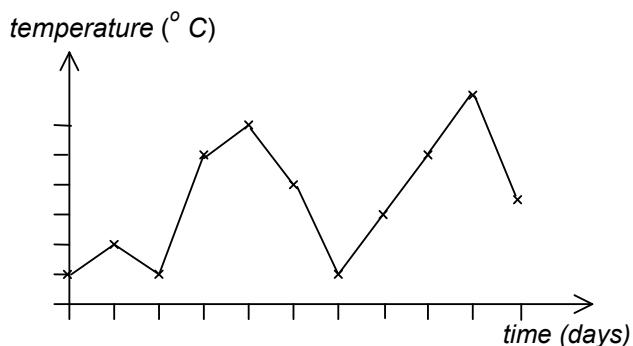
(i) *State space and time domain*

We would use a discrete time set (domain), say the non-negative integers $t = 0, 1, 2, \dots$ where t represents the number of days since 1 January 2006.

In practice we might only quote temperature to the nearest degree, so if the number of possible values was small we could use a discrete state space. Alternatively, we might prefer to use a continuous state space if for example we were recording temperatures to the nearest 0.1 of a degree and we thought values might range from -30°C to $+40^\circ\text{C}$.

(ii) *Sketch*

You will have seen displays of data like this in the newspapers or perhaps when you studied science or geography at school.



The lines joining the observed points are added to show the ‘shape’ of the data. They do not indicate that there are observations at these intermediate points, although they may be a good rough guide.

Daily rainfall

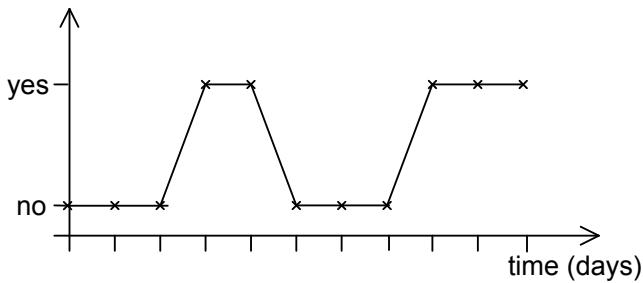
(i) ***State space and time domain***

We would use a discrete time set (domain), say the non-negative integers $t = 0, 1, 2, \dots$ where t represents the number of days since 1 January 2006.

The state space would be discrete consisting of two values: (yes: it rained) and (no: it didn't rain).

(ii) ***Sketch***

You can display the data in a similar way to the temperature data, but there will be more 'flat' sections in your plot.



The lines joining the observed points are added to show the 'shape' of the data. They do not indicate that there are observations at these intermediate points.

Injured cyclists

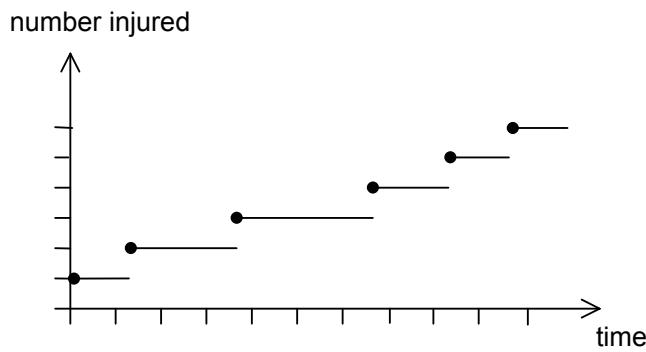
(i) ***State space and time domain***

In principle there will be a value for this random variable at every point in time. We would use a continuous time set (domain), say the non-negative numbers $t \geq 0$, where t represents time since 1 January 2006.

The state space would be discrete consisting of the non negative integers. Only if we thought that the number of possible values was very large would it be appropriate to approximate this discrete state space by a continuous one.

(ii) ***Sketch***

The display would be similar to the previous ones, except that here there is no need to use lines to link observations to show the ‘shape’. Because the time domain is continuous and the process ‘jumps’ when each event occurs we use ‘blobs’ to show which value should be used at the points of discontinuity.

***Solution 2.3***

There are two processes at work:

- the day to day transactions of the business.

These will tend to produce a ‘smoothly’ changing bank balance as there are lots of transactions and the majority are for relatively small amounts. We have a continuous state space and a continuous time domain.

- the very infrequent transactions which result from car sales.

These will produce big jumps in the bank balance. The state space will consist of a limited number of discrete values and the time domain will be continuous.

The combination of the two processes suggests that a mixed process would be appropriate.

Solution 2.4

The value of X_t is the previous value plus a random change of ± 1 . The value therefore depends very much on the previous value and so they are not independent. For example $P[X_{10} = 110] > 0$ but $P[X_{10} = 110 | X_1 = 99] = 0$.

They cannot be identically distributed either since, for example, the possible values that X_1 can take are just 99 or 101 whereas the possible values of X_2 are 98, 100 and 102; corresponding to two days of price falls, a one day fall and a one day rise and two days of rises.

Note that the unconditional variance of X_t increases as t increases since we are less certain about where the share price might be. (By unconditional here we mean that we are not conditioning on previous values, other than X_0 .)

Solution 2.5(i) **Probabilities**

$$\begin{aligned} \text{(a)} \quad P(X_2 = 100, X_4 = 103 | X_0 = 100) &= 2p(1-p) \times 0 \\ &= 0 \end{aligned}$$

because there are two ways of going from $X_0 = 100$ to $X_2 = 100$, each having a probability $p(1-p)$. It is impossible to go from $X_2 = 100$ to $X_4 = 103$ because there are not enough steps to produce an increment of 3.

$$\text{(b)} \quad P(X_2 = 100, X_4 = 102 | X_0 = 100) = (p(1-p) + (1-p)p)p^2 = 2(1-p)p^3.$$

(ii) ***Joint distribution***

To work out the joint distribution of X_2 and X_4 we would calculate each possibility as in (ii) above. To be transparent, we've included below a table of the probabilities for the various paths from time $t = 0$ through to $t = 4$, given $X_0 = 100$.

Path	Probability	Path	Probability
100 101 102 103 104	p^4	100 99 100 101 102	$p^3(1-p)$
100 101 102 103 102	$p^3(1-p)$	100 99 100 101 100	$p^2(1-p)^2$
100 101 102 101 102	$p^3(1-p)$	100 99 100 99 100	$p^2(1-p)^2$
100 101 102 101 100	$p^2(1-p)^2$	100 99 100 99 98	$p(1-p)^3$
100 101 100 101 102	$p^3(1-p)$	100 99 98 99 100	$p^2(1-p)^2$
100 101 100 101 100	$p^2(1-p)^2$	100 99 98 99 98	$p(1-p)^3$
100 101 100 99 100	$p^2(1-p)^2$	100 99 98 97 98	$p(1-p)^3$
100 101 100 99 98	$p(1-p)^3$	100 99 98 97 96	$(1-p)^4$

From this table we can derive the joint distribution of X_2 and X_4 :

$$\begin{aligned}
 P(X_2 = 98, X_4 = 96) &= (1-p)^4 & P(X_2 = 100, X_4 = 102) &= 2p^3(1-p) \\
 P(X_2 = 98, X_4 = 98) &= 2p(1-p)^3 & P(X_2 = 102, X_4 = 100) &= p^2(1-p)^2 \\
 P(X_2 = 98, X_4 = 100) &= p^2(1-p)^2 & P(X_2 = 102, X_4 = 102) &= 2p^3(1-p) \\
 P(X_2 = 100, X_4 = 98) &= 2p(1-p)^3 & P(X_2 = 102, X_4 = 104) &= p^4 \\
 P(X_2 = 100, X_4 = 100) &= 4p^2(1-p)^2
 \end{aligned}$$

Solution 2.6

What the student says is true, that to define a stochastic process you need to define the probability distributions of X_t for all $t \in J$. However, this does not give the full story.

You actually need to give information about how the X 's relate to one another at different times, as in general they will not be independent. For example, the past will often influence the future.

In order to define the full process we need to specify the joint distribution of $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ for all t_1, t_2, \dots, t_n in J and all integers n . This is often done indirectly.

Solution 2.7

$P(X_{10} = 10) = p^{10}$ but $P(X_2 = 10) = 0$. So the random walk is non-stationary. Note that in order to show something is non-stationary we only have to demonstrate that one particular requirement fails to hold.

Solution 2.8

$$\begin{aligned}\text{cov}(aX + bY, cW + d) \\ &= \text{cov}(cW, aX + bY) + \text{cov}(d, aX + bY) \\ &= \text{cov}(cW, aX) + \text{cov}(cW, bY) + \text{cov}(d, aX) + \text{cov}(d, bY) \\ &= ac \text{ cov}(W, X) + cb \text{ cov}(W, Y) + 0 + 0 \\ &= ac \text{ cov}(W, X) + cb \text{ cov}(W, Y)\end{aligned}$$

Solution 2.9

The first equality here uses the fact that we are given $X_s = x$ so that $X_t = X_t - X_s + x$. The second equality follows from the assumption that the increment $X_t - X_s$ is independent of past increments. Finally, the third equality uses the fact that $X_s = x$ again.

Solution 2.10

A random walk has independent increments, so the result is immediate from the property described at the end of Section 2.4.

Solution 2.11

The increments, $X_t - X_{t-1}$, depend on the current state X_{t-1} so the process does not have independent increments. It is Markov, however, since our knowledge of past values additional to the current value is irrelevant.

Note that independent increments imply the Markov property, but that the Markov property does not imply that the increments of the process are independent. The two properties are not equivalent.

Solution 2.12

The key to many results for continuous time stochastic processes is to realise that the random variables representing behaviour in non-overlapping time periods are independent. Here the non-overlapping time periods are $(0, t)$ and $(t, t+s)$.

(i) **Covariance of Poisson process**

$$\begin{aligned}\text{cov}(X(t), X(t+s)) &= \text{cov}(X(t), X(t) + (X(t+s) - X(t))) \\ &= \text{cov}(X(t), X(t)) + \text{cov}(X(t), X(t+s) - X(t)) \\ &= \text{var}(X(t)) + 0 \\ &= \lambda t\end{aligned}$$

since $X(t) \sim \text{Poisson}(\lambda t)$.

Solution 2.13

The word ‘symmetric’ is important as it denotes a particular process that is equally likely to ‘step’ upwards or downwards in its walk.

(i)(a) **Symmetric simple random walk**

A symmetric simple random walk on $\{ \dots -2, -1, 0, +1, +2, \dots \}$ is:

$$X_n = \sum_{j=1}^n Y_j$$

where the random variables Y_j are independent and identically distributed with common probability distribution:

$$P[Y_j = +1] = \frac{1}{2} \text{ and } P[Y_j = -1] = \frac{1}{2}$$

The initial distribution of the process is $P[X_0 = 0] = 1$.

(i)(b) **Compound Poisson process**

To specify a compound Poisson process we will need to say something about a Poisson process.

A Poisson process is a counting process $\{N_t, t \geq 0\}$, where N_t records the number of occurrences of a specified event in the time interval $(0, t)$.

The probability that a specified event occurs in the short time interval $(t, t+h)$ is λh where λ is the rate of the process, ie the expected number of events that will occur in a unit of time.

The rate of the process is constant over time, ie λ does not depend on t .

A compound Poisson process is:

$$X_t = \sum_{j=1}^{N_t} Y_j$$

where the random variables $\{Y_j; j = 1, 2, \dots\}$ are independent, identically distributed and (usually) continuous.

The initial distribution of the process is $P[Y_0 = 0] = 1$.

(ii) ***Classification of processes***

The symmetric simple random walk has a discrete state space consisting of the values, $\{ \dots -2, -1, 0, +1, +2, \dots \}$ and a discrete time domain consisting of the values $\{0, 1, 2, \dots\}$.

The compound Poisson process has a continuous state space. In general the state space will include all real numbers, but the exact values in the state space will depend on the values taken by the random variables, Y_j . The process has a continuous time domain consisting of the values $\{t, t \geq 0\}$.

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Chapter 3

Markov chains



Syllabus objectives

- (iii) Define and apply a Markov chain.
1. State the essential features of a Markov chain model.
 2. State the Chapman-Kolmogorov equations that represent a Markov chain.
 3. Calculate the stationary distribution for a Markov chain in simple cases.
 4. Describe a system of frequency-based experience rating in terms of a Markov chain and describe other simple applications.
 5. Describe a time-inhomogeneous Markov chain model and describe simple applications.
 6. Demonstrate how Markov chains can be used as a tool for modelling and how they can be simulated.

0 Introduction

Recall from Chapter 2 that the Markov property means that the future value of a process is independent of the past history and only depends on the current value. Any process satisfying the Markov property is a Markov process. The term *Markov chain* refers to Markov processes in discrete time and with a discrete state space.

1 An example of a Markov chain

As an example, consider the *no claims discount (NCD)* model run by motor insurance companies. A company might offer discounts of say 0%, 30% and 60% of the full premium.

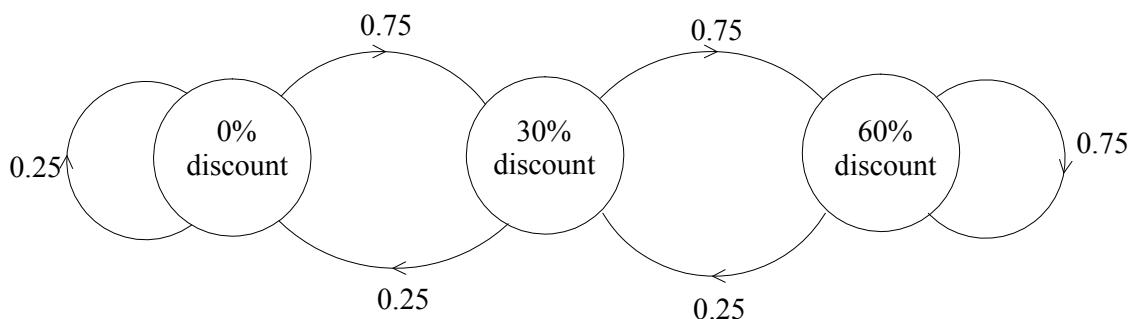
A policyholder's status is determined by the following rules:

- All new policyholders start at the 0% level.
- If no claim is made during the current year then you move up one discount level or remain at 60% level.
- If one or more claims are made you move down one level, or remain at 0% discount.

This can be modelled using a Markov chain on the state space $S = \{0\%, 30\%, 60\%\}$. When the policy is renewed each year the policyholder moves to another level or remains at the same level with various probabilities depending on the chance of making a claim. Assume the chance of claiming is independent of the current level and take say $P(\text{no claim}) = \frac{3}{4}$.

This is an example of a discrete time process that satisfies the Markov property (because the future only depends on your current level, and not on your history). Hence this is a Markov chain.

One way of representing a Markov chain is by its *transition graph*.



The states are represented by the circles, and each arrow represents a possible *transition*. Note that staying in a state for one time period is also a possible transition, so we need an arrow to show this. Next to each arrow is written the corresponding *transition probability*. The arrows at either end correspond to policyholders who remain in the 0% or 60% states. Policyholders in the 30% discount state will always move to another state because of the rules of the discount scheme.

The transition probability between two states in unit time is therefore given explicitly. These can be written in the form of a *transition matrix*:

$$P = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

The (i, j) th entry in the matrix, ie i th row and j th column, gives the probability of moving in *one step* from state i to state j . In order for this notation to make sense, we need the states to be labelled to correspond with the matrix entries. For example, the state “0%” would correspond to $i = 1$, the state “30%” would correspond to $i = 2$, and the state “60%” to $i = 3$.

Note that the sum of the entries in each row adds up to 1. This can be interpreted as saying that at any time, some transition into another state must occur or the process stays where it is. The sum of the probabilities of the mutually exclusive and exhaustive events that can happen when the process is in state i equals 1.

It turns out that to calculate transition probabilities over 2 steps, we use the matrix P^2 , and for three steps we use P^3 and so on. This greatly simplifies the analysis of such situations. This is the basic content of the *Chapman-Kolmogorov* equations, which we discuss in Section 2. The one-step transition probabilities will generally be given, and we will have to construct general transition probabilities by applying the Chapman-Kolmogorov equations.

There is an added complication however. In the stochastic process we've just looked at, the transition matrix did not depend on the current time. Such a process is said to be *time-homogeneous*. In general, however, we need to consider the possibility that even the one-step transition matrices can vary with time. It takes a while to get used to working with the matrix notation but the theory itself is not too difficult.

In Section 3 we concentrate on the time-homogeneous case, which simplifies things.

Section 5 contains several examples, including further discussion of the NCD model introduced above. We also consider random walks on both finite and infinite state spaces.

Section 6 looks at the long-term behaviour of the Markov chain. This is important. For example, in the NCD case above, we would expect after a while that the process would settle down, and that the same number of people would be in each discount level at any one time. This does not mean that each individual stays put, but that, although each individual moves around, the process as a whole reaches an equilibrium or stationary position.

Mathematically, the problem of finding the proportion of people who are in each state in the long run can easily be tackled using the transition matrices defined above. The problem reduces to solving a set of simultaneous equations.

Not all Markov chains have a single stationary distribution. Some chains may have no stationary distribution and some chains may have more than one stationary distribution. Chains with a single stationary distribution may be such that this distribution is never reached.

We will describe three classifications of Markov chains, and use these classifications to define categories so that all those chains in the same category have the same long run behaviour. For one category there will be a unique long-term stationary distribution that will be reached after a sufficient length of time. We will describe how to find this unique stationary distribution.

2 The Chapman-Kolmogorov equations

Recall from Chapter 2 that the term **Markov chain** is reserved for discrete-time Markov processes with a finite or countable state space S ; so a **Markov chain** is a sequence of random variables $X_0, X_1, \dots, X_n, \dots$ with the following property:

$$P[X_n = j | X_0 = i_0, X_1 = i_1, \dots, X_{m-1} = i_{m-1}, X_m = i] = P[X_n = j | X_m = i] \quad (3.1)$$

for all integer times $n > m$ and states $i_0, i_1, \dots, i_{m-1}, i, j$ in S .



Question 3.1

Why couldn't the definition of a *general* Markov process (which we looked at in Chapter 2) be given as above?

The **Markov property** (3.1) has the following interpretation: given the present state of the process $X_m = i$, the additional knowledge of the past is irrelevant for the calculation of the probability distribution of future values of the process.

Note that some knowledge of the past may be incorporated in X_m . It is the additional information contained in the earlier values of the process that does not provide any help in predicting the future behaviour of the process.

The conditional probabilities on the right-hand side of (3.1) are the key objects for the description of a Markov chain; we call them **transition probabilities**, and we denote them by:

$$P[X_n = j | X_m = i] = p_{ij}^{(m,n)}.$$

So $p_{ij}^{(m,n)}$ is the probability of being in state j at time n having been in state i at time m .

In particular, we can define the *one-step* transition probabilities:

$$P(X_{m+1} = j | X_m = i) = p_{ij}^{(m,m+1)}$$

These tell us in a probabilistic sense what will happen at the next step at any time m . These one-step transitions therefore describe the immediate future. The NCD transition matrix P was an example of this.

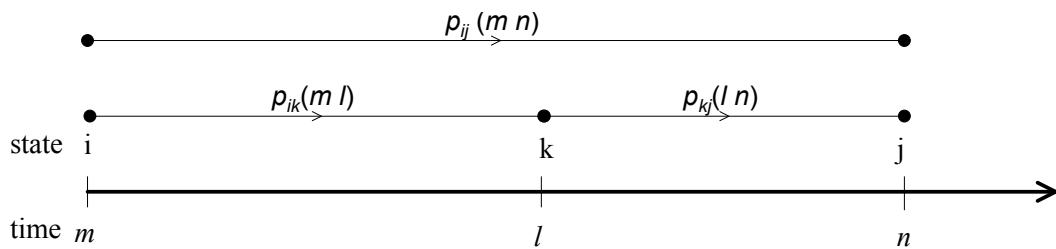
If we know all such probabilities, then intuitively we should be able to calculate any long-term transition probability, from time m to time $n > m$, by considering a sequence of such one-step transitions. This can be deduced from the following fundamental result.

The transition probabilities of a discrete-time Markov chain obey the *Chapman-Kolmogorov equations*:

$$p_{ij}^{(m,n)} = \sum_{k \in S} p_{ik}^{(m,l)} p_{kj}^{(l,n)}$$

for all states i, j in S and all integer times $m \leq l \leq n$.

Before giving a proof of this result we should be clear about its interpretation. We can split any path between times m and n into two parts by introducing some fixed intermediate time l . Assume at this time we are in some state k .



The associated transition probability for the whole of that path will then be the product of the transition probabilities for each part, namely $p_{ik}^{(m,l)} p_{kj}^{(l,n)}$. So $p_{ik}^{(m,l)} p_{kj}^{(l,n)}$ is the probability of going from state i at time m to state j at time n , going via state k at time l . The probabilities are “chained together”, which is the reason we call these processes *Markov chains*.

However, if we start in state i at time m , and finish in state j at time n , then in general there will be several possibilities for this intermediate state k . To take into account all of these different paths we must therefore sum over all the mutually exclusive and exhaustive possibilities. This gives us the right hand side of the above equation. Note that we have used the phrase “finish in” here rather than “go to”, because it could be that the transitions involve staying in the same state, ie no movement at all.

Although the above equations may appear to be rather daunting at first, it should be noted that they can be simplified vastly by considering the transition probability $p_{ij}^{(m,n)}$ as the i,j th entry of a *transition matrix* $P^{(m,n)}$. The above equations can then be written using matrix multiplication as $P^{(m,n)} = P^{(m,l)}P^{(l,n)}$. We discuss this approach in more detail in the next section.

First we derive the Chapman-Kolmogorov equations mathematically.

Proof

(Students should understand this proof, but they will not be expected to reproduce it in the examination.)

This is based on the Markov property (3.1) and on the law of total probability in its conditional form.

If $A_1, A_2, \dots, A_k, \dots$ form a complete set of disjoint events, ie:

$$\bigcup_{k=1}^{\infty} A_k = \Omega, \quad A_k \cap A_j = \emptyset, \quad k \neq j$$

then for any two events B, C :

$$P[B|C] = \sum_{k=1}^{\infty} P[B|C, A_k] P[A_k|C]$$

We now identify B with “ $X_n = j$ ”, C with “ $X_m = i$ ”, and A_k with “ $X_l = k$ ”.

Thus:

$$\begin{aligned} P[X_n = j | X_m = i] &= \sum_{k \in S} P[X_n = j | X_m = i, X_l = k] P[X_l = k | X_m = i] \\ &= \sum_{k \in S} P[X_n = j | X_l = k] P[X_l = k | X_m = i] \end{aligned}$$

using the Markov property (note $l > m$).

This is the stated result.

The Chapman-Kolmogorov equations allow us to calculate general transition probabilities in terms of the one-step transition probabilities $p_{ij}^{(n,n+1)}$.

For example, if we wish to calculate the two-step transition probabilities, we can take the intermediate time $l = m + 1$ and apply the equations. Once we have the two-step transition probabilities we can use them to calculate the three-step transitions and, by iterating the procedure, the transition probabilities of any order can be found.

Hence the distribution of a Markov chain is fully determined once the following are specified:

- the one-step transition probabilities $p_{ij}^{(n,n+1)}$
- the initial probability distribution $q_k = P[X_0 = k]$.

Indeed we can deduce from these the probability of any path:

$$P[X_0 = i_0, X_1 = i_1, \dots, X_n = i_n] = q_{i_0} p_{i_0 i_1}^{(0,1)} p_{i_1 i_2}^{(1,2)} \dots p_{i_{n-1} i_n}^{(n-1,n)}.$$

It is therefore convenient, where possible, to determine states in a manner that forms a Markov chain. The model in Section 5.2 illustrates this.

This is referring to the fact that a chain may be given in a form that isn't Markov. In these cases we can't apply the techniques described above to tackle the problem. However, it is sometimes possible to change the state space so that the process *is* given as a Markov chain. For example, this is the case for Model 5.2, which we will meet later in this chapter. Generally, when this can be done it will simplify the analysis of the problem.

3 Time-homogeneous Markov chains

A simplification occurs if the one-step transition probabilities are time-independent:

$$p_{ij}^{(n,n+1)} = p_{ij} \quad (3.2)$$

In this case, we say that the Markov chain is *time-homogeneous*.

It follows easily from (3.2) that general transition probabilities depend only on time differences:

$$P[X_{l+m} = j | X_m = i] = p_{ij}^{(l)} \quad (3.3)$$

This equation defines $p_{ij}^{(l)}$ to be $p_{ij}^{(m,l+m)}$. However, the definition only makes sense if the left-hand side is independent of m .

We refer to (3.3) as the *l -step transition probability*. For time-homogeneous Markov chains, the Chapman-Kolmogorov equations read:

$$p_{ij}^{(n-m)} = \sum_{k \in S} p_{ik}^{(l-m)} p_{kj}^{(n-l)}$$

This has a very simple interpretation. The *transition matrix* P of a time-homogeneous Markov chain is a square $N \times N$ matrix where N is the number of states in S (possibly infinite), with the elements P_{ij} being the one-step transition probabilities p_{ij} :

$$(P)_{ij} = p_{ij}$$

The *l -step transition probability* $p_{ij}^{(l)}$ can be obtained by calculating the entry (i, j) of the l -th power of the matrix P :

$$p_{ij}^{(l)} = (P^l)_{ij}$$

You will need to be familiar with matrices and matrix multiplication. (You may find it helpful to look at FAC if your memory is a bit rusty.) Recall that the (i, j) th entry in a matrix A is denoted by $(A)_{ij}$ (or just A_{ij}); the i refers to the row number, and j to the column.

Warning: expressions such as $(A)_{12}$ and $(A)_{31}$ represent numbers and not matrices. Similarly $(A)_{ij}$ is a number, namely the (i, j) th entry in the matrix A . It is A that is the matrix. However, some abuse of terminology does occur and occasionally you will see “the matrix $(A)_{ij}$ ” referred to.

Finally recall that powers are written in the same way as for ordinary numbers. For example, A^2 means AA , as writing two matrices side by side denotes matrix multiplication.

In the same way that we can think of the transition probabilities as the entries of a matrix, we can think of the distributions $P(X_n = i)$ as the entries of a row vector ($1 \times N$ matrix, N being the “size” of the state space, as above). In particular, the initial distribution X_0 will be given by the row vector with entries q_k . It will be denoted by \underline{X}_0 .

For a time-homogeneous Markov chain, we have seen that, $P(X_1 = i) = \sum_{k \in S} q_k p_{ki}$.

The distribution of X_1 can now also be viewed as a row vector, call it \underline{X}_1 , with i th entry $\sum_{k \in S} q_k p_{ki}$. Then the equation above can be rewritten in matrix form as:

$$\underline{X}_1 = \underline{X}_0 P$$

This is a shorthand notation for the original equation with summation over indices.

Returning to the previous warning on terminology, note that $\sum_{k \in S} q_k p_{ki} = \sum_{k \in S} p_{ki} q_k$.

This is because q_k and p_{ki} are real numbers, so the order in which we multiply them doesn't matter. However, in the shorthand notation, $\underline{X}_1 = \underline{X}_0 P$, the objects are matrices and the order does matter; the expression $\underline{X}_1 = P \underline{X}_0$ doesn't even make sense, as we cannot multiply a row vector on the left by an $N \times N$ matrix (unless N happens to equal 1). If instead of a row vector we have a column vector, say the transposed vector \underline{X}_0^T , then this order does make sense. For example, the following equation is valid $\underline{X}_1^T = P^T \underline{X}_0^T$.



Question 3.2

For a time-inhomogeneous process, the one-step transition matrices are dependent on time and so can be labelled $P^{(n,n+1)}$, where n refers to the time.

- (i) Write down a matrix equation representing the distribution of the random variable X_5 in terms of the initial distribution and transition matrices.
- (ii) How does this simplify for a time-homogeneous chain?

The normalisation condition $\sum_{j \in S} p_{ij} = 1$ holds for all i , ie each row of P must add up to one. More generally:

$$\sum_{j \in S} p_{ij}^{(I)} = 1 \text{ for all } i.$$

It is often revealing to draw the *transition graph* of a Markov chain: this is a diagram in which each state in S is represented as a node of the graph and an arrow is drawn from node i to node j whenever $p_{ij} > 0$, indicating that a direct transition from state i to state j is possible. The value of p_{ij} can be recorded above the arrow.

**Question 3.3**

Consider a Markov process with state space $S = \{0,1,2\}$ and transition matrix, P :

$$P = \begin{pmatrix} p & q & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ p - \frac{1}{2} & \frac{7}{10} & \frac{1}{5} \end{pmatrix}$$

- (i) What can you say about the values of p and q ?
- (ii) Calculate the transition probabilities $p_{ij}^{(3)}$.
- (iii) Draw the transition graph for the process represented by P .

4 Time-inhomogeneous Markov chains

For a time-inhomogeneous Markov chain, the transition probabilities cannot simply be denoted by p_{ij} because they will depend on the absolute values of time, rather than just the time difference.

For a time-inhomogeneous chain, one-step the transition probabilities are denoted by $p_{ij}^{(n,n+1)}$.

The value of “time” can be represented by many factors, for example the time of year, age or duration.

So for a time-inhomogeneous Markov chain, the probability of going from state i at time 0 to state j at time n is not necessarily the same as going from state i at time m ($m \neq 0$) to state j at time $m+n$, even though both time intervals are of length n . For a time-inhomogeneous chain, the transition probabilities depend not only on the length of the time interval, but also on when it starts.

5 Models

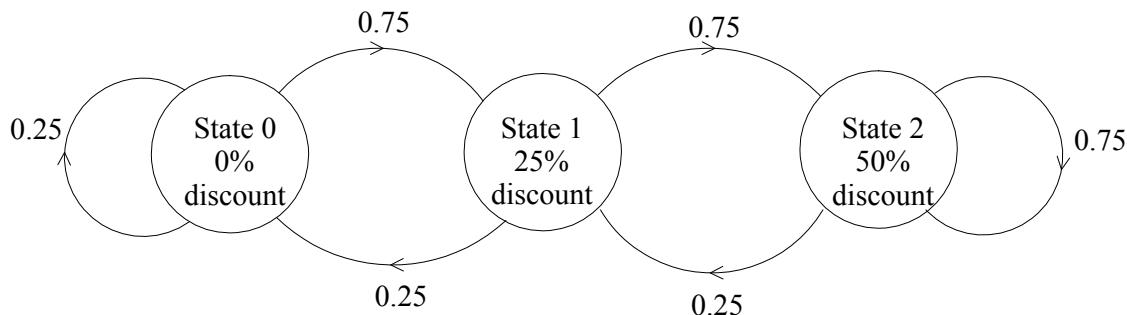
Throughout the rest of this chapter we shall refer to the models in the following sections by their section number. For example, Model 5.1 will mean the model in Section 5.1.

5.1 A simple model of a No Claims Discount (NCD) policy

The No Claims Discount system (NCD) in motor insurance whereby the premium charged depends on the driver's claim record is a prime application of Markov chains. We present two simple models and we suggest various possible improvements.

A motor insurance company grants its customers either no discount (state 0) or 25% discount (state 1) or 50% discount (state 2). A claim-free year results in a transition to the next higher state the following year (or in the retention of the maximum discount); similarly, a year with one claim or more causes a transition to the next lower state (or the retention of the zero discount status).

Under these rules, the discount status of a policyholder forms a Markov chain with state space $S = \{0, 1, 2\}$; if the probability of a claim-free year is $\frac{3}{4}$ the transition graph and transition matrix are:



The transition matrix is given by:

$$\mathbf{P} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

The probability of holding the maximum discount in year $n+3$ given that you do not qualify for any discount in year n is:

$$p_{0,2}^{(3)} = (\mathbf{P}^3)_{1,3} = \frac{9}{16}$$

Try not to get confused with the notation here! $p_{0,2}^{(3)}$ denotes the probability of going from state 0 to state 2 in 3 time steps. This is equal to $(P^3)_{1,3}$, the entry in the first row and the third column of the matrix P^3 (the cube of the matrix P). So you could calculate P^3 and identify the $(1,3)th$ entry. This turns out to be $\frac{9}{16}$. However, calculating the cube of P is time consuming and is not actually necessary given that we only need one probability and there aren't that many possible paths.

One alternative approach is the following. Given that you do not qualify for any discount in year n , you must be starting in state 0. We want the probability of ending in state 2 after three one-step transitions. This can only happen if our path is:

$$0 \rightarrow 0 \rightarrow 1 \rightarrow 2 \text{ with probability } \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4}$$

or:

$$0 \rightarrow 1 \rightarrow 2 \rightarrow 2 \text{ with probability } \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4}$$

The sum of these gives the result $\frac{9}{16}$. So we see that the matrix representation is a notation for the way we have always calculated probabilities.

However, if there are many possible paths, the above approach can be tedious. Probably the most efficient way to proceed with a problem like this is to work as follows.

Since we know that the distribution at time n is $(1, 0, 0)$, we can calculate the probability distribution at time $n+1$ by post-multiplying the vector $(1, 0, 0)$ by the transition matrix P . This gives:

$$\left(\frac{1}{4}, \frac{3}{4}, 0 \right)$$

Then post-multiplying $\left(\frac{1}{4}, \frac{3}{4}, 0 \right)$ by the transition matrix P , we get the probability distribution at time $n+2$, which is:

$$\left(\frac{1}{4}, \frac{3}{16}, \frac{9}{16} \right)$$

Post-multiplying this by P gives us the probability distribution at time $n+3$. However, since we just need the probability of being on maximum discount in year $n+3$, we only have to multiply the vector $\left(\frac{1}{4}, \frac{3}{16}, \frac{9}{16}\right)$ by the last column of P . We find that the probability distribution at time $n+3$ is of the form:

$$\left(*, *, \frac{9}{16}\right)$$

So the required probability is $\frac{9}{16}$.

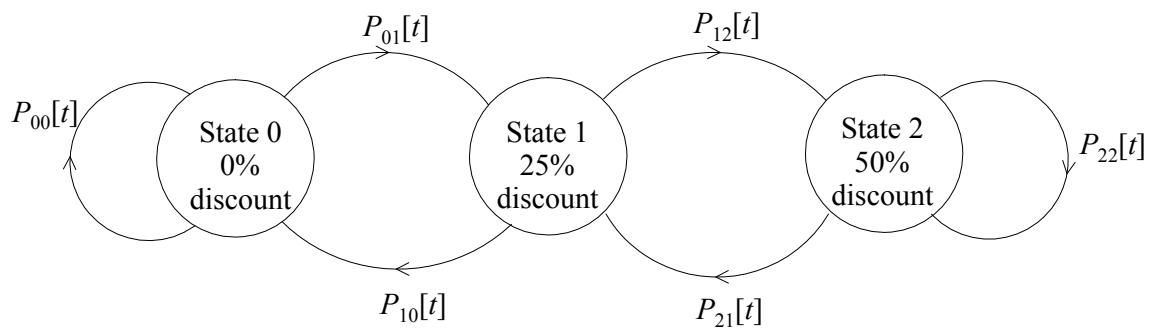


Question 3.4

Calculate the probability in the above model of starting with a discount level of 25% and ending up 4 years later at the same level.

Time-inhomogeneous model

For a time-inhomogeneous case of this model, the probability of an accident would be time-dependent to reflect changes in traffic conditions and possibly weather conditions. This could be due to general annual trends in the density of traffic and/or propensity to claim. The transition graph and matrix would then become:



and:

$$P[t] = \begin{bmatrix} P_{00}[t] & P_{01}[t] & P_{02}[t] \\ P_{10}[t] & P_{11}[t] & P_{12}[t] \\ P_{20}[t] & P_{21}[t] & P_{22}[t] \end{bmatrix}$$

5.2 Another model of an NCD policy

In the model in Section 5.1 there were 3 discount levels, namely 0%, 25% and 50%.

Modify the previous model as follows: there are now four levels of discount:

- 0 : no discount
- 1 : 25% discount
- 2 : 40% discount
- 3 : 60% discount.

The rules for moving up the discount scale are as before, but in the case of a claim during the current year, the discount status moves down one or two steps (if this is possible) according to whether or not the previous year was claim-free.

So your discount level next year depends on your claims for this year *and* last year.

Under these rules, the discount status X_n of a policyholder does not form a Markov chain on $S = \{0, 1, 2, 3\}$ because:

$$P[X_{n+1} = 0 | X_n = 2, X_{n-1} = 1] = 0$$

whereas:

$$P[X_{n+1} = 0 | X_n = 2, X_{n-1} = 3] > 0$$

Note that $P[X_{n+1} = 0 | X_n = 2, X_{n-1} = 1]$ is the probability of you being on 0% discount in year $n+1$ given that you were on 40% discount in year n *and* 25% discount in year $n-1$. This probability is zero since there was no claim in year $n-1$. (If there had been, you wouldn't have moved up to 40% discount.)

The Core Reading equations above show that the future value of the process depends not only on its current value, but also on the past. Earlier we commented that it is sometimes possible to transform a chain that isn't Markov into one that is. We do this by altering the state space.

To construct a Markov chain $\{Y_n, n = 0, 1, 2, \dots\}$, one needs to incorporate some information on the previous year into the state; in fact this is necessary only for state 2, which we split as:

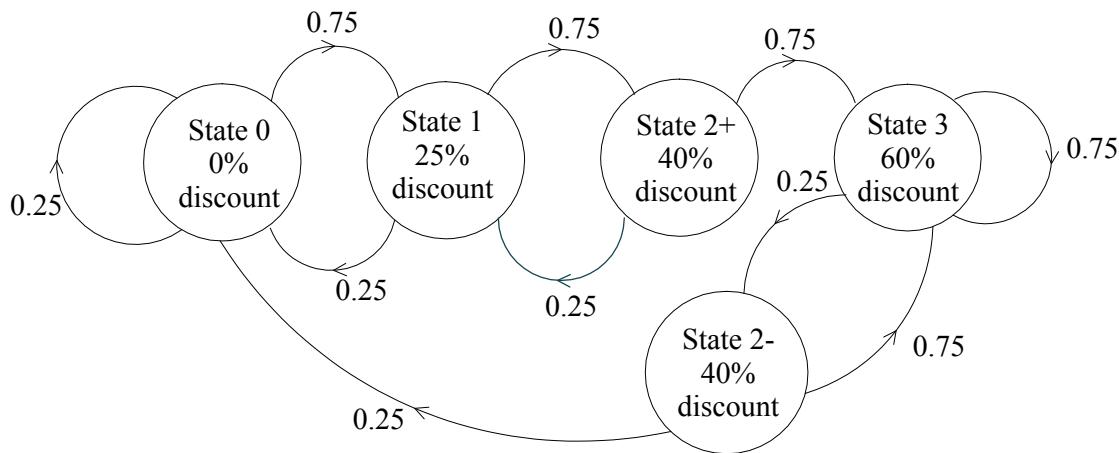
2^+ : 40% discount and no claim in the previous year

2^- : 40% discount and claim in the previous year.

We can see why it's only state 2 that needs to be split, by reasoning as follow:

- If you were on 0% discount this year and you made a claim, you'd stay on 0% discount next year.
- If you were on 25% discount this year and you made a claim, you'd go down to 0% discount next year.
- If you were on 40% discount this year and you made a claim, you'd go down to either 0% discount or to 25% discount next year, depending on whether you'd made a claim in the previous year.
- If you were on 60% discount this year, it would mean that you hadn't claimed in the previous year. So if you made a claim this year, you'd go down to 40% discount next year.

Assuming as before a probability of $\frac{3}{4}$ of no claim in any given year, we have a Markov chain on the state space $S' = \{0, 1, 2^+, 2^-, 3\}$ with transition graph:



The transition matrix is:

$$\mathbf{P} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} \\ \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

The probability of being at the 60% discount level in year $n+3$ given that you hold 25% in year n is:

$$p_{1,3}^{(3)} = (\mathbf{P}^3)_{2,5} = \frac{27}{64}$$

Question 3.5

If you start at 0% discount, what is the probability that you are on the maximum discount after 5 years?

Time-inhomogeneous model

This basic model is amenable to numerous improvements. For instance the accident probability can be made to depend on the discount status to reflect the influence of the latter on driver care. Also the accident probability can be time-dependent (leading to a time-inhomogeneous chain) to reflect changes in traffic conditions.

5.3 Simple random walk on $S = \{-2, -1, 0, 1, 2, \dots\}$

This is defined as $X_n = Y_1 + Y_2 + \dots + Y_n$ where the random variables Y_j (the steps of the walk) are mutually independent with the common probability distribution:

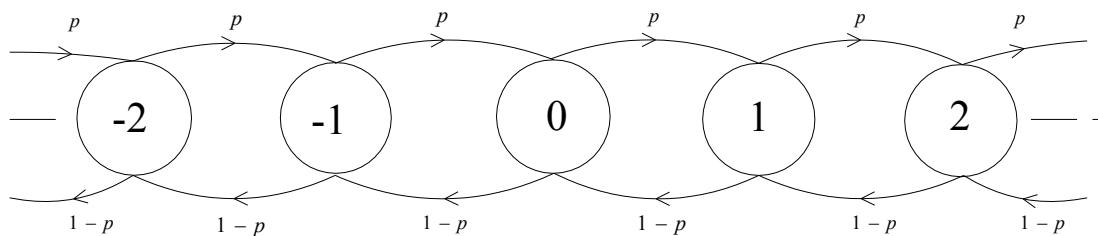
$$P[Y_j = 1] = p, \quad P[Y_j = -1] = 1 - p$$

The Markov property holds:

$$\begin{aligned} & P[X_n = j | X_1 = i_1, \dots, X_{m-1} = i_{m-1}, X_m = i] \\ &= P[X_m + Y_{m+1} + \dots + Y_n = j | X_1 = i_1, \dots, X_{m-1} = i_{m-1}, X_m = i] \\ &= P[Y_{m+1} + \dots + Y_n = j - i] = P[X_n = j | X_m = i] \end{aligned}$$

We also know this because the process has independent increments.

The transition graph and the transition matrix are infinite:



$$P = \left[\begin{array}{ccccccc} \ddots & \ddots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \ddots & \ddots & \ddots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 1-p & 0 & p & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1-p & 0 & p & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & 1-p & 0 & p \\ \cdots & \cdots & \cdots & \cdots & \cdots & 1-p & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \ddots \end{array} \right]$$

In order to get from i to j in n steps the random walk must make $u = \frac{1}{2}(n + j - i)$ steps in an upward direction, $n - u$ in a downward direction. Since the distribution of the number of upward jumps in n steps is Binomial with parameters n and p , the n -step transition probabilities can be calculated as:

$$p_{ij}^{(n)} = \begin{cases} \binom{n}{u} p^u (1-p)^{n-u} & \text{if } 0 \leq n + j - i \leq 2n \text{ and } n + j - i \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$



Question 3.6

Prove that the n -step transition probabilities for a simple random walk on the integers are indeed given by the formula above.

Hint: it is better to reason probabilistically rather than to compute P^n . A transition from i to j in n steps is equivalent to u positive steps and d negative steps with:

$$u - d = j - i, \quad u + d = n$$

Note that in addition to being time-homogeneous a simple random walk is also space-homogeneous:

$$p_{ij}^{(n)} = p_{i+r, j+r}^{(n)}$$

This means that only the time taken and the overall distance travelled (including minus sign if necessary) affect the transition probability. Exactly when and where they occur doesn't matter. So, for example, the probability of going from state 4 at time 4 to state -1 at time 11, will be the same as the probability of going from state 8 at time 3 to state 3 at time 10. In both cases we have moved 5 steps to the left in a time of 7 units.

The special case when $p = 1 - p = \frac{1}{2}$, is known as a *symmetric* random walk.

5.4 Simple random walk on $\{0, 1, 2, \dots, b\}$

This is similar to the previous model, except that **boundary conditions** have to be specified at 0 and b ; these will depend on the interpretation given to the chain. Commonly used boundary conditions include:

Reflecting boundary: $P[X_{n+1} = 1 | X_n = 0] = 1$

Absorbing boundary: $P[X_{n+1} = 0 | X_n = 0] = 1$

Mixed boundary:
$$\begin{cases} P[X_{n+1} = 0 | X_n = 0] = \alpha \\ P[X_{n+1} = 1 | X_n = 0] = 1 - \alpha \end{cases}$$

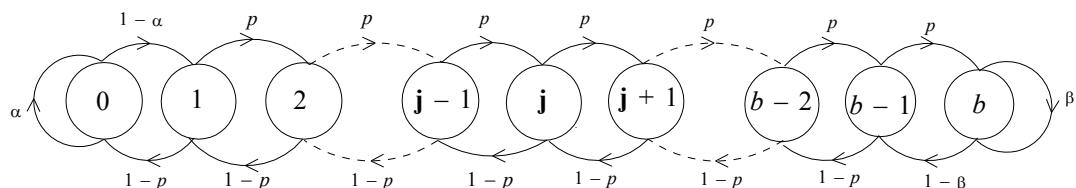
A random walk with absorbing boundary conditions at 0 and b can be used (for example) to describe the wealth of a gambler who will continue to gamble until either his fortune reaches a target b or his fortune hits 0 and he is ruined; in either case, reaching the boundary means staying there forever.



Question 3.7

For each of the types of boundary conditions above, explain why they have the names they do.

In the general case, with mixed boundary conditions, the transition graph is:



The transition matrix for this Markov chain is:

$$P = \begin{bmatrix} 0 & 1 & 2 & \dots & \dots & \dots & b-2 & b-1 & b \\ \alpha & 1-\alpha & \dots \\ 1-p & 0 & p & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & 1-p & 0 & p & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & 1-p & 0 & p & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \ddots & \ddots & \ddots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & 1-p & 0 & p & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & 1-p & 0 & p & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & 1-p & 0 & p \\ \dots & 1-\beta & \beta \end{bmatrix}$$

Reflecting and absorbing boundary conditions are obtained as special cases, taking α, β equal to 0 or 1.

The simple NCD model of Model 5.1 is another practical example of a bounded random walk.



Question 3.8

What are the boundary conditions for the NCD model of Model 5.1.?

5.5 A model of accident proneness

For a given driver, any period j is either accident free ($Y_j = 0$) or gives rise to exactly one accident ($Y_j = 1$).

The possibility of more than one accident in any time period is ignored for simplicity.

The probability of an accident in the next period is estimated using the driver's past record as follows (all variables y_j are either 0 or 1):

$$P[Y_{n+1} = 1 | Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n] = \frac{f(y_1 + y_2 + \dots + y_n)}{g(n)}$$

where f , g are two given increasing functions satisfying $0 \leq f(m) \leq g(m)$. Of course:

$$P[Y_{n+1} = 0 | Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n] = 1 - \frac{f(y_1 + y_2 + \dots + y_n)}{g(n)}$$



Question 3.9

- (i) Why do the functions f and g have to be increasing functions?
- (ii) Why do they satisfy the given inequalities?
- (iii) Interpret this model in the case where $f(m) = m$ and $g(n) = n$.

The dependence on the past record means that $Y_1, Y_2, \dots, Y_n, \dots$ does *not* have the Markov property (it depends on all previous values of Y_j). Consider, however, the cumulative number of accidents suffered by the driver:

$$X_n = \sum_{j=1}^n Y_j$$

This is a Markov chain with state space $S = \{0, 1, 2, \dots\}$.

It possesses the Markov property because:

$$P[X_{n+1} = 1+x_n | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]$$

$$= P[X_{n+1} = 1+x_n | Y_1 = x_1, Y_2 = x_2 - x_1, \dots, Y_n = x_n - x_{n-1}]$$

Since $\sum_{j=1}^n Y_j = X_n$, the condition $Y_1 = x_1, Y_2 = x_2 - x_1, \dots, Y_n = x_n - x_{n-1}$ is a

function only of X_n and hence:

$$P[X_{n+1} = 1+x_n | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]$$

$$= P[X_{n+1} = 1+x_n | X_n = x_n] = \frac{f(x_n)}{g(n)}$$

and this is independent of the values of x_1, x_2, \dots, x_{n-1} .

Note that the random variables Y_n , which are the increments of the chain $\{X_n\}$, are not independent.

Note that the chain is only time-homogeneous if $g(n)$ is constant (but this is neither very realistic nor useful).

If $g(n)$ is a constant, we may as well take it to be 1, since we can rescale the function $f(x_n)$ suitably. For example, $\frac{f(x_n)}{3} = \frac{1}{3}f(x_n)$. So we define a new function $f^*(x_n) = \frac{1}{3}f(x_n)$ and take $g(n) = 1$.



Example

Consider the following two drivers. The first is a 45-year old who has had two accidents in the last 20 years of motoring (both in the last year), and the other is an 18-year old who's had two accidents in the last year. For any time-homogeneous model as described above, with time period of one month say, the probabilities of the two drivers having an accident next month would be the same!

A more meaningful model should take into account the length of time over which your previous accidents have occurred. In the above example, this is the length of time you've been driving. So we should use a time-inhomogeneous model here.



Example

For time periods of one year, let $f(x_n) = 0.5 + x_n$ and $g(n) = 1 + n$ so that:

$$P[Y_{n+1} = 1 | X_n = x_n] = \frac{0.5 + x_n}{1 + n}$$

This example is rather crude but demonstrates some broad features.



Question 3.10

Consider the model in the example.

- (i) What is the probability of a driver who has had no accidents in their first year, having an accident in their second?
- (ii) What if they have had an accident every year for the last ten years?



Question 3.11

For the model in the example, what is the ij -th entry in the one-step transition matrix $\left(P^{(n,n+1)}\right)_{ij}$ of the Markov chain $\{X_n\}$?

6 The long-term distribution of a Markov chain

6.1 The stationary probability distribution

We say that $\pi_j, j \in S$ is a **stationary probability distribution** for a Markov chain with transition matrix P if the following conditions hold for all j in S :

- $\pi_j = \sum_{i \in S} \pi_i p_{ij}$ (3.4)
- $\pi_j \geq 0$
- $\sum_{j \in S} \pi_j = 1$

Note how (3.4) can be stated in the compact form $\pi = \pi P$ where π is viewed as a row vector.

The interpretation of (3.4) is that, if we take π as our *initial probability distribution*, that is to say $P[X_0 = i] = \pi_i$, then the distribution at time 1 is again given by π :

$$P[X_1 = j] = \sum_{i \in S} P[X_1 = j | X_0 = i] P[X_0 = i] = \sum_{i \in S} p_{ij} \pi_i = \pi_j$$

The same is true at all times $n \geq 1$, so that π is an *invariant probability distribution*; in fact the chain is then a **stationary process** in the sense of Chapter 2.

So if the chain ever reaches the distribution π at some time n , ie $P(X_n = i) = \pi_i$ for all values of i , then because the transition matrix sends π back to itself, $\pi = \pi P$, the distribution of X_t will be π for all subsequent times $t \geq n$. The statistical properties of the process do not change over time, so the chain is a stationary process.

In general a Markov chain need not have a stationary probability distribution, and if it exists it need not be unique. For instance no stationary probability distribution exists for Model 5.3, whereas in Model 5.4 uniqueness depends on the values of α, β . When the state space S is *finite*, the situation is simpler.



Important result

A Markov chain with a finite state space has at least one stationary probability distribution.

The proof of this is beyond the syllabus.

As an example, we will compute a stationary probability for NCD Model 5.2. The equations (3.4) read:

$$\begin{aligned}
 \pi_0 &= \frac{1}{4} \pi_0 + \frac{1}{4} \pi_1 + \frac{1}{4} \pi_{2-} \\
 \pi_1 &= \frac{3}{4} \pi_0 + \frac{1}{4} \pi_{2+} \\
 \pi_{2+} &= \frac{3}{4} \pi_1 \\
 \pi_{2-} &= \frac{1}{4} \pi_3 \\
 \pi_3 &= \frac{3}{4} \pi_{2+} + \frac{3}{4} \pi_{2-} + \frac{3}{4} \pi_3
 \end{aligned} \tag{3.5}$$

If you look at the transition matrix P you'll see that the coefficients in these equations correspond to the columns in the matrix.

This linear system is not linearly independent since adding up all the equations results in an identity (this is a general feature of equations $\pi = \pi P$ due to the property $\sum_{j \in S} p_{ij} = 1$). Because of this we can discard any one of the equations, say the last one.

To say that the above equations are not linearly independent means that any four of them will always rearrange to give the remaining one, which is therefore redundant. This will always be true for equations of the form $\pi = \pi P$ if P is a matrix whose rows sum to 1. As a result we may discard one of them – it doesn't matter which – and solve the remaining system. We usually discard the most complicated looking one.

We give an example of what we mean in the case of only two equations in two unknowns.



Example

Consider a Markov chain with only two states, $S = \{0, 1\}$, and transition matrix:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

Determine the stationary distribution(s) of this chain.

Solution

We want to solve $\pi P = \pi$, that is:

$$(\pi_0, \pi_1) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} = (\pi_0, \pi_1)$$

or explicitly:

$$\begin{aligned} \frac{1}{2}\pi_0 + \frac{1}{3}\pi_1 &= \pi_0 \\ \frac{1}{2}\pi_0 + \frac{2}{3}\pi_1 &= \pi_1 \end{aligned}$$

Taking all terms to the left-hand side in both equations we get:

$$\begin{aligned} -\frac{1}{2}\pi_0 + \frac{1}{3}\pi_1 &= 0 \\ \frac{1}{2}\pi_0 - \frac{1}{3}\pi_1 &= 0 \end{aligned}$$

These equations are equivalent (since they just have opposite signs). So either one of them can be discarded. Solving either one of them gives $\pi_1 = \frac{3}{2}\pi_0$.

We also require $\pi_0 + \pi_1 = 1$ and therefore we have the unique solution:

$$\pi_0 = \frac{2}{5} \quad \text{and} \quad \pi_1 = \frac{3}{5}$$

We can check the solution by substituting these values back into the discarded equation.

Note also that by linearity, any multiple of a solution of (3.5) is again a solution; uniqueness comes only as a result of the normalisation $\sum_{j \in S} \pi_j = 1$. For this

reason, it is good practice to solve for the components of π in terms of one of them (say π_1 here), which we will refer to as the working variable. The value of the working variable is determined at the last step by normalisation.

Once the value of the working variable has been established, the others can be deduced as well.

With respect to the last paragraph, note that, although uniqueness comes only as a result of the normalisation $\sum_{j \in S} \pi_j = 1$, this does not mean that uniqueness has to come at all.

It might be the case that even after applying the normalisation condition the solution is not unique. In addition, if the state space is not finite, then there may not be a stationary distribution at all.

We now summarise the method and apply it to the above example.

Step 1: Discard one of the equations. Here the first or the last one are obvious choices; delete the final one say.

Step 2: Select one of the π_j 's as working variable. Here $\pi_1, \pi_{2+}, \pi_{2-}$ or π_3 are reasonable choices; choose π_1 .

Step 3: Rewrite remaining equations in terms of the working variable.

$$3\pi_0 - \pi_{2-} = \pi_1 \quad (\text{a})$$

$$3\pi_0 + \pi_{2+} = 4\pi_1 \quad (\text{b})$$

$$\pi_{2+} = \frac{3}{4}\pi_1 \quad (\text{c})$$

$$4\pi_{2-} - \pi_3 = 0 \quad (\text{d})$$

Step 4: Solve the equations in terms of the working variable.

In general we might do this by Gaussian elimination (a general method for solving a system of linear equations), but here the equations are so simple that the solution can be read off if we take them in the right order:

$$\pi_{2+} = \frac{3}{4}\pi_1$$

We get this directly from (c), then substituting this into (b) gives:

$$\pi_0 = \frac{\pi_1}{3}(4 - \frac{3}{4}) = \frac{13}{12}\pi_1$$

Then substitute for π_0 in (a) to get:

$$\pi_{2-} = \pi_1(-1 + \frac{13}{4}) = \frac{9}{4}\pi_1$$

Finally:

$$\pi_3 = 9\pi_1$$

from (d).

Step 5: Solve for the working variable.

We now have:

$$\pi = \pi_1 \left(\frac{13}{12}, 1, \frac{3}{4}, \frac{9}{4}, 9 \right)$$

$$\sum_j \pi_j = \frac{\pi_1}{12}(13 + 12 + 9 + 27 + 108) = \frac{169}{12}\pi_1 = 1$$

$$\pi_1 = \frac{12}{169}$$

Step 6: Combine the results of the last two steps to obtain the solution.

$$\pi = \left(\frac{13}{169}, \frac{12}{169}, \frac{9}{169}, \frac{27}{169}, \frac{108}{169} \right)$$

It is good practice to use the equation discarded earlier to verify that the calculated solution is correct.

In summary, we were given a Markov chain with finite state space and we've derived a stationary distribution by solving the equations $\pi = \pi P$.

In the above example it has turned out that there is only one solution, but this won't always be the case. A sufficient condition, though not a necessary one, is given below. This requires the introduction of a further classification of Markov chains into those that are irreducible and those that are not.

The question of uniqueness of the stationary distribution is more delicate than existence; we shall consider only irreducible chains.



Irreducibility

A Markov chain is said to be irreducible if any state j can be reached from any other state i .

In other words, a chain is irreducible if, given any pair of states i, j there exists an integer n with $p_{ij}^{(n)} > 0$.

This is a property that can normally be judged from the transition graph alone.



Question 3.12

- (i) Is the process with the following transition matrix irreducible?

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

- (ii) What are the stationary distributions of the process?

The Markov chains of Models 5.1, 5.2 and 5.3 are irreducible; so is 5.4 except when either boundary is absorbing ($\alpha=1$ or $\beta=1$). Such absorbing states occur in many practical situations (eg ruin).



Question 3.13

Explain why a random walk with absorbing barriers is not irreducible.



Important result

An irreducible Markov chain with a finite state space has a unique stationary probability distribution.

The proof of this result is beyond the syllabus.



Question 3.14

Is the following process irreducible?

$$P = \begin{pmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ \frac{1}{10} & \frac{1}{5} & \frac{3}{5} & \frac{1}{10} \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \end{pmatrix}$$



Question 3.15

(i) Is the following process irreducible?

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

(ii) What are the stationary distributions?

It is common for Markov chains with infinite state spaces to have no stationary probability distribution, even if the chain is irreducible; this is the case for the simple random walk of Model 5.3.

6.2 The long-term behaviour of Markov chains

The importance of the stationary distribution comes from its connection with the long-term behaviour of a Markov chain. Under suitable conditions (to be made precise below), a Markov chain will “settle down” to its stationary distribution after a sufficiently long period of time.

The result is expressed in terms of the limits of probability distributions. You need to be familiar with “ $\lim_{n \rightarrow \infty}$ ” and convergence.

It is natural to expect the distribution of a Markov chain to tend to the invariant distribution π for large times. This is why the stationary distribution is so important: if the above convergence holds, $p_{ij}^{(n)}$ will be close to π_j for an overwhelming fraction of the time in the long run.

Certain phenomena complicate the above picture somewhat.



The period of a state

A state i is said to be *periodic* with period $d > 1$ if a return to i is possible only in a number of steps that is a multiple of d (ie $p_{ii}^{(n)} = 0$ unless $n = md$ for some integer m).

A state is said to be *aperiodic* if it is not periodic.

It is only for aperiodic states that $\lim_{n \rightarrow \infty} p_{ii}^{(n)}$ can exist.

One can check using the transition graphs that, in Models 5.1 and 5.2, all states are aperiodic whereas in Model 5.3, all states have period 2. Finally in Model 5.4 all states are aperiodic unless both α and β are either 0 or 1, ie unless $\alpha = 0$ or 1 and $\beta = 0$ or 1.

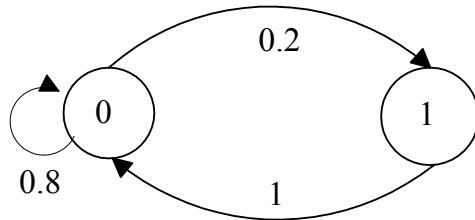
It is not necessarily the case that return to an aperiodic state is possible after an arbitrary number of steps, only that return is not constrained to be in a multiple of some number $d > 1$. So, in effect, the highest common factor of the return times for the state needs to be 1 in order for the state to be aperiodic.

Drawing a transition diagram might help you to decide whether or not a state is periodic. Note that if a state has an arrow back to itself, that state is aperiodic because a return to that state is possible in any number of steps. However, even if there is no arrow from the state back to itself, the state may still be aperiodic.



Example

Consider a time-homogeneous Markov chain on the state space $S = \{0,1\}$ with transition probabilities $p_{00} = 0.8$, $p_{01} = 0.2$, $p_{10} = 1$ and $p_{11} = 0$. The transition diagram is:



Since state 0 has an arrow back to itself, it is aperiodic.

A return to state 1 is not possible in 1 step, but is possible in 2, 3, 4, ... steps. The highest common factor of 2, 3, 4, ... is 1. So state 1 is aperiodic.

In fact, once we have decided that state 0 is aperiodic, we can say straight away that state 1 is also aperiodic if we know the following important result for irreducible chains.



Important result

If a Markov chain is irreducible all its states have the same period (or all are aperiodic).

This greatly simplifies the problem of finding periodicities for irreducible chains as only one state need be considered. Note that the chain in the example above *is* irreducible (every state can be reached from every other state).



Aperiodic Markov chains

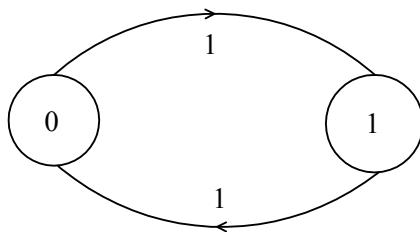
A Markov chain is said to be aperiodic if all its states are aperiodic.

In the next example we consider a Markov chain that is not aperiodic.



Example

Consider a time-homogeneous Markov chain on the state space $S = \{0,1\}$ with transition probabilities $p_{01} = 1$ and $p_{10} = 1$. The transition graph is:



The chain has a finite number of states. Each state can be reached from the other state so the chain is irreducible. So it must have a unique stationary distribution. It isn't difficult to see that this must be $(\frac{1}{2}, \frac{1}{2})$, ie an equal chance of being in either state. (See Question 3.16 below.)

From the diagram we see that a return to state 0 is possible only in an even number of steps. So the period of state 0 is 2. The same is true of state 1.

So the chain is not aperiodic and the process may not conform to the stationary distribution in the long term.

This tells us that, although the process has a stationary distribution, if the process doesn't start off in that distribution, then it will never reach it. The idea that a process can "start off in a distribution" might be confusing: any particular run of the process surely starts in a particular state? This is true. However, we can have a lack of information about which state that is. For example, we might only know that at time 0 there is a 10% chance that it's in state 0, and a 90% chance that it's in state 1. At time 1, these probabilities will be reversed, with a 90% chance of being in state 0, and only 10% chance of being in state 1. The process doesn't settle down to an equilibrium position.

An alternative way of thinking about this is to suppose we have a large number of independent copies of the Markov process running. To be concrete, let's assume that each independent process describes the state of a person, as in the case of policyholders following an NCD Markov chain. We can then picture the idea of the starting distribution in terms of numbers of people. For example, the 10%/90% split referred to above would correspond to 10% of the people starting in state 0, and 90% starting in state 1. Because of the transition probabilities, each person will change state. This will continue at each time step, so that the process never settles down to equilibrium.



Question 3.16

Show by solving the necessary matrix equation that the stationary distribution for the process in the example above is $(\frac{1}{2}, \frac{1}{2})$.



Example

Consider a time-homogeneous Markov chain on the state space $S = \{0,1\}$ with transition probabilities $p_{00} = p_{01} = p_{10} = p_{11} = \frac{1}{2}$. So no matter what state you are in, the chances of moving to the other state are fifty-fifty, and therefore so also are the chances of staying put.

Again the process is finite and irreducible, so a unique stationary distribution exists. Moreover, this stationary distribution is again $(\frac{1}{2}, \frac{1}{2})$.

However, in this case a sample path starting in state 0, say, can return to state 0 after an arbitrary number of steps, hence the state is aperiodic. Exactly the same is true of state 1.

Furthermore, $p_{00}^{(2)} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, since to go from 0 to 0 in 2 time steps, you either stay still for two time steps, or go from 0 to 1 and back to 0 again.

In fact, since $P^2 = P$, we must have $P^n = P$ by induction.

Therefore $p_{00}^{(n)} = \frac{1}{2}$ for all $n = 1, 2, 3, \dots$. It follows that $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{1}{2}$.

In contrast to the process in the previous example, this process does settle down to the equilibrium state. In fact this occurs after one time step. But, in general, this would take longer.

We will now state a result on convergence to the stationary probability distribution.



Important result

Let $p_{ij}^{(n)}$ be the n -step transition probability of an irreducible aperiodic Markov chain on a finite state space. Then for every i and j :

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j$$

where π is the stationary probability distribution.

Note how the above limit is independent of the starting state i . This proof is beyond the syllabus.

This result is saying that no matter what state i you are in, the probability of ending up in state j after a very long time, is the same as the probability of being in state j given the stationary distribution π . This is the same as saying that, after a very long time, the distribution is constant and equal to the stationary distribution.



Summary

As we have seen:

- a Markov chain with a finite state space has at least one stationary distribution
- an irreducible Markov chain with a finite state space has a unique stationary distribution
- an irreducible, aperiodic Markov chain with a finite state space will settle down to its unique stationary distribution in the long run.

7 Modelling using Markov chains

Using the principle of economy of effort, it is common to start the modelling process by attempting to fit a simple stochastic model, such as a Markov chain, to a set of observations. If tests show that this is inadequate, a more sophisticated model can be attempted at the next stage of the modelling process.

This section assumes that the model being fitted is time-homogeneous. The situation is generally more complicated when fitting a time-inhomogeneous model.

7.1 Estimating transition probabilities

The first thing to fix when setting up a Markov model is the state space. As shown by the example in Section 5.2, the state space which first springs to mind may not be the most suitable and may need some modification before a Markov model can be fitted.

Recall that the example referred to was the NCD system where it was required to split one of the discount levels into two, depending on where you came from.

Once the state space is determined, however, the Markov model must be fitted to the data by estimating the transition probabilities p_{ij} .

Denote by x_1, x_2, \dots, x_N the available observations and define:

- n_i as the number of times t ($1 \leq t \leq N - 1$) such that $x_t = i$;
- n_{ij} as the number of times t ($1 \leq t \leq N - 1$) such that $x_t = i$ and $x_{t+1} = j$.

Thus n_{ij} is the observed number of transitions from state i to j , n_i the observed number of transitions from state i .

The reason that the definition of n_i only allows t to go up to $N - 1$, rather than N , is so that it equals the number of chances for a transition out of state i , and not just the number of times it is in state i .

Then the best estimate of p_{ij} is $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$.

If a confidence interval is required for a transition probability, the fact that the conditional distribution of N_{ij} given N_i is $\text{Binomial}(N_i, p_{ij})$ means that a confidence interval may be obtained by standard techniques.

Finding confidence intervals for binomial parameters was covered in Subject CT3.

7.2 Assessing the fit

The next step is to ensure that the fit of the model to the data is adequate, or in other words to check that the Markov property seems to hold.

For a general Markov chain model a full verification of the Markov property would involve a great deal of work and a voluminous supply of data. In practice it is generally considered sufficient to look at triplets of successive observations.

Denote by n_{ijk} the number of times t ($1 \leq t \leq N - 2$) such that $x_t = i$, $x_{t+1} = j$ and $x_{t+2} = k$. If the Markov property holds we expect n_{ijk} to be an observation from a Binomial distribution with parameters n_{ij} and p_{jk} . A simple but effective test, therefore, is the chi-square goodness-of-fit test based on the test statistic:

$$\chi^2 = \sum_i \sum_j \sum_k \frac{(n_{ijk} - n_{ij}\hat{p}_{jk})^2}{n_{ij}\hat{p}_{jk}}$$

The number of degrees of freedom to use in such a situation is $r - 2q + s$, where:

s denotes the number of states i in the state space for which $n_i > 0$

q denotes the number of pairs (i, j) for which $n_{ij} > 0$, and

r denotes the number of triplets (i, j, k) for which $n_{ij}n_{jk} > 0$



Question 3.17

You have been given the following series of data from a 3-state process:

1,3,2,2,1,3,3,2,3,1,2,3,2,1,1,2,2,1,3,3

- (i) Calculate the values of n_i , n_{ij} and n_{ijk} .
- (ii) Estimate the one-step transition probabilities.
- (iii) Carry out the χ^2 -test and state what concerns you might have in this situation.

An additional method in frequent use for assessing goodness of fit is to run some simulations of the fitted chain and to compare graphs of the resulting trajectories with a graph of the process actually observed. This method often highlights deficiencies that are missed by the chi-square test. For example, given a sequence y_1, y_2, \dots, y_N of closing values of an exchange rate, one model which suggests itself is to let x_t be the nearest integer to $K \log y_t$ where K is a scaling constant of suitable magnitude, and to model x_1, x_2, \dots, x_N as a random walk, with transition probabilities:

$$p_{i,i+1} = \theta, \quad p_{i,i-1} = \phi, \quad p_{i,i} = 1 - \theta - \phi.$$

The parameters θ and ϕ can be estimated quite satisfactorily in practice, but a visual comparison of a simulated random walk with the observed trajectory of x tends to show that the real exchange rate remains relatively constant for long periods with occasional bursts of increased volatility, whereas the Markov chain model is incapable of simulating such behaviour.

7.3 Simulation

Simulating a time-homogeneous Markov chain is fairly straightforward, as the Markov property means that the conditional distribution of X_{t+1} given the history of X up until time t is only dependent on X_t .

If the state space of X is finite, there are only a limited number of distributions, all discrete, from which the program needs to be able to sample; these can be listed individually, along with instructions telling the program which distribution to use for each step.



Example

A two-state Markov chain has transition matrix:

$$\begin{matrix} & 0 & 1 \\ \begin{matrix} 0 \\ 1 \end{matrix} & \left[\begin{matrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{matrix} \right] \end{matrix}$$

If the process is in state 0 at time 0, explain how you would simulate a series of observations from this process.

Solution

Row 1 of the transition matrix is the conditional distribution of X_1 given that $X_0 = 0$.

We use Monte Carlo simulation to generate a simulated value for X_1 .

If the simulated value is 0, then we repeat the Monte Carlo simulation to obtain a simulated value of X_2 . If the simulated value is 1, then we use row 2 of the transition matrix, which is the conditional distribution of X_2 given that $X_1 = 1$, and simulate a value for X_2 .

This process is repeated to simulate additional values of the Markov chain.

Models that assume an infinite state space usually have a simple transition structure, often based on the distribution of the increments. The random walk, which has independent increments, is one such example; another might be a process which can only make transitions of the form $x \mapsto x+1$ or $x \mapsto x-1$, with respective probabilities θ_x and $1-\theta_x$.

Note that the second example is not a random walk because the increments (+1 or -1) are not identically distributed. The associated probabilities depend on the current state.

In addition to commercial simulation packages, which are able to simulate Markov chains without difficulty, even standard spreadsheet software can easily cope with the practical aspects of estimating transition probabilities and performing a simulation.

8 Exam-style question

Here is a past exam question on Markov chains.



Question 3.18

Subject CT4, September 2006, Questions A4

The credit-worthiness of debt issued by companies is assessed at the end of each year by a credit rating agency. The ratings are A (the most credit-worthy), B and D (debt defaulted). Historic evidence supports the view that the credit rating of a debt can be modelled as a Markov chain with one-year transition matrix:

$$X = \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix}$$

- (i) Determine the probability that a company currently rated A will never be rated B in the future. [2]
 - (ii)
 - (a) Calculate the second order transition probabilities of the Markov chain.
 - (b) Hence calculate the expected number of defaults within the next two years from a group of 100 companies, all initially rated A. [2]
 - (iii) Calculate the expected number of defaults for this investment manager over the next two years, given that the portfolio initially consists of 100 A-rated bonds. [2]
 - (iv) Comment on the suggestion that the downgrade trigger strategy will improve the return on the portfolio. [2]
- [Total 8]

This page has been left blank so that you can keep the chapter summaries together for revision purposes.



Chapter 3 Summary

Markov chains

A Markov process with a discrete time set and discrete state space is called a *Markov chain*.

Chapman-Kolmogorov equations

$$p_{ij}^{(m,n)} = \sum_{k \in S} p_{ik}^{(m,l)} p_{kj}^{(l,n)}$$

for all states i, j in S and all integer times $m \leq l \leq n$.

Time-homogeneous Markov chains

A simplification occurs if the one step transition probabilities are time independent:

$$p_{ij}^{(n,n+1)} = p_{ij}$$

Then the Chapman-Kolmogorov equations are:

$$p_{ij}^{(n-m)} = \sum_{k \in S} p_{ik}^{(l-m)} p_{kj}^{(n-l)}$$

The transition matrix

The matrix P above is a square $N \times N$ matrix, where n is the number of states in S . The entry in the i th row and j th column is p_{ij} .

In the time-homogeneous case, the l -step transition probability $p_{ij}^{(l)}$ (ie the probability of moving from state i to state j in exactly l steps) can be obtained by calculating the (i, j) th entry of the matrix P^l .

Random walks

Random walks are important examples of Markov chains. The increments of a random walk are IID. In other words, the values move up or down by completely random amounts at each step.

A *simple* random walk has step-sizes of ± 1 , ie:

$$P(X_n = x+1 | X_{n-1} = x) = p, \quad P(X_n = x-1 | X_{n-1} = x) = 1-p (= q)$$

In a simple *symmetric* random walk, $p = q = \frac{1}{2}$.

Stationary distributions

These probabilities must satisfy the vector equation $\pi = \pi\mathbf{P}$ with $\pi_i \geq 0$ and $\sum \pi_i = 1$.

A solution to these equations is called a *stationary distribution*.

Irreducible chains

A Markov chain is said to be irreducible if every state can be reached from every other state.

Periodicity

A state i is said to be periodic with period $d > 1$ if a return to state i is possible only in a number of steps that is a multiple of d .

A state is said to be aperiodic if it is not periodic.

A Markov chain is said to be aperiodic if all its states are aperiodic.

If a Markov chain is irreducible, all its states have the same period or are all aperiodic.

Long-term behaviour of Markov chains

A Markov chain with a finite state space has at least one stationary distribution.

An irreducible Markov chain with a finite state space has a unique stationary distribution.

An irreducible, aperiodic Markov chain with a finite state space will settle down to its unique stationary distribution in the long run.

Estimating transition probabilities

The transition probability p_{ij} is estimated by:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} = \frac{\text{number of transitions from state } i \text{ to state } j}{\text{number of transitions out of state } i}$$

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 3 Solutions

Solution 3.1

This definition is appropriate only when the time set is discrete. In general, it need not be.

Solution 3.2

(i) *Matrix equation*

With the given notation:

$$\underline{X}_5 = \underline{X}_0 P^{(0,1)} P^{(1,2)} P^{(2,3)} P^{(3,4)} P^{(4,5)}$$

(ii) *Simplification*

In the homogeneous case:

$$P^{(0,1)} = P^{(1,2)} = P^{(2,3)} = P^{(3,4)} = P^{(4,5)} = P$$

So:

$$\underline{X}_5 = \underline{X}_0 P^5$$

Solution 3.3(i) **Values of p and q**

Since each row must sum to one we have $p = \frac{3}{5}$ from row 3 and therefore $q = \frac{2}{5}$ from the first row.

(ii) **Transition probabilities**

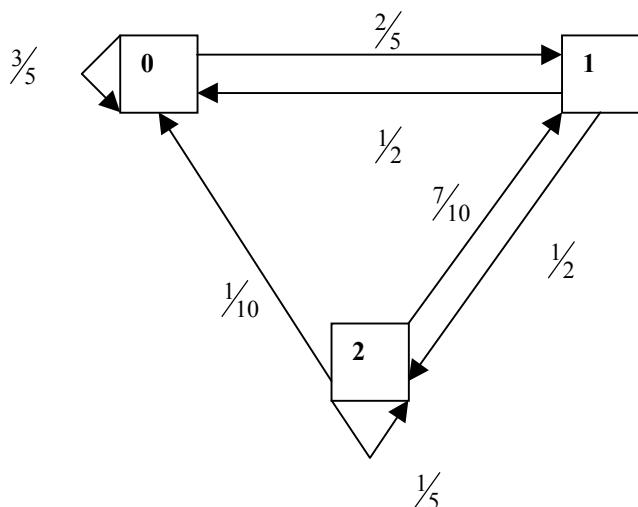
We use the fact that $p_{ij}^{(3)} = (P^3)_{ij}$:

$$\begin{aligned} P^3 &= \frac{1}{1,000} \begin{pmatrix} 6 & 4 & 0 \\ 5 & 0 & 5 \\ 1 & 7 & 2 \end{pmatrix}^3 = \frac{1}{1,000} \begin{pmatrix} 56 & 24 & 20 \\ 35 & 55 & 10 \\ 43 & 18 & 39 \end{pmatrix} \begin{pmatrix} 6 & 4 & 0 \\ 5 & 0 & 5 \\ 1 & 7 & 2 \end{pmatrix} \\ &= \frac{1}{1,000} \begin{pmatrix} 476 & 364 & 160 \\ 495 & 210 & 295 \\ 387 & 445 & 168 \end{pmatrix} = \begin{pmatrix} 0.476 & 0.364 & 0.160 \\ 0.495 & 0.210 & 0.295 \\ 0.387 & 0.445 & 0.168 \end{pmatrix} \end{aligned}$$

Note that you can work out powers of matrices either way round. For example, $P^2P = PP^2 = P^3$. In fact for any three matrices A, B, C we have $A(BC) = (AB)C$.

(iii) **Transition graph**

The transition graph is:



Solution 3.4

Repeated post-multiplication of the vector $(0, 1, 0)$ by the transition matrix P gives:

$$\begin{aligned}(0, 1, 0) &\rightarrow \left(\frac{1}{4}, 0, \frac{3}{4}\right) \\ &\rightarrow \left(\frac{1}{16}, \frac{3}{8}, \frac{9}{16}\right) \\ &\rightarrow \left(\frac{7}{64}, \frac{3}{16}, \frac{45}{64}\right) \\ &\rightarrow \left(*, \frac{33}{128}, *\right)\end{aligned}$$

So the required probability is $\frac{33}{128}$.

Alternatively, we can consider each possible sample path separately and sum the probabilities as follows.

The only paths are:

$$1 \rightarrow 0 \rightarrow 0 \rightarrow 0 \rightarrow 1 \quad \text{with probability} \quad \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{256}$$

$$1 \rightarrow 0 \rightarrow 1 \rightarrow 0 \rightarrow 1 \quad \text{with probability} \quad \frac{1}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{9}{256}$$

$$1 \rightarrow 0 \rightarrow 1 \rightarrow 2 \rightarrow 1 \quad \text{with probability} \quad \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{9}{256}$$

$$1 \rightarrow 2 \rightarrow 1 \rightarrow 0 \rightarrow 1 \quad \text{with probability} \quad \frac{3}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{9}{256}$$

$$1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 1 \quad \text{with probability} \quad \frac{3}{4} \times \frac{1}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{9}{256}$$

$$1 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 1 \quad \text{with probability} \quad \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{27}{256}$$

So adding up probabilities for each path we get $\frac{66}{256} = \frac{33}{128}$ as above.

Solution 3.5

We will order the states as $\{0, 1, 2^+, 2^-, 3\}$. The initial distribution is $(1, 0, 0, 0, 0)$.

Successive post-multiplication by the transition matrix:

$$P = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} \\ \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

gives:

$$\begin{aligned} (1, 0, 0, 0, 0) &\rightarrow \left(\frac{1}{4}, \frac{3}{4}, 0, 0, 0 \right) \\ &\rightarrow \left(\frac{1}{4}, \frac{3}{16}, \frac{9}{16}, 0, 0 \right) \\ &\rightarrow \left(\frac{7}{64}, \frac{21}{64}, \frac{9}{64}, 0, \frac{27}{64} \right) \\ &\rightarrow \left(\frac{7}{64}, \frac{15}{128}, \frac{63}{256}, \frac{27}{256}, \frac{27}{64} \right) \\ &\rightarrow \left(*, *, *, *, \frac{297}{512} \right) \end{aligned}$$

So the required probability is $\frac{297}{512}$.

Alternatively, we could calculate the (1,5)th entry in the matrix P^5 . In order to do this we first calculate the two matrices P^2 and P^3 :

$$P^2 = \frac{1}{16} \begin{pmatrix} 1 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 3 \\ 1 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}^2 = \frac{1}{16} \begin{pmatrix} 4 & 3 & 9 & 0 & 0 \\ 1 & 6 & 0 & 0 & 9 \\ 1 & 0 & 3 & 3 & 9 \\ 1 & 3 & 0 & 3 & 9 \\ 1 & 0 & 0 & 3 & 12 \end{pmatrix}$$

$$P^3 = \frac{1}{64} \begin{pmatrix} 1 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 0 & 3 \\ 1 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 4 & 3 & 9 & 0 & 0 \\ 1 & 6 & 0 & 0 & 9 \\ 1 & 0 & 3 & 3 & 9 \\ 1 & 3 & 0 & 3 & 9 \\ 1 & 0 & 0 & 3 & 12 \end{pmatrix} = \frac{1}{64} \begin{pmatrix} 7 & 21 & 9 & 0 & 27 \\ 7 & 3 & 18 & 9 & 27 \\ 4 & 6 & 0 & 9 & 45 \\ 7 & 3 & 9 & 9 & 36 \\ 4 & 3 & 0 & 12 & 45 \end{pmatrix}$$

So (1,5)th entry is:

$$\begin{aligned} \frac{1}{1,024} \sum_{k=1}^5 (P^3)_{1k} (P^2)_{k5} &= \frac{1}{1,024} (7 \times 0 + 21 \times 9 + 9 \times 9 + 0 \times 9 + 27 \times 12) \\ &= \frac{594}{1,024} = \frac{297}{512} \end{aligned}$$

Solution 3.6

Consider going from i to j in n steps. Let the number of positive steps be u and the number of negative steps be d . Because we make n steps in total, it follows that:

$$u + d = n$$

Because the net positive movement must equal the excess of positive steps over negative steps, we also have:

$$u - d = j - i$$

Solving these simultaneous equations we get:

$$u = \frac{1}{2}(n + j - i)$$

and:

$$d = n - u = n - \frac{1}{2}(n + j - i) = \frac{1}{2}(n - j + i)$$

Since u and d must be non-negative whole numbers, it must be the case that $n + j - i$ and $n + i - j$ are non-negative *even* numbers. We also know that $j - i \leq n$. Putting these together we have $0 \leq n + j - i \leq 2n$ and $n + j - i$ is even.

Now the order in which the positive and negative steps occur doesn't matter. There are $\binom{n}{u} = \binom{n}{\frac{1}{2}(n + j - i)}$ ways of choosing these. Each positive step occurs with probability p . Putting all this together we have:

$$p_{ij}^{(n)} = \begin{cases} \binom{n}{u} p^u (1-p)^d & \text{if } 0 \leq n + j - i \leq 2n \text{ and } n + j - i \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

Solution 3.7

One way of viewing a random walk is to picture a particle randomly moving from place to place.

- If the particle is absorbed by a state then the probability of it moving to another state is 0 and hence the terminology “absorbing state.”
- Similarly, if the particle is fully reflected or bounces back out with probability 1 then the state is called a reflecting state.
- The third type of boundary condition is a mix of the two. There is some chance the particle will be reflected and some that it will stay put. If there is a non-zero chance of being reflected then it will be reflected if we wait long enough.

Solution 3.8

In Model 5.1 the boundary conditions are:

$$P[X_{n+1} = 0 | X_n = 0] = \frac{1}{4}$$

$$P[X_{n+1} = 1 | X_n = 0] = \frac{3}{4}$$

$$P[X_{n+1} = 2 | X_n = 2] = \frac{3}{4}$$

$$P[X_{n+1} = 1 | X_n = 2] = \frac{1}{4}$$

.

Solution 3.9(i) ***Why functions are increasing***

$\frac{f(y_1 + y_2 + \dots + y_n)}{g(n)}$ is the probability of a driver having an accident in the next time period. You would expect this to be higher for a driver who has had more accidents in the last n time periods. In other words, for fixed n we expect the probability to be higher for larger $y_1 + y_2 + \dots + y_n$. This says that f is an increasing function. (We assume that both f and g are positive. If one was negative, the other would have to be negative too, since their ratio is positive. They would then both have to be decreasing. We can ignore this possibility without loss of generality.)

On the other hand, if two drivers have the same number of accidents, but one driver's occurred within a shorter time period, then we would expect this driver to have a higher probability of having another accident. In other words, if n is smaller for fixed value of $y_1 + y_2 + \dots + y_n$ then we expect the probability to be higher. In turn this says that g must be smaller. Thus g must also be an increasing function.

(ii) ***Inequalities***

In order that $\frac{f(y_1 + y_2 + \dots + y_n)}{g(n)}$ is a probability we must have:

$$0 \leq \frac{f(y_1 + y_2 + \dots + y_n)}{g(n)} \leq 1$$

The above inequality is equivalent to:

$$0 \leq f(y_1 + y_2 + \dots + y_n) \leq g(n)$$

In particular, for the “maximum” case when all the y 's are equal to 1, we have $y_1 + y_2 + \dots + y_n = n$, and we obtain $0 \leq f(n) \leq g(n)$ as required.

(iii) ***Special case***

In this case:

$$\frac{f(y_1 + \dots + y_n)}{g(n)} = \frac{1}{n}(y_1 + \dots + y_n)$$

So we're estimating the probability of a claim next year using the average number of years that had a claim in the past.

Solution 3.10

$$(i) \quad P[Y_2 = 1 | X_1 = 0] = \frac{0.5 + 0}{2} = 0.25$$

$$(ii) \quad P[Y_{11} = 1 | X_{10} = 10] = \frac{0.5 + 10}{11} = \frac{10.5}{11} = 0.955$$

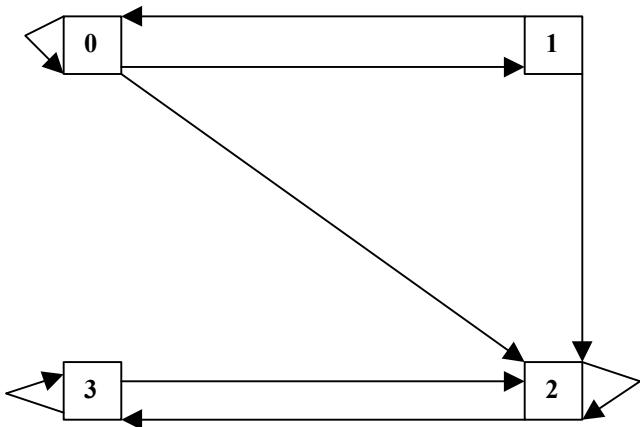
Solution 3.11

$$\left(P^{(n,n+1)} \right)_{ij} = P[X_{n+1} = j | X_n = i] = \begin{cases} \frac{0.5 + i}{1+n} & \text{if } j = i + 1 \\ \frac{0.5 + n - i}{1+n} & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

Solution 3.12

- (i)
- Is process irreducible?***

The transition graph looks like this:



From the transition graph we see that this process is not irreducible. For example, we cannot get from state 3 to state 1.

- (ii)
- Stationary distributions***

To calculate the stationary distributions we need to solve the matrix equation:

$$(\pi_0, \pi_1, \pi_2, \pi_3) \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\pi_0, \pi_1, \pi_2, \pi_3)$$

This is equivalent to:

$$\begin{aligned} 6\pi_0 + 8\pi_1 &= 12\pi_0 \\ 3\pi_0 &= 12\pi_1 \\ 3\pi_0 + 4\pi_1 + 8\pi_2 + 6\pi_3 &= 12\pi_2 \\ 4\pi_2 + 6\pi_3 &= 12\pi_3 \end{aligned}$$

We can ignore one equation, say the third one. Then rearranging we have:

$$\begin{aligned}-6\pi_0 + 8\pi_1 &= 0 \\ 3\pi_0 - 12\pi_1 &= 0 \\ 4\pi_2 - 6\pi_3 &= 0\end{aligned}$$

Choose π_1 as the working variable. From the second equation we have $\pi_0 = 4\pi_1$ which isn't consistent with the first equation unless $\pi_0 = \pi_1 = 0$. We must therefore change our working variable to π_2 or π_3 , it doesn't matter which we choose. The third equation gives $\pi_2 = \frac{3}{2}\pi_3$.

The solution is therefore $\frac{1}{5}(0, 0, 3, 2)$.

An alternative approach, is to note that if we enter state 2 or 3, then we can never return. The states 2 and 3 together can be regarded as an irreducible Markov process. So it must have a unique stationary distribution. In fact, this two-state process is basically the same as the one in the example on Page 29, so we know that $\frac{1}{5}(3, 2)$ is the stationary distribution by using the result obtained there. With respect to the full four-state process this distribution is $\frac{1}{5}(0, 0, 3, 2)$, since there is zero probability of being in states 0 or 1.

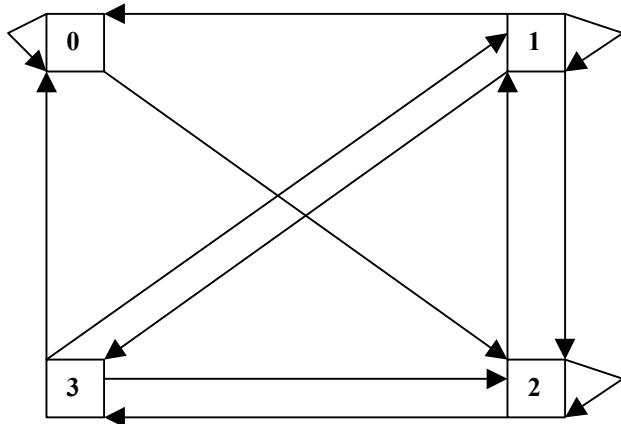
Since the four-state process is reducible, however, we cannot be sure (without checking) that there isn't another stationary distribution. This can be seen intuitively by noting that any other stationary distribution must involve state 0 or 1. But from either of these states we can get to either of the states 2 or 3, and we're then "trapped" in this "dead end" again, therefore there cannot be any other stationary distributions.

Solution 3.13

If the boundary is at state i then there can be no path from i to any other state. So a random walk with an absorbing boundary is not irreducible.

Solution 3.14

The process is irreducible, ie given any two states i and j there is a path from i to j . We can see this by inspection of the transition matrix or by drawing the transition graph.



Note that it is not necessary to include probabilities on the graph as we are only looking to see if there exists a path from i to j for any two states i and j .

Since the process is irreducible, it follows that there will be a unique stationary distribution.

We can also see that it is irreducible by calculating P^2 . It turns out that this has no zero entries, so that there is a non-zero probability of moving from any state, to any other. In addition, since it is P^2 we're talking about, these transitions must be possible in a maximum of 2 steps.

We wouldn't normally use this technique, however, as it's time consuming. In the general case, a process is irreducible if, for every pair of states i and j , there exists some power of P such that the (i, j) th entry is non-zero. This power might vary with i and j . It is much easier to trace out paths in the transition graph to see if there is a route from i to j . On the other hand, if you've already calculated P^2 say, then it is important to understand what it means to have non-zero entries (ie it means that it is irreducible).

Solution 3.15(i) ***Is process irreducible?***

We assume the states are labelled in direct correspondence with matrix indices, $\{1, 2, 3, 4, 5\}$.

This is not irreducible. Note that by the structure of the transition matrix:

- it is not possible to leave the states $\{1, 2\}$
- similarly it is not possible to leave the states $\{4, 5\}$
- it is not possible to leave the state 3.

(ii) ***Stationary distributions***

Each of these three subsets of states, $\{1, 2\}$, $\{3\}$ and $\{4, 5\}$ is effectively an irreducible Markov process on its own. Therefore each of these has a unique stationary distribution. In fact, using the results from the example on Page 29, we see that:

$$\frac{1}{5}(2, 3, 0, 0, 0) \text{ and } \frac{1}{5}(0, 0, 0, 3, 2)$$

are both stationary distributions.

The stationary distribution corresponding to state 3 will be:

$$(0, 0, 1, 0, 0)$$

Since these possible stationary distributions are all independent, we can have linear combinations of them. A general stationary distribution will therefore be of the form:

$$\frac{1}{5a+b+5c}(2a, 3a, b, 3c, 2c)$$

where a , b and c are arbitrary constants.

Alternatively we could solve the stationary condition $\pi P = \pi$ as before.

$$(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5) \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$$

This leads to the same result.

Solution 3.16

The transition matrix is given by $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. So for a stationary distribution we need to solve:

$$(\pi_0, \pi_1) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (\pi_0, \pi_1)$$

This is equivalent to $\pi_0 = \pi_1$. Together with the normalisation condition this gives $\pi_0 = \pi_1 = \frac{1}{2}$. This is unique, as you would expect for an irreducible Markov chain on a finite state space.

Solution 3.17(i) **Values**

The required values are:

$$\begin{aligned} \left(n_{1jk} \right) &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} & \left(n_{2jk} \right) &= \begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} & \left(n_{3jk} \right) &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ \left(n_{ij} \right) &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \\ 1 & 3 & 2 \end{pmatrix} & \left(n_i \right) &= \begin{pmatrix} 6 \\ 7 \\ 6 \end{pmatrix} \end{aligned}$$

n_{ijk} is based on 18 "triples" and n_{ij} is based on 19 "doubles".

(ii) **One-step transition probabilities**

The one-step transition probabilities are:

$$\left(p_{ij} \right) = \left(\frac{n_{ij}}{n_i} \right) = \begin{pmatrix} \frac{1}{6} & \frac{2}{6} & \frac{3}{6} \\ \frac{3}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{6} & \frac{3}{6} & \frac{2}{6} \end{pmatrix} = \begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ \frac{3}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$$

(iii) **Chi-squared test**

The test statistic is:

$$\chi^2 = \sum_{i,j,k} \frac{(n_{ijk} - n_{ij} \hat{p}_{jk})^2}{n_{ij} \hat{p}_{jk}} = 14.611$$

This has been derived using the spreadsheet shown below. We have calculated n_{ij} using the formula $n_{ij} = \sum_k n_{ijk}$. Note that this gives $n_{33} = 1$. The final 3,3 in the list of observations is not counted here because this is unable to give rise to an observation of the form 3,3,k.

There are 27 non-zero $n_{ij} \times n_{jk}$ values, 9 non-zero n_{ij} values and 3 non-zero n_i values. So the number of degrees of freedom is $r - 2q + s = 27 - 2 \times 9 + 3 = 12$.

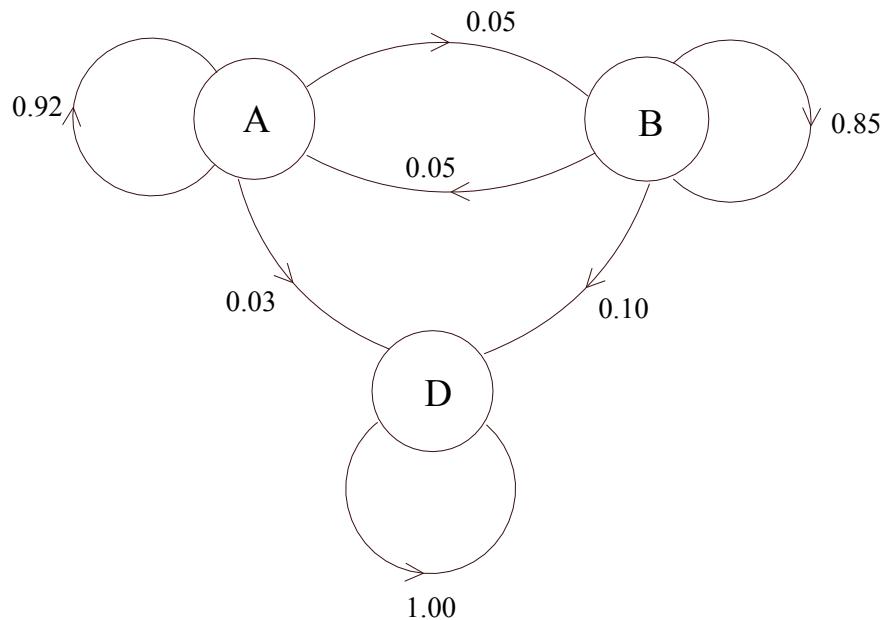
The critical value for χ^2_{12} is 21.03 at the 5% level, so there is no evidence to suggest the model is not a reasonable fit. We should be wary however, since the sample size is small. In fact the validity of the test is questionable if some of the expected frequencies are small, eg less than 5 and this is the case for all of the expected frequencies here.

ijk	n_j	n_{ij}	n_{jk}	\hat{p}_{jk}	$n_{ij} \hat{p}_{jk}$	n_{ijk}	$\frac{(n_{ijk} - n_{ij} \hat{p}_{jk})^2}{n_{ij} \hat{p}_{jk}}$
111	6	1	1	0.167	0.167	0	0.167
112	6	1	2	0.333	0.333	1	1.333
113	6	1	3	0.500	0.500	0	0.500
121	7	2	3	0.429	0.857	0	0.857
122	7	2	2	0.286	0.571	1	0.321
123	7	2	2	0.286	0.571	1	0.321
131	6	3	1	0.167	0.500	0	0.500
132	6	3	3	0.500	1.500	1	0.167
133	6	3	2	0.333	1.000	2	1.000
211	6	3	1	0.167	0.500	1	0.500
212	6	3	2	0.333	1.000	0	1.000
213	6	3	3	0.500	1.500	2	0.167
221	7	2	3	0.429	0.857	2	1.524
222	7	2	2	0.286	0.571	0	0.571
223	7	2	2	0.286	0.571	0	0.571
231	6	2	1	0.167	0.333	1	1.333
232	6	2	3	0.500	1.000	1	0.000
233	6	2	2	0.333	0.667	0	0.667
311	6	1	1	0.167	0.167	0	0.167
312	6	1	2	0.333	0.333	1	1.333
313	6	1	3	0.500	0.500	0	0.500
321	7	3	3	0.429	1.286	1	0.063
322	7	3	2	0.286	0.857	1	0.024
323	7	3	2	0.286	0.857	1	0.024
331	6	1	1	0.167	0.167	0	0.167
332	6	1	3	0.500	0.500	1	0.500
333	6	1	2	0.333	0.333	0	0.333
							14.611

Solution 3.18

- (i) **Probability A never rated B in the future**

We have the following transition diagram:



A company that is never rated B in the future will:

- (a) remain in State A for some period of time, and
- (b) then move to State D and remain there.

So we can sum over all future times at which the single transition from State A to State D can take place. This gives us the following expression:

$$0.03 + 0.92 \times 0.03 + (0.92)^2 \times 0.03 + (0.92)^3 \times 0.03 + \dots$$

This is an infinite geometric progression, whose sum is:

$$\frac{0.03}{1 - 0.92} = 0.375$$

So the probability that a company is never rated B in the future is 0.375.

(ii)(a) ***Second order transition probabilities***

The second order transition probabilities are given by:

$$\mathbf{X}^2 = \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.8489 & 0.0885 & 0.0626 \\ 0.0885 & 0.7250 & 0.1865 \\ 0 & 0 & 1 \end{pmatrix}$$

(ii)(b) ***Expected number of defaults***

The probability that a company rated A at time zero is in State D at time 2 is 0.0626. So the expected number of companies in this state out of 100 is 6.26.

(iii) ***Expected number of defaults***

For this manager we use the original matrix \mathbf{X} . After one year, the expected number of companies in each state will be:

$$(100 \ 0 \ 0) \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix} = (92 \ 5 \ 3)$$

If the five state B's are moved to State A and the process repeated, we have:

$$(97 \ 0 \ 3) \begin{pmatrix} 0.92 & 0.05 & 0.03 \\ 0.05 & 0.85 & 0.1 \\ 0 & 0 & 1 \end{pmatrix} = (89.24 \ 4.85 \ 5.91)$$

So the expected number of defaults by the end of the second year under this arrangement is 5.91.

(iv) ***Comment***

The downgrade trigger strategy will reduce the expected number of defaults, as we have seen. However, the return on the portfolio will also be a function of the yields on the debt. Companies rated B are likely to have bonds with a higher yield (because of the higher risk), so excluding these may in fact reduce the yield on the portfolio.

Also, the actual number of defaults may not match the expected number. The return depends on the actual progress of the portfolio, rather than the expected outcome.

There will also be a cost incurred when buying and selling bonds.

Chapter 4

The two-state Markov model



Syllabus objectives

- (v) Explain the concept of survival models.
- 8. Describe the two-state model of a single decrement and compare its assumptions with those of the random lifetime model (this model will be discussed in detail in Chapter 7).
- (vii) Derive maximum likelihood estimators for the transition intensities in models of transfers between states with piecewise constant transition intensities.
 - 1. Describe an observational plan in respect of a finite number of individuals observed during a finite period of time, and define the resulting statistics, including the waiting times.
 - 2. Derive the likelihood function for constant transition intensities in a Markov model of transfers between states given the statistics in (vii) 1.

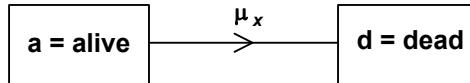
0 Introduction

In this chapter we consider a formulation of the problem in which we analyse the random process by which a life passes from one state (alive) to another (dead). The results are consistent with those that we obtain when we model a person's future lifetime as a continuous random variable. We will discuss this alternative model in Chapter 7.

The model discussed in this chapter is an example of a Markov jump process. These processes are discussed in Chapters 5 and 6. In this chapter we consider a simple two-state model. In Chapters 5 and 6 we will consider models with multiple states.

1 The two-state Markov model

The two-state model is illustrated in the figure below. There is an alive state and a dead state, with transitions in one direction only.



We define a transition probability ${}_t q_x$ where:

$${}_t q_x = P[\text{person in the dead state at age } x+t \mid \text{in the alive state at age } x]$$

and an occupancy or survival probability ${}_t p_x$ where:

$${}_t p_x = P[\text{person in the alive state at age } x+t \mid \text{in the alive state at age } x]$$

The probability that a life alive at a given age will be dead at any subsequent age is governed by the age-dependent transition intensity μ_{x+t} ($t \geq 0$), where:

$$\mu_{x+t} = \lim_{dt \rightarrow 0} \frac{dt q_{x+t}}{dt}$$

This is used again in Assumption 2 below.

Note that transition intensities are also sometimes called *forces of transition* or *transition rates*.

1.1 Assumptions underlying the model

There are three assumptions underlying the simple two-state model.



Assumption 1

The probabilities that a life at any given age will be found in either state at any subsequent age depend only on the ages involved and on the state currently occupied. This is the *Markov assumption*.

So, past events do not affect the probability of a future event.

In particular, the past history of an individual – for example, spells of sickness, occupation – is excluded from the model. If we knew these factors, we could:

- (a) treat each combination of factors as a separate model; in other words, *stratify* the problem; or
- (b) specify a model which took them into account; in other words, treat the problem as one of *regression*.

We will consider approach (b) in Chapter 9, *The Cox regression model*.



Assumption 2

For a short time interval of length dt :

$$dt q_{x+t} = \mu_{x+t} dt + o(dt) \quad (t \geq 0)$$

In other words, the probability of dying in a very short time interval dt is equal to the transition intensity multiplied by the time interval, plus a small correction term. This is equivalent to assuming that $dt q_{x+t} \approx \mu_{x+t} dt$.

Remember that a function $g(t)$ is said to be ‘ $o(t)$ ’ if $\lim_{t \rightarrow 0} \frac{g(t)}{t} = 0$, in other words if $g(t)$

tends to zero “faster” than t itself. Where we are not concerned about the precise form of $g(t)$, we can use the term $o(t)$ in an equation to denote any function that is $o(t)$.

For the purpose of inference, we restrict our attention to ages between x and $x+1$, and introduce a further assumption.



Assumption 3

μ_{x+t} is a constant μ for $0 \leq t < 1$.

Our investigation will consist of many observations of small segments of lifetimes, *ie* single years of age. Assumption 3 simplifies the model by treating the transition intensity as a constant for all individuals aged x last birthday. This does not mean that we believe that the transition intensity will increase by a discrete step when an individual reaches age $x+1$, although this is a consequence of the assumption.

**Question 4.1**

In the light of Assumption 3, how might we interpret μ for a particular year of age?

1.2 Comparison with other models

It is important to emphasise that this two-state model is not the same as the model based on the future lifetime random variable T_x , which is discussed in Chapter 7; we start with different assumptions. The model in Chapter 7 is formulated in terms of a random variable T representing future lifetime. The model in this chapter is in terms of a transition intensity between states. It is easy to impose some mild conditions under which the models are equivalent, but when we consider more than one decrement these two formulations lead in different directions.

We will consider models with more than one decrement in Chapters 5 and 6, including a simple multiple-state model with three states: healthy, sick and dead. In that particular model, a life in the healthy state can move to the sick state or the dead state. Similarly, a life in the sick state can move to the healthy state or the dead state.

For now, though, we will concentrate on the simple two-state model to derive some important results, many of which can be generalised to multiple-state models.

2 Probabilities

Since we have specified the model in terms of a transition intensity, we must see how to compute transition probabilities.

Consider the survival probability ${}_{t+dt}p_x$, and condition on the state occupied at age $x+t$, ie we consider separately the survival probabilities from age x to $x+t$ and from age $x+t$ to $x+t+dt$.

By the Markov assumption (Assumption 1), nothing else affects the probabilities of death or survival after age $x+t$.

$$\begin{aligned}
 {}_{t+dt}p_x &= {}_t p_x \times P[\text{Alive at } x+t+dt \mid \text{Alive at } x+t] \\
 &\quad + {}_t q_x \times P[\text{Alive at } x+t+dt \mid \text{Dead at } x+t] \\
 &= ({}_t p_x \times {}_{dt}p_{x+t}) + ({}_t q_x \times 0) \\
 &= {}_t p_x \times (1 - \mu_{x+t} dt + o(dt))
 \end{aligned}
 \tag{from Assumption 2}$$

Therefore:

$$\begin{aligned}
 \frac{\partial}{\partial t} {}_t p_x &= \lim_{dt \rightarrow 0^+} \frac{{}_{t+dt}p_x - {}_t p_x}{dt} \\
 &= -{}_t p_x \mu_{x+t} + \lim_{dt \rightarrow 0^+} \frac{o(dt)}{dt} \\
 &= -{}_t p_x \mu_{x+t}
 \end{aligned}
 \tag{4.1}$$

So:

$${}_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)$$



Question 4.2

Show that the solution of the differential equation $\frac{\partial}{\partial t} {}_t p_x = -{}_t p_x \mu_{x+t}$ is ${}_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)$.

A reminder of two techniques that can be used to solve first-order differential equations is given in Appendix 1 in Section 7 of this chapter.

Note that Assumption 3 has not been used so far.

We will get the same result in Chapter 7 when we formulate a survival model in terms of T , the lifetime distribution.

The important point is that it has been derived here strictly from the assumptions of the two-state model, and that the method is easily extended to models with more states. In the Markov framework, (4.1) is an example of the *Kolmogorov forward (differential) equations*. These are discussed in detail in Chapters 5 and 6.

3 Statistics

3.1 Definitions

Next we define our observations.

We suppose that we observe a total of N lives during some finite period of observation, between the ages of x and $x+1$.

We could suppose that lives were observed, or not, as a result of some random mechanism (not depending on any parameter of interest), but here we suppose that data are analysed retrospectively, so we regard N as a non-random quantity. We need not assume that we observe the N lives simultaneously, nor need we assume that we observe each life for the complete year of age. We do assume that all N lives are identical and statistically independent.

In reality no two lives are truly identical. Here we are using the word “identical” to refer to the fact that all the lives follow the same stochastic model of living and dying. So the lives will all have the same value of μ , but they won’t all die at the same time.



Definition

For $i = 1, \dots, N$ define:

$x + a_i$ to be the age at which observation of the i th life starts

$x + b_i$ to be the age at which observation of the i th life must cease if the life survives to that age.

$x + b_i$ will be either $x+1$, or the age of the i th life when the investigation ends, whichever is smaller.

For simplicity we consider Type I censoring.

This means that the value of b_i is known, when the period of observation starts at $x + a_i$. So b_i is a fixed number, and not a random variable. If we plan to observe a life from 52.25 until 52.75, then $b_i = 0.75$, but of course not all lives will survive to the end of the planned period of observation. To complete our model we will need another random variable that measures whether it was death before 52.75 or survival to 52.75, that ended the period of observation).

The approach can be extended to more realistic forms of censoring.

In other words, we could modify the derivation to allow for lives leaving the investigation at random times through decrements other than death.

In this case b_i would be a random variable. If we plan to observe a life from 52.25 until death or retirement, whichever event occurs first, then b_i is a random variable (and to complete our model we will need another random variable that measures whether it was death or retirement that ended the period of observation).



Definition

Define a random variable D_i as follows:

$$D_i = \begin{cases} 1 & \text{if the } i\text{th life is observed to die} \\ 0 & \text{if the } i\text{th life is not observed to die} \end{cases}$$

D_i is an example of an indicator random variable; it indicates the occurrence of death.

In the above definition, we are talking about whether or not the i th life dies during the observation period. D_i is the extra random variable that completes our model.



Question 4.3

In its simplest form, what is $E[D_i]$?



Definition

Define a random variable T_i as follows:

$$x + T_i = \text{the age at which observation of the } i\text{th life ends}$$

Notice that D_i and T_i are not independent, since:

$$D_i = 0 \Leftrightarrow T_i = b_i$$

i.e if no death has been observed, the life must have survived to $x + b_i$.

$$D_i = 1 \Leftrightarrow a_i < T_i < b_i$$

i.e an observed death must have occurred between $x + a_i$ and $x + b_i$.



Definition

It will often be useful to work with the time spent under observation, so define:

$$V_i = T_i - a_i$$

V_i is called the *waiting time*. It has a mixed distribution, with a probability mass at the point $b_i - a_i$.



Question 4.4

What is a mixed distribution?



Example

Observation of life i begins at exact age 82 years and 3 months. Observation will continue until the earlier of the life's 83rd birthday or death.

State the value or range of values taken by:

(i) a_i

(ii) b_i

(iii) V_i

Solution

(i) $a_i = 0.25$

(ii) $b_i = 1$

(iii) V_i is a random variable taking values between 0 and 0.75. V_i has a mixed distribution with a probability mass at 0.75.

3.2 Joint density function

The pair (D_i, V_i) comprise a **statistic**, meaning that the outcome of our observation is a sample (d_i, v_i) drawn from the distribution of (D_i, V_i) .

Let $f_i(d_i, v_i)$ be the joint distribution of (D_i, V_i) .

It is easily written down by considering the two cases $D_i = 0$ and $D_i = 1$.

If $D_i = 0$, no death has been observed and the life is known to have survived for the period of $(b_i - a_i)$ from $x + a_i$ to $x + b_i$.

If $D_i = 1$, the life is known to have survived for the period v_i ($0 < v_i < b_i - a_i$) from $x + a_i$ to $x + a_i + v_i$ before dying at age $x + a_i + v_i$.

Therefore, $f_i(d_i, v_i)$ has a distribution that is specified by the following expression, which is a combination of a probability mass (corresponding to $d_i = 0$) and a probability density (corresponding to $d_i = 1$).

$$f_i(d_i, v_i) = \begin{cases} b_i - a_i p_{x+a_i} & (d_i = 0) \\ v_i p_{x+a_i} \cdot \mu_{x+a_i+v_i} & (d_i = 1) \end{cases}$$

$$\begin{aligned} &= \begin{cases} \exp\left(-\int_0^{b_i - a_i} \mu_{x+a_i+t} dt\right) & (d_i = 0) \\ \exp\left(-\int_0^{v_i} \mu_{x+a_i+t} dt\right) \mu_{x+a_i+v_i} & (d_i = 1) \end{cases} \\ &= \exp\left(-\int_0^{v_i} \mu_{x+a_i+t} dt\right) \mu_{x+a_i+v_i}^{d_i} \end{aligned}$$

Now assume that μ_{x+t} is a constant μ for $0 \leq t < 1$ (this is the first time we have needed Assumption 3) and so $f_i(d_i, v_i)$ takes on the simple form:

$$f_i(d_i, v_i) = e^{-\mu v_i} \mu^{d_i}$$

We can then write down an expression for the joint probability function, provided that we can assume that the lifetimes of all the lives involved are statistically independent.

The joint probability function of all the (D_i, V_i) , by independence, is proportional to:

$$\begin{aligned} \prod_{i=1}^N e^{-\mu v_i} \mu^{d_i} &= e^{-\mu(v_1 + \dots + v_N)} \mu^{d_1 + \dots + d_N} \\ &= e^{-\mu v} \mu^d \end{aligned}$$

$$\text{where } d = \sum_{i=1}^N d_i \text{ and } v = \sum_{i=1}^N v_i .$$

In other words, define random variables D and V to be the total number of deaths and the total waiting time, respectively, and the joint probability function of all the (D_i, V_i) can be simply expressed in terms of D and V .

For a known transition intensity, we can calculate the likelihood of any combination of deaths and waiting time. However, in practice the value of the transition intensity is unknown. We use statistical inference to calculate the value of the transition intensity that is most plausible given the observed data, *ie* the maximum likelihood estimate of μ . This is the subject of the next section.



Question 4.5

Let the transition intensity μ equal 0.1. Using the information in the example on Page 10 calculate:

- (i) the probability function of D_i
- (ii) $E[D_i]$
- (iii) the probability density/mass function of V_i
- (iv) $E[V_i]$.

Check your answer to (iv) by applying a reasonableness test.

4 The maximum likelihood estimator

A reminder of maximum likelihood estimation is given in Appendix 2, which you'll find in Section 7 at the end of this chapter. You may wish to read that before continuing.

4.1 Maximising the likelihood function

The probability function immediately furnishes the likelihood for μ :

$$L(\mu; d, v) = e^{-\mu v} \mu^d$$

which yields the maximum likelihood estimate (MLE) for μ :

$$\hat{\mu} = d / v$$



Question 4.6

Prove that the likelihood is maximised by $\hat{\mu} = d / v$.

It is reassuring that the mathematical approach produces a result that is intuitive, ie that the maximum likelihood estimate of the hazard rate is the number of observed deaths divided by the total time for which lives were exposed to the hazard.

The measurement of the total time for which lives are exposed to the hazard is one of the fundamental techniques covered by this course. It enables accurate assessment of risks, from the probability of a policyholder dying to the probability of a claim under a motor insurance policy. The technical term is “exposed to risk” – we will study it in more detail in Chapter 11.

4.2 Properties of the maximum likelihood estimator

The estimate $\hat{\mu}$, being a function of the sample values d and v , can itself be regarded as a sample value drawn from the distribution of the corresponding estimator:

$$\tilde{\mu} = D / V$$

As usual we are using capital letters to denote random variables, and lower case letters to denote sample values.

So, the estimator $\tilde{\mu}$ is a random variable and the maximum likelihood estimate $\hat{\mu}$ is the observed value of that random variable.

It is important in applications to be able to estimate the moments of the estimator $\tilde{\mu}$, for example to compare the experience with that of a standard table. At least, we need to estimate $E[\tilde{\mu}]$ and $\text{var}[\tilde{\mu}]$.

In order to derive the properties of the estimator $\tilde{\mu}$ we will use two results that link the random variables D and V .

The following exact results are obtained:

$$E[D_i - \mu V_i] = 0 \quad (4.2)$$

$$\text{var}[D_i - \mu V_i] = E[D_i] \quad (4.3)$$

Note that the first of these can also be written as $E[D_i] = \mu E[V_i]$.

In the case that the $\{a_i\}$ and $\{b_i\}$ are known constants, this follows from integrating/summing the probability function of (D_i, V_i) over all possible events to obtain:

$$\int_0^{b_i-a_i} e^{-\mu v_i} \mu d v_i + e^{-\mu(b_i-a_i)} = 1 \quad (*)$$

and then differentiating with respect to μ , once to obtain the mean and twice to obtain the variance.

We will show how to use this to prove Result (4.2) above in a moment, but first we need to derive two other results. (The derivation of Result (4.3) is covered in the Question & Answer Bank.)



Question 4.7

Show that $E[D_i] = \int_0^{b_i-a_i} e^{-\mu t} \cdot \mu dt$

**Question 4.8**

Show that $E[V_i] = \int_0^{b_i - a_i} t \cdot e^{-\mu t} \cdot \mu \cdot dt + (b_i - a_i) e^{-\mu(b_i - a_i)}$

Proof of (4.2)

Differentiating (*) with respect to μ gives:

$$\int_0^{b_i - a_i} e^{-\mu v_i} \cdot \mu \cdot dv_i - \mu \int_0^{b_i - a_i} v_i \cdot e^{-\mu v_i} \cdot \mu \cdot dv_i - (b_i - a_i) e^{-\mu(b_i - a_i)} = 0$$

(Because the limits of the integrals don't depend on μ , this just involves differentiating the expressions inside the integral with respect to μ .)

Multiplying throughout by μ then gives:

$$\int_0^{b_i - a_i} e^{-\mu v_i} \cdot \mu \cdot \mu \cdot dv_i - \mu \left\{ \int_0^{b_i - a_i} v_i \cdot e^{-\mu v_i} \cdot \mu \cdot \mu \cdot dv_i + (b_i - a_i) e^{-\mu(b_i - a_i)} \right\} = 0$$

Using the results of Question 4.7 and Question 4.8, we can see that the first term is $E[D_i]$ and the expression in curly brackets is $E[V_i]$.

So $E[D_i] - \mu \cdot E[V_i] = 0$ as required.

4.3 The distribution of $\tilde{\mu}$

To find the asymptotic distribution of $\tilde{\mu}$, consider:

$$\frac{1}{N}(D - \mu V) = \frac{1}{N} \sum_{i=1}^N (D_i - \mu V_i)$$

We know that $E[D_i - \mu V_i] = 0$ and that $\text{var}[D_i - \mu V_i] = E[D_i]$.

So, by the Central Limit Theorem:

$$\frac{1}{N}(D - \mu V) \sim \text{Normal}\left(0, \frac{E[D]}{N^2}\right)$$

Recall that $\tilde{\mu} = \frac{D}{V}$. So:

$$\begin{aligned}\tilde{\mu} - \mu &= \frac{D}{V} - \mu \\ &= \frac{D - \mu V}{V} \\ &= \frac{N}{V} \left(\frac{D - \mu V}{N} \right)\end{aligned}$$

Then note that (not rigorously):

$$\lim_{N \rightarrow \infty} (\tilde{\mu} - \mu) = \lim_{N \rightarrow \infty} \frac{N}{V} \left(\frac{D}{N} - \frac{\mu V}{N} \right)$$

By the law of large numbers, $V/N \rightarrow E(V_i)$. Technically, this refers to “convergence in probability”. So asymptotically:

$$E(\tilde{\mu} - \mu) = \frac{1}{E(V_i)} E\left(\frac{D - \mu V}{N}\right) = 0$$

Also:

$$\begin{aligned}\text{var}(\tilde{\mu} - \mu) &= \text{var}\left[\left(\frac{D - \mu V}{N}\right) \times \frac{1}{E(V_i)}\right] \\ &= \frac{E(D)}{N^2 [E(V_i)]^2} \\ &= \frac{E[D]}{[E(V_1 + V_2 + \dots + V_N)]^2}\end{aligned}$$

because $E(V_1) = E(V_2) = \dots = E(V_N)$. Now since $V = \sum_{i=1}^N V_i$, we have:

$$\text{var}(\tilde{\mu} - \mu) = \frac{E[D]}{[E(V)]^2}$$

So:

$$(\tilde{\mu} - \mu) \sim \text{Normal}\left(0, \frac{E(D)}{[E(V)]^2}\right)$$

But we know that $E(D - \mu V) = 0$. So:

$$E(D) = \mu E(V)$$

and:

$$\mu = \frac{E(D)}{E(V)}$$

So asymptotically:

$$\tilde{\mu} \sim \text{Normal}\left(\mu, \frac{\mu}{E[V]}\right)$$

**Question 4.9**

A scientist identifies 1,282 newborn wildebeest and observes them during their first year of life on the savannah. The scientist wishes to calculate the constant transition intensity over this period covering all types of death, including natural causes and ending up as a tasty snack for passing carnivores.

If the true transition intensity is 0.18, what is the probability that the scientist observes a hazard rate in excess of 0.2?

5 Alternative approach

In this section we describe another way of obtaining the asymptotic distribution of $\tilde{\mu}$, the maximum likelihood estimator of μ .

We start as before from the likelihood function:

$$L = \mu^d e^{-\mu v}$$

The log-likelihood is then:

$$\log L = d \log \mu - \mu v$$

and differentiating this with respect to μ gives:

$$\frac{d \log L}{d \mu} = \frac{d}{\mu} - v$$

Setting this equal to 0 and solving for μ yields the required maximum likelihood estimate:

$$\hat{\mu} = \frac{d}{v}$$

We can check that this does maximise the likelihood, by examining the sign of the second derivative of the log-likelihood:

$$\frac{d^2 \log L}{d \mu^2} = -\frac{d}{\mu^2} < 0 \Rightarrow \text{max}$$

The corresponding maximum likelihood estimator is:

$$\tilde{\mu} = \frac{D}{V}$$

where D and V are random variables denoting the number of deaths and the total waiting time, respectively.

Now instead of deriving results for the expectation and variance of $D_i - \mu V_i$ as in Section 4.2, we can appeal to the properties of maximum likelihood estimators.

From Subject CT3, we know that maximum likelihood estimators are asymptotically normally distributed and are asymptotically unbiased. So $E(\tilde{\mu}) = \mu$. It just remains for us to find an expression for $\text{var}(\tilde{\mu})$. This is given by the Cramér-Rao lower bound:

$$\text{var}(\tilde{\mu}) = \frac{-1}{E\left(\frac{d^2 \log L}{d\mu^2}\right)} = \frac{-1}{E\left(\frac{-D}{\mu^2}\right)} = \frac{\mu^2}{E(D)}$$

So:

$$\tilde{\mu} \sim N\left(\mu, \frac{\mu^2}{E(D)}\right) \quad \text{asymptotically}$$

Note that this is consistent with the result in Section 4 since $E(D) = \mu E(V)$.

However, in practice, we will not know the exact variance, so we need to estimate it. This can be done by replacing μ by $\hat{\mu}$, and by replacing $E(D)$ by the observed number of deaths, d . That is:

$$\text{var}(\tilde{\mu}) \approx \frac{\hat{\mu}^2}{d}$$



Question 4.10

Show that this is equivalent to the approximation:

$$\text{var}(\tilde{\mu}) \approx \frac{\hat{\mu}}{v}$$

where v is the total observed waiting time.

6 Further comments

6.1 Applications of μ_x

Having estimated piecewise constant intensities over single years of age, we can use these (if required) to estimate the function μ_x as a smooth function of age (the process of smoothing is called graduation). For this purpose we usually assume that $\hat{\mu}$ estimates $\mu_{x+1/2}$.

This is a very important assumption, which we will use frequently throughout the course.

We can calculate any required probabilities from:

$${}_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)$$

using numerical methods if necessary.

The process of graduation is covered in Chapter 12, *Graduation and statistical tests*, and Chapter 13, *Methods of graduation*.

6.2 The “central exposed to risk”

In actuarial terminology, the observed waiting time at age x , which we have denoted v , is often called the **central exposed to risk** and is denoted E_x^c .

We consider this concept in more detail in Part 4 of the course. In the meantime you should be prepared to use either term.



Question 4.11

What are the three assumptions underlying the simple two-state model?

7 Appendices

7.1 Appendix 1 – solving first-order differential equations

In this appendix, we remind you of two methods that can be used to solve first-order differential equations. These are the separation method and the integrating factor method.

The separation method

The separation method can be used to solve equations of the form:

$$\frac{dy}{dx} = g(x)h(y)$$

where $g(x)$ is a function of x and $h(y)$ is a function of y . The variables are separated by rewriting the equation as:

$$\frac{dy}{h(y)} = g(x)dx$$

Each side is then integrated to obtain the solution. If we are given an initial condition or a boundary condition, this can be used to determine the value of the constant of integration.



Example

Solve the differential equation $\frac{dy}{dx} = (x+1)y$ subject to the initial condition $y(0) = 2$.

Solution

Separating the variables we obtain:

$$\frac{dy}{y} = (x+1)dx$$

Then integrating both sides gives:

$$\ln|y| = \frac{1}{2}x^2 + x + C$$

Note that C denotes a constant of integration.

Taking exponentials, this becomes:

$$y = e^{\frac{1}{2}x^2 + x + C} = Ae^{\frac{1}{2}x^2 + x}$$

where $A = e^C$.

Finally, using the initial condition:

$$y(0) = 2 \Rightarrow Ae^0 = 2 \Rightarrow A = 2 \Rightarrow y(x) = 2e^{\frac{1}{2}x^2 + x}$$

The integrating factor method

The integrating factor method can be used to solve equations of the form:

$$\frac{dy}{dx} + P(x)y = Q(x) \quad (*)$$

where $P(x)$ and $Q(x)$ are both functions of x . In the context of this course, y will usually denote some probability.

The first step is to calculate the integrating factor (IF):

$$IF = e^{\int P(x)dx}$$

Then multiply each term in $(*)$ by the integrating factor:

$$\frac{dy}{dx}e^{\int P(x)dx} + P(x)e^{\int P(x)dx}y = Q(x)e^{\int P(x)dx} \quad (**)$$

Now integrate both sides of $(**)$ with respect to x . The left-hand side will be:

$$y e^{\int P(x)dx} = y \times IF$$

(You can check this by applying the product rule for differentiation to the product $y \times IF$.) Finally, we divide through by IF to obtain an expression for y .

Example

Solve the differential equation $x \frac{dy}{dx} = 2x - (x+1)y$ subject to the condition $y(1) = 0$.

Solution

We first write the differential equation in the form $\frac{dy}{dx} + P(x)y = Q(x)$:

$$\frac{dy}{dx} + \left(\frac{x+1}{x}\right)y = 2$$

The integrating factor is given by:

$$\exp\left[\int\left(\frac{x+1}{x}\right)dx\right] = \exp\left[\int\left(1 + \frac{1}{x}\right)dx\right] = \exp[x + \ln x] = e^x e^{\ln x} = xe^x$$

NB: We don't have to bother about the constant of integration here. The constants will cancel out when we multiply every term by the integrating factor.

Multiplying both sides of the differential equation by the integrating factor gives:

$$\frac{dy}{dx} xe^x + \left(1 + \frac{1}{x}\right) xe^x y = 2xe^x$$

Integrating the left-hand side with respect to x gives:

$$y \times IF = yxe^x$$

Integrating the right-hand side with respect to x (using integration by parts), we obtain:

$$\int 2xe^x dx = 2xe^x - \int 2e^x dx = 2xe^x - 2e^x + C$$

Equating these gives:

$$\begin{aligned} yxe^x &= 2xe^x - 2e^x + C \\ \Rightarrow y &= 2 - \frac{2}{x} + \frac{C}{xe^x} \end{aligned}$$

Finally, from the condition $y(1) = 0$, we have:

$$0 = 2 - 2 + \frac{C}{e} \Rightarrow C = 0$$

So the required solution is:

$$y(x) = 2 - \frac{2}{x}$$

7.2 Appendix 2 – likelihoods

The models in Part 3 of the course are all fitted using the method of maximum likelihood. The first step in this method is to write down the likelihood function. You can think of the likelihood as being “the probability of getting the data that you’ve observed”.

Discrete case

If we observe n realisations of a *discrete* random variable X , and we let X_i denote the i th outcome, then the likelihood is given by:

$$L = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

where x_1, x_2, \dots, x_n are the observed values of the random variables X_1, X_2, \dots, X_n .

If X_1, X_2, \dots, X_n are independent (as is the usual assumption), then the likelihood can be written as:

$$L = P(X_1 = x_1) \times P(X_2 = x_2) \times \cdots \times P(X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$$



Example

Suppose that X is a Poisson random variable with mean λ . A single observation from this distribution is recorded. The observed value is 18. Write down an expression for the likelihood function.

Solution

The likelihood is given by:

$$L = P(X = 18) = \frac{e^{-\lambda} \lambda^{18}}{18!}$$



Example

Suppose that X is a Poisson random variable with mean λ . Two independent observations from this distribution are recorded. The observed values are 18 and 23. Write down an expression for the likelihood function.

Solution

The likelihood is given by:

$$L = P(X_1 = 18) \times P(X_2 = 23) = \frac{e^{-\lambda} \lambda^{18}}{18!} \times \frac{e^{-\lambda} \lambda^{23}}{23!} = C e^{-2\lambda} \lambda^{41}$$

where C is a constant, ie C is independent of λ .

Continuous case

If we observe n realisations of a *continuous* random variable X and we let X_i denote the i th outcome, then the likelihood is given by the joint PDF:

$$L = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$$

where x_1, x_2, \dots, x_n are the observed values of the random variables X_1, X_2, \dots, X_n .

If X_1, X_2, \dots, X_n are independent, then the likelihood can be written as:

$$L = f_{X_1}(x_1) \times f_{X_2}(x_2) \times \cdots \times f_{X_n}(x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$



Example

Suppose that X is an exponential random variable with parameter λ . A single observation from this distribution is recorded. The observed value is 18. Write down an expression for the likelihood function.

Solution

The likelihood is given by:

$$L = f_X(18) = \lambda e^{-18\lambda}$$



Example

Suppose that X is an exponential random variable with parameter λ . Two independent observations from this distribution are recorded. The observed values are 18 and 23. Write down an expression for the likelihood function.

Solution

The likelihood is given by:

$$L = f_{X_1}(18) \times f_{X_2}(23) = \lambda e^{-18\lambda} \times \lambda e^{-23\lambda} = \lambda^2 e^{-41\lambda}$$

Maximum likelihood estimation

Recall from Subject CT3, that to determine the maximum likelihood estimates of a set of parameters, you:

- write down the likelihood function L
- take logs and simplify the resulting expression
- partially differentiate the log-likelihood with respect to each parameter to be estimated
- set the derivatives equal to 0 and solve the equations simultaneously
- check that the resulting values of the parameters give you a maximum.

8 ***End of Part 1***

You have now completed Part 1 of the Subject CT4 Notes.

Review

Before looking at the Question and Answer Bank we recommend that you briefly review the key areas of Part 1, or maybe re-read the summaries at the end of Chapters 1 to 4.

Question and Answer Bank

You should now be able to answer the questions in Part 1 of the Question and Answer Bank. We recommend that you work through several of these questions now and save the remainder for use as part of your revision.

Assignments

On completing this part, you should be able to attempt the questions in Assignment X1.

Reminder

If you have not yet booked a tutorial, then maybe now is the time to do so.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.



Chapter 4 Summary

We can model mortality as a Markov process with two states (*alive* and *dead*) and a transition intensity μ_x .

Assumptions underlying the two-state model

From this simple model we can derive the following formula for the survival probability:

$${}_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)$$

This is consistent with the result that will be obtained in Chapter 7 (where future lifetime will be modelled as a continuous random variable).

The *waiting time* for a life is the time spent under observation. The observed waiting time is often called the *central exposed to risk*.

We can use the observed total waiting time and the observed number of deaths to estimate the underlying transition intensity. The estimation can be done using the method of maximum likelihood. To proceed, we have to consider the “probability” of getting the results we have observed from our mortality investigation.

Joint distribution of an observed sample (d_i, v_i)

$$f_i(d_i, v_i) = e^{-\mu v_i} \mu^{d_i}$$

Joint distribution of all observed samples from (D_i, V_i)

$$f\{(d_1, v_1), (d_2, v_2), \dots, (d_N, v_N)\} = e^{-\mu v} \mu^d$$

$$\text{where } d = \sum_{i=1}^N d_i \text{ and } v = \sum_{i=1}^N v_i .$$

The maximum likelihood estimate

$$\hat{\mu} = d / v$$

The maximum likelihood estimator

$$\tilde{\mu} = D / V$$

Asymptotically, $\tilde{\mu} \sim N\left(\mu, \frac{\mu}{E[V]}\right)$ or, equivalently, $\tilde{\mu} \sim N\left(\mu, \frac{\mu^2}{E[D]}\right)$.

We usually assume that the estimated transition intensity $\hat{\mu}$ estimates $\mu_{x+\frac{1}{2}}$.

Chapter 4 Solutions

Solution 4.1

In reality we would expect μ_{x+t} to vary for $0 \leq t < 1$, ie we would expect μ to be a smooth function rather than a step function. In this case, we might interpret the observed constant μ to be an estimate of $\mu_{x+1/2}$.

Solution 4.2

Separating the variables gives:

$$\frac{\frac{\partial}{\partial t} {}_t p_x}{{}_t p_x} = -\mu_{x+t}$$

But:

$$\frac{\frac{\partial}{\partial t} {}_t p_x}{{}_t p_x} = \frac{\partial}{\partial t} \ln {}_t p_x$$

So:

$$\frac{\partial}{\partial t} \ln {}_t p_x = -\mu_{x+t}$$

Integrating both sides of this equation with respect to t between the limits $t = 0$ and $t = n$ gives:

$$\int_0^n \frac{\partial}{\partial t} \ln {}_t p_x dt = [\ln {}_t p_x]_0^n = - \int_0^n \mu_{x+t} dt$$

or equivalently:

$$\ln {}_n p_x - \ln {}_0 p_x = - \int_0^n \mu_{x+t} dt$$

Now ${}_0 p_x = 1$, since this is the probability that a life aged x survives for at least 0 years, and $\ln {}_0 p_x = 0$.

Taking exponentials then gives:

$${}_n p_x = \exp \left[- \int_0^n \mu_{x+t} dt \right]$$

or equivalently:

$${}_t p_x = \exp \left[- \int_0^t \mu_{x+s} ds \right]$$

Solution 4.3

$$E[D_i] = 0 \times P[D_i = 0] + 1 \times P[D_i = 1] = P[D_i = 1]$$

i.e. $E[D_i]$ is simply the probability of a death being observed.

Solution 4.4

A mixed distribution is a mixture of a discrete distribution and a continuous distribution.

Solution 4.5

We will need to use the result from Section 2 that:

$${}_t p_x = \exp \left(- \int_0^t \mu_{x+s} ds \right) = \exp(-\mu \times t) = \exp(-0.1 \times t)$$

$$(i) \quad P[D_i = 0] = {}_{0.75} p_{82.25} = \exp(-0.1 \times 0.75) = 0.9277$$

$$P[D_i = 1] = 1 - 0.9277 = 0.0723$$

$$(ii) \quad E[D_i] = 0 \times 0.9277 + 1 \times 0.0723 = 0.0723$$

(iii) The probability density/mass function of V_i is:

$$\begin{aligned} f(v_i) &= \begin{cases} v_i p_{82.25} \cdot \mu_{82.25+v_i} & v_i < 0.75 \\ 0.75 p_{82.25} & v_i = 0.75 \end{cases} \\ &= \begin{cases} \exp(-0.1 \times v_i) \times 0.1 & v_i < 0.75 \\ \exp(-0.1 \times 0.75) & v_i = 0.75 \end{cases} \\ &= \begin{cases} 0.1 \exp(-0.1v_i) & v_i < 0.75 \\ 0.9277 & v_i = 0.75 \end{cases} \end{aligned}$$

(iv) $E[V_i] = \int_0^{0.75} t \cdot \exp(-0.1 \times t) \times 0.1 dt + 0.9277 \times 0.75$

Integrating by parts, the integral becomes:

$$\begin{aligned} &[-t \cdot \exp(-0.1 \times t)]_0^{0.75} - \int_0^{0.75} -\exp(-0.1 \times t) dt \\ &= -0.75 \times \exp(-0.075) - 10 [\exp(-0.1 \times t)]_0^{0.75} \\ &= -0.6958 - 10 \times [\exp(-0.075) - 1] \\ &= -0.6958 + 0.7226 = 0.0268 \end{aligned}$$

$$E[V_i] = 0.0268 + 0.9277 \times 0.75 = 0.7226$$

As expected, the answer to (iv) is just less than 0.75.

Solution 4.6

The likelihood is $L(\mu; d, v) = e^{-\mu v} \mu^d$. So $\log L = -\mu v + d \log \mu$.

Differentiating, $\frac{\partial}{\partial \mu} \log L = -v + \frac{d}{\mu}$.

Setting $-v + \frac{d}{\mu} = 0 \Rightarrow \hat{\mu} = \frac{d}{v}$, which is a maximum since $\frac{\partial^2}{\partial \mu^2} \log L = -\frac{d}{\mu^2} < 0$.

Solution 4.7

Since $D_i = 0$ if the life is not observed to die, we only need to consider the probability

of death occurring ($D_i = 1$), ie $E[D_i] = 1 \times \int_0^{b_i - a_i} {}_t p_{x+a_i} \mu_{x+a_i+t} dt$.

We are assuming a constant transition intensity μ , so ${}_t p_{x+a_i} = e^{-\mu t}$.

Hence $E[D_i] = \int_0^{b_i - a_i} e^{-\mu t} \mu dt$.

Solution 4.8

$E[V_i] = \int_0^{b_i - a_i} P[\text{life dies at time } t] \times t dt + (b_i - a_i) \times P[\text{life survives}]$

Hence $E[V_i] = \int_0^{b_i - a_i} t \cdot e^{-\mu t} \cdot \mu dt + (b_i - a_i) e^{-\mu(b_i - a_i)}$.

Solution 4.9

The expected waiting time for the i th animal, assuming a hazard rate of 0.18, is:

$$\begin{aligned} E[V_i] &= \int_0^1 t \cdot \exp(-0.18 \times t) \times 0.18 dt + \exp(-0.18) \times 1 \\ &= -\exp(-0.18) + \frac{1}{0.18} [1 - \exp(-0.18)] + \exp(-0.18) = 0.915165 \end{aligned}$$

The total expected waiting time is:

$$E[V] = 1,282 \times E[V_i] = 1,173.2$$

$$\text{So: } \tilde{\mu} \sim N\left(\mu, \frac{\mu}{E[V]}\right) \equiv N\left(0.18, \frac{0.18}{1173.2}\right) \equiv N\left(0.18, (0.01239)^2\right)$$

$$\text{And } P[\tilde{\mu} > 0.2] = 1 - \Phi\left(\frac{0.2 - 0.18}{0.01239}\right) = 1 - \Phi(1.61) = 1 - 0.9463 = 0.0537.$$

Solution 4.10

We have already seen that $\hat{\mu} = \frac{d}{v}$. So:

$$\frac{\hat{\mu}^2}{d} = \frac{\hat{\mu}}{d} \times \frac{d}{v} = \frac{\hat{\mu}}{v}$$

Solution 4.11

- (1) The probabilities that a life at any given age will be found in either state at any subsequent age depend only on the ages involved and on the state currently occupied. (This is the Markov property.)
- (2) $dt q_{x+t} = \mu_{x+t} dt + o(dt) \quad (t \geq 0)$
- (3) For each integer x , μ_{x+t} takes a constant value μ for $0 \leq t < 1$.

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Chapter 5

Time-homogeneous Markov jump processes



Syllabus objectives

- (iv) Define and apply a Markov process.
1. State the essential features of a Markov process model.
 2. Define a Poisson process, derive the distribution of the number of events in a given time interval, derive the distribution of inter-event times, and apply these results.
 3. Derive the Kolmogorov equations for a Markov process with time independent transition intensities.
 4. Solve the Kolmogorov equations in simple cases.
- (vii) Derive maximum likelihood estimators for the transition intensities in models of transfers between states with piecewise constant transition intensities.
2. Derive the likelihood function for constant transition intensities in a Markov model of transfers between states given the statistics in (vii) 1.
 3. Derive maximum likelihood estimators for the transition intensities in (vii) 2 and state their asymptotic distribution.

0 Introduction

In Chapter 4 we considered a simple two-state Markov model. In this chapter we will show how the model and the results can be extended to any number of states.

One important aspect of the simple two-state model was that transition was possible in one direction only, from *alive* to *dead*. In practice, we may wish to study a model in which transition between states is possible in both directions. This opens up the possibility of a life entering a particular state more than once.

An example of this is a model in which the states relate to marital status. Such a model might comprise five states – *single*, *married*, *divorced*, *widowed* and *dead*.



Question 5.1

Draw this model, showing clearly the possible transitions between the five states.

Some of the questions that we will consider in this chapter are:

- What is the probability of a life remaining in a particular state for a period of length t if there is more than one possible mode of decrement from that state?
- How can we use our observations to estimate the transition intensities?
- How can we calculate the probability of a particular decrement occurring based on our estimates of the transition intensities?

The results that we derive here form the basic building blocks of several actuarial techniques. Some financial applications of this theory are discussed in Subject CT5.

Much of the theory is analogous to that for Markov chains. The Chapman-Kolmogorov equations can be written in the same format for example. However, in discrete time we have the central notion of the one-step transition probabilities. In the continuous case there is no longer the same fundamental significance to the unit time interval, as we can consider time intervals of arbitrarily small length h . As is often the case with continuous variables, the natural thing to do is to consider limits as $h \rightarrow 0$. This leads to a reformulation of the Chapman-Kolmogorov equations as *differential equations*. Much of our time will be spent constructing and interpreting such differential equations, along with their *integral equation* analogues.

These differential and integral equations can be solved to give results for the transition probabilities in terms of the transition (probability) rates. All the versions of the equations will have the same solution for a particular model. For some models, one of the equations may be more straightforward to solve than the others. In the exam you will usually be guided towards a particular equation, rather than having to choose which of the four equations to use.

In this chapter we consider only time-homogeneous Markov jump process. These are processes in which the transition rates do not vary over time, so the transition probabilities $P(X_t = j | X_s = i)$ depend only on the length of the time interval, $t - s$.

1 Notation

Different authors tend to use different notation for the same quantities, and the Markov model is an example of this. Actuaries have often used notations derived from the standardised International Actuarial Notation, in which the “transition rate” is the force of mortality μ_x , and the corresponding probabilities are the life table probabilities $_t p_x$ and $_t q_x$. Moreover, the index x is generally understood to indicate age (eg age now, or age at policy inception) and the index t indicates duration since age x . Probabilists and statisticians tend to use different notations.

We met the notation $_t p_x$ and $_t q_x$ in Chapter 4. Recall that $_t p_x$ denotes the probability that a life aged x survives for at least another t years, and $_t q_x = 1 - _t p_x$ is the probability that a life aged x dies within the next t years.

The non-homogeneous (ie time-inhomogeneous) Markov model offers particularly rich, and potentially confusing, opportunities to invent different notations for the same quantities. To try to limit any such confusion, we make the following remarks.

1. We have written $p_{ij}(s,t)$ to mean the probability of the process being in state j at time t , conditional on being in state i at time $s \leq t$.

The traditional actuarial notation would reserve the symbol t for duration since time s , in which case the above probability would be expressed $p_{ij}(s,s+t)$. Just as likely, the life table symbol $_t p_s$ would be adapted, so that $p_{ij}(s,s+t)$ would be written as ${}_t p_s^{ij}$. Other variants, such as ${}_t p_{ij}(s)$, may be encountered.

For time-homogeneous processes, it is just the length of the time interval that is important, not when it starts. So $p_{ij}(0,t) = p_{ij}(s,s+t)$ for all s , and we will use the notation $p_{ij}(t)$ to denote this probability.

2. We have written $\mu_{ij}(s)$ to mean the transition rate from state i to state j at time s . Following the actuarial tradition, the time (or age) may be indicated by a subscript, so that the same rate may be written μ_s^{ij} .

For time-homogeneous processes, the transition rates are constant and we will denote these by μ_{ij} . You may also see the notation σ_{ij} instead of μ_{ij} used to denote the transition rate from state i to state j . In particular, the formulae given in the *Tables* use the σ_{ij} notation.

While a standard international actuarial notation was adopted for the life table and its derived quantities, the same is not true for the richer models needed to represent insurance contracts that depend on more than just being alive or dead. The actuarial reader must always be prepared to assimilate the notation that each particular author decides to use.

So the notation used in exam questions (and other questions) may not be the same as the notation used in this chapter. You should try to be flexible and accept whatever notation is given to you in a question. You should also try to stick to the notation given in a question when writing your answer to that question.

2 The Poisson process

The Poisson process is the simplest example of a Markov jump process in continuous time. In studying the Poisson process we shall encounter features which are of general applicability in this chapter.

2.1 Definition

The standard time-homogeneous Poisson process is a counting process in continuous time, $\{N_t, t \geq 0\}$, where N_t records the number of occurrences of some type of event within the time interval from 0 to t . The events of interest occur singly and may occur at any time.

In fact, we have already given a definition of a Poisson process with parameter λ in Chapter 2. Recall that it was defined to be a continuous-time process, starting at 0, with stationary independent increments, and for a time period t these increments follow a Poisson distribution with parameter λt . An alternative definition is given below and we will show that they are equivalent.

The probability that an event occurs during the short time interval from time t to time $t+h$ is approximately equal to λh for small h ; the parameter λ is called the rate of the Poisson process.

The Poisson process is very commonly used to model the occurrence of unpredictable incidents, such as car accidents or arrival of claims at an office.

The above definition should be made more precise if it is to be used for calculations. Formally, an integer-valued process $\{N_t, t \geq 0\}$, with filtration $\{F_t, t \geq 0\}$, is a Poisson process if:

$$\begin{aligned} P[N_{t+h} - N_t = 1 | F_t] &= \lambda h + o(h) \\ P[N_{t+h} - N_t = 0 | F_t] &= 1 - \lambda h + o(h) \\ P[N_{t+h} - N_t \neq 0, 1 | F_t] &= o(h) \end{aligned} \tag{5.1}$$

where the statement that $f(h) = o(h)$ as $h \rightarrow 0$ means $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$.

As may be seen from the definition, the increment $N_{t+h} - N_t$ of the Poisson process is independent of past values of the process and has a distribution which does not depend on t . It therefore follows that the Poisson process is a process with stationary, independent increments and, in addition, satisfies the Markov property.

It is far from obvious that the process defined above coincides with the Poisson process characterised in Chapter 2 as having independent stationary Poisson-distributed increments. That is one of the properties that we shall prove.



Important result

N_t is a Poisson random variable with mean λt . More generally, $N_{t+s} - N_s$ is a Poisson random variable with mean λt , independent of anything that has occurred before time s .

Proof

Define $p_j(t) = P(N_t = j)$, the probability that there have been exactly j events by time t . The proof will be complete if we can verify that, for each $j \geq 0$,

$$p_j(t) = \frac{e^{-\lambda t} (\lambda t)^j}{j!} \quad (5.2)$$

We will do this by setting up a differential equation and a boundary condition that $p_j(t)$ must satisfy. It will then be possible to check that the given expression does satisfy them.

For any $j > 0$, and for small positive h :

$$\begin{aligned} p_j(t+h) &= P(N_{t+h} = j) \\ &= P(N_t = j \text{ and } N_{t+h} = N_t) + P(N_t = j-1 \text{ and } N_{t+h} = N_t + 1) + o(h) \\ &= p_j(t)(1 - \lambda h) + p_{j-1}(t)\lambda h + o(h) \end{aligned}$$

Rearranging this equation, and letting $h \rightarrow 0$, we obtain, again for $j > 0$:

$$\frac{dp_j(t)}{dt} = -\lambda p_j(t) + \lambda p_{j-1}(t) \quad (5.3)$$

with initial condition $p_j(0) = 0$.

The same analysis yields, in the case $j = 0$:

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) \quad (5.4)$$

with $p_0(0) = 1$. It is now straightforward to verify that the suggested solution (5.2) satisfies both the differential equations (5.3) and (5.4) as well as the initial conditions.



Question 5.2

Verify that the function $p_j(t) = \frac{e^{-\lambda t} (\lambda t)^j}{j!}$ satisfies the equations given above.

In view of the fact that the increments of N are stationary and are independent of the past, this result may be generalised to a statement that $N_{t+s} - N_s$ is a Poisson random variable with mean λt , independent of anything that has occurred before time s .



Example

Explain how a Poisson process could be used to model motor insurance claims.

Solution

The events in this case could be occurrences of claims events (eg accidents, fires, thefts etc) or claims reported to the insurer. The parameter λ represents the average rate of occurrence of claims (eg 50 per day). The assumption that, in a sufficiently short time interval, there can be at most one claim is satisfied because we are working in continuous time. If there is a motorway pile-up, we can say that claims occurred at times 3:00, 3:01, 3:02 etc.

2.2 Sums of Poisson processes



Example

Claims are made to two insurance companies, A and B . The numbers of claims made to each are independent and follow Poisson processes with parameters λ_A (claims per day) and λ_B respectively. Show that the combined number of claims $(A+B)_t$ is a Poisson process with parameter $\lambda_A + \lambda_B$.

Solution

We will verify the three defining properties of a Poisson process that were given in Chapter 2.

Firstly, note that as both processes start at 0, trivially so does their sum. It remains to show that the increments are independent and stationary, and have a Poisson distribution with parameter $(\lambda_A + \lambda_B)t$.

Since the processes are independent of one another, it follows that their increments are independent of one another. These increments are Poisson with parameters $\lambda_A(t-s)$ and $\lambda_B(t-s)$. Their sum is therefore Poisson with parameter $(\lambda_A + \lambda_B)(t-s)$. They are therefore also stationary and independent. So we do have a Poisson process with parameter $\lambda_A + \lambda_B$.



Question 5.3

Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ be independent random variables. Prove that:

$$X + Y \sim \text{Poisson}(\lambda + \mu)$$

The previous example and question show that if we consider two independent Poisson processes with parameters λ and μ then the sum of the processes is Poisson with parameter $\lambda + \mu$. This conforms to intuition. For example, suppose that the arrivals of two different types of insurance claim follow a Poisson process, one at the rate of 5 per day the other at the rate of 6 per day. You would expect that arrivals of both followed a Poisson process at the rate of 11 per day. This is true as long as the processes are independent.

2.3 Thinning of Poisson processes

It is also useful to know that a Poisson process behaves in an intuitive way when considering the problem of *thinning* or *sampling*. Again, consider insurance claims arriving such that they follow a Poisson process with rate 10 per day. Then if one in every 10 claims is of a certain type, eg those over £10,000, the arrival of these will be Poisson with rate 1 per day. This assumes that such claims occur randomly within the arrivals of all claims. So every claim that arrives is over £10,000 with probability 0.1, independently of anything else. We have “thinned” the Poisson process.



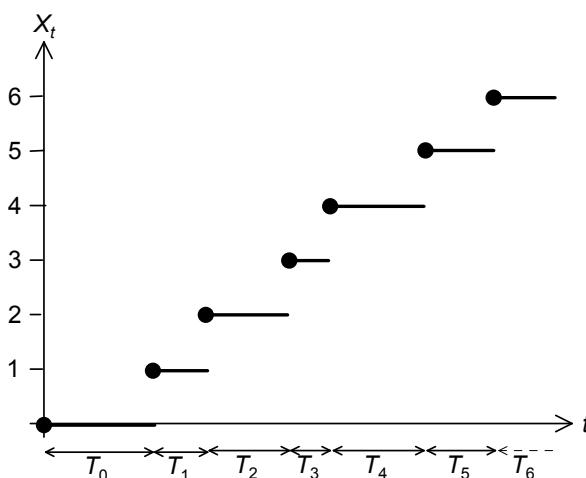
Question 5.4

An insurance company has two types of policy, A and B. Claims arriving under A follow a Poisson process with a rate of 5 per day. Claims arrive independently under B and follow a Poisson process with a rate of 3 per day. A randomly selected claim from A has a probability of $\frac{1}{5}$ of being over £10,000 while a randomly selected claim from B has probability $\frac{2}{3}$ of being over £10,000.

How many claims over £10,000 would you expect tomorrow?

2.4 Inter-event times

Since the Poisson process N_t changes only by unit upward jumps, its sample paths are fully characterised by the times at which the jumps take place. Denote by T_0, T_1, T_2, \dots the successive *inter-event times* (or holding times or inter-arrival times), a sequence of random variables.



Note that we choose (by convention) the sample paths of X_t to be right-continuous so that $X_{T_0} = 1, X_{T_0+T_1} = 2, \dots$.

So:

- $N_t = 0$ for values of t in the interval $[0, T_0)$
- $N_t = 1$ for values of t in the interval $[T_0, T_0 + T_1)$
- $N_t = 2$ for values of t in the interval $[T_0 + T_1, T_0 + T_1 + T_2)$

and so on. Because we have chosen the sample paths to be right-continuous, N_t is constant over intervals of the form $[a, b)$. If we had chosen the sample paths to be left-continuous, N_t would have been constant over intervals of the form $(a, b]$.



Important result

T_0, T_1, T_2, \dots is a sequence of independent exponential random variables, each with parameter λ .

Proof

$P(T_0 > t)$ is the probability that no events occur between time 0 and time t , which is also equal to $P(N_t = 0) = p_0(t) = e^{-\lambda t}$.

Now the distribution function of T_0 is $F(t) = P(T_0 \leq t) = 1 - e^{-\lambda t}$, $t > 0$, implying that T_0 is exponentially distributed.

Consider now the conditional distribution of T_1 given the value of T_0 .

$$\begin{aligned} P[T_1 > t | T_0 = s] &= P[N_{t+s} = 1 | T_0 = s] \\ &= P[N_{t+s} - N_s = 0 | T_0 = s] \\ &= P[N_{t+s} - N_s = 0] \\ &= p_0(t) \\ &= e^{-\lambda t} \end{aligned}$$

where the third equality reflects the independence of the increment $N_{t+s} - N_s$ from the past of the process (up to and including time s).

The above calculation proves two results at once: T_1 is independent of T_0 and has the same exponential distribution. The calculation can be repeated for T_2, T_3, \dots

In summary, all of the inter-event times are independent and are exponentially distributed with parameter λ . We will see shortly that for a time-homogeneous Markov jump process, the holding time in any given state is exponentially distributed.



Question 5.5

Reported claims from a certain group of policies follow a Poisson process with a uniform rate of 5 per day (and the insurer has a 24-hour hotline). Calculate:

- (i) the probability that there will be fewer than 2 claims reported on a given day
- (ii) the probability that another claim will be reported during the next hour
- (iii) the expected time before a claim comes in, if there haven't been any claims for over a week.

The exponential distribution of the holding times gives us a third definition of the Poisson process. We summarise these definitions below.



Summary of definitions of a Poisson process

Let $\{N_t\}_{t \geq 0}$ be an increasing, integer valued process starting at 0 (and continuous from the right). Let $\lambda > 0$. Then $\{N_t\}_{t \geq 0}$ is a Poisson process if any of the following three equivalent conditions hold:

- (1) $\{N_t\}_{t \geq 0}$ has stationary, independent increments and for each t , N_t has a Poisson distribution with parameter λt .
- (2) $\{N_t\}_{t \geq 0}$ is a Markov jump process with independent increments and transition probabilities over a short time period h given by:

$$P[N_{t+h} - N_t = 1 | F_t] = \lambda h + o(h)$$

$$P[N_{t+h} - N_t = 0 | F_t] = 1 - \lambda h + o(h)$$

$$P[N_{t+h} - N_t \neq 0, 1 | F_t] = o(h)$$

- (3) The holding times, T_0, T_1, \dots of $\{N_t\}_{t \geq 0}$ are independent exponential random variables with parameter λ and $N_{T_0+T_1+\dots+T_{n-1}} = n$.

There is also a fourth definition, which is given below. This is a restatement of (2) using the terminology of general Markov jump processes, which we will meet shortly. For completeness we will include it here, although you will have to wait for the definition of a general transition rate μ_{ij} .

- (4) $\{N_t\}_{t \geq 0}$ is a Markov jump process with independent increments and transition rates given by:

$$\mu_{ij} = \begin{cases} -\lambda & \text{if } j=i \\ \lambda & \text{if } j=i+1 \\ 0 & \text{otherwise} \end{cases}$$

3 Features of time-homogeneous Markov jump processes

We start with the definition of a Markov jump process.



Definition of a Markov jump process

A continuous-time Markov process X_t , $t \geq 0$ with a discrete (ie finite or countable) state space S is called a *Markov jump process*.

3.1 The Chapman-Kolmogorov equations

In this chapter consideration will be given to the *time-homogeneous case*, where probabilities $P(X_t = j | X_s = i)$ depend only on the length of the time interval, $t - s$.

The transition probabilities of the Markov jump process:

$$p_{ij}(t) = P(X_t = j | X_0 = i)$$

obey the Chapman-Kolmogorov equations:

$$p_{ij}(t+s) = \sum_{k \in S} p_{ik}(s) p_{kj}(t) \quad \text{for all } s, t > 0 \quad (5.5)$$

The derivation of the Chapman-Kolmogorov equations in continuous time is identical to the derivation in discrete time. See Chapter 3.

3.2 The transition matrix

Denoting by $P(t)$ the matrix with entries $p_{ij}(t)$, known as the *transition matrix*, Equation (5.5) reads:

$$P(t+s) = P(s)P(t) \quad \text{for all } s, t > 0$$

If we know the transition matrix $P(t)$ and the initial probability distribution $q_i = P(X_0 = i)$, we can find general probabilities involving the process X_t by using the Markov property.

For instance, when $0 < t_1 < t_2 < \dots < t_n$:

$$\begin{aligned} P[X_0 = i, X_{t_1} = j_1, X_{t_2} = j_2, \dots, X_{t_n} = j_n] \\ = q_i p_{ij_1}(t_1) p_{j_1 j_2}(t_2 - t_1) \dots p_{j_{n-1} j_n}(t_n - t_{n-1}) \end{aligned}$$

Adding over all states i gives:

$$\begin{aligned} P[X_{t_1} = j_1, X_{t_2} = j_2, \dots, X_{t_n} = j_n] \\ = \sum_{i \in S} q_i p_{ij_1}(t_1) p_{j_1 j_2}(t_2 - t_1) \dots p_{j_{n-1} j_n}(t_n - t_{n-1}) \end{aligned}$$

The above results are very much analogous to the results given in Section 2 of Chapter 3. If the derivation of them isn't clear then you should revise that material now.

3.3 Transition rates

For Markov chains we have the fundamental notion of the one-step transition probabilities. This is because Markov chains operate in discrete time. Together with the starting distribution, $P[X_0 = i] = q_i$, these fully determine the distribution of the chain. When we come to deal with Markov jump processes, however, we may consider transitions over arbitrarily small times, so that time steps of one unit are no longer of the same fundamental importance.

For continuous time we consider transition probabilities over a very short time interval of time h . Dividing by h expresses this as a probability of transition in unit time. Taking limits as h tends to 0 leads to the concept of a *transition rate*. We have seen this before in the two-state Markov model of Chapter 4. Recall that transition rates are also sometimes referred to as *transition intensities* or *forces of transition*.

These transition rates are the fundamental concept in continuous time; they are analogous to the one-step transition probabilities in the discrete case. Unlike probabilities, these transition rates can take values greater than 1 (as frequently happens with annual recovery rates). For example, if, on average, you spend half an hour in a particular state before leaving, then the transition rate out will be 2 per hour.

In order to differentiate the transition probabilities and avoid technical problems with the mathematics, we will make the following assumption.

We will assume that the functions $p_{ij}(t)$ are continuously differentiable. This is a large assumption to make; indeed, the full theory of Markov jump processes permits transition probabilities that do not satisfy this requirement. Such processes are called *irregular*. They are of little use in practical modelling, however, and the loss involved in restricting our attention to regular Markov processes is not significant for the purposes of this course.

Noting that:

$$p_{ij}(0) = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \quad (5.6)$$

the assumption of differentiability implies the existence of the following quantities:

$$\mu_{ij} = \frac{d}{dt} p_{ij}(t)|_{t=0} = \lim_{h \rightarrow 0} \frac{p_{ij}(h) - \delta_{ij}}{h}$$

μ_{ij} is the force of transition from state i to state j . Transition rates in time-homogeneous process do not vary over time. The function δ_{ij} in the expression above is known as the Kronecker delta.



Question 5.6

Explain Equation (5.6).

Equivalently, the following relations hold as $h \rightarrow 0$ ($h > 0$):

$$p_{ij}(h) = \begin{cases} h\mu_{ij} + o(h) & \text{if } i \neq j \\ 1 + h\mu_{ii} + o(h) & \text{if } i = j \end{cases} \quad (5.7)$$

The interpretation of the first line of (5.7) is simply that the probability of a transition from i to j during any short time interval $[s, s+h]$ is proportional to h ; hence the name *transition rate* or *transition intensity* given to μ_{ij} .

So the first line of (5.7) says that if i and j are different states, then the probability of going from state i to state j in a short time interval of length h is:

$$(h \times \text{the force of transition from state } i \text{ to state } j) + o(h)$$

This is analogous to Assumption 2 for the two-state Markov model in Chapter 4. We also assume that the probability of more than one transition in a short time interval of length h is $o(h)$.

Note finally that as a result of (5.7) $\mu_{ij} \geq 0$ for $i \neq j$, but $\mu_{ii} \leq 0$. In fact differentiating the identity $\sum_{j \in S} p_{ij}(t) = 1$ with respect to t at $t = 0$ yields:

$$\mu_{ii} = -\sum_{j \neq i} \mu_{ij}$$

Alternatively, you could argue as follows:

$$\mu_{ii} = \lim_{h \rightarrow 0} \frac{p_{ii}(h) - 1}{h} = \lim_{h \rightarrow 0} \frac{1 - \sum_{j \neq i} p_{ij}(h) - 1}{h} = -\sum_{j \neq i} \lim_{h \rightarrow 0} \frac{p_{ij}(h)}{h} = -\sum_{j \neq i} \mu_{ij}$$



Generator matrix

The generator matrix A of a Markov jump process is the matrix of transition rates. In other words, the i, j th entry of A is μ_{ij} .

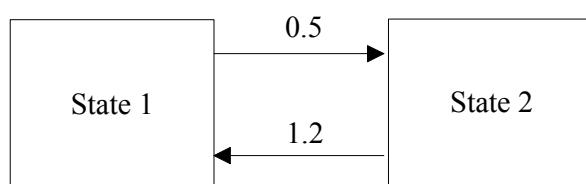
Hence each row of the matrix A has zero sum.

The relationship $\mu_{ii} = -\sum_{j \neq i} \mu_{ij}$ is often used as a working definition of μ_{ii} . The transition rate μ_{ii} is then defined as minus the sum of the transition rates out of state i .



Example

A Markov jump process has two states, with transition rates as shown in the diagram below.



Write down the generator matrix of this process.

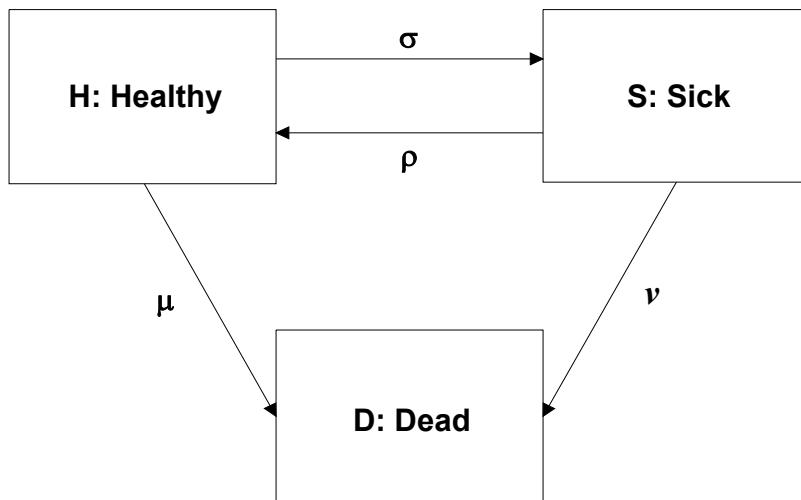
Solution

Taking the states in the order 1, 2, the generator matrix is:

$$\begin{pmatrix} -0.5 & 0.5 \\ 1.2 & -1.2 \end{pmatrix}$$

3.4 The time-homogeneous health-sickness-death model

Consider the following health-sickness-death (HSD) model with constant transition rates σ, ρ, μ, ν .



Note that a life may be in the healthy state or the sick state on a number of separate occasions before making the one-way transition to the dead state. Alternatively, a life may pass from the healthy state to the dead state without ever having been in the sick state.

Note that the transition intensity from sick to dead is represented by ν (nu, the 13th letter of the Greek alphabet), not v (the 22nd letter of the English alphabet).

**Question 5.7**

Using the notation for the HSD model shown above, what are the rates μ_{SH} , μ_{HH} and μ_{DD} ?

**Example**

Write down the generator matrix for the HSD model.

Solution

The generator matrix for the HSD model is:

$$A = \begin{pmatrix} -\sigma - \mu & \sigma & \mu \\ \rho & -\rho - \nu & \nu \\ 0 & 0 & 0 \end{pmatrix}$$

Here the order of the states has been taken to be H, S, then D (as usual).

Note that the rows sum to 0, which is consistent with our earlier equation. A common mistake is to think that the rate from dead to dead is 1. It isn't. The transition probability is 1, but the transition rate is the derivative, and hence the constant 1 goes to 0.

Another way to think of the last row of this matrix is as follows. You can't go from the dead state to the healthy state, so the force of transition from dead to healthy is 0. Similarly, the force of transition from the dead state to the sick state is 0. Each row of the generator matrix must sum to 0, so the DD entry must also be 0.

4 Kolmogorov's forward differential equations

Transition rates are of fundamental importance in that they characterise fully the distribution of Markov jump processes. In order to see this, substitute $t = h$ and $s = t$ in (5.5):

$$p_{ij}(t+h) = \sum_{k \in S} p_{ik}(t) p_{kj}(h) = p_{ij}(t) + h \sum_{k \in S} p_{ik}(t) \mu_{kj} + o(h)$$

The second equality follows from the relationship:

$$p_{jk}(h) = \begin{cases} h\mu_{jk} + o(h) & \text{if } k \neq j \\ 1 + h\mu_{jj} + o(h) & \text{if } k = j \end{cases}$$

This leads to the differential equation:

$$\frac{d}{dt} p_{ij}(t) = \sum_{k \in S} p_{ik}(t) \mu_{kj} \quad \text{for all } i, j \quad (5.8)$$

Note that you may see either lower case $p_{ij}(t)$ or upper case $P_{ij}(t)$ used to denote these transition probabilities.



Question 5.8

Derive this differential equation.

These are known as **Kolmogorov's forward equations**.



Kolmogorov's forward differential equations (time-homogeneous case)

These can be written in compact (ie matrix) form as:

$$\frac{d}{dt} P(t) = P(t) A$$

where A is the matrix with entries μ_{kj} .

Recall that A is often called the **generator matrix** of the Markov jump process.

Equipped with this general equation, we can write down specific equations for a given Markov jump process and solve them in simple cases.



Example

For the HSD model given in Section 3.4, state the forward differential equation for $p_{HH}(t)$.

Solution

By using the general forward equation as a template, we obtain:

$$\frac{d}{dt} p_{HH}(t) = p_{HH}(t) \mu_{HH} + p_{HS}(t) \mu_{SH} + p_{HD}(t) \mu_{DH}$$

Now substituting in for the transition rates, we have:

$$\frac{d}{dt} p_{HH}(t) = -p_{HH}(t)(\sigma + \mu) + p_{HS}(t)\rho$$

You need to be able to write down any forward equation, such as the last one, pretty quickly. This means not relying on the general template every time. You should be able to do this with a bit of practice if you notice the following. When writing down the equation for H to H above, we are including a term for each type of possible path; if you look at the transition graph and ask yourself “How can I go from H to H ?", the answer is that you can only be at H or S in the “middle”. The two types of paths to include are therefore $H \rightarrow H \rightarrow H$ or $H \rightarrow S \rightarrow H$. So we can consider the RHS as follows:

- start with the probability of going from H to H over an interval of length t (ie $p_{HH}(t)$) and multiply this by the force of transition that keeps you in H at time t (ie $-(\sigma + \mu)$)
- then add on the probability of going from H to S over an interval of length t (ie $p_{HS}(t)$) and multiply this by the force of transition that takes you from S to H at time t (ie ρ).



Question 5.9

Write down Kolmogorov's forward differential equation for the transition probability $p_{HS}(t)$.

Note that the equation for $p_{HS}(t)$ in Question 5.9 involves $p_{HH}(t)$ as well. This will also be unknown initially so this equation cannot be solved in its own right. The forward equations for all transitions $i \rightarrow j$ will often need to be constructed and solved as a set of *simultaneous* differential equations. Generally, writing down such sets of equations is straightforward, but solving them is much more difficult.

You will be expected to solve such equations in simple cases. In order to do this, you will usually be able to use one of two methods: separation of variables, and integrating factors. We gave a brief review of these in an appendix to Chapter 4. You may want to look at this again before continuing.



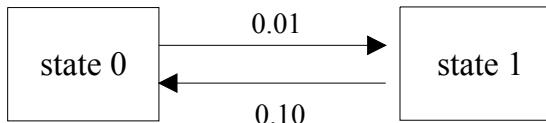
Example

A Markov jump process has two states, labelled state 0 and state 1. You are given $\mu_{01} = 0.01$ and $\mu_{10} = 0.10$.

Write down Kolmogorov's forward differential equation for the probability $p_{01}(t)$ and solve this differential equation to obtain an expression for $p_{01}(t)$.

Solution

The transition diagram for this process is as follows:



Kolmogorov's forward differential equation for $p_{01}(t)$ is:

$$\begin{aligned}\frac{d}{dt} p_{01}(t) &= p_{00}(t) \mu_{01} + p_{01}(t) \mu_{11} \\ &= 0.01 p_{00}(t) - 0.10 p_{01}(t)\end{aligned}$$

since μ_{11} is minus the (total) force of transition out of state 1.

Now, since there are only two states, we have:

$$p_{00}(t) = 1 - p_{01}(t)$$

So the differential equation can be written as:

$$\frac{d}{dt} p_{01}(t) = 0.01[1 - p_{01}(t)] - 0.10 p_{01}(t) = 0.01 - 0.11 p_{01}(t)$$

or equivalently:

$$\frac{d}{dt} p_{01}(t) + 0.11 p_{01}(t) = 0.01$$

This equation can be solved using the integrating factor method. The integrating factor in this case is $e^{0.11t}$. Multiplying every term in the previous equation by the integrating factor gives:

$$e^{0.11t} \frac{d}{dt} p_{01}(t) + 0.11 e^{0.11t} p_{01}(t) = 0.01 e^{0.11t}$$

Then integrating both sides with respect to t , we get:

$$\begin{aligned} e^{0.11t} p_{01}(t) &= \int 0.01 e^{0.11t} dt \\ &= \frac{1}{11} e^{0.11t} + C \end{aligned}$$

where C is a constant of integration. We can calculate the value of C using the initial condition $p_{01}(0) = 0$. This gives:

$$0 = \frac{1}{11} + C$$

So $C = -\frac{1}{11}$. Hence:

$$e^{0.11t} p_{01}(t) = \frac{1}{11} (e^{0.11t} - 1)$$

and dividing through by $e^{0.11t}$ gives:

$$p_{01}(t) = \frac{1}{11} (1 - e^{-0.11t})$$

If the state space S is finite, (5.8) gives for each fixed i a *finite* linear system of differential equations (in fact the index i enters only through the initial condition (5.6)). Accordingly, for given transition rates μ_{ij} , Equation (5.8) has a unique solution compatible with (5.6). For this reason Markov models are normally formulated simply by specifying their transition rates μ_{ij} .

5 Kolmogorov's backward differential equations

Substituting $s = h$ in (5.5) and proceeding as before, we obtain a different set of equations.



Kolmogorov's backward differential equations (time-homogeneous case)

These can be written in matrix form as:

$$\frac{d}{dt} \mathbf{P}(t) = \mathbf{A} \mathbf{P}(t)$$



Question 5.10

Derive this equation.



Example

For the time-homogeneous HSD model in Section 3.4, write down Kolmogorov's backward differential equation for $p_{HH}(t)$.

Solution

Using the general backward equation as a template, we obtain:

$$\frac{d}{dt} p_{HH}(t) = \mu_{HH} p_{HH}(t) + \mu_{HS} p_{SH}(t) + \mu_{HD} p_{DH}(t)$$

Now substituting in the transition rates, we have:

$$\frac{d}{dt} p_{HH}(t) = -(\sigma + \mu) p_{HH}(t) + \sigma p_{SH}(t)$$

Once again, you should be able to write these equations down without resorting to the general equation. We can think about the RHS of the equation above in the following way:

- start with the force of transition that keeps you in state H at the start (*ie* $-(\sigma + \mu)$) and multiply this by probability of going from H to H over an interval of length t (*ie* $p_{HH}(t)$)
- then add on the force of transition that takes you from H to S at time t (*ie* σ) multiplied by the probability of going from S to H over an interval of length t (*ie* $p_{SH}(t)$).



Question 5.11

Write down the backward equation for the transition probability $p_{HS}(t)$.

Under “normal” circumstances, the forward and the backward systems are equivalent; this is so in particular when the transition rates are bounded:

$$\sup_{i,j} |\mu_{ij}| < \infty \quad \forall t \geq 0$$

Here “sup” stands for “supremum”. Technically, this is the least upper bound of a set. With finite sets, this is the largest value in the set, and you could write “max” instead of “sup”. For example, the supremum of the set $\{0,1,2\}$ is 2, the same as the maximum value. The supremum of the set $(0,2)$ is also 2, since 2 is the smallest number that is greater than or equal to all the numbers in the set $(0,2)$.

However, when this condition fails, the backward system is of more fundamental importance.

The forward equations are more useful in numerical work for actuarial applications because we usually have an initial condition such as knowing that a policyholder is healthy when a policy is sold, thus we want equations that we can solve forwards in time from that starting point.

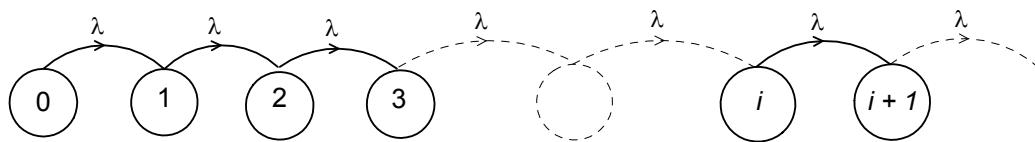
6 The Poisson process revisited

We already mentioned briefly at the end of Section 2 that the Poisson process can be formulated as a Markov jump process. We now revisit this idea.

Consider the Markov jump process with state space $S = \{0, 1, 2, \dots\}$ and transition rates:

$$\mu_{ij} = \begin{cases} -\lambda & \text{if } j = i \\ \lambda & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases}$$

The diagram representation is:



and the matrix A in Kolmogorov's equations is:

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ -\lambda & \lambda & 0 \\ & -\lambda & \lambda \\ 0 & & \ddots & \ddots \end{pmatrix}$$

This leads to the forward equations:

$$\begin{cases} p'_{i0}(t) = -\lambda p_{i0}(t) \\ p'_{ij}(t) = \lambda p_{j-1}(t) - \lambda p_{ij}(t), j > 0 \end{cases}$$

essentially identical to (5.3) and (5.4).

It is interesting also to consider the backward equations:

$$p'_{ij}(t) = -\lambda p_{ij}(t) + \lambda p_{i+1,j}(t)$$

which of course have the same solution as the forward equations despite looking dissimilar.

7 Holding times and occupancy probabilities

We shall see that time-homogeneous processes are an extension of the Poisson process. Rather than increasing in unit steps there can be transitions from each state to any other one. However, the way in which this occurs and the timing of the transitions has a lot in common with the Poisson process.

The exponential character of the holding times of the Poisson process is no accident: the *memoryless property*:

$$P[T > t + u | T > t] = P[T > u]$$

which characterises exponentially distributed random variables, is a necessary requirement for the holding times of time-homogeneous Markov processes.

Consider the first holding time $T_0 = \inf\{t : X_t \neq X_0\}$.

The infimum, inf, is the greatest lower bound of a set. The first holding time is therefore the length of time before the process first changes state.



Distribution of the first holding time

The first holding time of a time-homogeneous Markov jump process with transition rates μ_{ij} is exponentially distributed with parameter:

$$\lambda_i = -\mu_{ii} = \sum_{j \neq i} \mu_{ij}$$

Note that we are defining λ_i to be the total force of transition out of state i .

In other words:

$$P[T_0 > t | X_0 = i] = e^{-\lambda_i t}$$

We will also use the notation $p_{ii}^-(t)$ to denote the probability of remaining in a state i throughout a period of length t , so $p_{ii}^-(t) = e^{-\lambda_i t}$ also.

You may find the following proof of this result hard going. We give an alternative proof afterwards.

Proof

(Students should understand this proof, but they will not be expected to reproduce it in the examination.)

The proof is more complicated than in the case of the Poisson process because the fact that X may leave the state X_0 and return later means that it is no longer possible to claim that $\{T_0 > t\}$ is identical to the event $\{X_t = X_0\}$. Instead the event $\{T_0 > t\} = \{X_s = X_0, 0 \leq s \leq t\}$ is difficult to handle because it involves a continuum of times $0 \leq s \leq t$; so we approximate it by discretising time.

We divide the interval up into 2^n equal pieces. The number of pieces will therefore double each time n increases by 1.

Define the events:

$$B_n = \left\{ X_{\frac{kt}{2^n}} = X_0, k = 1, 2, \dots, 2^n \right\}$$

So B_n is the set of events for which:

$$X_{\frac{t}{2^n}} = X_{\frac{2t}{2^n}} = X_{\frac{3t}{2^n}} = \dots = X_{\frac{2^n t}{2^n}} = X_0$$

For the event B_n there must be no *overall* transition during any of the 2^n independent time periods of length $\frac{t}{2^n}$. (Transitions *during* these periods are allowed, as long as the process returns to its original state by the end of that time period.)

Now we can calculate:

$$P(B_n | X_0 = i) = \left(p_{ii} \left(\frac{t}{2^n} \right) \right)^{2^n} = \left(1 + \mu_{ii} \frac{t}{2^n} + o\left(\frac{1}{2^n}\right) \right)^{2^n}$$

This second inequality follows from the assumption that:

$$p_{ii}(h) = 1 + \mu_{ii} h + o(h)$$

for small h – see Section 3.3.

Here we are replacing h by $\frac{t}{2^n}$. However, since t is a constant, we can write $o\left(\frac{1}{2^n}\right)$ instead of $o\left(\frac{t}{2^n}\right)$. (An $o(h)$ function multiplied by a constant gives another $o(h)$ function.)

But note that $B_1 \supset B_2 \supset \dots \supset B_n \supset \dots$

This is because B_1 is the set of events for which:

$$X_{\frac{t}{2}} = X_t = X_0$$

B_2 is the set of events for which:

$$X_{\frac{t}{4}} = X_{\frac{2t}{4}} = X_{\frac{3t}{4}} = X_t = X_0$$

B_3 is the set of events for which:

$$X_{\frac{t}{8}} = X_{\frac{2t}{8}} = X_{\frac{3t}{8}} = \dots = X_{\frac{7t}{8}} = X_t = X_0$$

and so on. So B_2 lies wholly within B_1 , B_3 lies wholly within B_2 , etc.

In addition, $\{T_0 > t\} = \bigcap_{m=1}^{\infty} B_m$, so that:

$$\begin{aligned} P[T_0 > t | X_0 = i] &= P\left[\bigcap_{m=1}^{\infty} B_m \mid X_0 = i\right] \\ &= \lim_{n \rightarrow \infty} P\left[\bigcap_{m=1}^n B_m \mid X_0 = i\right] \\ &= \lim_{n \rightarrow \infty} P[B_n \mid X_0 = i] \end{aligned}$$

In fact, it follows directly from the above that $P[T_0 > t | X_0 = i] = \lim_{n \rightarrow \infty} P[B_n \mid X_0 = i]$.

We don't really need the two intermediate parts in the equation above.

We then have:

$$\begin{aligned}
 P[T_0 > t | X_0 = i] &= \lim_{n \rightarrow \infty} P[B_n | X_0 = i] \\
 &= \lim_{n \rightarrow \infty} \left(1 + \mu_{ii} \frac{t}{2^n} + o\left(\frac{1}{2^n}\right) \right)^{2^n} \\
 &= e^{\mu_{ii}t} \\
 &= e^{-\lambda_i t}
 \end{aligned}$$

The standard result $\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$ has been used to get the penultimate equation above.

Alternative proof

Alternatively, we can set up and solve the forward differential equation:

$$\frac{d}{dt} p_{ii}^-(t) = -p_{ii}^-(t) \lambda_i$$

which can be derived as follows.

From the Chapman-Kolmogorov equations we have:

$$p_{ii}^-(t+h) = p_{ii}^-(t) p_{ii}^-(h)$$

Suppose that h is small. Then:

$$p_{ii}^-(h) = 1 - \sum_{j \neq i} p_{ij}(h) + o(h)$$

where the $o(h)$ term denotes the probability of leaving state i and returning to it in a short time interval of length h .

Also, for small h and $j \neq i$:

$$p_{ij}(h) = h \mu_{ij} + o(h)$$

So:

$$p_{ii}^-(h) = 1 - \sum_{j \neq i} h \mu_{ij} + o(h)$$

and:

$$p_{ii}^-(t+h) = p_{ii}^-(t) \left(1 - \sum_{j \neq i} h \mu_{ij} \right) + o(h)$$

Rearranging gives:

$$\frac{p_{ii}^-(t+h) - p_{ii}^-(t)}{h} = -p_{ii}^-(t) \sum_{j \neq i} \mu_{ij} + \frac{o(h)}{h} = -p_{ii}^-(t) \lambda_i + \frac{o(h)}{h}$$

since we earlier defined $\lambda_i = -\mu_{ii} = \sum_{j \neq i} \mu_{ij}$.

By taking the limit as $h \rightarrow 0$, we get:

$$\frac{d}{dt} p_{ii}^-(t) = -p_{ii}^-(t) \lambda_i$$

This can be solved by separating variables (see the appendix to Chapter 4) to give $p_{ii}^-(t) = e^{-\lambda_i t}$.

Equivalently:

$$P[T_0 > t | X_0 = i] = e^{-\lambda_i t}$$

Note that, unlike the Poisson process, the first holding time depends on the initial state i . However, when given i , the holding time is then still exponentially distributed with parameter λ_i .



Question 5.12

What is the expected value of the first holding time for a time-homogeneous Markov jump process starting in state i ?

In the Poisson process the timing of the jumps is everything.

This is because the value of the process goes up by 1 at a time.

However, in general we must also characterise the state to which the process jumps; this is remarkably simple: the jump takes place from $X_0 = i$ to $X_{T_0} = j$ with probability proportional to the transition rate μ_{ij} and moreover X_{T_0} is independent of the holding time T_0 . In order to see this consider for $j \neq i$:

$$\begin{aligned} P[X_{t+h} = j, t < T_0 \leq t+h | X_0 = i] &= P[X_{t+h} = j, T_0 > t | X_0 = i] \\ &= P[X_{t+h} = j | X_0 = i, T_0 > t] P[T_0 > t | X_0 = i] \\ &= P[X_{t+h} = j | X_s = i, 0 \leq s \leq t] e^{-\lambda_i t} \\ &= p_{ij}(h) e^{-\lambda_i t} \end{aligned}$$

Now, divide by h and let $h \rightarrow 0$: the joint probability distribution/density of X_{T_0} and T_0 is, conditionally on $X_0 = i$, equal to:

$$\mu_{ij} e^{-\lambda_i t}$$

So it is the product of the density of the holding time $\lambda_i e^{-\lambda_i t}$ and of $\frac{\mu_{ij}}{\lambda_i}$.

This proves two results at once: the probability that the jump out of i is to state j is:

$$P[X_{T_0} = j | X_0 = i] = \frac{\mu_{ij}}{\lambda_i} \quad (j \neq i)$$

and moreover X_{T_0} is independent of T_0 .

These results are important and they are worth restating.



Probability that the process goes into state j when it leaves state i

Given that a time-homogeneous Markov jump process is currently in state i , the probability that it moves into state j when it leaves state i is given by:

$$\frac{\mu_{ij}}{\lambda_i} = \frac{\text{the force of transition from state } i \text{ to state } j}{\text{the total force of transition out of state } i}$$

Also, given a jump has occurred, the time at which it took place does not affect the probability of the jump being to a particular state.

As a result of the Markov property, the pattern is identical for successive jumps: after some state j is entered, the process stays there for an exponentially distributed time with parameter λ_j . It then jumps to state k with probability

$$\frac{\mu_{jk}}{\lambda_j}.$$

This is a key result for time-homogeneous Markov jump processes, so we will restate it.



Distribution of holding time random variables and occupancy probabilities

For a time-homogeneous Markov jump process, let W_i denote the holding time (or waiting time) in state i . Then:

$$W_i \sim \text{Exp}(\lambda_i)$$

where λ_i is the total force of transition out of state i .

So the probability of staying in state i for at least t time units (*ie* the occupancy probability for state i) is:

$$P(W_i > t) = p_{ii}^-(t) = e^{-\lambda_i t}$$

Note that the mean holding time of state j is $\frac{1}{\lambda_j}$; this is an important thing to remember when assigning numerical values to the transition rates.

So if, for example, the transition rate is 12 per hour, the mean holding time is 1/12 hour, *ie* 5 minutes.

8 Expected time to reach state k starting from state i

Let m_i denote the expected time for the process to reach state k given that it is currently in state i . Then m_i can be calculated using the recursive formula:

$$m_i = \frac{1}{\lambda_i} + \sum_{j \neq i, k} \frac{\mu_{ij}}{\lambda_i} m_j$$

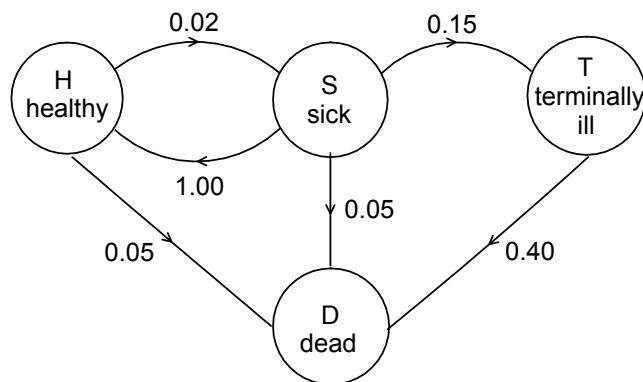
This formula is given on Page 38 of the *Tables*. Note that the *Tables* use the notation σ_{ij} instead of μ_{ij} to denote the force of transition from state i to state j .

The first term on the RHS is the expected holding time in state i . When the process leaves state i , the probability that it goes into state j is $\frac{\mu_{ij}}{\lambda_i}$, as we have just seen in Section 7. To get the second term on the RHS, we multiply this probability by the expected time to reach state k starting from state j , namely m_j , and sum over all possible values of j .



Example

Consider the following Health, Sickness, Death model with the addition of an extra “Terminally ill” state, T. The rates given are per year.



- (i) Calculate the expected holding time in state S.
- (ii) Calculate the probability that a sick life goes into state D when he leaves the sick state.
- (iii) Calculate the expected future lifetime of a healthy life.

Solution(i) ***Expected holding time in state S***

The total rate out is 1.2 so the expected holding time is $\frac{1}{1.2} = \frac{5}{6}$ years .

(ii) ***Probability that a sick life goes into state D when he leaves the sick state***

This is the proportion of the total rate out of S that goes to D, ie:

$$\frac{0.05}{1.2} = \frac{1}{24}$$

(iii) ***Expected future lifetime of a healthy life***

Let m_i denote the expected future lifetime of a life in state i . We have:

$$\lambda_H = 0.02 + 0.05 = 0.07$$

So:

$$m_H = \frac{1}{0.07} + \frac{0.02}{0.07} m_S + \frac{0.05}{0.07} m_D = \frac{100}{7} + \frac{2}{7} m_S$$

since $m_D = 0$. Also:

$$\lambda_S = 1.00 + 0.15 + 0.05 = 1.20$$

So:

$$m_S = \frac{1}{1.20} + \frac{1.00}{1.20} m_H + \frac{0.15}{1.20} m_T$$

But:

$$m_T = \frac{1}{0.40} = 2.5$$

So:

$$m_S = \frac{1}{1.20} + \frac{1.00}{1.20} m_H + \frac{0.15}{1.20} \times 2.5 = \frac{5}{6} m_H + \frac{110}{96}$$

and:

$$m_H = \frac{100}{7} + \frac{2}{7} m_S = \frac{100}{7} + \frac{2}{7} \left(\frac{5}{6} m_H + \frac{110}{96} \right)$$

This rearranges to give:

$$\left(1 - \frac{5}{21} \right) m_H = \frac{100}{7} + \frac{2}{7} \times \frac{110}{96} = \frac{2,455}{168}$$

i.e:

$$m_H = 19.18$$

So the expected future lifetime of a healthy life is 19.18 years.

9 The jump chain

If a Markov jump process is examined only at the times of its transitions, the resulting process, denoted $\{\hat{X}_n : n = 0, 1, \dots\}$, where \hat{X}_0 is the initial state, and for $n \geq 1$:

$$\hat{X}_n = X_{T_0 + T_1 + \dots + T_{n-1}}$$

is called the *jump chain* associated with X .

The jump chain is also sometimes called the embedded chain. It is the sequence of states that the process is observed to take. The time spent in each state is ignored.

The foregoing analysis shows that \hat{X}_n is independent of $T_0 + T_1 + \dots + T_{n-1}$, ie the time of the n th transition, and is also independent of anything that happened prior to the $(n-1)$ th transition: in fact, the distribution of \hat{X}_n depends only on \hat{X}_{n-1} . In other words, the jump chain possesses the Markov property and is a Markov chain in its own right.

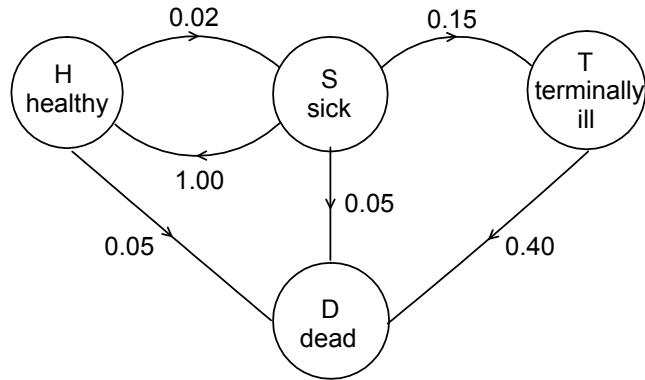
The only way in which the jump chain differs from a standard Markov chain is when the jump process $\{X_t, t \geq 0\}$ encounters an absorbing state. From that time on it makes no further transitions, implying that time stops for the jump chain. In order to deal with the jump chain entirely within the framework of Markov chains it is permissible to treat the absorbing state in the same way as for a Markov chain, so that transitions continue to occur but the chain remains in the same state after the transition.

Questions dealing solely with the sequence of states visited by the Markov jump process, such as “What is the probability that it visits state i_0 before it reaches the absorbing state?” or “Is state j visited infinitely often?”, can be answered equally well with reference to the jump chain, since the two processes take identical paths through the state space. The theory of Markov chains can therefore be employed to arrive at solutions to such questions. Questions dealing with the time taken to visit a state, however, are likely to have very different answers in the two cases and are only accessible using the theory of Markov jump processes.



Example

Consider the following Health, Sickness, Death model with the addition of an extra “Terminally ill” state, T. The rates given are per year.



Calculate the probability that a life in the sick state dies without ever recovering to the healthy state.

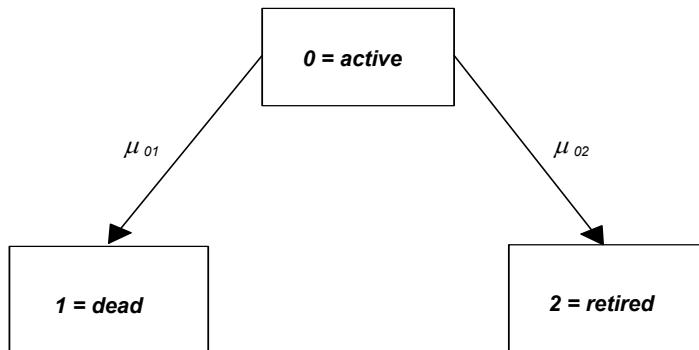
Solution

Here we can use the jump chain since the times are irrelevant. You need either to go straight to death at the next jump, or to T. If you go straight to T then you definitely die without recovering. The probability is therefore:

$$\frac{0.05 + 0.15}{1.2} = \frac{1}{6} \quad \text{or} \quad 1 - \frac{1.0}{1.2} = \frac{1}{6}$$

10 Application: a simple two-decrement model

In certain elementary cases, the solutions of the Kolmogorov equation can simply be written down, and the two-state model is often an intuitive guide. For example, consider the two-decrement model, in which the transition intensities are constant.



Note that the term *active* is usually applied to individuals in employment, in order to differentiate them from individuals who are healthy but who have retired.

We have:

$$p_{01}(x, x+t) = \frac{\mu_{01}}{\mu_{01} + \mu_{02}} \left[1 - e^{-(\mu_{01} + \mu_{02})t} \right]$$

$$p_{02}(x, x+t) = \frac{\mu_{02}}{\mu_{01} + \mu_{02}} \left[1 - e^{-(\mu_{01} + \mu_{02})t} \right]$$

Here the Core Reading is using the notation $p_{ij}(x, x+t)$ to denote the probability that a life is in state j at age $x+t$, given that he was in state i at age x . We could equally well have used the notation $p_{ij}(t)$ (since the transition probabilities depend only on the length of the time interval, t).



Question 5.13

Write down the Kolmogorov forward differential equations for $p_{01}(x, x+t)$ and $p_{02}(x, x+t)$. Hence derive the two equations above.

This is easily interpreted – the term in brackets is the probability of having left the active state, and the fraction gives the conditional probability of each decrement having occurred, given that one of them has occurred.

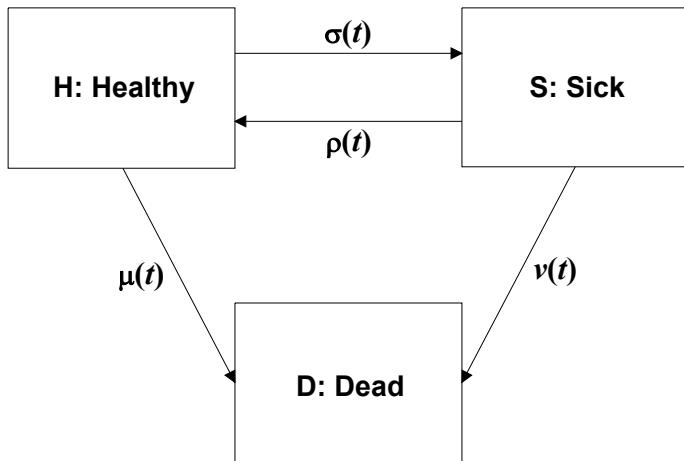
Note that $p_{00}^-(x, x+t) = e^{-(\mu_{01} + \mu_{02})t}$ is the probability that an active life aged x stays in the active state (state 0) up to age $x+t$.

However, in practice we do not always work with such simple models, or with constant transition intensities, and it is not possible to rely on solving the equations explicitly. Fortunately this does not matter; the Kolmogorov equations are quite simple to solve using ordinary numerical techniques.

The methods of solution are not part of the CT4 syllabus.

11 The maximum likelihood estimator in the general model

As we saw earlier in this chapter, the two-state model can be extended to any number of states, with arbitrary transitions between them, including increments and repeated transitions. Consider again the illness-death model, which has three states: healthy (H), sick (S) and dead (D):



The observations in respect of a single life are now:

- (a) the times between successive transitions; and
- (b) the numbers of transitions of each type.

If the transition intensities are constant, each spell of length t in the able or ill states contributes a factor of the form $e^{-(\mu+\sigma)t}$ or $e^{-(\nu+\rho)t}$ respectively to the likelihood, so it suffices to record the total waiting time spent in each state.

We saw in Section 7 that the probability of staying in state i for at least another t time units is $e^{-\lambda_i t}$, where λ_i denotes the total force of transition out of state i .



Question 5.14

A student asks: “Is it appropriate to assume that transition intensities are constant? Surely the chance of becoming sick or dying increases with age.” Comment on the student’s question.

We will now define some notation. Note that this is not standard notation, and it is quite cumbersome. You should be able to deal with whatever notation is used in a given situation. We will introduce some more general notation at the end of Section 11.1.



Definition

V_i = Waiting time of the i th life in the healthy state

W_i = Waiting time of the i th life in the sick state

S_i = Number of transitions healthy \rightarrow sick by the i th life

R_i = Number of transitions sick \rightarrow healthy by the i th life

D_i = Number of transitions healthy \rightarrow dead by the i th life

U_i = Number of transitions sick \rightarrow dead by the i th life

We also need to define totals $V = \sum_{i=1}^N V_i$ (and so on).

11.1 Maximum likelihood estimators

Using lower case symbols for the observed samples as usual, it is easily shown that the likelihood for the four parameters, μ, v, σ, ρ , given the data is proportional to:

$$L(\mu, v, \sigma, \rho) = e^{-(\mu+\sigma)v} e^{-(v+\rho)w} \mu^d v^u \sigma^s \rho^r$$

This result is obtained using a similar method to that for the two-state model, as set out in Chapter 4. The likelihood function $L(\mu, v, \sigma, \rho)$ for the i th life reflects:

- the probability of the life remaining in the healthy state for total time v_i and in the sick state for time w_i , giving the factors $e^{-(\mu+\sigma)v_i}$ and $e^{-(v+\rho)w_i}$ respectively
- the probability of the life making the relevant number of transitions between states, giving the factors $\mu^{d_i}, v^{u_i}, \sigma^{s_i}$ and ρ^{r_i} .

The likelihood factorises into functions of each parameter of the form $e^{-\mu v} \mu^d$:

$$\begin{aligned} ie \quad L(\mu, v, \sigma, \rho) &= e^{-(\mu+\sigma)v} e^{-(v+\rho)w} \mu^d v^u \sigma^s \rho^r \\ &= (e^{-\mu v} \mu^d) \times (e^{-\sigma v} \sigma^s) \times (e^{-v w} v^u) \times (e^{-\rho w} \rho^r) \end{aligned}$$

So the log-likelihood is:

$$\log L = -(\mu + \sigma)v - (v + \rho)w + d \log \mu + u \log v + s \log \sigma + r \log \rho$$

Differentiating this with respect to each of the four parameters gives:

$$\frac{\partial \log L}{\partial \mu} = -v + \frac{d}{\mu} \quad \frac{\partial \log L}{\partial v} = -w + \frac{u}{v}$$

$$\frac{\partial \log L}{\partial \sigma} = -v + \frac{s}{\sigma} \quad \frac{\partial \log L}{\partial \rho} = -w + \frac{r}{\rho}$$

Setting each of these derivatives equal to 0 and solving the resulting equations, we get:

$$\mu = \frac{d}{v} \quad v = \frac{u}{w} \quad \sigma = \frac{s}{v} \quad \rho = \frac{r}{w}$$

When there is more than one parameter to be estimated, the second order condition to check for maxima is that the Hessian matrix is negative definite, or equivalently, the eigenvalues of the Hessian matrix are all negative. The Hessian matrix is the matrix of second derivatives. So in this case we consider:

$$\begin{pmatrix} \frac{\partial^2 \ln L}{\partial \mu^2} & \frac{\partial^2 \ln L}{\partial \mu \partial v} & \frac{\partial^2 \ln L}{\partial \mu \partial \sigma} & \frac{\partial^2 \ln L}{\partial \mu \partial \rho} \\ \frac{\partial^2 \ln L}{\partial v \partial \mu} & \frac{\partial^2 \ln L}{\partial v^2} & \frac{\partial^2 \ln L}{\partial v \partial \sigma} & \frac{\partial^2 \ln L}{\partial v \partial \rho} \\ \frac{\partial^2 \ln L}{\partial \sigma \partial \mu} & \frac{\partial^2 \ln L}{\partial \sigma \partial v} & \frac{\partial^2 \ln L}{\partial \sigma^2} & \frac{\partial^2 \ln L}{\partial \sigma \partial \rho} \\ \frac{\partial^2 \ln L}{\partial \rho \partial \mu} & \frac{\partial^2 \ln L}{\partial \rho \partial v} & \frac{\partial^2 \ln L}{\partial \rho \partial \sigma} & \frac{\partial^2 \ln L}{\partial \rho^2} \end{pmatrix} = \begin{pmatrix} -\frac{d}{\mu^2} & 0 & 0 & 0 \\ 0 & -\frac{u}{v^2} & 0 & 0 \\ 0 & 0 & -\frac{s}{\sigma^2} & 0 \\ 0 & 0 & 0 & -\frac{r}{\rho^2} \end{pmatrix}$$

Since this is a negative definite matrix, the maximum likelihood estimates of μ, v, σ, ρ are:

$$\hat{\mu} = \frac{d}{v} \quad \hat{v} = \frac{u}{w} \quad \hat{\sigma} = \frac{s}{v} \quad \hat{\rho} = \frac{r}{w}$$

However, this goes beyond the scope of Subject CT4 and you would not be expected to check the Hessian matrix in the exam.

The corresponding maximum likelihood estimators are:

$$\tilde{\mu} = \frac{D}{V}, \quad \tilde{\nu} = \frac{U}{W}, \quad \tilde{\sigma} = \frac{S}{V}, \quad \tilde{\rho} = \frac{R}{W}$$

What we have just seen is a special case of a general result.



Estimating transition rates in a time-homogeneous Markov jump process

The maximum likelihood estimate of the transition rate μ_{ij} is:

$$\hat{\mu}_{ij} = \frac{n_{ij}}{t_i}$$

where n_{ij} is the number of transitions from state i to state j , and t_i is the total waiting time (or total holding time) in state i .



Example

During a large study into rates of sickness:

- 2,710 healthy lives fell sick and 2,490 sick lives recovered
- 70 healthy lives and 120 sick lives died.

For the whole group, the periods of health and sickness totalled 41,200 and 6,700 years.

Estimate the annual forces of transition between the states healthy, sick and dead.

Solution

The information tells us that:

$$t_H = 41,200 \quad t_S = 6,700$$

$$n_{HS} = 2,710 \quad n_{SH} = 2,490$$

$$n_{HD} = 70 \quad n_{SD} = 120$$

The MLEs of the transition intensities are therefore:

$$\hat{\sigma} = \frac{n_{HS}}{t_H} = \frac{2,710}{41,200} = 0.0658 \quad \hat{\rho} = \frac{n_{SH}}{t_S} = \frac{2,490}{6,700} = 0.3716$$

$$\hat{\mu} = \frac{n_{HD}}{t_H} = \frac{70}{41,200} = 0.0017 \quad \hat{\nu} = \frac{n_{SD}}{t_S} = \frac{120}{6,700} = 0.0179$$

11.2 Properties of the estimators

The asymptotic properties of these estimators follow from results similar to equations (1) and (2) in Chapter 4 (Section 4.2), and the fact that the random variables $(D_i - \mu V_i)$, $(U_i - \nu W_i)$, $(S_i - \sigma V_i)$, $(R_i - \rho W_i)$ are uncorrelated, that is:

$$E[(D_i - \mu V_i)(U_i - \nu W_i)] = 0 \text{ etc}$$

Recall that:

V_i = Waiting time of the i th life in the healthy state

W_i = Waiting time of the i th life in the sick state

S_i = Number of transitions healthy \rightarrow sick by the i th life

R_i = Number of transitions sick \rightarrow healthy by the i th life

D_i = Number of transitions healthy \rightarrow dead by the i th life

U_i = Number of transitions sick \rightarrow dead by the i th life

The estimators are not independent: D_i and U_i are both 0 or 1, but $D_i U_i \neq 1$, while (assuming that the i th life starts in the able state) $S_i = R_i$ or $R_i + 1$.

**Question 5.15**

Explain in words why:

- (i) D_i and U_i are both 0 or 1, but $D_i U_i \neq 1$
- (ii) (assuming that the i th life starts in the healthy state) $S_i = R_i$ or $R_i + 1$

The estimators are, however, asymptotically independent: the same argument as in the two-state model shows that:

- **the vector $(\tilde{\mu}, \tilde{v}, \tilde{\sigma}, \tilde{\rho})$ has an asymptotic multivariate Normal distribution;**
- **each component has a marginal asymptotic distribution of the same form as before:**

$$\tilde{\mu} \sim \text{Normal}\left(\mu, \frac{\mu}{E[V]}\right) \text{ etc}$$

- **asymptotically, the components are uncorrelated and so independent (being Normal).**

Recall from Subject CT3 that defining several random variables simultaneously on a sample space gives rise to a multivariate distribution.

**Question 5.16**

What are the marginal asymptotic distributions of \tilde{v} , $\tilde{\sigma}$ and $\tilde{\rho}$?

You should recall the following properties of a maximum likelihood estimator:

- it is asymptotically normally distributed
- it is asymptotically unbiased (*ie* if $\tilde{\theta}$ is an estimator of θ , then $E(\tilde{\theta}) = \theta$)
- asymptotically, its variance is equal to the Cramér-Rao lower bound (CRLB). The CRLB is given on Page 23 of the *Tables*.

So the maximum likelihood estimators $\tilde{\mu}_{ij}$ of the transition rates μ_{ij} all have the above properties. These results can be used to construct confidence intervals for the transition intensities or as the basis for hypothesis tests.

11.3 Calculating the total waiting time

The calculation of the estimates $\hat{\mu}$, etc, requires the total waiting time to be computed. This can be done exactly in some circumstances, but, if the exposure data are in census form, the simple census formulae in Chapter 11 provide estimates. Multiple state models are, therefore, especially well suited to the data available in many actuarial investigations.

In order to calculate total waiting time exactly, we would need to know the exact timing of each transition. This may not be possible in practice if full information is not available. Alternatively, it may be possible to perform the calculations but the process may be too time-consuming for it to be worthwhile.

A simpler approach to data collection is the census approach, in which a series of observations (“snapshots”) of a population is recorded, usually at regular intervals. Data in census form do not allow us to calculate waiting times exactly, but some simplifying assumptions allow us to calculate estimates quite accurately.

For example, we may observe 100 nonagenarians on 1 January 2006 and find that only 84 of these individuals are still alive at 1 January 2007. We could estimate the total waiting time if we were to assume that deaths occurred half way through the year on average. The accuracy of our estimated transition intensities would depend on the suitability of our assumption.

Actuarial investigations typically use the census approach to data collection, eg population studies from national censuses or the analysis of the experience of a pension scheme or a life insurance company as part of the regular valuation process. We will look at this area in more detail in Chapter 11, *Exposed to risk*.

12 Exam-style question

You should now be able to attempt the following past exam question.



Question 5.17

Subject 103, September 2004, Question 9

Vehicles in a certain country are required to be assessed every year for road-worthiness. At one vehicle assessment centre, drivers wait for an average of 15 minutes before the road-worthiness assessment of their vehicle commences. The assessment takes on average 20 minutes to complete. Following the assessment, 80% of vehicles are passed as road-worthy allowing the driver to drive home. A further 15% of vehicles are categorised as a “minor fail”; these vehicles require on average 30 minutes of repair work before the driver is allowed to drive home. The remaining 5% of vehicles are categorised as a “significant fail”; these vehicles require on average three hours of repair work before the driver can go home.

A continuous-time Markov model is to be used to model the operation of the vehicle assessment centre, with states W (waiting for assessment), A (assessment taking place), M (minor repair taking place), S (significant repair taking place) and H (travelling home).

- (i) Explain what assumption must be made about the distribution of the time spent in each state. [1]
 - (ii) Write down the generator matrix for this process. [2]
 - (iii)
 - (a) Use Kolmogorov's Forward Equations to write down differential equations satisfied by $p_{WM}(t)$ and by $p_{WA}(t)$.
 - (b) Verify that $p_{WA}(t) = 4e^{-t/20} - 4e^{-t/15}$ for $t \geq 0$, where t is measured in minutes.
 - (c) Derive an expression for $p_{WM}(t)$ for $t \geq 0$. [7]
 - (iv) Let T_i be the expected length of time (in minutes) until the vehicle can be driven home given that the assessment process is currently in state i .
 - (a) Explain why $T_W = 15 + T_A$.
 - (b) Derive corresponding equations for T_A , T_M and T_S .
 - (c) Calculate T_W . [4]
- [Total 14]

This page has been left blank so that you can keep the chapter summaries together for revision purposes.



Chapter 5 Summary

Poisson process

Let $\{N_t\}_{t \geq 0}$ be an increasing, integer valued process starting at 0 (and continuous from the right). Let $\lambda > 0$. Then $\{N_t\}_{t \geq 0}$ is a Poisson process if any of the following four equivalent conditions hold:

- (1) $\{N_t\}_{t \geq 0}$ has stationary, independent increments and for each t , N_t has a Poisson distribution with parameter λt .
- (2) $\{N_t\}_{t \geq 0}$ is a Markov jump process with independent increments and transition probabilities over a short time period h given by:

$$P[N_{t+h} - N_t = 1 | F_t] = \lambda h + o(h)$$

$$P[N_{t+h} - N_t = 0 | F_t] = 1 - \lambda h + o(h)$$

$$P[N_{t+h} - N_t \neq 0, 1 | F_t] = o(h)$$

- (3) The holding times, T_0, T_1, \dots of $\{N_t\}_{t \geq 0}$ are independent exponential random variables with parameter λ and $N_{T_0+T_1+\dots+T_{n-1}} = n$.
- (4) $\{N_t\}_{t \geq 0}$ is a Markov jump process with independent increments and transition rates given by:

$$\mu_{ij} = \begin{cases} -\lambda & \text{if } j = i \\ \lambda & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases}$$

Sums of Poisson processes

If we have two independent Poisson processes with parameters λ and μ , then the sum of the two processes is another Poisson process with parameter $\lambda + \mu$.

Thinning of a Poisson process

When the events in a Poisson process are of different types, each type occurring at random with a certain probability, the events of a particular type form a *thinned* process. The thinned process is also a Poisson process, with rate equal to the original rate multiplied by the probability for the type of event.

Inter-event times in a Poisson process

Suppose that T_0, T_1, T_2, \dots are the successive inter-event times or holding times in a Poisson process with parameter λ . Then T_0, T_1, T_2, \dots are independent $Exp(\lambda)$ random variables.

Markov jump processes

A Markov jump process is a stochastic process with a continuous time set and discrete state space that satisfies the Markov property.

Time-homogeneous Markov jump processes

A Markov jump process is said to be time-homogeneous if the transition probabilities $P(X_t = j | X_s = i)$ depend only on the length of the time interval, $t - s$. Then:

$$p_{ij}(t) = P(X_t = j | X_0 = i)$$

Chapman-Kolmogorov equations

$$p_{ij}(s+t) = \sum_k p_{ik}(s)p_{kj}(t) \quad \text{for all } s, t > 0$$

Transition rates

The transition rates (or transition intensities or forces of transition) for a time-homogeneous Markov jump process are given by:

$$\mu_{ij} = \frac{d}{dt} p_{ij}(t) \Big|_{t=0} = \lim_{h \rightarrow 0} \frac{p_{ij}(h) - p_{ij}(0)}{h}$$

This is equivalent to:

$$p_{ij}(h) = \begin{cases} h\mu_{ij} + o(h) & \text{if } i \neq j \\ 1 + h\mu_{ij} + o(h) & \text{if } i = j \end{cases}$$

for small values of h .

Generator matrix

The generator matrix is the matrix of transition rates. It is usually denoted by A . Each row of the generator matrix sums to zero since $\mu_{ii} = -\sum_{j \neq i} \mu_{ij}$.

Backward and forward differential equations (time-homogeneous case)

$$\text{Forward: } \frac{d}{dt} p_{ij}(t) = \sum_{k \in S} p_{ik}(t) \mu_{kj}$$

$$\frac{d}{dt} P(t) = P(t)A \quad (\text{matrix form})$$

$$\text{Backward: } \frac{d}{dt} p_{ij}(t) = \sum_{k \in S} \mu_{ik} p_{kj}(t)$$

$$\frac{d}{dt} P(t) = AP(t) \quad (\text{matrix form})$$

Holding time random variables

For a time-homogeneous Markov jump process, let T_i denote the holding time in state i . Then $T_i \sim \text{Exp}(\lambda_i)$, where λ_i is the total force of transition out of state i . The expected holding time in state i is $\frac{1}{\lambda_i}$.

Occupancy probabilities

The probability of remaining in state i for at least t time units is:

$$P(T_i > t) = p_{ii}^-(t) = e^{-\lambda_i t}$$

Probability that the process goes into state j when it leaves state i

Given that a time-homogeneous Markov jump process is currently in state i , the probability that it moves into state j when it leaves state i is $\frac{\mu_{ij}}{\lambda_i}$.

Expected time to reach a given state

To find the expected time to reach a given state, state k , starting from state i , apply the following formula recursively:

$$m_i = \frac{1}{\lambda_i} + \sum_{j \neq i, k} \frac{\mu_{ij}}{\lambda_i} m_j$$

This formula is on Page 38 of the Tables.

Jump chains

The jump chain (or embedded chain) of a Markov jump process is the sequence of states that the process enters. The time spent in each state is ignored. The jump chain is a Markov chain in its own right.

Estimating transition rates

The maximum likelihood estimate of the transition rate μ_{ij} , $i \neq j$, is:

$$\hat{\mu}_{ij} = \frac{n_{ij}}{t_i}$$

where n_{ij} is the number of transitions from state i to state j , and t_i is the total waiting time (or total holding time) in state i .

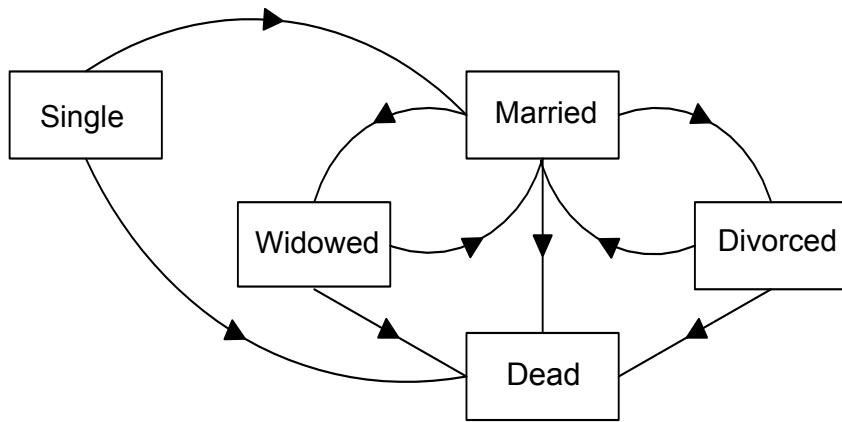
The maximum likelihood estimate of the transition rate μ_{ii} is $\hat{\mu}_{ii} = -\sum_{j \neq i} \hat{\mu}_{ij}$.

The maximum likelihood estimator of μ_{ij} has the following properties:

- it is asymptotically normally distributed
- it is asymptotically unbiased (*ie* if $\tilde{\theta}$ is an estimator of θ , then $E(\tilde{\theta}) = \theta$)
- asymptotically, its variance is given by the Cramér-Rao lower bound (CRLB). The CRLB is given on Page 23 of the *Tables*.

Chapter 5 Solutions

Solution 5.1



Solution 5.2

If $j = 0$ then $p_0(t) = e^{-\lambda t}$ and $p'_0(t) = -\lambda e^{-\lambda t} = -\lambda p_0(t)$ as required.

Otherwise:

$$p_j(t) = \frac{e^{-\lambda t} (\lambda t)^j}{j!}$$

and:

$$p'_j(t) = -\lambda \frac{e^{-\lambda t} (\lambda t)^j}{j!} + j\lambda (\lambda t)^{j-1} \frac{e^{-\lambda t}}{j!} = -\lambda p_j(t) + \lambda p_{j-1}(t)$$

Note also that:

$$p_j(0) = \frac{e^{-\lambda 0} (\lambda 0)^j}{j!} = 0 \text{ if } j > 0 \text{ and } p_0(0) = 1$$

Hence the boundary conditions are also satisfied.

Solution 5.3

If we use the convolution approach, we have:

$$\begin{aligned}
 P[X + Y = k] &= \sum_{i=0}^k P[X = i, Y = k - i] \\
 &= \sum_{i=0}^k P[X = i]P[Y = k - i] \quad \text{by independence} \\
 &= \sum_{i=0}^k \frac{e^{-\lambda} \lambda^i}{i!} \frac{e^{-\mu} \mu^{k-i}}{(k-i)!} \\
 &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \lambda^i \mu^{k-i} \\
 &= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k \quad \text{by the binomial expansion}
 \end{aligned}$$

This is the probability function for the $Poi(\lambda + \mu)$ distribution.

Alternatively, we can use probability generating functions:

$$\begin{aligned}
 G_{X+Y}(t) &= G_X(t)G_Y(t) \quad \text{by independence} \\
 &= e^{\lambda(t-1)}e^{\mu(t-1)} \quad \text{using the formulae for PGFs} \\
 &= e^{(\lambda+\mu)(t-1)}
 \end{aligned}$$

Since this is the same as the PGF of Poisson($\lambda + \mu$), we can apply the uniqueness property of PGFs to give the required result.

Solution 5.4

You need to calculate the Poisson parameter for claims over £10,000. This is the sum of the parameters for claims over £10,000 from each of A and B. By the “thinning rule”, claims under A that are over £10,000 arrive as a Poisson process with rate $\frac{1}{5} \times 5 = 1$ per day. Similarly under B we get $\frac{2}{3} \times 3 = 2$. Therefore, overall, we have a Poisson process with parameter $1 + 2 = 3$. So the mean number of claims is 3 per day.

Solution 5.5

(i) **Fewer than 2 claims in a day**

The number of reported claims in an interval of t days will have a Poisson distribution with parameter $5t$. For $t = 1$ we therefore have:

$$P[X \leq 1] = \frac{e^{-5} 5^0}{0!} + \frac{e^{-5} 5^1}{1!} = 6e^{-5} = 0.0404$$

(ii) **Another claim in the next hour**

Here $t = \frac{1}{24}$ and we want:

$$P[X \geq 1] = 1 - P[X = 0] = 1 - e^{-\frac{5}{24}} = 0.1881$$

Alternatively, we could use the waiting time until the next report. This has an exponential distribution with parameter 5. Therefore again we have:

$$P[T \leq \frac{1}{24}] = F_T\left(\frac{1}{24}\right) = 1 - e^{-\frac{5}{24}} = 0.1881$$

(iii) **Expected time until the next claim**

The waiting time has the lack of memory property, so the time before another claim comes in is independent of the time since the last one. The expected time is therefore the expected value of the exponential distribution, which in this case is 0.2 days.

In real life, the assumptions of a uniform rate and independence may not be valid. If there haven't been any claims reported for a week this may be because of a "blockage" in the system (eg bank holidays or telephone malfunction) and there may well be a "catch-up" effect the next day.

Solution 5.6

$p_{ij}(0)$ is the probability of simultaneously being in state i and state j at time 0. This will be 1 if $i = j$ but 0 otherwise.

Solution 5.7

$$\mu_{SH} = \rho, \quad \mu_{HH} = -(\sigma + \mu), \quad \mu_{DD} = 0$$

Solution 5.8

Using the Core Reading we have:

$$p_{ij}(t+h) = p_{ij}(t) + h \sum_{k \in S} p_{ik}(t) \mu_{kj} + o(h)$$

Rearranging this we have:

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \sum_{k \in S} p_{ik}(t) \mu_{kj} + \frac{o(h)}{h}$$

Taking the limit as $h \rightarrow 0$ gives the desired result:

$$\frac{d}{dt} p_{ij}(t) = \sum_{k \in S} p_{ik}(t) \mu_{kj}$$

Alternatively, we could have started with the Chapman-Kolmogorov equations:

$$p_{ij}(t) = \sum_{k \in S} p_{ik}(h) p_{kj}(t-h)$$

Differentiating this with respect to t gives:

$$\frac{d}{dt} p_{ij}(t) = \sum_{k \in S} p_{ik}(h) \frac{d}{dt} p_{kj}(t-h)$$

Finally we can evaluate this at $h = t$ to get:

$$\frac{d}{dt} p_{ij}(t) = \sum_{k \in S} p_{ik}(t) \mu_{kj}$$

Solution 5.9

$$\begin{aligned}\frac{d}{dt} p_{HS}(t) &= p_{HH}(t)\mu_{HS} + p_{HS}(t)\mu_{SS} + p_{HD}(t)\mu_{DS} \\ &= p_{HH}(t)\sigma - p_{HS}(t)(\rho + \nu)\end{aligned}$$

Solution 5.10

As suggested, we substitute $s = h$ into (5.5) to get:

$$p_{ij}(t+h) = \sum_{k \in S} p_{ik}(h) p_{kj}(t)$$

Since:

$$p_{ik}(h) = h \mu_{ik} + o(h) \text{ for } k \neq i$$

and:

$$p_{ii}(h) = 1 - \sum_{k \neq i} p_{ik}(h) = 1 - h \sum_{k \neq i} \mu_{ik} + o(h) = 1 + h \mu_{ii} + o(h)$$

we have:

$$\begin{aligned}p_{ij}(t+h) &= \sum_{k \neq i} h \mu_{ik} p_{kj}(t) + (1 + h \mu_{ii}) p_{ij}(t) + o(h) \\ &= p_{ij}(t) + h \sum_{k \in S} \mu_{ik} p_{kj}(t) + o(h)\end{aligned}$$

If we then take the $p_{ij}(t)$ term to the left-hand side, divide by h and then take the limit $h \rightarrow 0$ we get the differential equation:

$$\frac{d}{dt} p_{ij}(t) = \sum_{k \in S} \mu_{ik} p_{kj}(t) \quad \text{for all } i, j$$

or, equivalently:

$$\frac{d}{dt} P(t) = AP(t)$$

Solution 5.11

$$\begin{aligned}\frac{d}{dt} p_{HS}(t) &= \mu_{HH} p_{HS}(t) + \mu_{HS} p_{SS}(t) + \mu_{HD} p_{DS}(t) \\ &= -(\sigma + \mu) p_{HS}(t) + \sigma p_{SS}(t)\end{aligned}$$

Solution 5.12

The first holding time in i has an exponential distribution with parameter λ_i . The average time is therefore $\frac{1}{\lambda_i}$.

Solution 5.13

The Kolmogorov forward equations for this two-decrement model are:

$$\frac{\partial}{\partial t} p_{01}(x, x+t) = p_{00}(x, x+t) \mu_{01}$$

and:

$$\frac{\partial}{\partial t} p_{02}(x, x+t) = p_{00}(x, x+t) \mu_{02}$$

Since it is impossible to leave the active state and subsequently return to it, $p_{00}(x, x+t) = p_{00}^-(x, x+t)$. So:

$$p_{00}(x, x+t) = \exp\left(-\int_0^t (\mu_{01} + \mu_{02}) ds\right) = e^{-(\mu_{01} + \mu_{02})t}$$

So the Kolmogorov equations can be written as:

$$\frac{\partial}{\partial t} p_{01}(x, x+t) = \mu_{01} e^{-(\mu_{01} + \mu_{02})t}$$

and:

$$\frac{\partial}{\partial t} p_{02}(x, x+t) = \mu_{02} e^{-(\mu_{01} + \mu_{02})t}$$

Integrating the first of these equations with respect to t gives:

$$p_{01}(x, x+t) = -\frac{\mu_{01}}{\mu_{01} + \mu_{02}} e^{-(\mu_{01} + \mu_{02})t} + C$$

where C is a constant of integration.

Since $p_{01}(x, x) = 0$, it follows that:

$$C = \frac{\mu_{01}}{\mu_{01} + \mu_{02}}$$

So:

$$p_{01}(x, x+t) = \frac{\mu_{01}}{\mu_{01} + \mu_{02}} \left(1 - e^{-(\mu_{01} + \mu_{02})t}\right)$$

Similarly, integrating the second equation and using the initial condition $p_{02}(x, x) = 0$, we obtain:

$$p_{02}(x, x+t) = \frac{\mu_{02}}{\mu_{01} + \mu_{02}} \left(1 - e^{-(\mu_{01} + \mu_{02})t}\right)$$

Solution 5.14

Yes, the chance of becoming sick or dying does usually increase with age. But remember how we are studying the transition intensities – we are observing a large number of lives simultaneously over a narrow age interval, *ie* between ages x and $x+1$. Whilst we are confining our study to such intervals, it may be appropriate to assume that transition intensities over these intervals are constant.

What is harder to justify is the assumption that the transition rate for recovery is constant. In real life this will vary significantly with the duration of sickness.

Solution 5.15

- (i) A life must be in one of two states at the point of death. The life may be in the healthy state ($D_i = 1$) or it may be in the sick state ($U_i = 1$). It cannot be in both states, so $D_i U_i \neq 1$. (In fact, $D_i U_i$ always equals zero!)
- (ii) A life starts in the able state. If it is in the healthy state at the point of death, then it must have made the same number of transitions from healthy to sick as from sick to healthy ($S_i = R_i$). If it is in the sick state at the point of death, then it must have made one more transition from healthy to sick than it did from sick to healthy, in which case $S_i = R_i + 1$.

Solution 5.16

$$\tilde{\nu} \sim \text{Normal}\left(\nu, \frac{\nu}{E[W]}\right) \quad \tilde{\sigma} \sim \text{Normal}\left(\sigma, \frac{\sigma}{E[V]}\right) \quad \tilde{\rho} \sim \text{Normal}\left(\rho, \frac{\rho}{E[W]}\right)$$

Solution 5.17

- (i) ***Assumption about the time spent in each state***

With a continuous-time Markov jump process the times spent in each state are exponentially distributed.

- (ii) ***Generator matrix***

If we measure times in minutes, the generator matrix (with zeros omitted) is:

$$\begin{array}{ccccc} & W & A & M & S & H \\ W & \left[\begin{array}{ccccc} -\frac{1}{15} & \frac{1}{15} & 0 & 0 & 0 \\ 0 & -\frac{1}{20} & \frac{3}{400} & \frac{1}{400} & \frac{1}{25} \\ 0 & 0 & -\frac{1}{30} & 0 & \frac{1}{30} \\ 0 & 0 & 0 & -\frac{1}{180} & \frac{1}{180} \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] & & & \end{array}$$

If you work in hours, all these entries need to be multiplied by 60.

(iii)(a) **Kolmogorov forward differential equations**

The general formula for the Kolmogorov forward differential equation in the time-homogeneous case is:

$$\frac{d}{dt} p_{ij}(t) = \sum_k p_{ik}(t) \mu_{kj}$$

Applying this, with $i = W$ and $j = M$, we get:

$$\frac{d}{dt} p_{WM}(t) = p_{WA}(t) \mu_{AM} + p_{WM}(t) \mu_{MM} = \frac{3}{400} p_{WA}(t) - \frac{1}{30} p_{WM}(t)$$

Similarly:

$$\frac{d}{dt} p_{WA}(t) = \frac{1}{15} p_{WW}(t) - \frac{1}{20} p_{WA}(t)$$

(iii)(b) **Verify the formula for $p_{WA}(t)$**

In order to check that the formula given in the question satisfies the differential equation just stated, we first need a formula for $p_{WW}(t)$. Since there are no arrows entering state W , $p_{WW}(t)$ is the same as $p_{\overline{WW}}(t)$, which we can work out as:

$$p_{WW}(t) = p_{\overline{WW}}(t) = \exp\left(-\int_0^t \mu_{WA} ds\right) = e^{-t/15}$$

If we substitute the formula given in the question for $p_{WA}(t)$ into the Kolmogorov equation, we get:

$$LHS = \frac{d}{dt} p_{WA}(t) = \frac{d}{dt} \left(4e^{-t/20} - 4e^{-t/15} \right) = -\frac{1}{5} e^{-t/20} + \frac{4}{15} e^{-t/15}$$

and:

$$\begin{aligned} RHS &= \frac{1}{15} p_{WW}(t) - \frac{1}{20} p_{WA}(t) \\ &= \frac{1}{15} e^{-t/15} - \frac{1}{20} \left(4e^{-t/20} - 4e^{-t/15} \right) = -\frac{1}{5} e^{-t/20} + \frac{4}{15} e^{-t/15} \end{aligned}$$

So the differential equation is satisfied.

We also need to check the boundary condition. Substituting $t = 0$ into the formula given, we get:

$$p_{WA}(0) = 4e^0 - 4e^0 = 0$$

This is the correct value since the process cannot move from state W to state A in zero time.

(iii)(c) ***Derive an expression for $p_{WM}(t)$***

We can now use the formula for $p_{WA}(t)$ from part (iii)(b) in conjunction with the first differential equation from part (iii)(a) to find a formula for $p_{WM}(t)$. We have:

$$\begin{aligned} \frac{d}{dt} p_{WM}(t) &= \frac{3}{400} p_{WA}(t) - \frac{1}{30} p_{WM}(t) \\ &= \frac{3}{400} (4e^{-t/20} - 4e^{-t/15}) - \frac{1}{30} p_{WM}(t) \\ &= \frac{3}{100} (e^{-t/20} - e^{-t/15}) - \frac{1}{30} p_{WM}(t) \end{aligned}$$

We can solve this using an integrating factor. We first need to rearrange it in the form:

$$\frac{d}{dt} p_{WM}(t) + \frac{1}{30} p_{WM}(t) = \frac{3}{100} (e^{-t/20} - e^{-t/15})$$

The integrating factor is:

$$\exp\left(\int \frac{1}{30} dt\right) = e^{t/30}$$

Multiplying through by the integrating factor, we get:

$$e^{t/30} \frac{d}{dt} p_{WM}(t) + \frac{1}{30} e^{t/30} p_{WM}(t) = \frac{3}{100} (e^{-t/60} - e^{-t/30})$$

So:

$$\frac{d}{dt} [e^{t/30} p_{WM}(t)] = \frac{3}{100} (e^{-t/60} - e^{-t/30})$$

Now we can integrate to get:

$$e^{t/30} p_{WM}(t) = \frac{3}{100} (-60e^{-t/60} + 30e^{-t/30}) + c = -\frac{9}{5}e^{-t/60} + \frac{9}{10}e^{-t/30} + c$$

When $t = 0$, this becomes:

$$0 = -\frac{9}{5} + \frac{9}{10} + c = -\frac{9}{10} + c \Rightarrow c = \frac{9}{10}$$

So we have:

$$e^{t/30} p_{WM}(t) = -\frac{9}{5}e^{-t/60} + \frac{9}{10}e^{-t/30} + \frac{9}{10}$$

Dividing through by the integrating factor gives us the final answer:

$$p_{WM}(t) = -\frac{9}{5}e^{-t/20} + \frac{9}{10}e^{-t/15} + \frac{9}{10}e^{-t/30}$$

(iv)(a) ***Explain why $T_W = 15 + T_A$***

If a vehicle is currently in state W, it will wait 15 minutes on average in that state before moving to state A (definitely), after which it will wait on average a further time T_A before it can be driven home. So the average time T_W before it can be driven home is $15 + T_A$.

(iv)(b) ***Equations for T_A , T_M and T_S***

Using similar logic, a vehicle in state A will wait 20 minutes on average in that state before moving either to state H (with probability 0.8) or to state M (with probability 0.15) or to state S (with probability 0.05). So the corresponding equation is:

$$T_A = 20 + 0.8 \times 0 + 0.15T_M + 0.05T_S$$

Since we know that $T_M = 30$ and $T_S = 180$, this gives $T_A = 33.5$ minutes.

(iv)(c) ***Calculate T_W***

Using the equation from part (iii)(a), we find that:

$$T_W = 15 + T_A = 15 + 33.5 = 48.5 \text{ minutes}$$

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Chapter 6

Time-inhomogeneous Markov jump processes



Syllabus objectives

- (iv) *Define and apply a Markov process.*
1. *State the essential features of a Markov process model.*
 3. *Derive the Kolmogorov equations for a Markov process with time/age dependent transition intensities.*
 4. *Solve the Kolmogorov equations in simple cases.*
 5. *Describe simple survival models, sickness models and marriage models in terms of Markov processes and describe other simple applications.*
 6. *State the Kolmogorov equations for a model where the transition intensities depend not only on age/time, but also on the duration of stay in one or more states.*
 7. *Describe sickness and marriage models in terms of duration dependent Markov processes and describe other simple applications.*
 8. *Demonstrate how Markov jump processes can be used as a tool for modelling and how they can be simulated.*

0 **Introduction**

In this chapter we discuss time-inhomogeneous Markov jump processes. The transition probabilities $P(X_t = j | X_s = i)$ for a time-inhomogeneous process depend not only on the length of the time interval $[s,t]$, but also on the times s and t when it starts and ends. This is because the transition rates for a time-inhomogeneous process vary over time.

We start by discussing the important features of time-inhomogeneous processes. Then just as we did for time-homogeneous processes in Chapter 5, we study the forward and backward Kolmogorov differential equations and occupancy probabilities. We then introduce the integrated forms of the Kolmogorov equations, and we look at some applications. Finally, we cover some modelling techniques for Markov jump processes and describe how parameters should be estimated.

1 Features of time-inhomogeneous Markov jump processes

1.1 Chapman-Kolmogorov equations

The more general continuous-time Markov process $\{X_t, t \geq 0\}$ has transition probabilities:

$$p_{ij}(s, t) = P[X_t = j | X_s = i] \quad (s \leq t)$$

which obey a version of the *Chapman-Kolmogorov equations*, written in matrix form as:

$$P(s, t) = P(s, u)P(u, t) \quad \text{for all } s < u < t$$

or equivalently:

$$p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, u) p_{kj}(u, t) \quad \text{for all } u, s < u < t$$

Again, you will see both upper and lower case P 's used to denote these transition probabilities.

1.2 Transition rates

Proceeding as in the time-homogeneous case, we obtain:

$$p_{ij}(s, s+h) = \begin{cases} h \mu_{ij}(s) + o(h) & \text{if } i \neq j \\ 1 + h \mu_{ii}(s) + o(h) & \text{if } i = j \end{cases}$$

Equivalently, we have:

$$\mu_{ij}(s) = \left[\frac{\partial}{\partial t} p_{ij}(s, t) \right]_{t=s} = \lim_{h \rightarrow 0} \frac{p_{ij}(s, s+h) - p_{ij}(s, s)}{h} = \lim_{h \rightarrow 0} \frac{p_{ij}(s, s+h) - \delta_{ij}}{h}$$

We see that the only difference between this case and the time-homogeneous case studied earlier is that the transition rates $\mu_{ij}(s)$ are allowed to change over time.

2 Kolmogorov's forward differential equations

Kolmogorov's forward equations can be derived.

Start with the Chapman-Kolmogorov equations as given above:

$$p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, u) p_{kj}(u, t) \quad \text{for } s \leq u \leq t$$

To obtain the differential form we can differentiate with respect to t :

$$\frac{\partial}{\partial t} p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, u) \frac{\partial}{\partial t} p_{kj}(u, t) \quad \text{for } s \leq u \leq t$$

Since this is valid for any u in the given range, we can set $u = t$ afterwards (or take the limit as $u \rightarrow t$ from below):

$$\frac{\partial}{\partial t} p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, t) \left[\frac{\partial}{\partial t} p_{kj}(u, t) \right]_{u=t}$$

We recognise the term in square brackets as the transition rate $\mu_{kj}(t)$ so that the differential form is:

$$\frac{\partial}{\partial t} p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, t) \mu_{kj}(t)$$

This result is given in Section 7.1 on Page 38 of the Formulae in the *Tables*. However, in the *Tables*, the notation $\sigma_{kj}(t)$ rather than $\mu_{kj}(t)$ is used to denote the force of transition from state k to state j at time t .

Alternatively, to derive this result we could start with the equation:

$$p_{ij}(s, t+h) = \sum_{k \in S} p_{ik}(s, t) p_{kj}(t, t+h)$$

If we suppose that h is small, then:

$$p_{kj}(t, t+h) = h \mu_{kj}(t) + o(h) \quad \text{for } j \neq k$$

and: $p_{jj}(t, t+h) = 1 + h \mu_{jj}(t) + o(h)$

So:

$$\begin{aligned} p_{ij}(s, t+h) &= \sum_{k \neq j} p_{ik}(s, t) h \mu_{kj}(t) + p_{ij}(s, t) (1 + h \mu_{jj}(t)) + o(h) \\ &= p_{ij}(s, t) + \sum_{k \in S} p_{ik}(s, t) h \mu_{kj}(t) + o(h) \end{aligned}$$

Rearranging gives:

$$\frac{p_{ij}(s, t+h) - p_{ij}(s, t)}{h} = \sum_{k \in S} p_{ik}(s, t) \mu_{kj}(t) + \frac{o(h)}{h}$$

and letting $h \rightarrow 0$, we obtain:

$$\frac{\partial}{\partial t} p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, t) \mu_{kj}(t)$$



Kolmogorov's forward differential equations (time-inhomogeneous case)

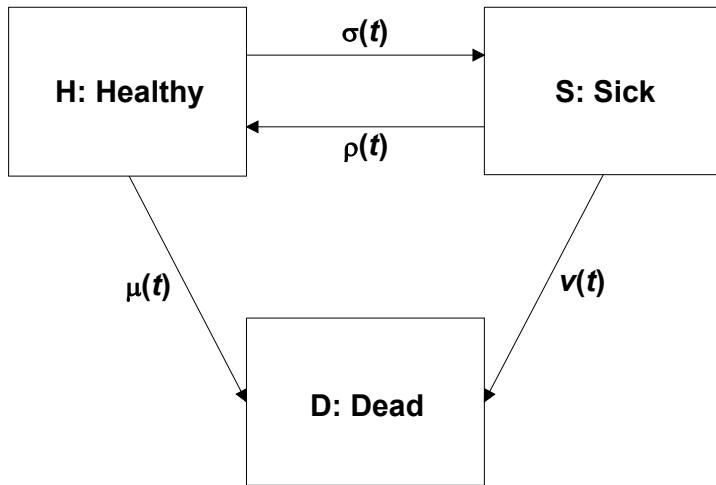
Written in matrix form these are:

$$\frac{\partial}{\partial t} P(s, t) = P(s, t) A(t)$$

where $A(t)$ is the matrix with entries $\mu_{ij}(t)$.

2.1 Time-inhomogeneous HSD model

We met the HSD model in Chapter 5. In Chapter 5 we assumed that the transition rates were constant. However, in this chapter we will assume that they vary over time. The transition diagram is shown below.



Question 6.1

Write down Kolmogorov's forward differential equation for $p_{HD}(s,t)$.

2.2 Non-standard forward equations

We may also want to construct differential equations for probabilities other than transition probabilities of the form $p_{ij}(s,t)$, in particular, the probability of remaining in state i throughout the time period (s,t) . This will be denoted by $p_{ii}(s,t)$. This is different from $p_{ii}(s,t)$ since the latter allows the possibility of leaving state i during the period. In these cases the standard forward equation is not applicable and we may have to go back to first principles, as in the following example.

**Example**

Derive the differential equation $\frac{\partial}{\partial t} p_{SS}(s, t) = -(\rho(t) + \nu(t)) p_{SS}(s, t)$.

Solution

Consider $p_{SS}(s, t+h)$ and condition on the state at time t . Analogous to the Chapman-Kolmogorov equations we have:

$$p_{SS}(s, t+h) = p_{SS}(s, t) p_{SS}(t, t+h)$$

However, during the short interval h , the process either remains in S , moves from S to H or moves from S to D . We assume here that the probability of more than one move is very small (represented by the $o(h)$ term) so that:

$$p_{SS}(t, t+h) + p_{SH}(t, t+h) + p_{SD}(t, t+h) + o(h) = 1$$

Since we know that $p_{SH}(t, t+h) = h\rho(t) + o(h)$ and $p_{SD}(t, t+h) = h\nu(t) + o(h)$ this gives $p_{SS}(t, t+h) = 1 - h(\rho(t) + \nu(t)) + o(h)$. Therefore:

$$p_{SS}(s, t+h) = p_{SS}(s, t)(1 - h(\rho(t) + \nu(t))) + o(h)$$

This can be rearranged to give:

$$\frac{p_{SS}(s, t+h) - p_{SS}(s, t)}{h} = -p_{SS}(s, t)(\rho(t) + \nu(t)) + \frac{o(h)}{h}$$

Finally, taking the limit as $h \rightarrow 0$ we obtain the differential equation:

$$\frac{\partial}{\partial t} p_{SS}(s, t) = -p_{SS}(s, t)(\rho(t) + \nu(t))$$

**Question 6.2**

Derive the differential equation $\frac{\partial}{\partial t} p_{HH}(s, t) = -(\sigma(t) + \mu(t)) p_{HH}(s, t)$.

The equations in the above example and in Question 6.2 can be solved by the standard technique of separating variables, which is discussed in the appendix to Chapter 4. For example, dividing both sides of the latter equation by $p_{\overline{HH}}(s, t)$ gives:

$$\frac{1}{p_{\overline{HH}}(s, t)} \frac{\partial}{\partial t} p_{\overline{HH}}(s, t) = -(\sigma(t) + \mu(t))$$

which can be written in the form:

$$\frac{\partial}{\partial t} \ln p_{\overline{HH}}(s, t) = -(\sigma(t) + \mu(t))$$

or equivalently, changing the variable from t to u :

$$\frac{\partial}{\partial u} \ln p_{\overline{HH}}(s, u) = -(\sigma(u) + \mu(u))$$

Integrating both sides with respect to u between the limits of s and t then gives:

$$[\ln p_{\overline{HH}}(s, u)]_{u=s}^{u=t} = - \int_s^t (\sigma(u) + \mu(u)) du$$

However, since $p_{\overline{HH}}(s, s) = 1$ and $\ln 1 = 0$, this simplifies to:

$$\ln p_{\overline{HH}}(s, t) = - \int_s^t (\sigma(u) + \mu(u)) du$$

Taking exponentials, we obtain the result:

$$p_{\overline{HH}}(s, t) = \exp\left(- \int_s^t (\sigma(u) + \mu(u)) du\right)$$

Assuming the integral can be evaluated, we have an explicit expression for $p_{\overline{HH}}(s, t)$.

3 Occupancy probabilities

We have just seen that $p_{\overline{HH}}(s, t) = \exp\left(-\int_s^t (\sigma(u) + \mu(u)) du\right)$. This result can be generalised to give an expression for the probability of staying in a general state i between time s and time t (otherwise known as the occupancy probability for state i).



Occupancy probabilities for time-inhomogeneous Markov jump processes

For a time-inhomogeneous Markov jump process:

$$p_{ii}^-(s, t) = \exp\left(-\int_0^{t-s} \lambda_i(s+u) du\right) = \exp\left(-\int_s^t \lambda_i(u) du\right)$$

So the probability that a process in state i at time s remains in state i until at least time t is given by:

$$\begin{aligned} & \exp\left(-\int_0^{t-s} (\text{total force of transition out of state } i \text{ at time } s+u) du\right) \\ &= \exp\left(-\int_s^t (\text{total force of transition out of state } i \text{ at time } u) du\right) \end{aligned}$$

Note that if the transition rates are constant (*ie* the process is time-homogeneous), the occupancy probabilities simplify to:

$$p_{ii}^-(s, t) = e^{-\lambda_i(t-s)}$$

and the holding time in state i has an $\text{Exp}(\lambda_i)$ distribution. We saw this result in Chapter 5.

4 Kolmogorov's backward differential equations

Just as we did in the time-homogeneous case, we should be able to derive and to write down Kolmogorov's backward differential equations in the time-inhomogeneous case.



Kolmogorov's backward differential equations (time-inhomogeneous case)

The matrix form of Kolmogorov's backward equations is:

$$\frac{\partial}{\partial s} P(s, t) = -A(s)P(s, t)$$

It is still the case that:

$$\mu_{ii}(s) = - \sum_{j \neq i} \mu_{ij}(s)$$

Hence each row of the matrix $A(s)$ has zero sum.

Written in non-matrix form the equations are:

$$\frac{\partial}{\partial s} p_{ij}(s, t) = - \sum_{k \in S} \mu_{ik}(s) p_{kj}(s, t)$$

This result is given in Section 7.1 on Page 38 of the *Tables*.

There are a couple of particular points to note here:

- We are now differentiating with respect to s rather than t .
- There is a minus sign on the RHS.

We can derive the backward differential equations as follows. Start with the Chapman-Kolmogorov equation:

$$p_{ij}(s - h, t) = \sum_{k \in S} p_{ik}(s - h, s) p_{kj}(s, t)$$

If we suppose that h is small, then for $k \neq i$:

$$p_{ik}(s-h, s) = h \mu_{ik}(s-h) + o(h)$$

and:

$$p_{ii}(s-h, s) = 1 + h \mu_{ii}(s-h) + o(h)$$

So:

$$\begin{aligned} p_{ij}(s-h, t) &= \sum_{k \neq i} h \mu_{ik}(s-h) p_{kj}(s, t) + (1 + h \mu_{ii}(s-h)) p_{ij}(s, t) + o(h) \\ &= p_{ij}(s, t) + \sum_{k \in S} h \mu_{ik}(s-h) p_{kj}(s, t) + o(h) \end{aligned}$$

Rearranging gives:

$$\frac{p_{ij}(s-h, t) - p_{ij}(s, t)}{h} = \sum_{k \in S} \mu_{ik}(s-h) p_{kj}(s, t) + \frac{o(h)}{h}$$

or equivalently:

$$\frac{p_{ij}(s, t) - p_{ij}(s-h, t)}{h} = - \sum_{k \in S} \mu_{ik}(s-h) p_{kj}(s, t) + \frac{o(h)}{h}$$

Letting $h \rightarrow 0$ we obtain:

$$\frac{\partial}{\partial s} p_{ij}(s, t) = - \sum_{k \in S} \mu_{ik}(s) p_{kj}(s, t)$$

Alternatively, we could derive this result by showing that $\mu_{ij}(s) = - \left[\frac{\partial}{\partial s} p_{ij}(s, t) \right]_{t=s}$

and then differentiating the Chapman-Kolmogorov equation:

$$p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, u) p_{kj}(u, t)$$

with respect to s .

**Question 6.3**

Show that $\mu_{ij}(s) = -\left[\frac{\partial}{\partial s} p_{ij}(s,t)\right]_{t=s}$ and hence derive the backward equations.

We now look at a couple of examples of backward differential equations based on the time-inhomogeneous HSD model.

**Example**

Write down the Kolmogorov backward differential equation for $p_{HH}(s,t)$.

Solution

Using the general backward differential equation as a template, we obtain:

$$\frac{\partial}{\partial s} p_{HH}(s,t) = -[\mu_{HH}(s)p_{HH}(s,t) + \mu_{HS}(s)p_{SH}(s,t) + \mu_{HD}(s)p_{DH}(s,t)]$$

Now substituting in for the transition rates, we have:

$$\begin{aligned}\frac{\partial}{\partial s} p_{HH}(s,t) &= -[-(\sigma(s) + \mu(s))p_{HH}(s,t) + \sigma(s)p_{SH}(s,t)] \\ &= (\sigma(s) + \mu(s))p_{HH}(s,t) - \sigma(s)p_{SH}(s,t)\end{aligned}$$

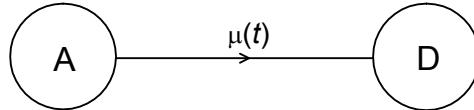
**Question 6.4**

Write down the Kolmogorov backward differential equation for $p_{HS}(s,t)$.

The general theory of time-inhomogeneous Markov jump processes is rather too complicated to fall within the scope of the current syllabus, but the methods used can be illustrated by means of a number of practical examples.

5 Example – a two-state model

Consider the following *survival model*: transition from the alive state A to the dead state D takes place at rate $\mu_{AD}(t)$, which has been abbreviated to $\mu(t)$ here, since it is the only transition in the model.



In other words $\mu(t)$ is the force of mortality.

The theory of mortality functions, including the force of mortality, is discussed in detail in Chapter 7. The two-state model was discussed in Chapter 4, where we developed the results without using the generator matrix.

Since $A(t) = \begin{pmatrix} -\mu(t) & \mu(t) \\ 0 & 0 \end{pmatrix}$ the forward equations give:

$$\frac{\partial}{\partial t} p_{AA}(s, t) = -p_{AA}(s, t) \mu(t)$$

The solution corresponding to the initial condition $p_{AA}(s, s) = 1$ is:

$$p_{AA}(s, t) = e^{-\int_s^t \mu(x) dx}$$

Since $P_{AA}(s, t) = P_{AA}(s, t)$, this result should be familiar from Section 2.2.



Question 6.5

Write down the backward differential equation for $p_{AA}(s, t)$ and show that the solution of this equation is also $p_{AA}(s, t) = e^{-\int_s^t \mu(x) dx}$.

Equivalently the probability for an individual aged s to survive for a further period of length at least w is:

$$w p_s = p_{AA}(s, s + w) = e^{-\int_s^{s+w} \mu(x) dx} = e^{-\int_0^w \mu(s+y) dy} \quad (6.1)$$

Recall that ${}_w p_s$ denotes the probability that a person now aged s is still alive in w years' time.

**Question 6.6**

- (i) A life aged 60 has a constant force of mortality of 0.01 pa . Calculate the probability that he survives to age 70.
- (ii) Calculate the probability that a 25-year old with a constant force of mortality of 0.01 pa survives to age 35.
- (iii) Comment on your answers.

This illustrates the need for time-dependent rates in mortality and many other actuarial models: a constant force of mortality μ would give rise to an age-independent survival probability ${}_w p_s$, an absurd result.

6 Residual holding times

As it stands, (6.1) is peculiar to the specific survival model under consideration; however, if properly reinterpreted it is but an instance of a general formula.

We have seen this already in Section 3, where we discussed occupancy probabilities. The general result is:

$$p_{ii}^-(s, t) = \exp\left(-\int_0^{t-s} \lambda_i(s+u) du\right) = \exp\left(-\int_s^t \lambda_i(u) du\right)$$

For a general Markov jump process, X_t , define the *residual holding time* R_s as the (random) amount of time between s and the next jump:

$$\{R_s > w, X_s = i\} = \{X_u = i, s \leq u \leq s + w\}$$

The residual holding time at time s is the amount of time after time s for which the process stays in the current state. The Core Reading equation above says that for the residual holding time at time s to be greater than w , given that we are in state i at time s , the process must stay in state i for all times u between s and $s+w$.

Formula (6.1) gives the probability that $R_s > w$ given that the state at time s is A. In general one can prove:

$$P[R_s > w | X_s = i] = e^{-\int_s^{s+w} \lambda_i(t) dt} \quad (6.2)$$

by following the same steps as the proof in Section 7 of Chapter 5.

Note that $P[R_s > w | X_s = i] = p_{ii}^-(s, s+w)$.

Moreover, the characterisation of the state:

$$X_s^+ = X_{s+R_s}$$

to which the jump takes place is similar to the time-homogeneous case:

$$P[X_s^+ = j | X_s = i, R_s = w] = \frac{\mu_{ij}(s+w)}{\lambda_i(s+w)} \quad (6.3)$$

Let's restate this result in words:



Probability that the process goes into state j when it leaves state i

Given that the process is in state i at time s and it stays there until time $s+w$, the probability that it moves into state j when it leaves state i at time $s+w$ is:

$$\frac{\mu_{ij}(s+w)}{\lambda_i(s+w)} = \frac{\text{the force of transition from state } i \text{ to state } j \text{ at time } s+w}{\text{the total force out of state } i \text{ at time } s+w}$$



Question 6.7

Show that, for $w > 0$, the density function for the random variable $R_s | X_s = i$ is given by:

$$f_{R_s|X_s=i}(w) = \lambda_i(s+w) \exp\left(-\int_s^{s+w} \lambda_i(t) dt\right)$$

We now have information on both the time that transitions take place, and to which states the transitions are made. By combining these we can calculate general transition probabilities, as described below. To do this we condition on both residual holding times (a continuous variable, therefore we use integration over probability densities) and current states (discrete, therefore we use summation over probabilities).

7 Integrated form of the Kolmogorov backward equations

The above is more than a neat picture for the behaviour of Markov jump processes: it is also a powerful computational tool. Indeed, conditioning on R_s and X_s^+ we have using the law of total probability:

$$\begin{aligned} p_{ij}(s, t) &= P[X_t = j | X_s = i] \\ &= \sum_{l \neq i} \int_0^{t-s} e^{-\int_s^{s+w} \lambda_l(u) du} \mu_{il}(s+w) P[X_t = j | X_s = i, R_s = w, X_s^+ = l] dw \end{aligned}$$

and therefore:

$$p_{ij}(s, t) = \sum_{l \neq i} \int_0^{t-s} e^{-\int_s^{s+w} \lambda_l(u) du} \mu_{il}(s+w) p_{lj}(s+w, t) dw \quad (6.4)$$

provided $j \neq i$.

This is the *integrated form of the backward equation*, as can be checked by differentiation with respect to s .

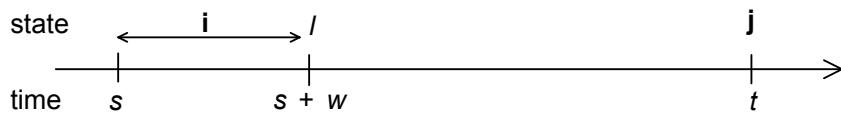
The formula may look intimidating but it conforms to intuition: since $j \neq i$, the process must jump out of i at some stage. By (6.2), the first jump after time s takes place at $s+w$ with probability density:

$$\lambda_i(s+w) e^{-\int_s^{s+w} \lambda_i(u) du}$$

We saw this in Question 6.7.

Backward integral equations always focus on the time on the first transition. Here we are thinking about the first transition occurring at time $s+w$.

By (6.3) the process jumps to l at time $s+w$ with probability $\frac{\mu_{il}(s+w)}{\lambda_i(s+w)}$. It then remains to effect a transition from l to j over the remaining time period $[s+w, t]$:



We can also reason as follows. The expression:

$$\exp\left(-\int_s^{s+w} \lambda_i(u) du\right) \mu_{il}(s+w) p_{lj}(s+w, t)$$

can be considered as the product of three terms:

- the probability of remaining in state i from time s to time $s+w$
- then making a transition to state l at time $s+w$
- and finally going from state l to state j between time $s+w$ and time t .

To take into account the possible values of w we integrate with respect to w from $w=0$ to $t-s$, and to take into account all possible intermediate states we sum over all possible values of $l \neq i$.

Note also that the exponential term in the expression above can also be written as $p_{ii}^-(s, s+w)$, so the backward integral equation can also be written as:

$$p_{ij}(s, t) = \sum_{l \neq i} \int_0^{t-s} p_{ii}^-(s, s+w) \mu_{il}(s+w) p_{lj}(s+w, t) dw$$

for $j \neq i$.



Question 6.8

Check that you obtain the backward equations by differentiating with respect to s .

Note that Equation (6.4) only gives a relationship between transition probabilities; to find them explicitly you still need to solve the equations.

When $i = j$ there is an additional term $e^{-\int_s^t \lambda_i(u) du}$ because the process can remain in state i throughout $[s, t]$.

Once again, we can write $e^{-\int_s^t \lambda_i(u) du}$ as $p_{ii}^-(s, t)$. So the integrated form of the backward equation for $p_{ii}(s, t)$ is:

$$p_{ii}(s, t) = \sum_{l \neq i} \int_0^{t-s} p_{ii}^-(s, s+w) \mu_{il}(s+w) p_{li}(s+w, t) dw + p_{ii}^-(s, t)$$

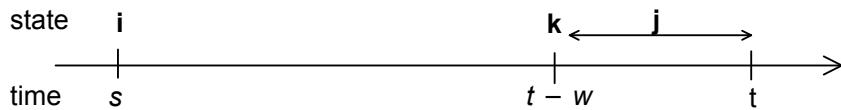
8 Integrated form of the Kolmogorov forward equations

If instead of considering the first jump after s one focuses on the last jump before t , one can obtain an intuitive derivation of the *integrated form of the forward equations*.

In the backward equation we thought about the time of the first transition as being $s + w$. For the forward equation we think about the time of the last transition as being $t - w$.

The forward equation when $i \neq j$:

$$p_{ij}(s, t) = \sum_{k \neq j} \int_0^{t-s} p_{ik}(s, t-w) \mu_{kj}(t-w) e^{-\int_{t-w}^t \lambda_j(u) du} dw \quad (6.5)$$



Again this can be written as:

$$p_{ij}(s, t) = \sum_{k \neq j} \int_0^{t-s} p_{ik}(s, t-w) \mu_{kj}(t-w) p_{jj}(t-w, t) dw$$

for $j \neq i$.

The terms in the integral are:

- the probability of going from state i to state k between time s and time $t-w$
- then making a transition from state k to state j at time $t-w$
- and staying in state j from time $t-w$ to time t .

Integrating over all possible values of w , namely 0 to $t-s$, and adding over all intermediate states $k \neq j$, we get the forward integral equation.

The forward integral equation for $p_{ii}(s, t)$ is:

$$p_{ii}(s, t) = \sum_{k \neq i} \int_0^{t-s} p_{ik}(s, t-w) \mu_{ki}(t-w) p_{ii}^-(t-w, t) dw + p_{ii}^-(s, t)$$

Here we've added on an extra term at the end to cover the possibility that the process stays in state i throughout the interval $[s, t]$.

Because of this intuitive interpretation it shouldn't be too difficult to write out the general backward and forward equations in integrated form. If you are asked to do so in specific cases then the general formula could serve either as a template, or as a check.

Derivation of the integrated form of the forward equations

For a full justification of this equation one needs to appeal to the properties of the current holding time C_t , namely the time between the last jump and t :

$$\{C_t \geq w, X_t = j\} = \{X_u = j, t - w \leq u \leq t\}$$

or the length of time that the process has been in the current state.

Alternatively, we can derive the integrated form of the forward equations by solving the differential form, as below. This method could also be used to derive the integrated form of the backward equations.

Start with the forward equations:

$$\frac{\partial}{\partial t} p_{ij}(s, t) = \sum_k p_{ik}(s, t) \mu_{kj}(t)$$

We are aiming for an expression that gives $p_{ij}(s, t)$ in terms of the other transition probabilities, so we first rewrite the forward equation in the form:

$$\frac{\partial}{\partial t} p_{ij}(s, t) - p_{ij}(s, t) \mu_{jj}(t) = \sum_{k \neq j} p_{ik}(s, t) \mu_{kj}(t)$$

Also note that $\mu_{jj}(t) = -\sum_{k \neq j} \mu_{jk}(t) = -\lambda_j(t)$ so that:

$$\frac{\partial}{\partial t} p_{ij}(s, t) + p_{ij}(s, t) \lambda_j(t) = \sum_{k \neq j} p_{ik}(s, t) \mu_{kj}(t)$$

This can be solved using the integrating factor method. Here the integral of the coefficient of $p_{ij}(s, t)$, $\lambda_j(t)$, with respect to t is a definite integral rather than an explicit function of t . The integral of the coefficient is the definite integral $\int_s^t \lambda_j(u) du$ and the integrating factor is $e^{\int_s^t \lambda_j(u) du}$. We therefore multiply by this integrating factor to give:

$$e^{\int_s^t \lambda_j(u) du} \frac{\partial}{\partial t} p_{ij}(s, t) + e^{\int_s^t \lambda_j(u) du} p_{ij}(s, t) \lambda_j(t) = e^{\int_s^t \lambda_j(u) du} \sum_{k \neq j} p_{ik}(s, t) \mu_{kj}(t)$$

The whole point of integrating factors is to note that the LHS can be simplified by noticing it is the result of the “product rule”:

$$\frac{\partial}{\partial t} \left[p_{ij}(s, t) e^{\int_s^t \lambda_j(u) du} \right] = e^{\int_s^t \lambda_j(u) du} \sum_{k \neq j} p_{ik}(s, t) \mu_{kj}(t)$$

We now change the variable from t to v and integrate both sides of the equation with respect to v between the limits of s and t . The LHS integrates to:

$$\begin{aligned} \left[p_{ij}(s, v) e^{\int_s^v \lambda_j(u) du} \right]_{v=s}^{v=t} &= p_{ij}(s, t) e^{\int_s^t \lambda_j(u) du} - p_{ij}(s, s) e^{\int_s^s \lambda_j(u) du} \\ &= p_{ij}(s, t) e^{\int_s^t \lambda_j(u) du} - p_{ij}(s, s) \end{aligned}$$

Therefore, the equation becomes:

$$p_{ij}(s, t) e^{\int_s^t \lambda_j(u) du} = p_{ij}(s, s) + \int_s^t e^{\int_s^v \lambda_j(u) du} \sum_{k \neq j} p_{ik}(s, v) \mu_{kj}(v) dv$$

The main integral on the right-hand side can be changed with the substitution $v = t - w$. Then:

$$dv = -dw$$

$$v = s \Rightarrow w = t - s$$

and:

$$v = t \Rightarrow w = 0$$

So we have:

$$p_{ij}(s, t) e^{\int_s^t \lambda_j(u) du} = p_{ij}(s, s) + \int_0^{t-s} e^{\int_s^{t-w} \lambda_j(u) du} \sum_{k \neq j} p_{ik}(s, t-w) \mu_{kj}(t-w) dw$$

Finally, we note that $p_{ij}(s, s) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$, and multiply both sides by $e^{-\int_s^t \lambda_j(u) du}$ to get:

$$p_{ij}(s, t) = \delta_{ij} e^{-\int_s^t \lambda_j(u) du} + \int_0^{t-s} e^{-\int_s^t \lambda_j(u) du} e^{\int_s^{t-w} \lambda_j(u) du} \sum_{k \neq j} p_{ik}(s, t-w) \mu_{kj}(t-w) dw$$

This simplifies to:

$$\begin{aligned} p_{ij}(s, t) &= \delta_{ij} e^{-\int_s^t \lambda_j(u) du} + \sum_{k \neq j} \int_0^{t-s} p_{ik}(s, t-w) \mu_{kj}(t-w) e^{-\int_{t-w}^t \lambda_j(u) du} dw \\ &= \delta_{ij} p_{ii}^-(s, t) + \sum_{k \neq j} \int_0^{t-s} p_{ik}(s, t-w) \mu_{kj}(t-w) p_{jj}^-(t-w, t) dw \end{aligned}$$

So we have:

$$p_{ij}(s, t) = \sum_{k \neq j} \int_0^{t-s} p_{ik}(s, t-w) \mu_{kj}(t-w) p_{jj}^-(t-w, t) dw \quad \text{for } i \neq j$$

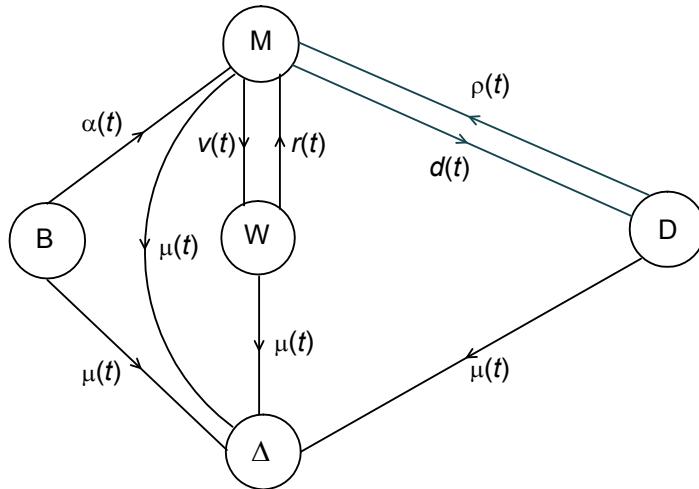
and:

$$p_{ii}(s, t) = \sum_{k \neq i} \int_0^{t-s} p_{ik}(s, t-w) \mu_{ki}(t-w) p_{ii}^-(t-w, t) dw + p_{ii}^-(s, t)$$

9 Applications

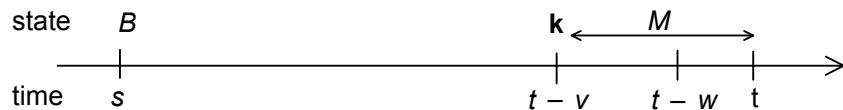
9.1 Marriage

Describe the marital status of an individual as one of the following: bachelor (never married) (B), married (M), widowed (W), divorced (D), dead (Δ). We can define a Markov jump process on the state space {B, M, W, D, Δ } as illustrated below:



In the above, the deaths rate $\mu(t)$ has been taken to be independent of the marital status for simplicity. This model can be studied exactly as the HSD example. For instance, the probability of being married at time t and of having been so for at least w given that you are a bachelor at time s is (assuming $w < t - s$):

$$\begin{aligned} P[X_t = M, C_t > w | X_s = B] &= \int_w^{t-s} [p_{BB}(s, t-v)\alpha(t-v) + p_{BW}(s, t-v)r(t-v) \\ &\quad + p_{BD}(s, t-v)\rho(t-v)] e^{-\int_{t-v}^t (\mu(u)+v(u)+d(u))du} dv \end{aligned}$$



where k is any of the states leading to M, namely B, W and D.

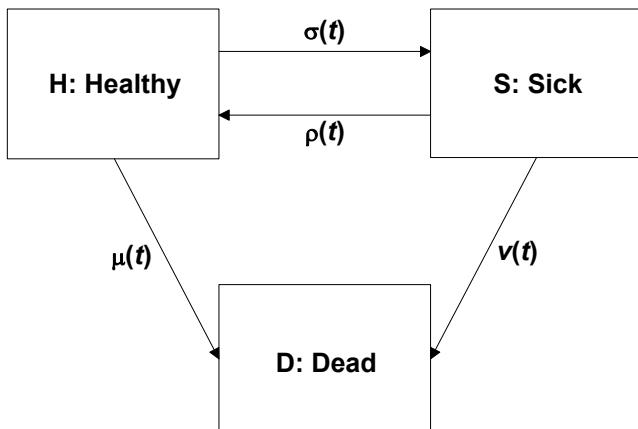
The integral on the RHS of this equation can also be written as:

$$\int_w^{t-s} \left[p_{BB}(s, t-v) \alpha(t-v) + p_{BW}(s, t-v) r(t-v) + p_{BD}(s, t-v) \rho(t-v) \right] p_{MM}(t-v, t) dv$$

9.2 Sickness and death

Here we return to the HSD model with time-dependent forces of transition.

Describe the state of a person as “healthy”, “sick” or “dead”. For given time-dependent (ie age-dependent) transition rates, we can construct a Markov jump process with state space {H, S, D}:



The matrix $A(t)$ in Kolmogorov's equations is:

$$A(t) = \begin{pmatrix} -\sigma(t) - \mu(t) & \sigma(t) & \mu(t) \\ \rho(t) & -\rho(t) - \nu(t) & \nu(t) \\ 0 & 0 & 0 \end{pmatrix}$$

In particular:

$$\lambda_H(t) = \sigma(t) + \mu(t), \quad \lambda_S(t) = \rho(t) + \nu(t) \quad \text{and} \quad \lambda_D = 0$$

Remember that λ_i denotes the total force of transition out of state i .

The easiest probabilities to calculate are those of remaining continuously healthy or continuously sick over $[s, t]$. Using (6.2) these are:

$$P[R_s > t - s | X_s = H] = e^{-\int_s^t (\sigma(u) + \mu(u)) du} \quad (6.6)$$

and:

$$P[R_s > t - s | X_s = S] = e^{-\int_s^t (\rho(u) + \nu(u)) du}$$

Note that these can also be denoted as $p_{\overline{HH}}(s, t)$ and $p_{\overline{SS}}(s, t)$, respectively. They are not the same as $p_{HH}(s, t)$ and $p_{SS}(s, t)$, which include the possibility of changing state one or more times during the interval (but returning so as to be in the original state at time t).



Question 6.9

Describe and evaluate $P[R_s > t - s | X_s = D]$.

The above equations can be used to give actual numerical values for the respective probabilities, assuming you can calculate the integrands.

However, solving Kolmogorov's equations (*ie* evaluating transition probabilities) will not be possible in the general case of non-constant transition rates. Numerical methods do however exist that can give approximate solutions, but these are beyond the scope of the CT4 syllabus.

We can also write down the integrated form of Kolmogorov's equations as below. Note that although this gives an expression for each transition probability, it does so only in terms of other *unknown* transition probabilities. In order to obtain actual transition probabilities we would still need to solve these equations.

Transition probabilities can be related to each other as in (6.4) and (6.5). For instance:

$$p_{HS}(s, t) = \int_0^{t-s} p_{SS}(s+w, t) \sigma(s+w) e^{-\int_s^{s+w} (\sigma(u) + \mu(u)) du} dw$$



This is the integrated form of the backward equation for $P_{HS}(s, t)$, which can also be written as:

$$p_{HS}(s, t) = \int_0^{t-s} p_{\overline{HH}}(s, s+w) \sigma(s+w) p_{SS}(s+w, t) dw$$

Remember that in the backward equation we can think about the time of the first transition as being $s+w$, and for the forward equation we can think about the time of the last transition as being $t-w$.



Question 6.10

Give the forward version of the above equation.



Example

Write down the integrated form of the backward equation for $p_{HD}(s, t)$.

Solution

In the backward equations we need to consider the two mutually exclusive events that the first transition from H is either to S, or to D:

$$p_{HD}(s, t) = \int_0^{t-s} p_{\overline{HH}}(s, s+w) \sigma(s+w) p_{SD}(s+w, t) dw \\ + \int_0^{t-s} p_{\overline{HH}}(s, s+w) \mu(s+w) dw$$

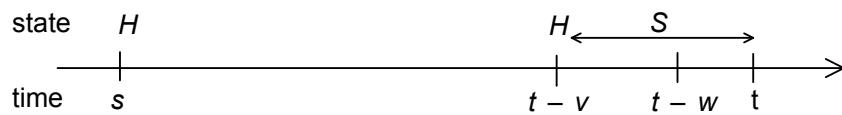
where we've used the fact that $p_{DD}(s+w, t) = 1$.

**Question 6.11**

Write down the integrated form of the backward equation for $p_{SH}(s, t)$.

Extra conditions on residual or current holding times can be handled without difficulty. Consider for instance the probability of being sick at time t and of having been so for at least w , given that you are healthy at time s . This is:

$$P[X_t = S, C_t > w | X_s = H] = \int_w^{t-s} p_{HH}(s, t-v) \sigma(t-v) e^{-\int_v^t (\rho(u)+\nu(u))du} dv$$



This equation can also be written as:

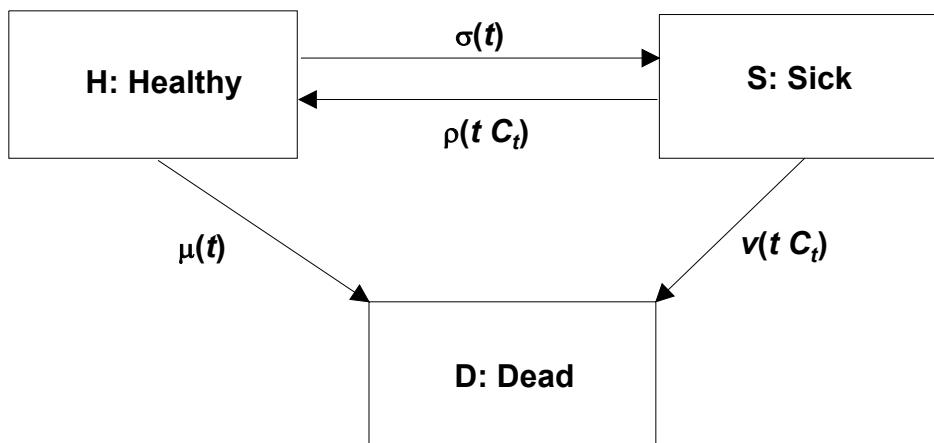
$$P[X_t = S, C_t > w | X_s = H] = \int_w^{t-s} p_{HH}(s, t-v) \sigma(t-v) p_{SS}(t-v, t) dv$$

9.3 Sickness and death with duration dependence

In Section 9.2, the Markov property implies that:

$$P[X_t = H \mid X_s = S, C_s = w] = P[X_t = H \mid X_s = S]$$

In other words, the duration of your current illness has no bearing on your future health prospects. In order to remove this undesirable feature, we modify the model by allowing the rates of transition out of S to depend on the current holding time C_t :



Question 6.12

Why do you think this model hasn't made the transitions from the healthy state dependent on the holding time?

This appears to take us outside the scope of this unit, as the value of C_t must now be incorporated into the state, so that the state space is not countable (ie discrete as opposed to continuous) any more.

However, the framework from above can still be used provided that there is careful conditioning on the relevant current holding time.

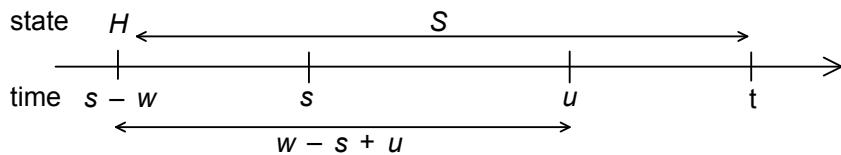
In fact, since the transition rates σ and μ do not depend on C_t the probability of remaining continuously healthy during $[s, t]$ is given by (6.6) as before.

This is because the overall rate out of H is $\sigma(t) + \mu(t)$ so that:

$$p_{\overline{HH}}(s, t) = \exp\left(-\int_s^t (\sigma(u) + \mu(u)) du\right) = \exp\left(-\int_0^{t-s} (\sigma(s+u) + \mu(s+u)) du\right)$$

This is unaffected by the fact that recovery rates and mortality rates for sick people depend on how long you have been sick.

On the other hand, to calculate the probability of remaining continuously sick during $[s, t]$ given a current illness period $[s-w, s]$, one needs to update the values of ρ and ν as the illness progresses:



$$P[X_t = S, R_s > t-s | X_s = S, C_s = w] = e^{-\int_s^t (\rho(u, w-s+u) + \nu(u, w-s+u)) du}$$

If there were no duration dependence, this expression would simplify to:

$$p_{\overline{SS}}(s, t) = \exp\left(-\int_s^t (\rho(u) + \nu(u)) du\right) = \exp\left(-\int_0^{t-s} (\rho(s+u) + \nu(s+u)) du\right)$$

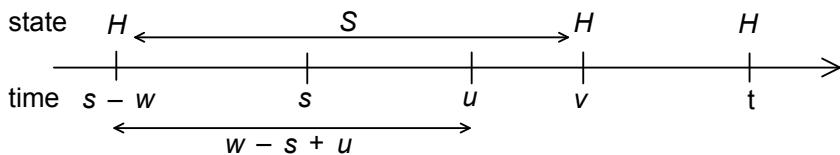
However, when there is duration dependence, it must be taken into account whenever ρ and ν occur. In the integral, u denotes a time between s and t . For a given u , the duration is the length of time the person has been sick. Since at time s the duration is w , the duration at time u must be $w+(u-s)$. This explains the $w-s+u$ terms in the probability above. The Core Reading equation above can also be written as:

$$p_{\overline{S_w S}}(s, t) = \exp\left(-\int_0^{t-s} (\rho(s+u, w+u) + \nu(s+u, w+u)) du\right)$$

The subscript of w in the $p_{\overline{S_w S}}(s, t)$ term indicates that the life has already been sick for w years at time s . So if the life stays in the sick state up to time $s+u$, he will then have duration of sickness $w+u$. You may find this form of the expression easier to work with. If you make the substitution $r = s+u$, you can show that the two versions are equivalent.

As a final example, the probability of being healthy at time t given that you are sick at time s with current illness duration w can be written as:

$$\begin{aligned} p_{S_w H}(s, t) &= P[X_t = H \mid X_s = S, C_s = w] \\ &= \int_s^t e^{-\int_u^v (\rho(u, w-s+u) + \nu(u, w-s+u)) du} \rho(v, w-s+v) p_{HH}(v, t) dv \end{aligned}$$



Again, this is the same as the formula without duration dependence, but with the transition rates ρ and ν modified as necessary. We can also write this expression as:

$$p_{S_w H}(s, t) = \int_s^t p_{\overline{S}_w S}(s, v) \rho(v, w-s+v) p_{HH}(v, t) dv$$

or, to be consistent with the formulation we have been using so far for backward integral equations:

$$p_{S_w H}(s, t) = \int_0^{t-s} p_{\overline{S}_w S}(s, s+v) \rho(s+v, w+v) p_{HH}(s+v, t) dv$$

This is saying that we remain sick throughout the time period s to $s+v$, then make a transition to healthy at time $s+v$ and duration of sickness $w+v$, and finally go from healthy at time $s+v$ to healthy at time t , though you may be sick in between these times.



Question 6.13

Write down an integral expression for the probability of being sick at time t , having been so for at least w years, given that you were healthy at time s .

**Question 6.14**

Consider again the marriage model in Section 9.1, only now assume that the transition rate $d(t)$ depends on the current holding time. (So the chance of divorce depends on how long you have been married.) Write down expressions for the probability that:

- (i) a bachelor remains a bachelor throughout a period $[s, t]$
- (ii) a person who gets married at time $s - w$ and remains married throughout $[s - w, s]$, continues to be married throughout $[s, t]$
- (iii) a person is married at time t and has been so for at least time w , given that he/she was divorced at time $s < t - w$.

10 Modelling and simulation

This section is similar to the modelling and simulation section of Chapter 3 that dealt with Markov chains.

Modelling is discussed first. We deal with Poisson models initially, including time-inhomogeneous Poisson processes (as introduced below), before proceeding to more general homogeneous processes, and finally dealing with inhomogeneous processes.

A short discussion of simulation is then given.

10.1 Time-homogeneous Poisson process models

A time-homogeneous Poisson process has a single parameter λ . The estimation of this parameter given a collection of data is straightforward.



Example

An insurance office observes that m claims arrive in a total of T time units. If the company decides that a Poisson process model is appropriate, the most suitable estimate for λ would appear to be $\hat{\lambda} = m/T$. This intuitive estimate is confirmed by more formal procedures such as maximum likelihood estimation.

You should be familiar with maximum likelihood estimation from Subject CT3.

Having estimated the parameter, all that remains is to test goodness of fit.

Note that it is a basic assumption here that a Poisson process is appropriate. If this is not a reasonable assumption then the fit may not be very good. So, if a goodness of fit test gives a result that would lead you to reject the null hypothesis, then an alternative model may be appropriate.

The test is carried out as follows.

Divide the total time T into k equal intervals. If the Poisson process model fits, the number of claims arriving in the k intervals should form a sequence of independent Poisson variates, each with mean $\lambda T/k$. There are two things to test here:

- (1) whether the distribution is Poisson and
- (2) whether the observations are independent.

A standard χ^2 goodness-of-fit test can be employed to determine whether the Poisson distribution fits.

Assuming that the fit is adequate, independence is probably best tested against the alternative that there is some form of serial dependence, implying that, for example, $\text{cov}(N_{(i+1)T/k} - N_{iT/k}, N_{iT/k} - N_{(i-1)T/k})$ is non-zero.

So the monthly claims arriving may not be uncorrelated with the previous months, for example.

Tests for serial correlation are covered in Chapter 12. Time series models are covered in Subject CT6.

10.2 Time-inhomogeneous Poisson process models

In some classes of business, such as insurance against storm damage, the intensity of arrival of claims may vary predictably with time, in the sense that the insurer can tell in advance that some time intervals will have more claim arrivals than other intervals of equal length. A suitable model here is the *time-inhomogeneous Poisson process*, for which the arrival rate of claims is a function $\lambda(t)$. In the given example λ will be periodic, with a period of one year.

It is impractical to attempt to estimate the value of $\lambda(t)$ separately for each value of t . A common procedure is to divide the whole time period up into pieces of a suitable size and to estimate the arrival rate separately for each piece.

Since t is continuous you cannot hope to have enough data to make the former estimation procedure statistically significant. The same applies if the pieces are too small.

Thus data for a whole year may be divided into months, giving 12 estimated claim arrival rates. Tests of goodness of fit should be carried out for each month separately, but tests for serial correlation should use the whole data set at once.

For example, if you had several years of monthly data, then you could think of all the January months together as a time-homogeneous Poisson process with a certain fixed parameter λ . This could be tested for goodness of fit separately.

On the other hand, when testing for serial correlation, you would need to test, for example, whether one month is correlated with the previous month, on average. You would therefore be using the whole data set at once.

10.3 Time-homogeneous Markov models

The structural analysis of Section 7 of Chapter 5 is of paramount importance when it comes to modelling continuous-time Markov jump processes. Recall that each visit to any given state i is of exponential duration with mean $1/\lambda_i$ and is independent of the durations of previous visits to that state and of the destination after the next jump. Further, the probability that the next transition is to state j is μ_{ij}/λ_i .

This suggests that it is feasible to separate the two estimation procedures.

- First the λ_i may be estimated: look at the data for the durations of visits to state i and let $1/\hat{\lambda}$ be equal to the sample mean of this collection.
- Next proceed as in Chapter 3: let n_i be the number of completed visits to state i , n_{ij} the number of direct transitions from state i to state j , and set $\hat{p}_{ij} = n_{ij} / n_i$. Since p_{ij} is equal to μ_{ij}/λ_i , a sensible estimator for μ_{ij} is $\hat{\mu}_{ij} = \hat{\lambda}_i \hat{p}_{ij}$.

We have already seen in Section 11 of Chapter 5 that the transition rate μ_{ij} is estimated by:

$$\hat{\mu}_{ij} = \frac{\text{number of transitions from state } i \text{ to state } j}{\text{total holding time in state } i}$$

This formula is equivalent to the one in the second bullet point above.

Also, λ_i can be estimated by $\hat{\lambda}_i = -\sum_{j \neq i} \hat{\mu}_{ij}$.

Tests for goodness of fit are more problematical, if only because there is a vast collection of possible alternative hypotheses. It is reasonable to test whether the visits to a given state really are exponentially distributed: a χ^2 goodness-of-fit test will do this. It is also reasonable to test whether the jump chain really does exhibit the Markov property: see Chapter 3 for a discussion. But there are other implications of the Markov structure that should be tested and the procedure is not always clear.

For example, to derive a formal test as to whether the destination of a jump is independent of the duration of the previous holding time we would need to do something like this:

- look at all visits to state i and classify them as long-duration, medium-duration or short-duration
- for each duration category, estimate the transition probabilities of the jump chain separately, giving estimates $\hat{p}_{ij}^{(L)}$, $\hat{p}_{ij}^{(M)}$ and $\hat{p}_{ij}^{(S)}$
- determine whether the differences between the sets of estimated transition probabilities are significant.

However, it is by no means clear what test statistic could be employed or what its distribution might be. In practice the investigation of this question would be accomplished graphically: for each visit to state i , plot a point on a graph whose x -coordinate represents the duration of the visit, y -coordinate the destination of the next jump. If a pattern appears, reject the assumption of independence.

Other tests, such as testing whether the first visit to a given state is significantly longer than subsequent visits, are also best treated graphically.

10.4 Time-inhomogeneous Markov models

The structural decomposition of the time-homogeneous Markov model does not apply to the time-inhomogeneous case. The estimation of time-dependent transition rates, such as the force of mortality or age-dependent rate of recovery from sickness, is best treated within the context of the particular model being studied.

10.5 Simulation

In order to simulate a process, random values of the random variables that are involved must be produced.

There are two approaches to the task of simulating a time-homogeneous Markov jump process. The first is an approximate method and the second exact.

Approximate method

Divide time into very short intervals of width h , say, where $\mu_{ij}h$ is much smaller than 1 for each i and j . The transition matrix $P(h)$ of the Markov chain has entries approximately given by:

$$p_{ij}^*(h) = \delta_{ij} + h\mu_{ij}$$

Using the techniques of Chapter 3 we may simulate a discrete-time Markov chain $\{Y_n, n \geq 0\}$ with these transition probabilities, then write $X_t = Y_{[t/h]}$.

For example, if $h = \frac{1}{100}$, we would simulate $\{Y_n, n \geq 0\}$ and define $X_t = Y_{[100t]}$, ie:

$$X_t = \begin{cases} Y_0 & \text{for } 0 \leq t < 0.01 \\ Y_1 & \text{for } 0.01 \leq t < 0.02 \\ Y_2 & \text{for } 0.02 \leq t < 0.03 \\ \vdots & \vdots \end{cases}$$

This simplistic method is not very satisfactory, as its long-term distribution may differ significantly from that of the process being modelled.

Since the probabilities being used are not exact, the errors introduced accumulate as time passes. In the long run they may be significant.

A much improved version of this method is available, which uses the exact transition probabilities $p_{ij}(h)$ instead of $p_{ij}^*(h)$, but this naturally requires that the exact probabilities be calculated in advance. General techniques for such calculations, where not covered by this chapter, are beyond the scope of the syllabus.

Exact method

This takes advantage of the structural decomposition of the jump process. First simulate the jump chain of the process as a Markov chain with transition probabilities $p_{ij} = \mu_{ij} / \lambda_i$. Once the path $\{\hat{X}_n : n = 0, 1, \dots\}$ has been generated, the holding times $\{T_n : n = 0, 1, \dots\}$ are a sequence of independent exponential random variables, T_n having rate parameter given by $\lambda_{\hat{X}_n}$.



Example

We will describe how to use the exact method to simulate a sample path for a Health-Sickness-Death model with generator matrix:

$$\begin{matrix} H & S & D \\ \begin{bmatrix} -0.5 & 0.4 & 0.1 \\ 0.6 & -0.8 & 0.2 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

We assume that a policyholder begins in the healthy state.

Solution

The transition matrix of the Markov jump chain is:

$$\begin{matrix} H & S & D \\ \begin{bmatrix} 0 & \frac{0.4}{0.5} = 0.80 & \frac{0.1}{0.5} = 0.20 \\ \frac{0.6}{0.8} = 0.75 & 0 & \frac{0.2}{0.8} = 0.25 \\ 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

Each probability is the ratio of the force between the two states and the total of the forces on paths leaving the initial state. Once the process enters state D it remains there for ever.

The holding time in the healthy state is $\text{Exp}(0.5)$ and the holding time in the sick state is $\text{Exp}(0.8)$.

The first step is to simulate the states occupied by the Markov jump chain.

Row 1 of the transition matrix is the conditional distribution of X_1 given that $X_0 = H$. We use Monte Carlo simulation to generate a simulated value for X_1 .

If the simulated value is D , then the simulation of the sample path is complete because the process never leaves state D . If the simulated value is S , then we use row 2 of the transition matrix, which is the conditional distribution of X_2 given that $X_1 = S$, to simulate a value for X_2 .

This process is repeated to simulate additional values of the Markov jump chain.

The second step is to simulate the holding times corresponding to the states in the simulated Markov jump chain.

The holding times for each occupancy of state H will be simulated from an $\text{Exp}(0.5)$ distribution. We use Monte Carlo simulation to generate these values. The same method can be used to generate $\text{Exp}(0.8)$ random variables for each holding time in state S .

By adding up the holding times to match the states simulated from the Markov jump chain, we will obtain the simulated times at which the Markov process jumps between states.

Time-inhomogeneous processes

Given the transition rates of a time-inhomogeneous Markov chain and given the state X_t at time t , it is in principle possible to determine the density function of the time until the next transition and the destination of the next jump: see Section 9 for examples. This means that standard simulation techniques can be deployed to generate an exact simulation of the process.

In practice, however, such a procedure is cumbersome in the extreme, unless the number of states is very small, and a more usual approach is to use the approximate method outlined above. The exact transition probabilities $p_{ij}(t, t+h)$ will seldom be to hand, meaning that the less satisfactory approximate values $p_{ij}^*(t, t+h) = \delta_{ij} + h\mu_{ij}(t)$ must be used instead. The method is acceptable for short-term simulations but is unreliable in the long term.

As above, the errors introduced will accumulate so that the long-term simulation is not acceptable.

11 Exam-style questions

You should now be able to attempt the following past exam questions.



Question 6.15

Subject CT4, September 2005 , Question A7

A time-inhomogeneous Markov jump process has state space {A, B} and the transition rate for switching between states equals $2t$, regardless of the state currently occupied, where t is time.

The process starts in state A at $t = 0$.

- (i) Calculate the probability that the process remains in state A until at least time s . [2]
 - (ii) Show that the probability that the process is in state B at time T , and that it is in the first visit to state B, is given by $T^2 \times e^{-T^2}$. [3]
 - (iii)
 - (a) Sketch the probability function given in (ii).
 - (b) Give an explanation of the shape of the probability function.
 - (c) Calculate the time at which it is most likely that the process is in its first visit to state B. [6]
- [Total 11]

As we said in Section 1 of Chapter 5, there are many different notations that can be used for transition probabilities and transition rates. In many past exam questions you will see the notation the following notation used:

- ${}_t p_x^{ij}$ denotes the probability that a life in state i at age x is in state j at age $x + t$
- ${}_t \bar{p}_x^{ii}$ denotes the probability that a life in state i at age x remains in state i until at least age $x + t$.

Here is a question that uses this notation. The question also uses the notation μ_{x+t}^{ij} to denote the force of transition from state i to state j at age $x + t$.

**Question 6.16****Subject 104, April 2003, Question 10**

An illness-death model has three states:

- 1 = healthy
- 2 = sick
- 3 = dead

- (i) Draw and label a diagram showing the three states and the transition intensities between them. [2]
- (ii) Show, from first principles, that in this illness-death model:

$$\frac{\partial}{\partial t} {}_t p_x^{12} = {}_t p_x^{11} \mu_{x+t}^{12} - {}_t p_x^{12} \mu_{x+t}^{21} - {}_t p_x^{12} \mu_{x+t}^{23} \quad [6]$$

[Total 8]

12 End of Part 2

You have now completed Part 2 of the Subject CT4 Notes.

Review

Before looking at the Question and Answer Bank we recommend that you briefly review the key areas of Part 2, or maybe re-read the summaries at the end of Chapters 5 and 6.

Question and Answer Bank

You should now be able to answer the questions in Part 2 of the Question and Answer Bank. We recommend that you work through several of these questions now and save the remainder for use as part of your revision.

Assignments

On completing this part, you should be able to attempt the questions in Assignment X2.

Reminder

If you have not yet booked a tutorial, then maybe now is the time to do so.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.



Chapter 6 Summary

Chapman-Kolmogorov equations

$$p_{ij}(s, t) = \sum_k p_{ik}(s, u) p_{kj}(u, t)$$

Here u is any intermediate time between s and t (possibly equal to s or t) that is convenient for the calculation.

Transition rates

$$\mu_{ij}(s) = \left[\frac{d}{dt} p_{ij}(s, t) \right]_{t=s} = \lim_{h \rightarrow 0} \frac{p_{ij}(s, s+h) - p_{ij}(s, s)}{h}$$

This is equivalent to:

$$p_{ij}(s, s+h) = \begin{cases} h\mu_{ij}(s) + o(h) & \text{if } i \neq j \\ 1 + h\mu_{ii}(s) + o(h) & \text{if } i = j \end{cases}$$

for small h .

Generator matrix

The generator matrix is the matrix of transition rates $\mu_{ij}(t)$. It is usually denoted by $A(t)$. Each row of the generator matrix $A(t)$ sums to zero since $\mu_{ii}(t) = -\sum_{j \neq i} \mu_{ij}(t)$.

Backward and forward differential equations (time-inhomogeneous case)

$$\text{Forward: } \frac{\partial}{\partial t} p_{ij}(s, t) = \sum_k p_{ik}(s, t) \mu_{kj}(t)$$

$$\frac{\partial}{\partial t} P(s, t) = P(s, t) A(t) \quad (\text{matrix form})$$

$$\text{Backward: } \frac{\partial}{\partial s} p_{ij}(s, t) = - \sum_k \mu_{ik}(s) p_{kj}(s, t)$$

$$\frac{\partial}{\partial s} P(s, t) = -A(s)P(s, t) \quad (\text{matrix form})$$

Occupancy probabilities

The probability of remaining in state i throughout the interval (s, t) is:

$$p_{ii}^-(s, t) = \exp\left(-\int_0^{t-s} \lambda_i(s+u) du\right) = \exp\left(-\int_s^t \lambda_i(u) du\right)$$

where $\lambda_i(u)$ is the total force of transition out of state i at time u .

Probability that the process goes into state j when it leaves state i

Given that the process is in state i at time s and it stays there until time $s+w$, the probability that it moves into state j when it leaves state i at time $s+w$ is:

$$\frac{\mu_{ij}(s+w)}{\lambda_i(s+w)} = \frac{\text{the force of transition from state } i \text{ to state } j \text{ at time } s+w}{\text{the total force out of state } i \text{ at time } s+w}$$

Backward and forward integral equations

$$\text{Backward: } p_{ij}(s, t) = \sum_{l \neq i} \int_0^{t-s} p_{ii}^-(s, s+w) \mu_{il}(s+w) p_{lj}(s+w, t) dw \quad i \neq j$$

The backward equation is obtained by considering the timing and nature of the first jump after time s . The duration spent in this initial state before jumping to another state (state k , say) is denoted by w . The integral reflects the three stages involved:

- (1) remaining in state i from time s to time $s+w$
- (2) jumping from state i to state k at time $s+w$
- (3) moving from state k at time $s+w$ to state j at time t (possibly visiting other states along the way).

We then consider the possible values of w to obtain limits of 0 and $t-s$ for the integral, and we sum over all possible intermediate states k .

When $i = j$, the equation is:

$$p_{ii}(s, t) = \sum_{l \neq i} \int_0^{t-s} p_{ii}^-(s, s+w) \mu_{il}(s+w) p_{li}(s+w, t) dw + p_{ii}^-(s, t)$$

The extra term here is to account for the possibility of staying in state i from time s to time t .

$$\text{Forward: } p_{ij}(s, t) = \sum_{k \neq j} \int_0^{t-s} p_{ik}(s, t-w) \mu_{kj}(t-w) p_{jj}^-(t-w, t) dw \quad i \neq j$$

The forward equation is obtained by considering the timing and nature of the last jump before time t . The duration then spent in this final state (state k , say) before time t is denoted by w . The integral reflects the three stages involved:

- (1) moving from state i at time s to state k at time $t-w$ (possibly visiting other states along the way)
- (2) jumping from state k to state j at time $t-w$
- (3) remaining in state j from time $t-w$ to time t .

We then consider the possible values of w to obtain limits of 0 and $t-s$ for the integral, and sum over all possible intermediate states k .

When $i = j$, the equation is:

$$p_{ii}(s, t) = \sum_{k \neq i} \int_0^{t-s} p_{ik}(s, t-w) \mu_{ki}(t-w) p_{ii}^-(t-w, t) dw + p_{ii}^-(s, t)$$

The extra term here is to account for the possibility of staying in state i from time s to time t .

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 6 Solutions

Solution 6.1

$$\begin{aligned}\frac{\partial}{\partial t} p_{HD}(s,t) &= p_{HH}(s,t)\mu_{HD}(t) + p_{HS}(s,t)\mu_{SD}(t) + p_{HD}(s,t)\mu_{DD}(t) \\ &= p_{HH}(s,t)\mu(t) + p_{HS}(s,t)v(t)\end{aligned}$$

Solution 6.2

We consider $p_{\overline{HH}}(s, t+h)$ and condition on the state at time t . In this case no transition out of H is possible so that:

$$p_{\overline{HH}}(s, t+h) = p_{\overline{HH}}(s, t)p_{\overline{HH}}(t, t+h)$$

However, during the short interval h , the process either remains in H , changes from H to S or changes from H to D . We assume here that the probability of more than one change is very small so that:

$$p_{\overline{HH}}(t, t+h) + p_{HS}(t, t+h) + p_{HD}(t, t+h) + o(h) = 1$$

Since we know that $p_{HS}(t, t+h) = h\sigma(t) + o(h)$ and $p_{HD}(t, t+h) = h\mu(t) + o(h)$ this gives:

$$p_{\overline{HH}}(t, t+h) = 1 - h(\sigma(t) + \mu(t)) + o(h)$$

Therefore:

$$p_{\overline{HH}}(s, t+h) = p_{\overline{HH}}(s, t)[1 - h(\sigma(t) + \mu(t))] + o(h)$$

This can be rearranged to give:

$$\frac{p_{\overline{HH}}(s, t+h) - p_{\overline{HH}}(s, t)}{h} = -p_{\overline{HH}}(s, t)(\sigma(t) + \mu(t)) + \frac{o(h)}{h}$$

Letting $h \rightarrow 0$ then gives:

$$\frac{\partial}{\partial t} p_{\overline{HH}}(s, t) = -p_{\overline{HH}}(s, t)(\sigma(t) + \mu(t))$$

Solution 6.3

The definition of the transition rates is such that:

$$p_{ij}(s, s+h) = \delta_{ij} + h\mu_{ij}(s) + o(h)$$

or equivalently:

$$p_{ij}(s-h, s) = \delta_{ij} + h\mu_{ij}(s-h) + o(h)$$

Rearranging this gives:

$$\mu_{ij}(s-h) = \frac{p_{ij}(s-h, s) - \delta_{ij} - o(h)}{h}$$

Now taking the limit of both sides as $h \rightarrow 0$ and noting that $p_{ij}(s, s) = \delta_{ij}$, we get:

$$\mu_{ij}(s) = -\lim_{h \rightarrow 0} \frac{p_{ij}(s-h, s) - p_{ij}(s, s) - o(h)}{-h} = -\left[\frac{\partial}{\partial s} p_{ij}(s, t) \right]_{t=s}$$

Now differentiating both sides of the Chapman-Kolmogorov equation:

$$p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, u) p_{kj}(u, t)$$

with respect to s gives:

$$\frac{\partial}{\partial s} p_{ij}(s, t) = \sum_{k \in S} \left(\frac{\partial}{\partial s} p_{ik}(s, u) \right) p_{kj}(u, t)$$

Setting $u = s$ we obtain:

$$\frac{\partial}{\partial s} p_{ij}(s, t) = \sum_{k \in S} \left[\frac{\partial}{\partial s} p_{ik}(s, u) \right]_{u=s} p_{kj}(s, t) = -\sum_{k \in S} \mu_{ik}(s) p_{kj}(s, t)$$

using the result we derived above. In shorthand (*ie* matrix) notation this becomes:

$$\frac{\partial}{\partial s} P(s, t) = -A(s)P(s, t)$$

Solution 6.4

$$\begin{aligned}\frac{\partial}{\partial s} p_{HS}(s,t) &= -[\mu_{HH}(s)p_{HS}(s,t) + \mu_{HS}(s)p_{SS}(s,t) + \mu_{HD}(s)p_{SD}(s,t)] \\ &= -[-(\sigma(s) + \mu(s))p_{HS}(s,t) + \sigma(s)p_{SS}(s,t)] \\ &= (\sigma(s) + \mu(s))p_{HS}(s,t) - \sigma(s)p_{SS}(s,t)\end{aligned}$$

Solution 6.5

The backward differential equation for $p_{AA}(s,t)$ is:

$$\frac{\partial}{\partial s} p_{AA}(s,t) = -[-\mu(s)p_{AA}(s,t)] = \mu(s)p_{AA}(s,t)$$

Separating the variables gives:

$$\frac{\frac{\partial}{\partial s} p_{AA}(s,t)}{p_{AA}(s,t)} = \mu(s)$$

Now changing the variable from s to x , we have:

$$\frac{\partial}{\partial x} \ln p_{AA}(x,t) = \mu(x)$$

Now integrating with respect to x between the limits of $x = s$ and $x = t$ gives:

$$[\ln p_{AA}(x,t)]_{x=s}^{x=t} = \int_s^t \mu(x) dx$$

But $p_{AA}(t,t) = 1$ and $\ln 1 = 0$. So we have:

$$-\ln p_{AA}(s,t) = \int_s^t \mu(x) dx$$

Shifting the minus sign on to the LHS and taking exponentials gives the required result:

$$p_{AA}(s,t) = \exp\left(-\int_s^t \mu(x) dx\right)$$

Solution 6.6

(i)
$${}_{10}P_{60} = \exp\left(-\int_{60}^{70} 0.01 dt\right) = e^{-10 \times 0.01} = 0.905$$

- (ii) The probability ${}_{10}P_{25}$ will be the same as ${}_{10}P_{60}$.
- (iii) This is unrealistic for healthy lives. We would need to use an age-dependent mortality rate.

Solution 6.7

The density function for any distribution can be obtained from the distribution function by differentiation. In this case the distribution function is:

$$P[R_s \leq w | X_s = i] = 1 - P[R_s > w | X_s = i] = 1 - \exp\left(-\int_s^{s+w} \lambda_i(u) du\right)$$

Differentiation of the exponential is straightforward leaving the remaining problem of differentiating the integral with respect to w .

We can differentiate integrals with constant limits by “differentiating under the integral sign”. However, when the limits depend on the variable of integration we get extra terms:

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(x, v) dv = b'(x) f(x, b(x)) - a'(x) f(x, a(x)) + \int_{a(x)}^{b(x)} \left(\frac{\partial}{\partial x} f(x, v) \right) dv$$

In the case above we have:

$$\frac{\partial}{\partial w} \int_s^{s+w} \lambda_i(u) du = \lambda_i(s+w)$$

An alternative approach is the following. Suppose that we knew that if you integrate the function $\lambda_i(u)$ you get the function $\Lambda_i(u)$ say. Then the integral would equal:

$$[\Lambda_i(u)]_s^{s+w} = \Lambda_i(s+w) - \Lambda_i(s)$$

If we differentiate this with respect to w we would get $\lambda_i(s+w)$, since $\lambda_i(u)$ is the derivative of $\Lambda_i(u)$, and the $\Lambda_i(s)$ term doesn't contain any w 's.

Using the formula derived by either of the methods given above, it follows that:

$$\begin{aligned} \frac{\partial}{\partial w} \left(1 - \exp\left(-\int_s^{s+w} \lambda_i(u) du\right) \right) &= -\left(\frac{\partial}{\partial w} \left(-\int_s^{s+w} \lambda_i(u) du \right) \right) \exp\left(-\int_s^{s+w} \lambda_i(u) du\right) \\ &= \lambda_i(s+w) \exp\left(-\int_s^{s+w} \lambda_i(u) du\right) \end{aligned}$$

Solution 6.8

For $i \neq j$ we start by taking derivatives:

$$\frac{\partial}{\partial s} p_{ij}(s, t) = \frac{\partial}{\partial s} \left[\sum_{k \neq i} \int_0^{t-s} e^{-\int_s^{s+w} \lambda_i(u) du} \mu_{ik}(s+w) p_{kj}(s+w, t) dw \right]$$

We need to apply the formula for differentiating integrals:

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(x, y) dy = b'(x) f(x, b(x)) - a'(x) f(x, a(x)) + \int_{a(x)}^{b(x)} \left(\frac{\partial}{\partial x} f(x, y) \right) dy$$

However, it is easier to first apply a substitution in the integral. Let $v = s + w$. Then $dw = dv$ and:

$$\frac{\partial}{\partial s} p_{ij}(s, t) = \frac{\partial}{\partial s} \left[\sum_{k \neq i} \int_s^t e^{-\int_s^v \lambda_i(u) du} \mu_{ik}(v) p_{kj}(v, t) dv \right]$$

In the formula for differentiating integrals, we are replacing x by s and y by v , so we have:

$$f(x, y) = f(s, v) = \sum_{k \neq i} e^{-\int_s^v \lambda_i(u) du} \mu_{ik}(v) p_{kj}(v, t)$$

Also, since $a(x) = a(s) = s$ and $b(x) = b(s) = t$:

$$a'(s) = 1$$

$$b'(s) = 0$$

and:

$$f(x, a(x)) = f(s, s) = \sum_{k \neq i} e^{-\int_s^s \lambda_i(u) du} \mu_{ik}(s) p_{kj}(s, t) = \sum_{k \neq i} \mu_{ik}(s) p_{kj}(s, t)$$

So:

$$\begin{aligned}
 \frac{\partial}{\partial s} p_{ij}(s, t) &= -\sum_{k \neq i} \mu_{ik}(s) p_{kj}(s, t) + \sum_{k \neq i} \int_s^t \frac{\partial}{\partial s} \left(e^{-\int_s^v \lambda_i(u) du} \mu_{ik}(v) p_{kj}(v, t) \right) dv \\
 &= -\sum_{k \neq i} \mu_{ik}(s) p_{kj}(s, t) + \sum_{k \neq i} \int_s^t \left(\frac{\partial}{\partial s} e^{-\int_s^v \lambda_i(u) du} \right) \mu_{ik}(v) p_{kj}(v, t) dv \\
 &= -\sum_{k \neq i} \mu_{ik}(s) p_{kj}(s, t) + \sum_{k \neq i} \int_s^t \lambda_i(s) e^{-\int_s^v \lambda_i(u) du} \mu_{ik}(v) p_{kj}(v, t) dv \\
 &= -\sum_{k \neq i} \mu_{ik}(s) p_{kj}(s, t) + \lambda_i(s) \left\{ \sum_{k \neq i} \int_s^t e^{-\int_s^v \lambda_i(u) du} \mu_{ik}(v) p_{kj}(v, t) dv \right\} \\
 &= -\sum_{k \neq i} \mu_{ik}(s) p_{kj}(s, t) - \mu_{ii}(s) p_{ij}(s, t) \\
 &= -\sum_k \mu_{ik}(s) p_{kj}(s, t)
 \end{aligned}$$

In the fourth line we have identified the expression in curly brackets with $p_{ij}(s, t)$. We have applied the differentiation of an integral rule in both the first and third lines. The last equation is the backward equation as required.

Solution 6.9

This is the probability that if you are dead at time s then you will remain dead at least until time t , which is 1!

Solution 6.10

$$p_{HS}(s, t) = \int_0^{t-s} p_{HH}(s, t-w) \sigma(t-w) p_{SS}(t-w, t) dw$$

The terms in this integral are:

- the probability of going from healthy at time s to healthy at time $t-w$
- then doing a transition from healthy to sick at time $t-w$
- and finally staying in the sick state from time $t-w$ to time t .

Integrating over all possible values of w gives the integrated form of the forward equation.

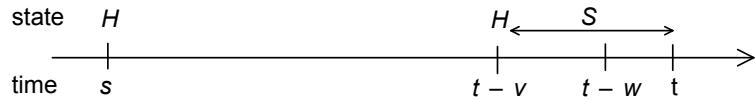
Solution 6.11

$$p_{SH}(s, t) = \int_0^{t-s} p_{SS}(s, s+w) \rho(s+w) p_{HH}(s+w, t) dw$$

Solution 6.12

With most illnesses sick people tend to follow a fairly predictable pattern of either recovering in roughly so many weeks, or getting worse and dying after such-and-such a time. So the probabilities of recovery and death will have a fairly definite pattern as a function of duration (for a specified illness, at least). With healthy people on the other hand, although they will in general become more likely to fall sick or die as they get older, there is no particular reason to expect a 40-year-old's chance of falling sick to depend on how long it is since (s)he was last sick, although it might do if some people are "sickly" by nature.

Also, serious or long-term illness (which will be the main concern of an insurer) is relatively infrequent, so few people make the transition from healthy to sick (which might be defined to be "off work for at least 6 months").

Solution 6.13

$$\begin{aligned} P[X_t = S, C_t > w | X_s = H] \\ &= \int_w^{t-s} p_{HH}(s, t-v) \sigma(t-v) p_{\overline{S_0S}}(t-v, t) dv \end{aligned}$$

where:

$$\begin{aligned} p_{\overline{S_0S}}(t-v, t) &= \exp\left(-\int_{t-v}^t (\rho(u, u+v-t) + \nu(u, u+v-t)) du\right) \\ &= \exp\left(-\int_0^v (\rho(r+t-v, r) + \nu(r+t-v, r)) dr\right) \end{aligned}$$

The integral in v is from w because we are told that the current holding time at time t is at least w . For the integral in u , at time u the current holding time is $u - (t-v) = u + v - t$. You can show that the integral in u and the integral in r are equivalent by making the substitution $r = u + v - t$.

Solution 6.14

- (i) **Probability of staying in the bachelor state**

This is unaffected by the time dependence. So:

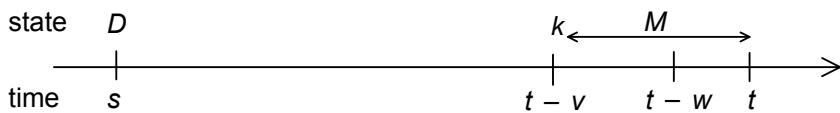
$$\begin{aligned} P[X_u = B, \forall u, s < u < t] &= \exp\left(-\int_s^t (\alpha(u) + \mu(u)) du\right) \\ &= \exp\left(-\int_0^{t-s} (\alpha(s+u) + \mu(s+u)) du\right) \end{aligned}$$

- (ii) **Probability of staying in the married state**

$$\begin{aligned} p_{\overline{M_w M}}(s, t) &= \exp\left(-\int_s^t (d(u, u-s+w) + v(u) + \mu(u)) du\right) \\ &= \exp\left(-\int_0^{t-s} (d(s+u, w+u) + v(s+u) + \mu(s+u)) du\right) \end{aligned}$$

- (iii) **Transition probability**

The diagram for this situation is as follows:



There are two possibilities for the state k , namely D and W, and both of these gives a contribution.

$$\begin{aligned} P[X_t = M, C_t > w | X_s = D] \\ = \int_w^{t-s} [p_{DD}(s, t-v) \rho(t-v) + p_{DW}(s, t-v) r(t-v)] p_{\overline{M_0 M}}(t-v, t) dv \end{aligned}$$

The integral in v is from w because we are told that the current holding time at time t is at least w . The subscript of 0 on the M shows that the life is newly married at time $t - v$.

Solution 6.15

- (i) **Probability that the process remains in state A until at least time s**

The probability that the process remains in state A until at least time s , given that it started in state A at time 0, is:

$$\begin{aligned} p_{\overline{AA}}(0, s) &= \exp\left(-\int_0^s \mu_{AB}(t) dt\right) \\ &= \exp\left(-\int_0^s 2t dt\right) \\ &= \exp\left(\left[-t^2\right]_0^s\right) \\ &= e^{-s^2} \end{aligned}$$

- (ii) **Proof**

The probability that the process is in state B at time T and that it is in the first visit to state B can be expressed in integral form as follows:

$$\int_0^T p_{\overline{AA}}(0, s) \mu_{AB}(s) p_{\overline{BB}}(s, T) ds$$

This expression is constructed using the following reasoning:

- Pick a point in time between 0 and T , call it s , and assume that the process stays in state A up to time s . This gives us the $p_{\overline{AA}}(0, s)$ term.
- Now suppose that there is a transition from state A to state B at time s . This gives us the $\mu_{AB}(s)ds$ term.
- Then we need the process to stay in state B from time s to time T . This gives us the $p_{\overline{BB}}(s, T)$ term.
- Finally, we integrate over all possible times s when the first transition could happen, ie from $s = 0$ up to T .

Now, from part (i) we know that:

$$p_{\overline{AA}}(0, s) = e^{-s^2}$$

Also:

$$\mu_{AB}(s) = 2s$$

and:

$$\begin{aligned} p_{\overline{B}}(s, T) &= \exp\left(-\int_s^T \mu_{BA}(t) dt\right) \\ &= \exp\left(-\int_s^T 2t dt\right) \\ &= \exp\left(\left[-t^2\right]_s^T\right) \\ &= e^{-(T^2 - s^2)} \end{aligned}$$

So the probability that the process is in state B at time T , and it is in the first visit to state B, is:

$$\int_0^T e^{-s^2} 2s e^{-(T^2 - s^2)} ds = e^{-T^2} \int_0^T 2s ds = e^{-T^2} \left[s^2\right]_0^T = T^2 e^{-T^2}$$

as required.

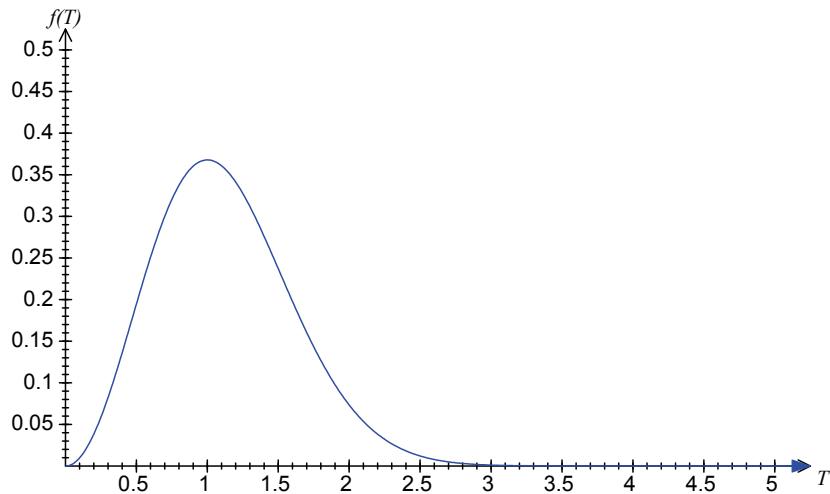
(iii)(a) ***Sketch of the probability function***

The function $f(T) = T^2 e^{-T^2}$ will tend to 0 as $T \rightarrow \infty$ because the exponential term will dominate the polynomial term. Also $f(0) = 0$. Differentiating f we get:

$$f'(T) = 2Te^{-T^2} - 2T^3 e^{-T^2} = 2Te^{-T^2} (1 - T^2)$$

This derivative is equal to 0 when $T = 1$. (We are only considering positive values of T here.) These calculations should help you to sketch the graph.

The function $f(T) = T^2 e^{-T^2}$ is shown below:

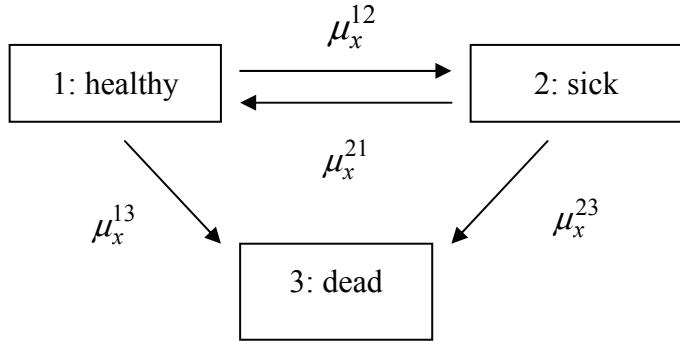


(iii)(b) ***Explanation of the shape of the probability function***

The graph increases at first, due to the increasing force of transition out of state A. It then reaches a peak and starts to decrease because the increasing force of transition out of state B means that the process is less likely to still be in its first visit to state B.

(iii)(c) ***Time at which it is most likely that the process is in its first visit to state B***

From our calculations in part (i)(a), and the graph above, we see that the time at which it is most likely that the process is in its first visit to state B is time 1.

Solution 6.16(i) **Diagram of three-state model**(ii) **Derivation of partial differential equation**

Consider the interval from age x to age $x + t + h$. By the Markov property, we have:

$${}_{t+h} p_x^{12} = {}_t p_x^{11} {}_h p_{x+t}^{12} + {}_t p_x^{12} {}_h p_{x+t}^{22}$$

However, using the assumption about the transition rates, we can write:

$${}_h p_{x+t}^{12} = h \mu_{x+t}^{12} + o(h)$$

$$\text{and: } {}_h p_{x+t}^{22} = 1 - {}_h p_{x+t}^{21} - {}_h p_{x+t}^{23} = 1 - h \mu_{x+t}^{21} - h \mu_{x+t}^{23} + o(h)$$

So:

$${}_{t+h} p_x^{12} = {}_t p_x^{11} {}_h p_{x+t}^{12} + {}_t p_x^{12} \left(1 - h \mu_{x+t}^{21} - h \mu_{x+t}^{23} \right) + o(h)$$

We can rearrange this equation to get:

$$\frac{{}_{t+h} p_x^{12} - {}_t p_x^{12}}{h} = {}_t p_x^{11} \mu_{x+t}^{12} - {}_t p_x^{12} (\mu_{x+t}^{21} + \mu_{x+t}^{23}) + \frac{o(h)}{h}$$

Finally, letting $h \rightarrow 0$ gives:

$$\frac{\partial}{\partial t} {}_t p_x^{12} = {}_t p_x^{11} \mu_{x+t}^{12} - {}_t p_x^{12} (\mu_{x+t}^{21} + \mu_{x+t}^{23})$$

Flashcards

D O N O T
F O R G E T

CT4 Ch 04: The two-state Markov model 9

Given that the likelihood function for μ is:

$$L(\mu) = \mu^d e^{-\mu v}$$

derive the maximum likelihood estimate of μ .

© IFE: 2010 □□□□

Maximum likelihood estimate of μ

The log-likelihood function for μ is:

$$\ln L(\mu) = d \ln \mu - \mu v$$

Differentiating with respect to μ and setting the derivative equal to 0:

$$\frac{d \ln L}{d \mu} = \frac{d}{\mu} - v = 0 \Rightarrow \hat{\mu} = \frac{d}{v}$$

The second derivative of the log-likelihood is:

$$\frac{d^2 \ln L}{d \mu^2} = -\frac{d}{\mu^2} < 0 \Rightarrow \text{max}$$

So the maximum likelihood estimate of μ is $\hat{\mu} = \frac{d}{v}$.

© IFE: 2010

Use our Flashcards to cover the key points of the course that most students like to commit to memory.



Order at www.ActEd.co.uk/estore



www.ActEd.co.uk/forums

Chapter 7

Survival models



Syllabus objectives

- (v) Explain the concept of survival models.
1. Describe the model of lifetime or failure time from age x as a random variable.
 2. State the consistency condition between the random variable representing lifetimes from different ages.
 3. Define the distribution and density functions of the random future lifetime, the survival function, the force of mortality or hazard rate, and derive relationships between them.
 4. Define the actuarial symbols $_t p_x$ and $_t q_x$ and derive integral formulae for them.
 5. State the Gompertz and Makeham laws of mortality.
 6. Define the curtate future lifetime from age x and state its probability function.
 7. Define the expected value and variance of the complete and curtate expected future lifetimes and derive expressions for them. Define the symbols e_x and \mathring{e}_x and derive an approximate relation between them.

0 **Introduction**

In this chapter we will meet the first model of random lifetimes, in which we treat the future lifetime of an individual as a continuous random variable. From this simple starting point we will derive many useful results that are the building blocks of actuarial work relating to human mortality. Later in the course (Chapters 8, 9 and 10) we will consider alternative models.

We will mostly study the lifetime model in the context of human mortality. However, the theory can equally be applied to other problems, such as:

- Analysing the lengths of time that surviving individuals hold insurance policies. Here mortality is replaced by “withdrawal”.
- Analysing the lengths of time that surviving individuals remain healthy. Here mortality is replaced by “sickness”.

1 A simple model of survival

1.1 Future lifetime

The starting point for a simple mathematical model of survival is the observation that the future lifetime of a person (called a “life” in actuarial work) is not known in advance. Further, we observe that lifetimes range from 0 to in excess of 100 years. A natural assumption therefore is that the future lifetime of a given life is a random variable.



Assumption

The future lifetime of a new-born person is a random variable, denoted T , which is continuously distributed on an interval $[0, \omega]$ where $0 < \omega < \infty$.

The maximum age ω is called the *limiting age*.

Typical values of ω for practical work are in the range 100–120. The possibility of survival beyond age ω is excluded by the model for convenience and simplicity.

When Jeanne Calment died in France on 4 August 1997, she was 122 years and 164 days old. According to the *Guinness Book of Records*, this is the highest authenticated age ever recorded.

The greatest age to which any man has ever lived is 120 years 237 days in the case of Shigechiyo Izumi. He was born on 29 June 1865 and was recorded as a 6-year-old in Japan's first census of 1871. He died on 21 February 1986.

Centenarians surviving beyond their 113th year are extremely rare. As of 2006, nobody has lived to celebrate their 123rd birthday.



Definition

$F(t) = P [T \leq t]$ is the distribution function of T .

$S(t) = P [T > t] = 1 - F(t)$ is the survival function of T .

$S(t)$ is known as the survival function of T because it represents the probability of a new-born person surviving to age t .

In insurance contexts, we will not be dealing with new-born babies, so we need to extend the notation to deal with older individuals.

We often need to deal with ages greater than zero. To meet this need, we define T_x to be the future lifetime after age x , of a life who survives to age x , for $0 \leq x \leq \omega$. Note that $T_0 = T$.



Definition

$$F_x(t) = P [T_x \leq t] \quad (0 \leq x \leq \omega)$$

is the distribution function of T_x .

$$S_x(t) = P [T_x > t] = 1 - F_x(t) \quad (0 \leq x \leq \omega)$$

is the survival function of T_x .

For example, the probability that a 40-year old dies before his/her 100th birthday is given by the function $F_{40}(60)$.



Question 7.1

What does the function $S_{29}(36)$ represent?

For consistency with T , the distribution function of the random variable T_x ($0 \leq x \leq \omega$) must satisfy the following relationships:

$$\begin{aligned} F_x(t) &= P [T_x \leq t] \\ &= P [T \leq x + t \mid T > x] \\ &= \frac{F(x+t) - F(x)}{S(x)} \end{aligned}$$

This expression comes from the definition of conditional probabilities. $P(A \mid B)$ represents the probability of event A given that event B has occurred. You should be familiar with the result:

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

So:

$$P(T \leq x + t | T > x) = \frac{P(x < T \leq x + t)}{P(T > x)} = \frac{F(x+t) - F(x)}{S(x)}$$

1.2 Probabilities of death and survival

We now introduce the notation used by actuaries for probabilities of death and survival.



Definition

$${}_t q_x = F_x(t)$$

$${}_t p_x = 1 - {}_t q_x = S_x(t)$$

So, ${}_{60}q_{40}$ represents the probability that a 40-year old dies before his/her 100th birthday and ${}_5 p_{37}$ represents the probability that a 37-year old lives for at least another 5 years.



Question 7.2

Which is bigger, ${}_5 p_{34}$ or ${}_7 p_{33}$?

It is convenient in much actuarial work to use a time unit of one year. When this is the case, so that $t = 1$, we omit the “ t ” from these probabilities. That is, we define:

$$q_x = {}_1 q_x \quad \text{and} \quad p_x = {}_1 p_x$$

q_x and ${}_t q_x$ are called *rates of mortality*.



Summary

- ${}_t q_x$ is the probability that a life now aged x dies within t years
- q_x is the probability that a life now aged x dies within 1 year
- ${}_t p_x$ is the probability that a life now aged x is still alive after t years
- p_x is the probability that a life now aged x is still alive after 1 year

1.3 The force of mortality μ_x

A quantity that plays a central role in a survival model is the force of mortality (which is more widely known as the hazard rate in statistics).



Definition

We denote the force of mortality at age x ($0 \leq x < \omega$) by μ_x , and define it as:

$$\mu_x = \lim_{h \rightarrow 0^+} \frac{1}{h} \times P [T \leq x + h \mid T > x]$$

We will always suppose that the limit exists.

The interpretation of μ_x is very important.

The force of mortality μ_x is an *instantaneous* measure of mortality at age x . It is the continuous equivalent of the discrete quantity q_x .

The probability $P [T \leq x + h \mid T > x]$ is (from the definitions above) $F_x(h) = h q_x$.

For small h , we can ignore the limit and write:

$$h q_x \approx h \cdot \mu_x$$

In other words, the probability of death in a short time h after age x is roughly proportional to h , the constant of proportionality being μ_x .

The intuitive way of thinking of the force of mortality is in terms of the expected number of deaths in a very large population. The expected number of deaths during a short time interval of length h years in a very large population consisting of n individuals aged exactly x is $n \times \mu_x \times h$.

For example, we could estimate the value of μ_{50} by taking a very large group of people, all aged exactly 50, and counting how many died during the next hour. We could then work out the proportion of the group that had died, and express this as an annual rate by multiplying by 24×365 . This figure would give the value of μ_{50} (very nearly).

**Question 7.3**

Give 4 reasons why this calculation would not lead to an exact value for μ_{50} .

We could actually have defined the force of mortality in two ways: either by thinking in terms of a new-born baby (as we did at the start of this section) or by thinking in terms of a person who has already reached the age in question.

**Definition**

For $x \geq 0$ and $t > 0$, we could define the force of mortality μ_{x+t} in two ways:

$$(1) \quad \mu_{x+t} = \lim_{h \rightarrow 0^+} \frac{1}{h} \times P [T \leq x + t + h \mid T > x + t]$$

$$(2) \quad \mu_{x+t} = \lim_{h \rightarrow 0^+} \frac{1}{h} \times P [T_x \leq t + h \mid T_x > t]$$

It is an easy exercise to show from the definitions that these are equal. We will often use μ_{x+t} for a fixed age x and $0 \leq t < \omega - x$.

1.4 Survival probabilities

The definition of $S_x(t)$ leads to an important relationship:

$$S_x(t) = P [T_x > t] = P [T > x + t \mid T > x]$$

$$= \frac{P [T > x + t]}{P [T > x]} = \frac{S(x+t)}{S(x)}$$

which can be expressed in actuarial notation as:

$${}_t p_x = \frac{x+t p_0}{x p_0}$$

Therefore, for any age x and for $s > 0, t > 0$:

$${}_{s+t} p_x = \frac{x+s+t p_0}{x p_0} = \frac{x+s p_0}{x p_0} \times \frac{x+s+t p_0}{x+s p_0} = {}_s p_x \times {}_t p_{x+s}$$

Similarly,

$${}_{s+t}p_x = {}_t p_x \times {}_s p_{x+t}$$

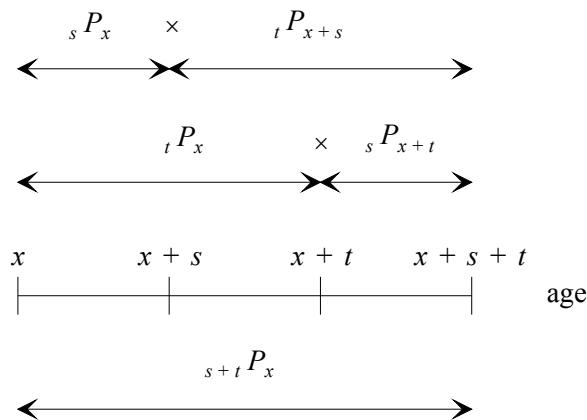
In words, the probability of surviving for time $(s + t)$ after age x is given by multiplying:

1. the probability of surviving for time s , and
2. the probability of then surviving for a further time t

or by multiplying:

1. the probability of surviving for time t , and
2. the probability of *then* surviving for a further time s .

This is illustrated below:



(As you would expect, the order in which we consider the two periods is irrelevant.)

The consistency condition referred to in syllabus item (v)2. simply refers to the results in this section.

1.5 The probability density function of T_x

The distribution function of T_x is $F_x(t)$, by definition. We also want to know its probability density function (pdf).

Denote this by $f_x(t)$, and recall that:

$$f_x(t) = \frac{d}{dt} F_x(t)$$

Then:

$$\begin{aligned} f_x(t) &= \frac{d}{dt} P[T_x \leq t] \\ &= \lim_{h \rightarrow 0^+} \frac{1}{h} \times (P[T_x \leq t+h] - P[T_x \leq t]) \\ &= \lim_{h \rightarrow 0^+} \frac{P[T \leq x+t+h | T > x] - P[T \leq x+t | T > x]}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{P[T \leq x+t+h] - P[T \leq x] - (P[T \leq x+t] - P[T \leq x])}{S(x) \times h} \\ &= \lim_{h \rightarrow 0^+} \frac{P[T \leq x+t+h] - P[T \leq x+t]}{S(x) \times h} \end{aligned}$$

Now multiply and divide by $S(x+t)$ and we have:

$$\begin{aligned} f_x(t) &= \frac{S(x+t)}{S(x)} \times \lim_{h \rightarrow 0^+} \frac{1}{h} \frac{P[T \leq x+t+h] - P[T \leq x+t]}{S(x+t)} \\ &= S_x(t) \times \lim_{h \rightarrow 0^+} \frac{1}{h} P[T \leq x+t+h | T > x+t] \\ &= S_x(t) \times \mu_{x+t} \end{aligned}$$

Or, in actuarial notation, for a fixed age x between 0 and ω :

$$f_x(t) = {}_t p_x \mu_{x+t} \quad (0 \leq t < \omega - x)$$

This is one of the most important results concerning survival models.

1.6 Summary

Let's summarise the model we have introduced.



Summary

T_x is the (random) future lifetime after age x .

It is, by assumption, a continuous random variable taking values in $[0, \omega - x]$.

Its distribution function is $F_x(t) = {}_t q_x$.

Its probability density function is $f_x(t) = {}_t p_x \mu_{x+t}$.

The force of mortality is interpreted by the approximate relationship ${}_h q_x \approx h \cdot \mu_x$ (for small h).

The survival functions $S_x(t)$ or ${}_t p_x$ satisfy the relationship:

$${}_{s+t} p_x = {}_s p_x \times {}_t p_{x+s} = {}_t p_x \times {}_s p_{x+t} \quad (\text{for any } s > 0, t > 0)$$

1.7 Life table functions

A life table is a table showing the expected number that will survive to each age in a hypothetical group of lives. For the English Life Tables No15 (see Pages 68 and 69 of the *Tables*), the table starts at age 0 with 100,000 lives. l_x denotes the expected number of lives at age x and d_x denotes the expected number of deaths between the ages of x and $x+1$. Note that:

$$d_x = l_x - l_{x+1}$$

$$p_x = \frac{l_{x+1}}{l_x}$$

$$q_x = 1 - p_x = 1 - \frac{l_{x+1}}{l_x} = \frac{l_x - l_{x+1}}{l_x} = \frac{d_x}{l_x}$$

$${}_t p_x = \frac{l_{x+t}}{l_x}$$

Values are tabulated only for integer ages. If we require a value at a non-integer age, we must make an assumption about how mortality varies between integer ages. For example, we could make one of the following assumptions:

- deaths occur uniformly between integer ages
- the force of mortality is constant between integer ages
- the Balducci assumption holds.

The Balducci assumption states that:

$${}_{1-t}q_{x+t} = (1-t)q_x$$

for integer ages x and $0 \leq t \leq 1$.

This result is given on Page 33 of the *Tables* and will be used in Chapter 10.



Example

Below is an extract from English Life Table 15 (Males):

Age, x	l_x
58	88,792
59	87,805

Estimate $l_{58.25}$ assuming a uniform distribution of deaths between exact ages 58 and 59.

Solution

There are:

$$88,792 - 87,805 = 987$$

deaths expected between the ages of 58 and 59. Assuming that these are uniformly distributed throughout the year of age, the number of deaths expected between the ages of 58 and 58.25 is:

$$\frac{987}{4} = 246.75$$

So the expected number of lives at age 58.25 is:

$$88,792 - 246.75 = 88,545.25$$

Note that, although the same number of people are dying each quarter, under the uniform distribution of deaths assumption, the surviving population at the start of each quarter is decreasing. So this assumption implies that the force of mortality is increasing over the year of age (58,59).

Note also the following useful formula.



Uniform distribution of deaths assumption

If deaths are uniformly distributed between the ages of x and $x+1$, it follows that:

$${}_t q_x = t q_x$$

for $0 \leq t \leq 1$.



Question 7.4

Prove this result.

1.8 Initial and central rates of mortality

q_x is called an *initial rate of mortality*, because it is the probability that a life alive at exact age x (the initial time) dies before exact age $x+1$.

An alternative often used (especially in demography) is the *central rate of mortality*, denoted m_x .



Definition

$$m_x = \frac{q_x}{\int_0^1 {}_t p_x dt}$$

The quantity m_x is the probability of dying between exact ages x and $x+1$ per person-year lived between exact ages x and $x+1$; the denominator $\int_0^1 t p_x dt$ is interpreted as the expected amount of time spent alive between ages x and $x+1$ by a life alive at age x , and the numerator is the probability of that life dying between exact ages x and $x+1$.

There is another (perhaps more intuitive) interpretation of m_x . We will see later in this chapter that q_x can be represented by the integral $q_x = \int_0^1 t p_x \mu_{x+t} dt$, which means that we can rewrite the equation in the definition of m_x as:

$$m_x = \frac{\int_0^1 t p_x \mu_{x+t} dt}{\int_0^1 t p_x dt}$$

So we see that m_x is a weighted average of the force of mortality over the next year of age. The weighting factors are the survival probabilities.

Note that m_x is a measure of the rate of mortality over the year from exact age x to exact age $x+1$, whereas the force of mortality μ_x is a measure of the *instantaneous* rate of mortality at exact age x .

m_x is useful when the aim is to project numbers of deaths, given the number of lives alive in age groups; this is one of the basic components of a population projection. In practice the age groups used in population projection are often broader than one year, so the definition of m_x has to be suitably adjusted.

In this course, we consider how to estimate the mortality of a particular population using data from an investigation. Historically, actuaries tended to use the data to estimate m_x rather than μ_x or q_x .

m_x was estimated by statistics of the form:

$$\frac{\text{Number of deaths}}{\text{Total time spent alive and at risk}}$$

called “occurrence-exposure rates”. More recently, these statistics have been used to estimate the force of mortality rather than m_x , because in that context they have a solid basis in terms of a probabilistic model. However, if μ_{x+t} is a constant, μ , between ages x and $x+1$, then:

$$m_x = \frac{q_x}{\int_0^1 t p_x \mu dt} = \frac{\int_0^1 t p_x \mu dt}{\int_0^1 t p_x dt} = \mu$$

So there is still a close connection.

It is important that you understand the relationship between these three measures of mortality.



Question 7.5

“ m_x can never be less than q_x .” True or false?

2 *Expected future lifetime*

2.1 *Complete expectation of life*



Definition

The expected future lifetime after age x , which is referred to by demographers as the expectation of life at age x , is defined as $E[T_x]$. It is denoted \mathring{e}_x .

The symbol \mathring{e}_x is read as “ e -circle- x ” and is tabulated in some of the actuarial tables we will be using.

The next bit of “algebra” uses the following result, which we will see again in Section 3.3:

$${}_t p_x \mu_{x+t} = f_x(t) = \frac{\partial}{\partial t} F_x(t) = \frac{\partial}{\partial t} {}_t q_x = \frac{\partial}{\partial t} (1 - {}_t p_x) = -\frac{\partial}{\partial t} {}_t p_x$$

By definition:

$$\begin{aligned}\mathring{e}_x &= \int_0^{\omega-x} t \cdot {}_t p_x \mu_{x+t} dt \\ &= \int_0^{\omega-x} t \left(-\frac{\partial}{\partial t} {}_t p_x \right) dt \\ &= -\left[t {}_t p_x \right]_0^{\omega-x} + \int_0^{\omega-x} {}_t p_x dt \quad (\text{integrating by parts}) \\ &= \int_0^{\omega-x} {}_t p_x dt\end{aligned}$$

since the term in square brackets is zero for both $t = 0$ and $t = \omega - x$.



Question 7.6

Turn to ELT15 (Males) in your *Tables* and find the complete expectation of life \mathring{e}_x for the following male lives:

- (a) a new-born baby
- (b) a 21-year old actuarial student
- (c) a 70-year old pensioner



Question 7.7

Explain why your answer to (b) is not equal to your answer to (a) less the 21 years already lived, ie why $\mathring{e}_0 \neq 21 + \mathring{e}_{21}$.

2.2 Curtate expectation of life

To define the *curtate expectation of life*, we first need to define K_x , the *curtate future lifetime of a life age x* .



Definition

The curtate future lifetime of a life age x is:

$$K_x = [T_x]$$

where the square brackets denote the integer part. In words, K_x is equal to T_x rounded down to the integer below.

So, the curtate future lifetime K_x of a life aged exactly x is the whole number of years lived after age x .

Clearly K_x is a discrete random variable, taking values on the integers:

$$0, 1, 2, \dots [\omega - x]$$

The probability distribution of K_x is easy to write down using the definitions of Section 1 of this chapter.

$$\begin{aligned} P[K_x = k] &= P[k \leq T_x < k+1] \\ &= P[k < T_x \leq k+1] \quad (*) \\ &= {}_k p_x \cdot q_{x+k} \end{aligned}$$

We also use the symbol ${}_k|q_x$ to represent $P(K_x = k)$. It is read as “ k deferred q_x ”, and you can think about this as deferring the event of death until the year that begins in k years from now.

Note that switching the inequalities at step (*) requires an assumption about T_x . It is enough to suppose that $F_x(t)$ is continuous in t . We will not discuss this further here.



Definition

We now define the *curtate expectation of life*, denoted e_x , by:

$$e_x = E[K_x]$$

Then:

$$\begin{aligned} e_x &= \sum_{k=0}^{[\omega-x]} k \cdot {}_k p_x \cdot q_{x+k} \\ &= {}_1 p_x \cdot q_{x+1} \\ &\quad + {}_2 p_x \cdot q_{x+2} + {}_2 p_x \cdot q_{x+2} \\ &\quad + {}_3 p_x \cdot q_{x+3} + {}_3 p_x \cdot q_{x+3} + {}_3 p_x \cdot q_{x+3} \\ &\quad + \dots \end{aligned}$$

$$\begin{aligned} &= \sum_{k=1}^{[\omega-x]} \sum_{j=k}^{[\omega-x]} {}_j p_x \cdot q_{x+j} \quad (\text{summing columns}) \\ &= \sum_{k=1}^{[\omega-x]} {}_k p_x \end{aligned}$$

The last step follows since $\sum_{j=k}^{[\omega-x]} {}_j p_x \cdot q_{x+j}$ represents the probability of dying at *any* time after age $x+k$, which can be written more simply as ${}_k p_x$.



Question 7.8

Look up the AM92 mortality table and find the curtate expectation of life e_x for the following lives:

- (a) a 21-year old actuarial student
- (b) a 70-year old pensioner



Question 7.9

Show algebraically that $e_x = p_x(1 + e_{x+1})$.

2.3 The relationship between \mathring{e}_x and e_x

We have two simple formulae:

$$\mathring{e}_x = \int_0^{\omega-x} {}_t p_x \, dt$$

$$e_x = \sum_{k=1}^{[\omega-x]} {}_k p_x$$

The complete and curtate expectations of life are related by the approximate equation:

$$\mathring{e}_x \approx e_x + \frac{1}{2}$$

To see this, define $J_x = T_x - K_x$ to be the random lifetime after the highest integer age to which a life age x survives.

Approximately, $E[J_x] = \frac{1}{2}$, but $E[T_x] = E[K_x] + E[J_x]$ so $\mathring{e}_x \approx e_x + \frac{1}{2}$ as stated.

**Question 7.10**

What assumption is made when stating that $E[J_x] = \frac{1}{2}$?

**Question 7.11**

Using ELT15 (Males) mortality, find the curtate expectation of life e_x for:

- (a) a new-born baby
- (b) a 21-year old actuarial student
- (c) a 70-year old pensioner

2.4 Future lifetimes – variance

It is easy to write down the variances of the complete and curtate future lifetimes:

$$\text{var}[T_x] = \int_0^{\omega-x} t^2 {}_t p_x \mu_{x+t} dt - \bar{e}_x^2$$

$$\text{var}[K_x] = \sum_{k=0}^{[\omega-x]} k^2 {}_k p_x q_{x+k} - \bar{e}_x^2$$

but these do not simplify neatly as the expected values do.

It is not particularly useful to know the variance of future lifetimes. However, it is useful to be able to find the variance of financial functions (eg the profits from a life insurance policy or the cost of providing a benefit from a pension scheme) based on future lifetimes. This information would enable us to quantify the likely variation in profits etc.

2.5 ***Uses of the expectation of life***

The expectation of life is often used as a measure of the standard of living and health care in a given country.

Here are some examples of average life expectancy at birth for males in different countries (2009):

35-40	Angola, Zambia
40-45	Afghanistan, Malawi
45-50	Nigeria, Rwanda, South Africa, Zimbabwe
50-55	Cameroon, Ethiopia, Uganda
55-60	Bangladesh, Ghana, Haiti, Kenya, Russia
60-65	Botswana, Burma, Guyana, Pakistan, Yemen
65-70	Brazil, Guatemala, India
70-75	Barbados, China, Serbia
75-80	Australia, Japan, New Zealand, USA, most Western European countries

The data come from the CIA World Factbook.

3 Some important formulae

3.1 Introduction

In this section we give two important formulae, one for tq_x and one for tp_x .

These formulae will provide a useful link between tq_x , tp_x and μ_x .

3.2 A formula for tq_x

The first follows from the result that $f_x(t) = tp_x \mu_{x+t}$. We have:

$$tq_x = F_x(t) = \int_0^t f_x(s) ds = \int_0^t s p_x \mu_{x+s} ds$$

This formula is easy to interpret. For each time s , between 0 and t , the integrand is the product of:

- (i) $s p_x$, the probability of surviving to age $x + s$, and
- (ii) $\mu_{x+s} ds$, which is approximately equal to $ds q_{x+s}$, the probability of dying just after age $x + s$.

Since it is impossible to die at more than one time, we simply add up, or in the limit integrate, all these mutually exclusive probabilities.

This result is not usually used to calculate tq_x from tp_x and μ_x since if we knew tp_x we could calculate tq_x directly. However, the result does allow us to derive a very important relationship between tp_x and μ_x .

3.3 A formula for tp_x

The formula for tp_x follows from the solution of the following equation:

$$\frac{\partial}{\partial s} s p_x = -\frac{\partial}{\partial s} s q_x = -f_x(s) = -s p_x \mu_{x+s}$$

(You will see why we have used s as the variable in a moment.)

To solve this, note that:

$$\frac{\partial}{\partial s} \log_s p_x = \frac{\frac{\partial}{\partial s} s p_x}{s p_x}$$

so that the above equation can be rewritten as:

$$\frac{\partial}{\partial s} \log_s p_x = -\mu_{x+s}$$

We used s (not t) as the variable so that we can integrate this relationship between the limits of 0 and t without causing confusion.

Hence:

$$\int_0^t \frac{\partial}{\partial s} \log_s p_x ds = - \int_0^t \mu_{x+s} ds + c$$

where c is some constant of integration.

Actually, we don't need to include the constant of integration here because we've put limits on both integrals. So c will turn out to be zero.

The left hand side is:

$$[\log_s p_x]_0^t = \log_t p_x \quad (\text{since } {}_0 p_x = 1)$$

so taking exponentials of both sides gives:

$${}_t p_x = \exp \left\{ - \int_0^t \mu_{x+s} ds + c \right\}$$

Now since ${}_0 p_x = 1$, we must have $c = 0$ (since $e^0 = 1$), so finally:

$${}_t p_x = \exp \left\{ - \int_0^t \mu_{x+s} ds \right\}$$

3.4 Summary

To summarise, we have derived the following *very important* results.



Integral expressions

$${}_t q_x = \int_0^t {}_s p_x \mu_{x+s} ds \quad (7.1)$$

$${}_t p_x = \exp \left\{ - \int_0^t \mu_{x+s} ds \right\} \quad (7.2)$$

4 Simple parametric survival models

Several survival models are in common use in which the random variable denoting future lifetime has a distribution expressed in terms of a small number of parameters. Perhaps the simplest is the exponential model, in which the hazard is constant:

$$\mu_x = \mu$$

It follows from (7.2) above that in the exponential model:

$${}_t p_x = S_x(t) = \exp \left\{ - \int_0^t \mu ds \right\} = \exp \left\{ - [\mu s]_0^t \right\} = \exp(-\mu t)$$

and hence that:

$${}_t q_x = 1 - {}_t p_x = 1 - \exp(-\mu t)$$



Question 7.12

If μ_x takes the constant value 0.001 between ages 25 and 35, calculate the probability that a life aged exactly 25 will survive to age 35.



Question 7.13

If μ_x takes the constant value 0.025 at all ages, calculate the age x for which ${}_x p_0 = 0.5$. What does this age represent?



Question 7.14

Given that $e_{50} = 30$ and $\mu_{50+t} = 0.005$ for $0 \leq t \leq 1$, what is the value of e_{51} ?

A simple extension to the exponential model is the Weibull model, in which the survival function $S_x(t)$ is given by the two-parameter formula:

$$S_x(t) = \exp[-\alpha t^\beta] \quad (7.3)$$

Note that:

$$S_x(t) = 1 - F_x(t)$$

where $F_x(t) = P(T_x \leq t)$. The distribution function of the Weibull distribution is given on Page 15 of the *Tables*.

Since:

$$\mu_{x+t} = -\frac{\partial}{\partial t} \log[S_x(t)]$$

we see that:

$$\mu_{x+t} = -\frac{\partial}{\partial t} [-\alpha t^\beta] = -[-\alpha \beta t^{\beta-1}] = \alpha \beta t^{\beta-1}$$

Different values of the parameter β can give rise to a hazard that is monotonically increasing or monotonically decreasing as t increases, or in the specific case where $\beta = 1$, a hazard that is constant, since if $\beta = 1$:

$$\alpha \beta t^{\beta-1} = \alpha \cdot 1 \cdot t^0 = \alpha$$

This can be seen also from the expression for $S_x(t)$ (7.3), from which it is clear that, when $\beta = 1$, the Weibull model is the same as the exponential model.

5 The Gompertz and Makeham laws of mortality

The Gompertz and Makeham laws of mortality are two further examples of parametric survival models. They can be expressed as follows:



Definition

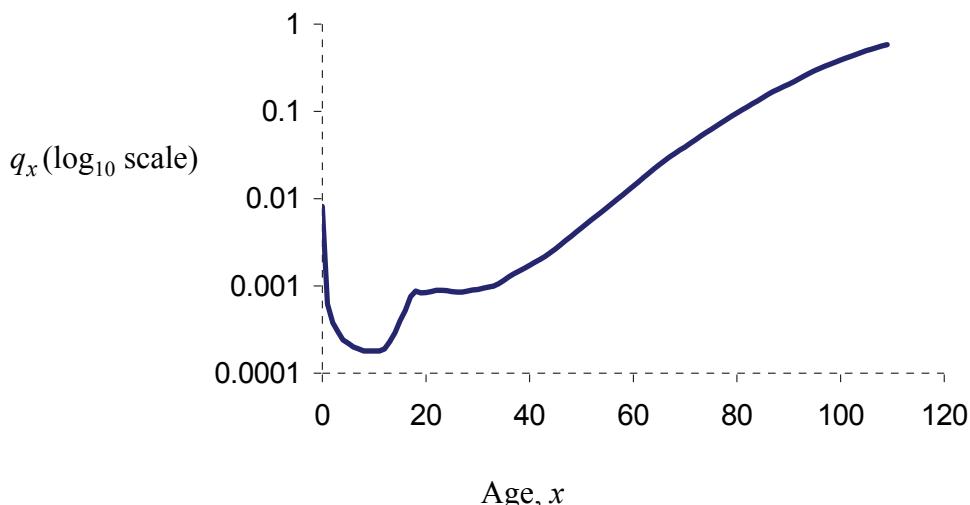
Gompertz' Law: $\mu_x = B c^x$

Makeham's Law: $\mu_x = A + B c^x$

Gompertz' Law is an exponential function, and it is often a reasonable assumption for middle ages and older ages.

Makeham's Law incorporates a constant term, which is sometimes interpreted as an allowance for accidental deaths, not depending on age.

The rationale behind the laws is based on an observation made by Benjamin Gompertz in the early 1800s. When μ_x is plotted on a logarithmic scale against age, the graph often appears to follow a straight line for a large part of the age range. You can see this from the graph of q_x in the diagram below.



5.1 Calculating the parameter values

If a life table is known to follow Gompertz' Law, the parameters B and c can be determined given the values of μ_x at any two ages. In the case of a life table following Makeham's Law, the parameters A, B and c can be determined given the values of μ_x at any three ages.

Question 7.15

For a force of mortality μ_x that is known to follow Gompertz' Law, calculate the parameters B and c if $\mu_{50} = 0.017609$ and $\mu_{55} = 0.028359$.

5.2 Survival probabilities

Survival probabilities $t p_x$ can be found using

$$t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right)$$

Gompertz' Law

In the case of Gompertz' Law:

$$t p_x = g^{c^x(c^t - 1)}$$

where:

$$g = \exp\left(\frac{-B}{\log c}\right)$$

Note that in this section we use the term "log" to mean natural log, ie \log_e .

Proof

We know that ${}_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right) = \exp\left(-\int_0^t Bc^{x+s} ds\right)$.

We can write c^{x+s} as $c^x e^{s \log c}$, so that the integral becomes:

$$\int_0^t Bc^x e^{s \log c} ds = \frac{Bc^x}{\log c} [e^{s \log c}]_0^t = \frac{Bc^x}{\log c} [c^s]_0^t = \frac{Bc^x}{\log c} (c^t - 1)$$

If we introduce the auxiliary parameter g defined by $\log g = -B/\log c$, the value of the integral is $-\log g \cdot c^x (c^t - 1)$ and we find that:

$${}_t p_x = \exp\left[\log g \cdot c^x (c^t - 1)\right] = (e^{\log g})^{c^x (c^t - 1)} = g^{c^x (c^t - 1)}$$

Makeham's Law

In the case of Makeham's Law:

$${}_t p_x = s^t g^{c^x (c^t - 1)}$$

where:

$$g = \exp\left(\frac{-B}{\log c}\right) \text{ and } s = \exp(-A)$$

We will use these laws in Chapter 13, *Methods of Graduation*.



Question 7.16

Show that, if mortality experience conforms to Gompertz' Law, then:

$$-\log(-\log p_x) = \log\left[\frac{\log c}{B(c-1)}\right] - x \log c$$

Suggest how this property could be used.


Question 7.17

A mortality table, which obeys Gompertz' Law for older ages, has:

$$\mu_{70} = 0.025330 \text{ and } \mu_{90} = 0.126255$$

Find the probability that a life aged 60 will survive for 20 years.

Gompertz-Makeham family

More generally, we can model the force of mortality using one of the Gompertz-Makeham family of curves. This family consists of functions of the form:

$$\begin{aligned} GM(r,s) = & \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \cdots + \alpha_r t^{r-1} \\ & + \exp\left\{\alpha_{r+1} + \alpha_{r+2} t + \alpha_{r+3} t^2 + \cdots + \alpha_{r+s} t^{s-1}\right\} \end{aligned}$$

where $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{r+s}$ are constants which do not depend on t .

This form of the Gompertz-Makeham family of curves is the one that is used most widely. However you should be aware that it does not match the form given on page 32 of the *Tables*.

The form given in the *Tables* is:

$$\mu_x = GM(r,s) = poly_1(t) + \exp\{poly_2(t)\}$$

where t is a linear function of x and $poly_1(t)$ and $poly_2(t)$ are polynomials of degree r and s respectively.

6 Exam-style question

You should now be able to attempt the following exam-style question.



Question 7.18

Subject 104, September 2003, Question 6

- (i) T_x denotes the future lifetime of a life currently aged x . Write down the probability density function of T_x . [1]
- (ii) Using your answer to (i), show that:
- $\frac{\partial}{\partial s} \log_s p_x = -\mu_{x+s}$, and
 - $_t p_x = \exp \left\{ - \int_0^t \mu_{x+s} ds \right\}$. [4]
- (iii) In a certain population, the force of mortality is given by:

$$\begin{array}{ll} \mu_x & \\ 60 < x \leq 70 & 0.01 \\ 70 < x \leq 80 & 0.015 \\ x > 80 & 0.025 \end{array}$$

Calculate the probability that a life aged exactly 65 will die between exact ages 80 and 83. [3]

[Total 8]

Chapter 7 Summary

Modelling mortality

We can model mortality by assuming that future lifetime is a continuous random variable taking values between 0 and some limiting age ω . From this starting point, we can calculate probabilities of survival ($_t p_x$) and death ($_t q_x$) for an individual aged x over a period of t years.

Definitions of probabilities of death and survival

$$_t q_x = F_x(t) = P [T_x \leq t]$$

$$_t p_x = 1 - _t q_x = S_x(t) = 1 - F_x(t) = P [T_x > t]$$

$$_{t+s} p_x = _t p_x \times _s p_{x+t} = _s p_x \times _t p_{x+s}$$

Force of mortality

The force of mortality μ_x is the *instantaneous rate of mortality* at age x . It is defined by the equation:

$$\mu_x = \lim_{h \rightarrow 0^+} \frac{1}{h} \times P [T \leq x + h \mid T > x]$$

We also have the following results about μ_x :

$$\mu_x = \lim_{h \rightarrow 0^+} \frac{1}{h} \times {}_h q_x \quad \text{so } {}_h q_x \approx h \cdot \mu_x \text{ (for small } h\text{)}$$

$$_t q_x = \int_0^t {}_s p_x \mu_{x+s} ds \quad {}_t p_x = \exp \left\{ - \int_0^t \mu_{x+s} ds \right\}$$

Life table functions

l_x is the expected number of survivors at age x

d_x is the expected number of deaths between the ages of x and $x+1$

The central rate of mortality

The central rate of mortality is given by:

$$m_x = \frac{q_x}{\int_0^1 {}_t p_x dt} = \frac{\int_0^1 {}_t p_x \mu_{x+t} dt}{\int_0^1 {}_t p_x dt}$$

It is a weighted average of the force of mortality between the ages of x and $x+1$.

The complete future lifetime random variable T_x

The PDF of T_x is given by:

$$f_x(t) = \frac{d}{dt} F_x(t) = {}_t p_x \mu_{x+t} \quad (0 \leq t \leq \omega - x)$$

The expected value of T_x , sometimes called the *complete expectation of life*, is:

$$\overset{\circ}{e}_x = E[T_x] = \int_0^{\omega-x} {}_t p_x dt$$

The variance of T_x is:

$$\text{var}[T_x] = \int_0^{\omega-x} t^2 {}_t p_x \mu_{x+t} dt - \overset{\circ}{e}_x^2$$

The curtate future lifetime random variable K_x

K_x is defined to be the integer part of T_x

The probability function of K_x is given by:

$$P(K_x = k) = {}_k p_x q_{x+k} = {}_{k|} q_x$$

The expected value of K_x , sometimes called the *curtate expectation of life*, is:

$$e_x = E[K_x] = \sum_{k=1}^{[\omega-x]} k p_x$$

The variance of K_x is:

$$\text{var}[K_x] = \sum_{k=0}^{[\omega-x]} k^2 p_x q_{x+k} - e_x^2$$

If deaths occur on average halfway between birthdays, then:

$$\overset{\circ}{e}_x \approx e_x + \frac{1}{2}$$

Exponential model

In the exponential model, the hazard rate (or force of mortality) is constant. So:

$${}_t p_x = e^{-\mu t}$$

Weibull model

In the Weibull model:

$${}_t p_x = \exp(-\alpha t^\beta)$$

$$\mu_{x+t} = \alpha \beta t^{\beta-1}$$

Different values of β can give rise to a hazard that is monotonically increasing or decreasing. In the case when $\beta = 1$, the Weibull model is the same as the exponential model.

Gompertz' law

$$\mu_x = B c^x$$

$${}_t p_x = g^{c^x(c'-1)} \quad \text{where} \quad g = \exp\left(\frac{-B}{\log c}\right)$$

Makeham's law

$$\mu_x = A + B c^x$$

$${}_t p_x = s^t g^{c^x(c'-1)} \quad \text{where} \quad g = \exp\left(\frac{-B}{\log c}\right) \text{ and } s = \exp(-A)$$

Both Gompertz' and Makeham's laws include an exponential term, which makes them particularly useful for middle and older ages.

Gompertz-Makeham family

The Gompertz-Makeham family consists of curves of the form:

$$\begin{aligned} GM(r,s) &= \alpha_1 + \alpha_2 t + \alpha_3 t^2 + \cdots + \alpha_r t^{r-1} \\ &\quad + \exp\left\{\alpha_{r+1} + \alpha_{r+2} t + \alpha_{r+3} t^2 + \cdots + \alpha_{r+s} t^{s-1}\right\} \end{aligned}$$

where $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{r+s}$ are constants which do not depend on t .

Chapter 7 Solutions

Solution 7.1

$S_{29}(36)$ represents the probability of an individual currently aged 29 reaching 65, *i.e.* living for at least another 36 years. (More importantly for me, it represents the probability that I will live long enough to draw my state pension!)

Solution 7.2

The probability of surviving from age 33 to 40 must be less than the probability of surviving from 34 to 39 since the first survival period includes the second, as well as the additional risk of dying between ages 33 and 34 and between ages 39 and 40.

Hence ${}_5 p_{34} > {}_7 p_{33}$.

Solution 7.3

Reasons why it would not be exact include:

1. The actual number of deaths we observe will differ from the expected number because of statistical fluctuations.
2. People die in “whole units”, so there will be rounding errors because of this discreteness.
3. We have ignored leap years. Assuming 365.25 days in a year would give a more “accurate” answer.
4. We have used a period of 1 hour. The force of mortality is an *instantaneous* measure, so we need to take the limit of 1 hour, 1 minute, 1 second ...

As well as these theoretical reasons, there are practical reasons why we could not do this. For example, there are only around 2,000 babies born each day in the whole of the UK. So, if we take being “exactly 50” to mean having your 50th birthday on the day in question, we will have somewhat less than 2,000 people in our group (because some people will have died before age 50). Since this is a relatively small number of people, it is very unlikely that any of them at all would die during the next hour.

Solution 7.4

Assuming that deaths are uniformly distributed between exact ages x and $x+1$, we have (by linear interpolation):

$$l_{x+t} = (1-t)l_x + tl_{x+1}$$

for $0 \leq t \leq 1$. So:

$$t q_x = 1 - \frac{l_{x+t}}{l_x} = 1 - \frac{(1-t)l_x + tl_{x+1}}{l_x} = \frac{tl_x - tl_{x+1}}{l_x} = t(1 - p_x) = t q_x$$

Solution 7.5

True. The denominator $\int_0^1 t p_x dt \leq 1$, so $m_x \geq q_x$ as stated.

Solution 7.6

(a) $\overset{\circ}{e}_0 = 73.413$ (years)

(b) $\overset{\circ}{e}_{21} = 53.497$

(c) $\overset{\circ}{e}_{70} = 11.187$

Solution 7.7

The relationship $\overset{\circ}{e}_0 = 21 + \overset{\circ}{e}_{21}$ is only true if the probability of dying before age 21 is zero. Although the probability of an individual dying before this age is quite low, it is greater than zero. Therefore $\overset{\circ}{e}_0 < 21 + \overset{\circ}{e}_{21}$.

Solution 7.8

(a) $e_{21} = 57.481$ (years)

(b) $e_{70} = 13.023$

Solution 7.9

$$\begin{aligned}
 e_x &= \sum_{k=1}^{[\omega-x]} {}_k p_x = {}_1 p_x + \sum_{k=2}^{[\omega-x]} {}_k p_x \\
 &= {}_1 p_x + \sum_{j=1}^{[\omega-x-1]} {}_1 p_x \cdot {}_j p_{x+1} = {}_1 p_x \left(1 + \sum_{j=1}^{[\omega-x-1]} {}_j p_{x+1} \right) = {}_1 p_x (1 + e_{x+1})
 \end{aligned}$$

Intuitively, this is saying that the life expectancy for a person now aged x is one year more than their life expectancy when they reach age $x+1$, *provided that* they do survive to age $x+1$.

Solution 7.10

The most natural implicit assumption is that $E[J_x] = \frac{1}{2}$ if deaths are assumed to occur uniformly within each year of age, ie the same number of people die each day during any particular year of age.

Solution 7.11

$$(a) \quad e_0 \equiv \overset{\circ}{e}_0 - 0.5 = 72.913 \text{ (years)}$$

$$(b) \quad e_{21} \equiv \overset{\circ}{e}_{21} - 0.5 = 52.997$$

$$(c) \quad e_{70} \equiv \overset{\circ}{e}_{70} - 0.5 = 10.687$$

Solution 7.12

$${}_{10} p_{25} = \exp \left(- \int_0^{10} 0.001 dt \right) = e^{-0.01} = 0.99005$$

In the UK, this assumption about mortality may be a reasonable approximation over this age range.

Solution 7.13

$${}_x p_0 = \exp(-0.025x) \Rightarrow 0.5 = \exp(-0.025x) \Rightarrow x = -\frac{\log_e 0.5}{0.025} = 27.726$$

If we observe a group of individuals from a chosen start date, the age x for which ${}_x p_0 = 0.5$ represents the time at which half of those from the original group are expected to be alive. It is equivalent to the concept of a radioactive *half-life* in physics. It is also the *median* of T , the lifetime of a new-born baby.

This mortality assumption would *not* be a reasonable approximation for any human population!

Solution 7.14

We can calculate the value of e_{51} using the formula from Solution 7.9:

$$e_{50} = p_{50} (1 + e_{51})$$

Since the force of mortality is constant between the ages of 50 and 51:

$$p_{50} = e^{-0.005}$$

So:

$$30 = e^{-0.005} (1 + e_{51}) \Rightarrow e_{51} = 29.15$$

Solution 7.15

$$\frac{\mu_{55}}{\mu_{50}} = \frac{Bc^{55}}{Bc^{50}} = c^5 \Rightarrow c = \left(\frac{0.028359}{0.017609} \right)^{1/5}$$

Therefore $c = 1.1$ and it is simple to calculate that $B = 0.00015$.

Solution 7.16

We can start from the formula $p_x = \exp\left(-\int_0^1 \mu_{x+t} dt\right)$

Taking logs and changing signs gives:

$$-\log p_x = \int_0^1 \mu_{x+t} dt = \int_0^1 Bc^{x+t} dt = Bc^x \int_0^1 c^t dt$$

Integrating by writing the integrand as an exponential function:

$$-\log p_x = Bc^x \int_0^1 e^{t \log c} dt = Bc^x \left[\frac{e^{t \log c}}{\log c} \right]_0^1 = \frac{Bc^x(c-1)}{\log c}$$

Taking logs and changing signs again gives:

$$-\log(-\log p_x) = -\log B - x \log c - \log(c-1) + \log \log c = \log \left[\frac{\log c}{B(c-1)} \right] - x \log c$$

How could this result be used?

The RHS is a linear function of x . This means that we can estimate the parameters B and c by fitting a straight line to a graph of $-\log(-\log p_x)$ plotted as a function of x .

The slope of the line will be $-\log c$ from which c can be estimated.

The intercept on the y -axis will be $\log \left[\frac{\log c}{B(c-1)} \right]$ from which B can be estimated.

Solution 7.17

If the table follows the Gompertz law then $\mu_x = Bc^x$ and:

$$\frac{\mu_{90}}{\mu_{70}} = \frac{Bc^{90}}{Bc^{70}} = c^{20} = \frac{0.126255}{0.025330}$$

$$c = \left(\frac{0.126255}{0.025330} \right)^{\frac{1}{20}} = 1.083629$$

So:

$$B = 0.126255 \times (1.083629)^{-90} = 9.16196 \times 10^{-5}$$

and:

$$g = \exp \left\{ -\frac{B}{\log_e c} \right\} = \exp \left\{ -\frac{9.16196 \times 10^{-5}}{\log_e 1.083629} \right\} = 0.998860$$

Then:

$${}_{20}p_{60} = g^{c^{60}(c^{20}-1)} = (0.998860)^{493.4052}$$

$$= 0.56959$$

$$= 0.570$$

Solution 7.18(i) **Probability density function**

$$f(t) = {}_t p_x \mu_{x+t}, \quad t \geq 0$$

(ii) **Proof**

To prove (a) we can start from:

$${}_s p_x = P[T_x \geq s] = 1 - P[T_x < s] = 1 - \int_0^s f(t)dt$$

Now differentiate with respect to s and use the result from part (i):

$$\frac{\partial}{\partial s} {}_s p_x = \frac{\partial}{\partial s} \left\{ 1 - \int_0^s f(t)dt \right\} = 0 - f(s) = - {}_s p_x \mu_{x+s}$$

So:

$$\frac{\partial}{\partial s} \log {}_s p_x = \frac{1}{{}_s p_x} \frac{\partial}{} {}_s p_x = - \mu_{x+s}$$

We can then derive (b) by integrating over the time interval $(0, t)$:

$$\begin{aligned} [\log {}_s p_x]_0^t &= - \int_0^t \mu_{x+s} ds \\ \log {}_t p_x - \log 1 &= - \int_0^t \mu_{x+s} ds \end{aligned}$$

$$\text{So: } {}_t p_x = \exp \left\{ - \int_0^t \mu_{x+s} ds \right\}$$

(iii) **Probability**

We need to calculate:

$$\begin{aligned} {}_{15|3} q_{65} &= {}_{15} p_{65} \times {}_3 q_{80} \\ &= {}_5 p_{65} \times {}_{10} p_{70} \times (1 - {}_3 p_{80}) \end{aligned}$$

Using the formula derived in part (ii)(b):

$${}_5 p_{65} = e^{-5(0.01)} = e^{-0.05}$$

$${}_{10} p_{70} = e^{-10(0.015)} = e^{-0.15}$$

and ${}_3 p_{80} = e^{-3(0.025)} = e^{-0.075}$

So we get:

$${}_{15|3} q_{65} = e^{-0.05} \times e^{-0.15} \times (1 - e^{-0.075}) = 0.0592$$

Chapter 8

Estimating the lifetime distribution function



Syllabus objectives

- (vi) *Describe estimation procedures for lifetime distributions.*
1. *Describe the various ways in which lifetime data might be censored.*
 2. *Describe the estimation of the empirical survival function in the absence of censoring, and what problems are introduced by censoring.*
 3. *Describe the Kaplan-Meier (or product limit) estimate of the survival function in the presence of censoring, compute it from typical data and estimate its variance.*
 4. *Describe the Nelson-Aalen estimate of the cumulative hazard rate in the presence of censoring, compute it from typical data and estimate its variance.*

0 Introduction

In Chapter 7 we introduced T , the continuous random variable representing future lifetime. In this chapter, we will see how to use observations from an investigation to obtain an empirical estimate (*i.e.* one based on observation) of the distribution function, $F(t) = P(T \leq t)$. We will derive the statistical properties of our estimate so that we can measure its variance and construct confidence intervals. We will also need to bear in mind that data may be incomplete in practice.

The Core Reading refers to the decrement of interest as “death”. (“Decrement” here means a method of leaving a population.) The models can easily be extended to the analysis of any decrements, *e.g.* sickness or mechanical breakdown.

1 Questions of inference

We now turn to statistical inference. Given some mild conditions on the distribution of T , we can obtain all information by estimating $F(t)$, $S(t)$, $f(t)$ or μ_t for all $t \geq 0$.

In other words, we can derive $F(t)$, $S(t)$, $f(t)$ and μ_t from any one of these items.



Question 8.1

State the fundamental relationships that link $F(t)$, $S(t)$, $f(t)$ and μ_t .

1.1 Estimating the lifetime distribution

The simplest experiment would be to observe a large number of new-born lives. The proportion alive at age $t > 0$ would furnish an estimate of $S(t)$. The estimate would be a step function, and the larger the sample the closer to a smooth function we would expect it to be. For use in applications it could be smoothed further.

For example, a life insurance company would prefer to base its premium calculations on a smooth estimate to ensure that the premiums change gradually from one age to the next without sudden jumps.

We need not assume that T is a member of any parametric family; this is a *non-parametric* approach to estimation. You will recognise this as the empirical distribution function of T .

Under a *non-parametric* approach, we make no prior assumptions about the shape or form of the distribution. Under a *parametric* approach, we assume that the distribution belongs to a certain family (eg normal or exponential) and use the data to estimate the appropriate parameters (eg mean and variance).

Statistical results can be derived *theoretically* (from first principles) or *empirically* (from observation). Since we cannot calculate T from first principles, we will use data to calculate the empirical distribution function $F(t)$.

Clearly, there are some practical problems:

- **Even if a satisfactory group of lives could be found, the experiment would take about 100 years to complete.**
- **The observational plan requires us to observe the deaths of all the lives in the sample. In practice many would be lost to the investigation, for one reason or another, and to exclude these from the analysis might bias the result. The statistical term for this problem is *censoring*. All we know in respect of some lives is that they died after a certain age.**

So censoring results in the loss of data. Depending on the nature of the censoring mechanism, it can also result in the introduction of bias into the mortality rates. This would occur if informative censoring were present – see below.

An “observational plan” is just the framework for a mortality investigation. Amongst other things, it will specify the start and end date of the investigation and the category (or categories) of lives to be included in the study.



Question 8.2

Although this experiment would take a very long time to complete, it would be worthwhile if the results were useful. Why do you think that the results would not be useful in practice?



Question 8.3

For what reasons might lives be “lost to the investigation” if we are carrying out:

- (a) a national investigation into the rate of death from natural causes, and
- (b) a study of the mortality of life insurance policyholders?

In medical statistics, where the lifetimes are often shorter, non-parametric estimation is very important.

In this chapter we show how the experiment above can be amended to allow for censoring. Otherwise, we must use a different observational plan, and base inference on data gathered over a shorter time, eg 3 or 4 years.

A consequence is that we no longer observe the same cohort throughout their joint lifetimes, so we might not be sampling from the same distribution. It might be sensible to widen the model assumption, so that the mortality of lives born in year y is modelled by a random variable T^y , for example. In practice we usually divide the investigation up into single years of age. We return to investigations like these in Chapter 10.

Observing lives between (say) integer ages x and $x+1$, and limiting the period of investigation, are also forms of censoring. Censoring might still occur at unpredictable times – by lapsing a life policy, for example – but survivors will certainly be lost to observation at a known time, either on attaining age $x+1$ or when the investigation ends.

2 Censoring mechanisms

Data are *censored* if we do not know the exact values of each observation but we do have information about the value of each observation in relation to one or more bounds. For example, we may know that an individual's lifetime exceeded 20 years because the individual was still alive at age 20 when the investigation closed, but we have no further information about the remaining lifetime.

Censoring is the key feature of survival data (indeed survival analysis might be defined as the analysis of censored data) and the mechanisms that give rise to censoring play an important part in statistical inference. Some of the most common assumptions are these (they are not all mutually exclusive):

Right censoring

Data are right censored if the censoring mechanism cuts short observations in progress. An example is the ending of a mortality investigation before all the lives being observed have died. Persons still alive when the investigation ends are right censored – we know only that their lifetimes exceed some value.

Right censoring also occurs when:

- life insurance policyholders surrender their policies
- active lives of a pension scheme retire
- endowment assurance policies mature.

Left censoring

Data are left censored if the censoring mechanism prevents us from knowing when entry into the state that we wish to observe took place. An example arises in medical studies in which patients are subject to regular examinations. Discovery of a condition tells us only that the onset fell in the period since the previous examination; the time elapsed since onset has been left censored.

Left censoring occurs, for example:

- when estimating functions of exact age and you don't know the exact date of birth
- when estimating functions of exact policy duration and you don't know the exact date of policy entry
- when estimating functions of the duration since onset of sickness and you don't know the exact date of becoming sick.

Note that left censoring does not occur when we are estimating functions of exact age and we have no information from before the start of the investigation period, or before the entry date of a policy, *etc.* These are examples of *left truncation* and do *not* affect your ability to measure the exact duration of individuals from their dates of birth.

Interval censoring

Data are interval censored if the observational plan only allows us to say that an event of interest fell within some interval of time. An example arises in actuarial investigations, where we might know only the calendar year of death. Both right and left censoring can be seen as special cases of interval censoring.

Further examples of interval censoring include the following situations:

- when you only know the calendar year of withdrawal
- when estimating functions of exact age and you only know that deaths were aged “ x nearest birthday” at the date of death
- when you know the calendar *date* of death and you know the calendar *year* of birth. This is an example of left censoring (and therefore interval censoring). Another way of viewing this situation is to say that we actually have data grouped by “age next birthday at the 1 January prior to death”. Since we only know that the lifetime falls within a certain range, this is an example of interval censoring.

In actuarial investigations, right-censoring is the most common form of censoring encountered.

Random censoring

If censoring is random, then the time C_i (say) at which observation of the i th lifetime is censored is a random variable. The observation will be censored if $C_i < T_i$ where T_i is the (random) lifetime of the i th life. In such a situation, censoring is said to be random.

Random censoring arises when individuals may leave the observation by a means other than death, and where the time of leaving is not known in advance.

Examples of random censoring include:

- life insurance withdrawals
- emigration from a population
- members of a company pension scheme may leave voluntarily when they move to another employer.

Random censoring is a special case of right censoring.

The case in which the censoring mechanism is a second decrement of interest gives rise to multiple decrement models.

For example, suppose that lives can leave a pension scheme through death, age retirement or withdrawal. We can estimate the rates of decrement for all three causes of decrement by using a multiple decrement model. You will meet multiple decrement models in Subject CT5.

Informative and non-informative censoring

Censoring is non-informative if it gives no information about the lifetimes $\{T_i\}$.

This just means that the mortality of the lives that remain in the at-risk group is the same as the mortality of the lives that have been censored.

In the case of random censoring, the independence of each pair T_i, C_i is sufficient to ensure that the censoring is non-informative. Informative censoring is more difficult to analyse, essentially because the resulting likelihoods cannot usually be factorised.

Recall from Subject CT3 that when we are dealing with events that are statistically independent, the likelihood function representing all the events is simply the product of the likelihood functions for each individual event. This greatly simplifies the mathematics required in the analysis.

Examples of informative censoring include:

- Withdrawal of life insurance policies, because these are likely to be in better average health than those who do not withdraw. So the mortality rates of the lives that remain in the at-risk group are likely to be higher than the mortality rates of the lives that surrendered their policies.
- Ill-health retirements from pension schemes, because these are likely to be in worse than average health than the continuing members. So the mortality rates of those who remain in the pension scheme are likely to be lower than the mortality rates of the lives that left through ill-health retirement.

An example of non-informative censoring is:

- the end of the investigation period (because it affects all lives equally, regardless of their propensity to die at that point).

Type I censoring

If the censoring times $\{C_i\}$ are known in advance (a degenerate case of random censoring) then the mechanism is called “Type I censoring”.

Type I censoring is therefore another special case of right censoring. Type I censoring occurs, for example:

- when estimating functions of exact age and you stop following individuals once they have reached their 60th birthday
- when lives retire from a pension scheme at normal retirement age (if normal retirement age is a pre-determined exact age)
- when estimating functions of policy duration and you only observe individuals up to their 10th policy anniversary
- when measuring functions of duration since having a particular medical operation and you only observe people for a maximum of 12 months from the date of their operation.

Lives censored at the end of an investigation period might also be considered as an example of Type I censoring. However, Type I censoring normally refers to cases where observation ceases at a predetermined exact age.

Type II censoring

If observation is continued until a predetermined number of deaths has occurred, then “Type II censoring” is said to be present. This can simplify the analysis, because then the number of events of interest is non-random.

An example of Type II censoring is:

- when a medical trial is ended after 100 lives on a particular course of treatment have died.

It is obvious that the observational plan is likely to introduce censoring of some kind, and consideration should be given to the effect on the analysis in specifying the observational plan. Censoring might also depend on the results of the observations to date. For example, if strong enough evidence accumulates during the course of a medical experiment, the investigation might be ended prematurely, so that the better treatment can be extended to all the subjects under study, or the inferior treatment withdrawn.

In actuarial investigations, right censoring is the most common form of censoring encountered.

Many actuarial investigations are characterised by a combination of random and Type I censoring, for example, in life office mortality studies where policies rather than lives are observed, and observation ceases either when a policy lapses (random censoring) or at some predetermined date marking the end of the period of investigation (Type I censoring).

Type I and Type II censoring are most frequently met with in the design of medical survival studies.



Question 8.4

An investigation is carried out into the lifestyle of male accountants. A group of 10,000 accountants is selected at random on 1 January 2001. Each member of the sample group supplies detailed personal information as at 1 January 2001 including name, address, date of birth and marital status. The same information is collected as at each 1 January in the years 2002, 2003, 2004 and 2005. The investigation closes in 2005.

A PhD student wishes to use the data from this investigation for her thesis on the mortality of married men. Describe the ways in which the available data for this investigation are censored.

3 **The Kaplan-Meier (product-limit) model**

3.1 **Introduction**

In this section we develop the empirical distribution function to allow for censoring, ie the distribution function derived from the data.

We will consider lifetimes as a function of time t without mention of a starting age x . The following could be applied equally to new-born lives, to lives aged x at outset, or to lives with some property in common at time $t = 0$, for example diagnosis of a medical condition. Medical studies are often based on time since diagnosis or time since the start of treatment, and if the patient's age enters the analysis it is usually as an explanatory variable in a regression model.

For example, we may be interested in measuring mortality amongst patients suffering from a highly virulent tropical disease. The future lifetime of a sufferer will depend on many factors. The age of the patient may be an important factor (eg the rate of deterioration may be quicker amongst older patients) but it may not be the sole determinant. It may be appropriate to model the lifetime as starting at the time of diagnosis. (In actuarial terminology, "duration" is the dominant factor here.)

We will look at regression models in Chapter 9.

Although the notation in this section looks quite complicated, you will see that the numerical calculations are quite intuitive.

3.2 **Assumptions and notation**

Suppose we observe a population of n lives in the presence of non-informative right censoring, and suppose we observe m deaths.

By assuming that the type of censoring present is non-informative, we are assuming that the mortality of those lives remaining in the group under observation is not systematically higher or lower than the mortality of the lives that have been censored.

If informative censoring is present and we ignore it, then the resulting estimates of the distribution and survival functions will be biased.

If informative censoring is present and we allow for it, then the lifetimes and censoring times will no longer be independent. This means that the likelihood function, which is made up of joint probabilities and probability density functions, can no longer be written as the product of simple probabilities and the resulting algebra will be very complicated.

So we proceed with the assumption that any censoring present is non-informative. However, you should bear in mind that the results of any model are only as reliable as the assumptions on which the model is based.

We now define the rest of the notation.

Let $t_1 < t_2 < \dots < t_k$ be the ordered times at which deaths were observed. We do not assume that $k = m$, so more than one death might be observed at a single failure time, eg two or more lives may die on the same day.

Suppose that d_j deaths are observed at time t_j ($1 \leq j \leq k$) so that $d_1 + d_2 + \dots + d_k = m$.

Observation of the remaining $n - m$ lives is censored. In other words, we don't try to track some of these remaining lives throughout the investigation.

Suppose that c_j lives are censored between times t_j and t_{j+1} ($0 \leq j \leq k$), where we define $t_0 = 0$ and $t_{k+1} = \infty$ to allow for censored observations after the last observed failure time; then $c_0 + c_1 + \dots + c_k = n - m$.

So, c_j represents the number of lives that are removed from the investigation between times t_j and t_{j+1} for a reason other than the decrement we are investigating.

The Kaplan-Meier estimator of the survivor function adopts the following conventions.

- (a) **The hazard of experiencing the event is zero at all durations except those where an event actually happens in our sample.**
- (b) **The hazard of experiencing the event at any particular duration, t_j , when an event takes place is equal to $\frac{d_j}{n_j}$, where d_j is the number of individuals experiencing the event at duration t_j and n_j is the risk set at that duration (that is, the number of individuals still at risk of experiencing the event just prior to duration t_j).**

So if we observed 2 deaths out of 10 lives at risk, the hazard would be equal to $\frac{2}{10}$.

- (c) **Persons that are censored are removed from observation at the duration at which censoring takes place, save that persons who are censored at a duration where events also take place are assumed to be censored immediately after the events have taken place (so that they are still at risk at that duration).**

In other words, if any of the individuals are observed to be censored at the same time as one of the deaths, the convention is to treat the censoring as if it happened shortly afterwards, *ie* the deaths are assumed to have occurred first.

We will use this notation for the rest of the chapter. Let's look at a simple example to illustrate how it can be applied to a practical situation.



Example

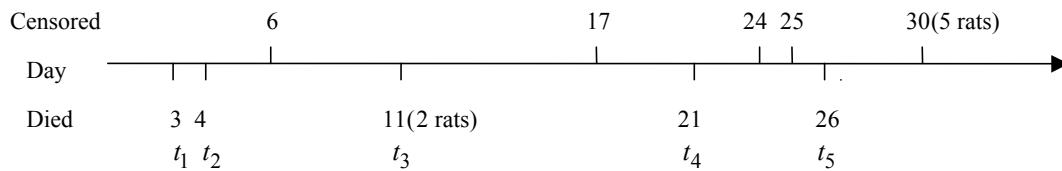
A group of 15 laboratory rats are injected with a new drug. They are observed over the next 30 days. The following events occur:

Day	Event
3	Rat 4 dies from effects of drug.
4	Rat 13 dies from effects of drug.
6	Rat 7 gnaws through bars of cage and escapes.
11	Rats 6 and 9 die from effects of drug.
17	Rat 1 killed by other rats.
21	Rat 10 dies from effects of drug.
24	Rat 8 freed during raid by animal liberation activists.
25	Rat 12 accidentally freed by journalist reporting earlier raid.
26	Rat 5 dies from effects of drug.
30	Investigation closes. Remaining rats hold street party.

How would this information be represented in the notation described above?

Solution

The timeline below illustrates the situation. (The numbers refer to the days on which the events occurred, not to the number of rats.)



Number of lives under investigation, $n = 15$

Number of drug-related rat deaths observed, $m = 6$

Note that the death on day 17 is not directly related to the effects of the drug, so it should not be treated in the same way. Like an escape, it is an example of random right censoring.

Number of times at which deaths were observed: $k = 5$

Times at which deaths are observed: $t_1 = 3, t_2 = 4, t_3 = 11, t_4 = 21, t_5 = 26$

Number of deaths observed at each failure time: $d_1 = 1, d_2 = 1, d_3 = 2, d_4 = 1, d_5 = 1$

Number of lives that didn't die because of the drug: $n - m = 15 - 6 = 9$

Number of lives censored: $c_0 = 0, c_1 = 0, c_2 = 1, c_3 = 1, c_4 = 2, c_5 = 5$

and note that $\sum_{j=0}^k c_j = n - m$

Number of lives alive and at risk at time t_i : $n_1 = 15, n_2 = 14, n_3 = 12, n_4 = 9, n_5 = 6$



Question 8.5

A chef specialising in the manufacture of fluffy meringues uses a *Whiskmatic* disposable electric kitchen implement. The *Whiskmatic* is rather unreliable and often breaks down, so the chef is in the habit of replacing the implement in use at a given time, shortly before an important social function or after making the 1,000th fluffy meringue with that implement.

The following times until mechanical failure (no asterisk) or replacement whilst in working order (asterisk) were observed (measured in days of use):

17, 13, 15*, 7*, 21, 18*, 5, 18, 6*, 22, 19*, 15, 4, 11, 14*, 18, 10, 10, 8*, 17

Define n, m, k, t_j, d_j, c_j and n_j for these data, assuming that censoring occurs just after the failures were observed.

Effectively, what we are doing is partitioning duration into very small intervals such that at the vast majority of such intervals no events occur. There is no reason to suppose, given the data that we have, that the risk of the event happening is anything other than zero at those intervals where no events occur. We have no evidence *in our data* to suppose anything else.

For those very small intervals in which events do occur, we suppose that the hazard is constant (ie piecewise exponential) within each interval, but that it can vary between intervals.

Recall that, if $\mu_{x+t} = \mu$, the survival function is given by:

$$S_x(t) = {}_t p_x = e^{-\mu t}$$

So the Core Reading means that the survival function is exponential over each short interval over which the force of mortality (or hazard) is constant.

We estimate the hazard within the interval containing event time t_j as:

$$\hat{\lambda}_j = \frac{d_j}{n_j}$$

Of course, effectively this formula is being used for all the other intervals as well, but as $d_j = 0$ in all these intervals, the hazard will be zero.

It is possible to show that this estimate arises as a maximum likelihood estimate. The likelihood of the data can be written:

$$\prod_{j=1}^k \lambda_j^{d_j} (1-\lambda_j)^{n_j-d_j}$$

Note that you don't need to be able to show where this comes from.

This is proportional to a product of independent binomial likelihoods, so that the maximum is attained by setting:

$$\hat{\lambda}_j = \frac{d_j}{n_j} \quad (1 \leq j \leq k)$$



Question 8.6

Show that the maximum likelihood estimates are $\hat{\lambda}_j = \frac{d_j}{n_j}$ for $j = 1, 2, \dots, k$.

(You may assume that the resulting estimates are maxima.)

3.3 Extending the force of mortality to discrete distributions

It is convenient to extend to discrete distributions the definition of force of mortality (or hazard) given in Chapter 7 for continuous distributions.

Definition

Suppose $F(t)$ has probability masses at the points t_1, t_2, \dots, t_k .

Then define:

$$\lambda_j = P[T = t_j | T \geq t_j] \quad (1 \leq j \leq k) \quad (8.1)$$

This is called the *discrete hazard function*.

(We use the symbol λ to avoid confusion with the usual force of mortality.)

Intuitively, you can think of λ_j as the probability that a given individual dies on day t_j , given that they were still alive at the start of that day.

Question 8.7

Butterflies of a certain species have short lives. After hatching, each butterfly experiences a lifetime defined by the following probability distribution:

Lifetime (days)	Probability
1	0.10
2	0.30
3	0.25
4	0.20
5	0.15

Calculate λ_j for $j = 1, 2, \dots, 5$ (to 3 decimal places) and sketch a graph of the discrete hazard function.

3.4 Calculating the Kaplan-Meier estimate of the survival function

If we assume that T has a discrete distribution then:

$$1 - F(t) = \prod_{t_j \leq t} (1 - \lambda_j)$$

Since $1 - F(t) = S(t)$, we can estimate the survival function using the formula:

$$\hat{S}(t) = \prod_{t_j \leq t} (1 - \hat{\lambda}_j)$$

This is the **Kaplan-Meier estimator**. To compute the Kaplan-Meier estimate of the survivor function, $\hat{S}(t)$, we simply multiply the survival probabilities within each of the intervals up to and including duration t .

The survival probability at time t_j is estimated by:

$$1 - \hat{\lambda}_j = \frac{n_j - d_j}{n_j} = \frac{\text{number of survivors}}{\text{number at risk}}$$

So the probability of survival at time t is estimated by:

$$\hat{S}(t) = \prod_{t_j \leq t} \frac{n_j - d_j}{n_j}$$

Because the Kaplan-Meier estimate involves multiplying up survival probabilities, it is sometimes called the *product limit estimate*. In effect, we choose finer and finer partitions of the time axis, and estimate $(1 - F(t))$ as the product of the probabilities of surviving each sub-interval. Then, with the above definition of the discrete force of mortality (8.1), we obtain the Kaplan-Meier estimate as the mesh of the partition tends to zero. This is the origin of the name “product-limit” estimate, by which the Kaplan-Meier estimate is sometimes known.

Note that the Kaplan-Meier estimate of the survivor function is constant after the last duration at which an event is observed to occur. It is not defined at durations longer than the duration of the last censored observation.

Only those at risk at the observed lifetimes $\{t_j\}$ contribute to the estimate. It follows that it is unnecessary to start observation on all lives at the same time or age; the estimate is valid for data truncated from the left, provided the truncation is non-informative in the sense that entry to the study at a particular age or time is independent of the remaining lifetime. (Note that left truncation is not the same as left censoring.)



Question 8.8

What is the difference between left censoring and left truncation?

Now let's take a look at a numerical example to see how the estimation actually works in practice.



Example

Using the data from the observation of laboratory rats, calculate the Kaplan-Meier estimate of $F(t)$.

Solution

The calculation of the Kaplan-Meier estimate is set out in the table:

j	t_j	d_j	n_j	$\hat{\lambda}_j = d_j / n_j$	$(1 - \hat{\lambda}_j)$	$1 - \prod_{k=1}^j (1 - \hat{\lambda}_k)$
1	3	1	15	0.06667	0.93333	0.06667
2	4	1	14	0.07143	0.92857	0.13333
3	11	2	12	0.16667	0.83333	0.27778
4	21	1	9	0.11111	0.88889	0.35803
5	26	1	6	0.16667	0.83333	0.46503

From the final column in the table, the Kaplan-Meier estimate of $F(t)$ is:

$$\hat{F}(t) = \begin{cases} 0 & \text{for } 0 \leq t < 3 \\ 0.06667 & \text{for } 3 \leq t < 4 \\ 0.13333 & \text{for } 4 \leq t < 11 \\ 0.27778 & \text{for } 11 \leq t < 21 \\ 0.35803 & \text{for } 21 \leq t < 26 \\ 0.46503 & \text{for } t \geq 26 \end{cases}$$



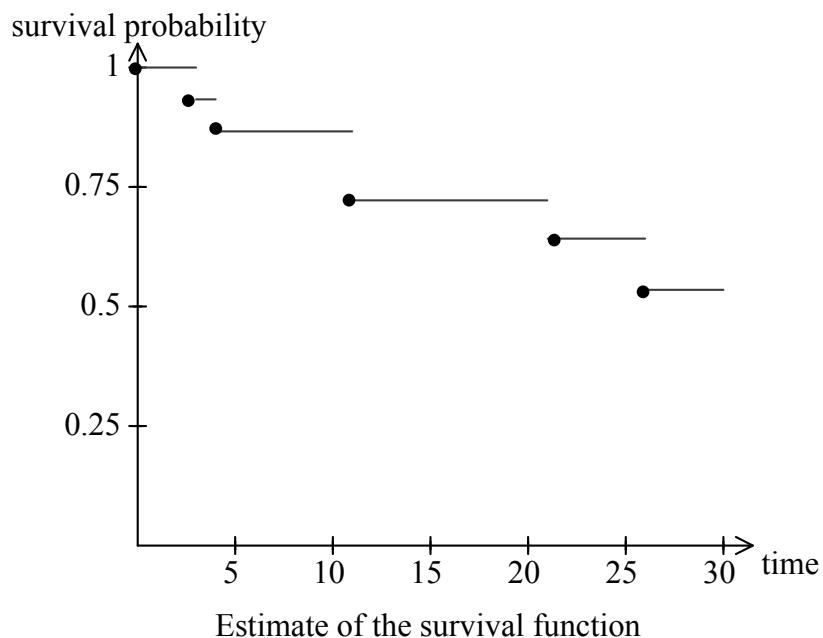
Question 8.9

Why does the value of $\hat{F}(t)$, our estimate of the distribution function, never reach 1?

3.5 A graphical approach

Rather than using a table and formulae, you could carry out the Kaplan-Meier calculations using the following graphical approach. Here we will estimate the survival function $S(t)$ and then use the relationship $F(t) = 1 - S(t)$ to obtain $\hat{F}(t)$.

The graph of $\hat{S}(t)$ will be a step function, starting at 1 and stepping down every time a death occurs. The graph of $\hat{S}(t)$ for the rats' data is given below.



To specify $\hat{S}(t)$, we need to work out the height of each of the steps.

We know that $\hat{S}(t)$ starts at 1 and remains constant until the first death, which occurs at time 3. So:

$$\hat{S}(t) = 1 \text{ for } 0 \leq t < 3$$

Just before time 3, there were 15 rats under observation. One rat died at time 3. Given that a death occurred at time 3, the probability of any given rat surviving past time 3 is $\frac{14}{15}$.

**Question 8.10**

What does this figure of $\frac{14}{15}$ correspond to in Kaplan-Meier notation?

As the next death does not occur until time 4, we have:

$$\hat{S}(t) = \frac{14}{15} = 0.93333 \text{ for } 3 \leq t < 4$$

One more rat died at time 4. There were 14 rats under observation just before time 4. So the probability that any given rat, which was alive just before time 4, still being alive at time 4 is $\frac{13}{14}$. If we treat survival in non-overlapping time intervals as independent, then the probability that any given rat is still alive at time 4 is:

$$\begin{aligned} P(\text{does not die at time 3}) \times P(\text{does not die at time 4}) &= \frac{14}{15} \times \frac{13}{14} \\ &= \frac{13}{15} \\ &= 0.86667 \end{aligned}$$

As the next death does not occur until time 11, it follows that:

$$\hat{S}(t) = \frac{13}{15} = 0.86667 \text{ for } 4 \leq t < 11$$

Just before time 11, there were 12 rats under observation (since one of the 13 still alive at time 4 was censored at time 6). Two rats died at time 11. So the probability that any given rat, which was alive just before time 11, is still alive at time 11 is $\frac{10}{12}$. Furthermore, the probability that any given rat is still alive at time 11 is:

$$\frac{14}{15} \times \frac{13}{14} \times \frac{10}{12} = \frac{13}{18} = 0.72222$$

As the next death does not occur until time 21, it follows that:

$$\hat{S}(t) = \frac{13}{18} = 0.72222 \text{ for } 11 \leq t < 21$$

Continuing in this way, we obtain:

$$\hat{S}(t) = \begin{cases} 1 & \text{for } 0 \leq t < 3 \\ \frac{14}{15} & \text{for } 3 \leq t < 4 \\ \frac{13}{15} & \text{for } 4 \leq t < 11 \\ \frac{13}{18} & \text{for } 11 \leq t < 21 \\ \frac{52}{81} & \text{for } 21 \leq t < 26 \\ \frac{130}{243} & \text{for } 26 \leq t \leq 30 \end{cases}$$

Hence:

$$\hat{F}(t) = \begin{cases} 0 & \text{for } 0 \leq t < 3 \\ \frac{1}{15} & \text{for } 3 \leq t < 4 \\ \frac{2}{15} & \text{for } 4 \leq t < 11 \\ \frac{5}{18} & \text{for } 11 \leq t < 21 \\ \frac{29}{81} & \text{for } 21 \leq t < 26 \\ \frac{113}{243} & \text{for } 26 \leq t \leq 30 \end{cases}$$



Question 8.11

Calculate the Kaplan-Meier estimate of $F(t)$ for the *Whiskmatic* data.

4 Comparing lifetime distributions

Since Kaplan-Meier estimates are often used to compare the lifetime distributions of two or more populations – for example, in comparing medical treatments – their statistical properties are important. Approximate formulae for the variance of $\tilde{F}(t)$ are available.



Question 8.12

What is the difference between an estimate and an estimator?

Since we are trying to estimate an entire function rather than a single parameter, note that the variance of $\tilde{F}(t)$ is itself a function of t , not just a single figure.

Greenwood's formula (proof not required):

$$\text{var}[\tilde{F}(t)] \approx (1 - \hat{F}(t))^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}$$

is reasonable over most t , but might tend to underestimate the variance in the tails of the distribution.

This formula is given on Page 33 of the *Tables*.

So, we have a measure of the approximate variance of our maximum likelihood estimate of $F(t)$. We can use this information to calculate confidence intervals at given values of t .



Question 8.13

Using Greenwood's formula, estimate $\text{var}[\tilde{F}(16)]$ for the *Whiskmatic* data. Explain how you would use this figure to calculate a confidence interval.

5 ***The Nelson-Aalen model***

The Kaplan-Meier is not the only non-parametric approach to calculating the empirical distribution function $\hat{F}(t)$. Like the Kaplan-Meier estimator, the Nelson-Aalen estimator is based on an assumption of non-informative censoring. So knowing when individuals are censored must not give us any extra information about their future lifetimes.

However, instead of using the $\hat{\lambda}_j$ values to estimate the survival probabilities via the Kaplan-Meier formula:

$$\hat{S}(t) = 1 - \hat{F}(t) = \prod_{t_j \leq t} (1 - \hat{\lambda}_j)$$

we use them to estimate the integrated (or cumulative) hazard function.

5.1 ***The integrated hazard function***

An alternative non-parametric approach is to estimate the integrated hazard, which is defined by Λ (capital λ):

$$\Lambda_t = \int_0^t \mu_s ds + \sum_{t_j \leq t} \lambda_j$$

where the integral deals with the continuous part of the distribution and the sum with the discrete part. (Since this methodology was developed by statisticians, the term “integrated hazard” is in universal use, and “integrated force of mortality” is almost never seen.)

The estimate of Λ_t can then be used to estimate $S(t)$ and $F(t)$. The integrated hazard is a function of t and is sometimes also written as $\Lambda(t)$.

To help you see where Λ_t comes from, consider the probability of surviving one year in two populations. Suppose that the hazard operates continuously in the first population, so that:

$$p_0^{(1)} = \exp \left[- \int_0^1 \mu_s ds \right]$$

Suppose also that the hazard operates discretely in the second population, at age $x + \frac{1}{2}$ say. Then:

$$p_0^{(2)} = 1 - \lambda_{\frac{1}{2}}$$

where $\lambda_{x+\frac{1}{2}}$ is the expected proportion of people dying at exact age $x + \frac{1}{2}$.

If both types of hazard were to occur in the same population, then the total survival probability is:

$$p_0 = p_0^{(1)} \times p_0^{(2)} = \exp \left[- \int_0^1 \mu_s \, ds \right] \times (1 - \lambda_{\frac{1}{2}})$$

If we extend this analysis to t years and assume that we have discrete hazards λ_j operating at exact ages t_j , then:

$$S(t) = {}_t p_0 = \exp \left[- \int_0^t \mu_s \, ds \right] \times \prod_{t_j \leq t} (1 - \lambda_j)$$

Now, using the approximation $e^x \approx 1 + x$ for small x , we have:

$$S(t) = \exp \left[- \int_0^t \mu_s \, ds \right] \times \prod_{t_j \leq t} e^{-\lambda_j}$$

$$= \exp \left[- \int_0^t \mu_s \, ds - \sum_{t_j \leq t} \lambda_j \right]$$

$$= \exp[-\Lambda_t]$$

As λ_j is the proportion of people dying at exact age t_j , we can estimate λ_j using $\hat{\lambda}_j = \frac{d_j}{n_j}$. Empirically (*ie* in real life), hazards (such as death) that we theorise as operating continuously, can only occur discretely. So the continuous part of Λ , disappears and we are left with $\hat{\Lambda}_t = \sum_{t_j \leq t} \frac{d_j}{n_j}$ as our estimate of the integrated hazard.

5.2 Calculating Nelson-Aalen estimates

The Nelson-Aalen estimate of the integrated hazard is:

$$\hat{\Lambda}_t = \sum_{t_j \leq t} \frac{d_j}{n_j}$$

The Nelson-Aalen estimate of the survival function is therefore:

$$\hat{S}(t) = \exp[-\hat{\Lambda}_t]$$

and the Nelson-Aalen estimate of the distribution function is:

$$\hat{F}(t) = 1 - \exp[-\hat{\Lambda}_t]$$



Question 8.14

Calculate the Nelson-Aalen estimate of $F(t)$ for the *Whiskmatic* data.

Corresponding to Greenwood's formula for the variance of the Kaplan-Meier estimator, there is a formula for the variance of the Nelson-Aalen estimator:

$$\text{var}[\tilde{\Lambda}_t] \approx \sum_{t_j \leq t} \frac{d_j(n_j - d_j)}{n_j^3}$$

Note that this formula gives the variance of the integrated hazard, not the variance of $\tilde{F}(t)$. It is given on Page 33 of the *Tables*.



Question 8.15

Calculate $\text{var}[\tilde{\Lambda}_{16}]$ for the *Whiskmatic* data.

5.3 Relationship between the Kaplan-Meier and Nelson-Aalen estimates

The connection between the Kaplan-Meier and Nelson-Aalen estimates is discussed below.

The Kaplan-Meier estimate can be approximated in terms of $\hat{\Lambda}_t$.

Recall that the Kaplan-Meier estimate is:

$$\hat{F}_t = 1 - \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

To avoid confusion between the Kaplan-Meier and Nelson-Aalen estimates, we will now denote:

- the Kaplan-Meier estimate of the distribution function by $\hat{F}_{KM}(t)$, and
- the Nelson-Aalen estimate of the distribution function by $\hat{F}_{NA}(t)$.

Then:

$$\hat{F}_{KM}(t) = 1 - \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

Using the approximation $e^x \approx 1 + x$ for small x , and replacing x by $-\frac{d_j}{n_j}$, we have:

$$\begin{aligned}\hat{F}_{KM}(t) &\approx 1 - \exp\left(-\sum_{t_j \leq t} \frac{d_j}{n_j}\right) \\ &= 1 - \exp(-\hat{\Lambda}_t) \\ &= \hat{F}_{NA}(t)\end{aligned}$$

In some respects, the Kaplan-Meier model can be viewed as a non-parametric analogue of the Binomial model (which is discussed in Chapter 10). This is because they both directly provide maximum likelihood estimates of the probability of dying within a specified period of time.

Similarly, the Nelson-Aalen model can be viewed as a non-parametric analogue of the Poisson model (also discussed in Chapter 10). Both models provide maximum likelihood estimates of the hazard function. Survival or death probabilities can then be estimated indirectly, by virtue of the equation:

$${}_n p_x = \exp\left(-\int_0^n \lambda_{x+t} dt\right)$$

6 Parametric estimation of the survival function

An alternative approach to estimating the survival function proceeds as follows:

- assume a functional form for the survival function $S(t)$
- express $S(t)$ and the hazard $h(t)$ in terms of the parameters of the chosen function
- estimate the parameters by maximum likelihood.

Unless the functional form chosen is very simple, estimation will involve the solution of several simultaneous equations and must be done iteratively.

Possible simple functional forms include the exponential and Weibull distributions, and the Gompertz' law, which are all described in Chapter 7.

For the exponential distribution:

$$S(t) = P(T \geq t) = e^{-\mu t}$$

and for the Weibull distribution:

$$S(t) = \exp[-\alpha t^\beta]$$

These can be obtained from the formulae for distribution functions given in the *Tables* using the result $S(t) = 1 - F(t)$. Gompertz' law, which states that:

$$\mu_x = Bc^x$$

is also given in the *Tables* (Page 32).

For many processes, such as human mortality, it turns out that no simple functional form can describe human mortality at all ages. However, for estimation purposes this is not a problem, since we can divide the age range into small sections, estimate the chosen function for each section (the parameters for each section will be different) and then 'chain' the sections together to create a life table for the whole age (or duration) range with which we are concerned (see Section 6.2 below).

A life table is a discrete survival model used in life assurance. The life table function l_x is the number of lives expected to survive to age x out of a group of l_0 newborn lives. Life tables are studied in detail in Subject CT5.

6.1 Maximum likelihood estimation

We illustrate maximum likelihood estimation by considering the exponential hazard, which has one parameter, μ .

In other words, we are considering the case when the future lifetime random variable T has an exponential distribution with parameter μ . This is equivalent to assuming that the force of mortality is constant.

Consider only the single year of age between exact ages x and $x+1$.

We follow a sample of n independent lives from exact age x until the first of the following things happens:

- (a) their death between exact ages x and $x+1$
- (b) they withdraw from the investigation between exact ages x and $x+1$
- (c) their $(x+1)$ th birthday.

The Core Reading means that each life stops being observed when the earliest of the 3 events described above happens to that life.

Cases (b) and (c) are treated as censored at either the time of withdrawal, or exact age $x+1$ respectively.

Assume that the hazard of death (or force of mortality) is constant between ages x and $x+1$ and takes the unknown value μ . We ask the question: what is the most likely value of μ given the data in our investigation? Assume that we measure duration in years since a person's x th birthday.

Consider first those lives in category (a), who die before exact age $x+1$. Suppose there are k of these.

Take the first of these, and suppose that he or she died at duration t_1 . Given only the data on this life, the value of μ that is most likely is the value that maximises the probability that he or she actually dies at duration t_1 .

The probability that Life 1 will actually die at duration t_1 is equal to $f(t_1)$, where $f(t)$ is the probability density function of T . So the value of μ that we need is the value that maximises $f(t_1)$.

For the exponential distribution:

$$f(t_1) = \mu e^{-\mu t_1}$$

However, in the investigation, we have more than one life that died. Suppose a second life died at duration t_2 . The probability of this happening is $f(t_2)$, and the joint probability that Life 1 died at duration t_1 and Life 2 died at duration t_2 is $f(t_1)f(t_2)$. Given just these two lives, the value of μ we need will be that which maximises $f(t_1)f(t_2)$.

If we now consider all the k lives that died, then the value of μ we want is that which maximises:

$$\prod_{\text{all lives which died}} f(t_i)$$

This product is the probability of observing the data we actually did observe.

It can also be written as:

$$\prod_{\text{deaths}} \mu e^{-\mu t_i} = \mu^k \exp\left(-\mu \sum_{i=1}^k t_i\right)$$

Note that we are summing from $i = 1$ to k because we are assuming that there are k deaths.

But what of the lives that were censored? Their experience must also be taken into account.

Consider the first censored life, and suppose he or she was censored at duration t_{k+1} . All we know about this person is that he or she was still alive at duration t_{k+1} . The probability that a life will still be alive at duration t_{k+1} is $S(t_{k+1})$.

We are using a subscript of $k + 1$ because we are assuming there are k deaths and we are labelling the first censored life as Life $k + 1$.

For the exponential distribution, we have:

$$S(t_{k+1}) = e^{-\mu t_{k+1}}$$

Considering all the censored lives, the probability of observing the data we do observe is:

$$\prod_{\text{all censored lives}} S(t_i)$$

This can also be written as:

$$\prod_{\text{all censored lives}} e^{-\mu t_i} = \exp\left(-\mu \sum_{i=k+1}^n t_i\right)$$

since we have n lives altogether and the censored ones are those labelled Life $k+1$ up to Life n .

Now, putting the deaths and the censored cases together, we can write down the probability of observing all the data we actually observe – both censored lives and those that died. This probability is:

$$\prod_{\text{all censored lives}} S(t_i) \prod_{\text{all lives which died}} f(t_i)$$

This is called the likelihood of the data.

For exponential hazard model, the likelihood function can also be written as:

$$\begin{aligned} L &= \mu^k \exp\left(-\mu \sum_{i=1}^k t_i\right) \exp\left(-\mu \sum_{i=k+1}^n t_i\right) \\ &= \mu^k \exp\left(-\mu \sum_{i=1}^n t_i\right) \end{aligned}$$



Question 8.16

Derive the form of the likelihood function assuming that the future lifetime random variable follows the Weibull distribution with parameters α and β .

The maximum likelihood estimate of the parameter μ , which we denote by $\hat{\mu}$, is the value that maximises this likelihood.

To obtain $\hat{\mu}$, define a variable δ_i such that:

$\delta_i = 1$ if life i died

$\delta_i = 0$ if life i was censored

Then, in the general case, the likelihood can be written:

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

Now, since:

$$f(t) = S(t)h(t)$$

or equivalently:

$$f(t) = {}_t p_x \mu_{x+t}$$

the likelihood can also be written:

$$L = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i)$$

We now substitute the chosen functional form into this equation to express the likelihood in terms of the parameter μ .

This produces:

$$L = \prod_{i=1}^n \mu^{\delta_i} \exp(-\mu t_i)$$

This is equivalent to the expression:

$$L = \mu^k \exp\left(-\mu \sum_{i=1}^n t_i\right)$$

given above, bearing in mind that we have observed k deaths out of the sample of n lives.

Noting that whatever value of μ maximises L will also maximise the logarithm of L , we first take the logarithm of L :

$$\log L = \sum_{i=1}^n \delta_i \log \mu - \sum_{i=1}^n \mu t_i$$

We differentiate this with respect to μ to give:

$$\frac{\partial \log L}{\partial \mu} = \frac{\sum_{i=1}^n \delta_i}{\mu} - \sum_{i=1}^n t_i$$

Setting this equal to zero produces:

$$\frac{\sum_{i=1}^n \delta_i}{\mu} = \sum_{i=1}^n t_i$$

so that:

$$\hat{\mu} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i}$$

or equivalently:

$$\hat{\mu} = \frac{k}{\sum_{i=1}^n t_i}$$

We can check that this is a maximum by noting that:

$$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{\sum_{i=1}^n \delta_i}{\mu^2}$$

This must be negative, as both numerator and denominator are necessarily positive (unless we have no deaths at all in our data, in which case the maximum likelihood estimate of the hazard is 0).

Since $\sum_{i=1}^n \delta_i$ is just the total number of deaths in our data, and $\sum_{i=1}^n t_i$ is the total

time that the lives in the data are exposed to the risk of death, our maximum likelihood estimate of the force of mortality (or hazard) is just deaths divided by exposed to risk, which is intuitively sensible.

This is the same estimate for μ as the one we obtained from the two-state Markov model in Chapter 4. In that chapter we used the notation v to represent the total

waiting time. Here we have $v = \sum_{i=1}^n t_i$.

For parametric distributions with more than one parameter, maximum likelihood estimation of the parameters involves the solution of simultaneous equations, the number of simultaneous equations being equal to the number of parameters to be estimated. These equations often require iterative methods.

6.2 Using the estimates for different age ranges

If we repeat the exercise for other years of age, we can obtain a series of estimates for the different hazards in each year of age.

Suppose that the maximum likelihood estimate of the constant force during the single year of age from x to $x+1$ is $\hat{\mu}_x$. Then the probability that a person alive at exact age x will still be alive at exact age $x+1$ is just $\hat{S}_x(1)$. Given the constant force, the maximum likelihood estimate of the survival function at time 1 is:

$$\hat{S}_x(1) = \exp(-\hat{\mu}_x)$$

To work out the probability that a person alive at exact age x will survive to exact age $x+2$ we note that this probability is equal to:

$$\hat{S}_x(1) \hat{S}_{x+1}(1) = \exp(-\hat{\mu}_x) \exp(-\hat{\mu}_{x+1})$$

Therefore:

$$\hat{S}_x(1) \hat{S}_{x+1}(1) = \hat{S}_x(2) = \exp[-(\hat{\mu}_x + \hat{\mu}_{x+1})]$$

In general, therefore:

$$\hat{S}_x(m) = {}_m \hat{p}_x = \exp\left(-\sum_{j=0}^{m-1} \hat{\mu}_{x+j}\right)$$

By ‘chaining’ together the probabilities in this way, we can evaluate probabilities over any relevant age range.



Question 8.17

If $\hat{\mu}_{60} = 0.01$, $\hat{\mu}_{61} = 0.02$ and $\hat{\mu}_{62} = 0.03$, estimate the values of p_{60} , ${}_2 p_{60}$ and ${}_3 p_{60}$.

7 Exam-style questions

You should now be able to attempt the following past exam questions.



Question 8.18

Subject 104, September 2000, Question 10

The following data relate to 12 patients who had an operation that was intended to correct a life-threatening condition, where time 0 is the start of the period of the investigation:

Patient number	Time of operation (in weeks)	Time observation ended (in weeks)	Reason observation ended
1	0	120	Censored
2	0	68	Death
3	0	40	Death
4	4	120	Censored
5	5	35	Censored
6	10	40	Death
7	20	120	Censored
8	44	115	Death
9	50	90	Death
10	63	98	Death
11	70	120	Death
12	80	110	Death

You can assume that censoring was non-informative with regard to the survival of any individual patient.

- (i) Compute the Nelson-Aalen estimate of the cumulative hazard function, $\Lambda(t)$, where t is the time since having the operation. [6]
 - (ii) Using the results of part (i), deduce an estimate of the survival function for patients who have had this operation. [2]
 - (iii) Estimate the probability of a patient surviving for at least 70 weeks after undergoing the operation. [1]
- [Total 9]

**Question 8.19****Subject CT4, April 2007, Question 8**

A medical study was carried out between 1 January 2001 and 1 January 2006, to assess the survival rates of cancer patients. The patients all underwent surgery during 2001 and then attended 3-monthly check-ups throughout the study.

The following data were collected:

For those patients who died during the study exact dates of death were recorded as follows:

<i>Patient</i>	<i>Date of surgery</i>	<i>Date of death</i>
A	1 April 2001	1 August 2005
B	1 April 2001	1 October 2001
C	1 May 2001	1 March 2002
D	1 September 2001	1 August 2003
E	1 October 2001	1 August 2002

For those patients who survived to the end of the study:

<i>Patient</i>	<i>Date of surgery</i>
F	1 February 2001
G	1 March 2001
H	1 April 2001
I	1 June 2001
J	1 September 2001
K	1 September 2001
L	1 November 2001

For those patients with whom the hospital lost contact before the end of the investigation:

<i>Patient</i>	<i>Date of surgery</i>	<i>Date of last check-up</i>
M	1 February 2001	1 August 2003
N	1 June 2001	1 March 2002
O	1 September 2001	1 September 2005

- (i) Explain whether and where each of the following types of censoring is present in this investigation:
- (a) type I censoring
 - (b) interval censoring; and
 - (c) informative censoring.
- [3]
- (ii) Calculate the Kaplan-Meier estimate of the survival function for these patients.
State any assumptions that you make.
- [7]
- (iii) Hence estimate the probability that a patient will die within 4 years of surgery.
- [1]
- [Total 11]



Chapter 8 Summary

Estimating the future lifetime distribution

As we saw in Chapter 7, we can derive many useful functions from the lifetime distribution $F(t) = P[T \leq t]$. However, $F(t)$ is typically unknown and must be estimated from data.

A *non-parametric* approach is one in which we do not pre-constrain the form of the distribution function before analysing the data.

Censored data

Data for some lives may be *censored*. The main types of censoring (which are not mutually exclusive) are:

- right censoring
- left censoring
- interval censoring
- random censoring
- informative censoring
- non-informative censoring
- Type I censoring
- Type II censoring.

Censored data must be accounted for in the likelihood function. They tend to make the maximisation procedure more complicated.

The Kaplan-Meier model

The *Kaplan-Meier* (or *product-limit*) estimate $\hat{F}_{KM}(t)$ of the lifetime distribution is a step function with jumps at each observed death. It is calculated with reference to the number and timing of deaths and the number of lives alive at each point. To calculate $\hat{F}_{KM}(t)$, we first need to estimate the discrete hazard function.

Discrete hazard function

The discrete hazard function is defined by:

$$\lambda_j = P\left[T = t_j \mid T \geq t_j \right] \quad (1 \leq j \leq k)$$

where t_j denotes the j th observed lifetime. It is estimated by:

$$\hat{\lambda}_j = \frac{d_j}{n_j} \quad (1 \leq j \leq k)$$

Formula for the Kaplan-Meier estimate of the distribution function

$$\hat{F}(t) = 1 - \prod_{t_j \leq t} \left(1 - \hat{\lambda}_j\right)$$

Variance of the Kaplan-Meier estimator

We can estimate the variance of the Kaplan-Meier estimate so that we can compare lifetime distributions of two or more populations and construct confidence intervals.

Greenwood's formula

$$\text{var}\left[\tilde{F}(t)\right] \approx \left(1 - \hat{F}(t)\right)^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}$$

The Nelson-Aalen model

An alternative non-parametric approach is the *Nelson-Aalen* method. For this method we need to estimate the *integrated hazard*.

The integrated (or cumulative) hazard

The integrated hazard is given by:

$$\Lambda_t = \int_0^t \mu_s ds + \sum_{t_j \leq t} \lambda_j$$

If we know (or can estimate) the integrated hazard function, then we can obtain (an estimate of) the distribution function using the result:

$$F(t) = 1 - S(t) = 1 - e^{-\Lambda_t}$$

The Nelson-Aalen estimate of the integrated hazard

$$\hat{\Lambda}_t = \sum_{t_j \leq t} \frac{d_j}{n_j}$$

The Nelson-Aalen estimate of the distribution function

$$\hat{F}_{NA}(t) = 1 - e^{-\hat{\Lambda}_t}$$

The variance of the Nelson-Aalen estimator of the integrated hazard

$$\text{var}(\hat{\Lambda}_t) \approx \sum_{t_j \leq t} \frac{d_j (n_j - d_j)}{n_j^3}$$

Parametric estimation of the survival function

The survival function can also be estimated by assuming that the future lifetime random variable belongs to a particular family of distributions and estimating the parameters of the distribution using maximum likelihood. The general likelihood function is of the form:

$$\prod_{\text{censored lives}} S(t_i) \prod_{\text{deaths}} f(t_i)$$

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 8 Solutions

Solution 8.1

The fundamental results from Chapter 7 are:

$$S(t) = 1 - F(t), \quad f(t) = \frac{d}{dt} F(t), \quad f(t) = S(t)\mu_t, \quad \mu_t = -\frac{S'(t)}{S(t)}$$

Solution 8.2

A well-designed experiment could provide detailed information on the lifetimes of a large cohort of individuals. (A “cohort” is just a collection of individuals whose progress we follow as a group.) However, this information may only be useful as a retrospective measure of mortality patterns. This is because the level and shape of mortality rates would probably have changed significantly over time.

Such an experiment would therefore not provide a clear indication of future levels of mortality (or even very recent levels), which is the information that we are most interested in. For example, 100 years ago people in industrialised countries were dying of diseases that are no longer significant today.

Solution 8.3

- (a) If we are interested only in natural causes of death, some lives will be “lost” to the investigation through accidents, crime, terrorism, suicide *etc.*

Even if we are interested in all causes of death, we may lose track of some lives through data collection problems, *eg* changes of address or emigration.

- (b) With life office policyholders the main reason for “losing” people is when policyholders cancel their policies and withdraw from the group.

Solution 8.4

There will be *left censoring* of all lives that change marital status from single (or divorced or widowed) to married during the investigation. We only know that the change of status occurred since the previous set of information was collected.

There will be *interval censoring* if the exact date of death is unknown, eg if only the calendar year of death is known.

There will be *random censoring* of all lives that change marital status from married to divorced or widowed, or give up accountancy, and consequently no longer qualify as participants of the mortality investigation. There will also be random censoring of all lives from whom data cannot be collected.

There will be *right censoring* of all lives that survive until the end of the investigation in 2005.

Solution 8.5

The original observations were:

17, 13, 15*, 7*, 21, 18*, 5, 18, 6*, 22, 19*, 15, 4, 11, 14*, 18, 10, 10, 8*, 17

These can be re-ordered to obtain:

4, 5, 6*, 7*, 8*, 10, 10, 11, 13, 14*, 15, 15*, 17, 17, 18, 18, 18*, 19*, 21, 22

Number of “lives” under investigation, $n = 20$

Number of *Whiskmatic* failures observed, $m = 13$

Number of times at which failures were observed: $k = 10$

Times at which failures are observed:

$t_1 = 4, t_2 = 5, t_3 = 10, t_4 = 11, t_5 = 13, t_6 = 15, t_7 = 17, t_8 = 18, t_9 = 21, t_{10} = 22$

Number of failures observed at each failure time:

$d_1 = 1, d_2 = 1, d_3 = 2, d_4 = 1, d_5 = 1, d_6 = 1, d_7 = 2, d_8 = 2, d_9 = 1, d_{10} = 1$

Number of remaining lives = $n - m = 20 - 13 = 7$

Number of lives censored:

$c_0 = 0, c_1 = 0, c_2 = 3, c_3 = 0, c_4 = 0, c_5 = 1, c_6 = 1, c_7 = 0, c_8 = 2, c_9 = 0, c_{10} = 0$

and note that $\sum_{j=0}^k c_j = n - m$

Number of lives alive and at risk at t_i :

$n_1 = 20, n_2 = 19, n_3 = 15, n_4 = 13, n_5 = 12, n_6 = 10, n_7 = 8, n_8 = 6, n_9 = 2, n_{10} = 1$

Solution 8.6

The log-likelihood is:

$$\ln L = \sum_{j=1}^k \left[d_j \ln \lambda_j + (n_j - d_j) \ln (1 - \lambda_j) \right]$$

Differentiating with respect to λ_1 :

$$\frac{\partial \ln L}{\partial \lambda_1} = \frac{d_1}{\lambda_1} - \frac{n_1 - d_1}{1 - \lambda_1}$$

Setting this equal to 0:

$$\begin{aligned} \frac{d_1}{\lambda_1} = \frac{n_1 - d_1}{1 - \lambda_1} &\Rightarrow d_1 - d_1 \lambda_1 = n_1 \lambda_1 - d_1 \lambda_1 \\ &\Rightarrow d_1 = n_1 \lambda_1 \\ &\Rightarrow \lambda_1 = \frac{d_1}{n_1} \end{aligned}$$

We are told to assume this is a maximum. So we have $\hat{\lambda}_1 = \frac{d_1}{n_1}$ and it similarly follows

that $\hat{\lambda}_j = \frac{d_j}{n_j}$ for $j = 2, 3, \dots, k$.

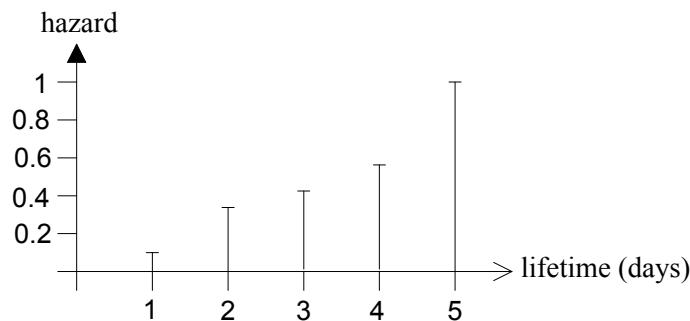
Solution 8.7

$$\lambda_j = P[T = t_j | T \geq t_j] = \frac{P[T = t_j]}{P[T \geq t_j]}$$

$$\lambda_1 = \frac{0.1}{1} = 0.100 \quad \lambda_2 = \frac{0.3}{0.9} = 0.333 \quad \lambda_3 = \frac{0.25}{0.6} = 0.417$$

$$\lambda_4 = \frac{0.2}{0.35} = 0.571 \quad \lambda_5 = \frac{0.15}{0.15} = 1.000$$

A graph of the discrete hazard function is given below.



Solution 8.8

Left censoring occurs when the exact time of entry into a particular state is unknown. All that is known about the time of entry is that it occurred before a particular date. This means that you don't know exactly when to start counting duration from.

Left truncation occurs when only the events (*eg* deaths) that happen after a particular time are observed.

Examples of left censoring include the following situations:

- when estimating functions of exact age and you don't know the exact date of birth;
- when estimating functions of exact policy duration and you don't know the exact date of policy entry;
- when estimating functions of the duration since onset of sickness and you don't know the exact date of becoming sick.

Examples of left censoring do *not* include:

- when estimating functions of exact age and we "lose" the information from before the start of the investigation period, or before the entry date of a policy, *etc.* These are examples of *left truncation* and do *not* affect your ability to measure the exact duration of individuals from their dates of birth.

Solution 8.9

Our estimate of the distribution function never reaches 1 because some rats were alive at the end of the investigation. The estimate will only ever reach 1 if we design an experiment in which observation continues until the last life dies.

Solution 8.10

The figure of $\frac{14}{15}$ corresponds to $1 - \hat{\lambda}_1$.

Solution 8.11

The original observations were:

17, 13, 15*, 7*, 21, 18*, 5, 18, 6*, 22, 19*, 15, 4, 11, 14*, 18, 10, 10 ,8*, 17

These can be re-ordered to obtain:

4, 5, 6*, 7*, 8*, 10, 10 , 11, 13, 14*, 15, 15*, 17, 17, 18, 18, 18*, 19*, 21, 22

The following table shows the rest of the calculation:

j	t_j	d_j	n_j	$\hat{\lambda}_j = d_j / n_j$	$(1 - \hat{\lambda}_j)$	$1 - \prod_{k=1}^j (1 - \hat{\lambda}_k)$
1	4	1	20	0.05000	0.95000	0.0500
2	5	1	19	0.05263	0.94737	0.1000
3	10	2	15	0.13333	0.86667	0.2200
4	11	1	13	0.07692	0.92308	0.2800
5	13	1	12	0.08333	0.91667	0.3400
6	15	1	10	0.10000	0.90000	0.4060
7	17	2	8	0.25000	0.75000	0.5545
8	18	2	6	0.33333	0.66667	0.7030
9	21	1	2	0.50000	0.50000	0.8515
10	22	1	1	1.00000	0.00000	1.0000

So the Kaplan-Meier estimate of $F(t)$ is:

$$\hat{F}_{KM}(t) = \begin{cases} 0 & \text{for } 0 \leq t < 4 \\ 0.0500 & \text{for } 4 \leq t < 5 \\ 0.1000 & \text{for } 5 \leq t < 10 \\ 0.2200 & \text{for } 10 \leq t < 11 \\ 0.2800 & \text{for } 11 \leq t < 13 \\ 0.3400 & \text{for } 13 \leq t < 15 \\ 0.4060 & \text{for } 15 \leq t < 17 \\ 0.5545 & \text{for } 17 \leq t < 18 \\ 0.7030 & \text{for } 18 \leq t < 21 \\ 0.8515 & \text{for } 21 \leq t < 22 \\ 1 & \text{for } t \geq 22 \end{cases}$$

Note that we are estimating an entire function, not just a single parameter value.

Alternatively, you could have used the intuitive graphical approach here.

Solution 8.12

An estimator is a random variable. So its value depends on the outcome of some experiment, and it has a statistical distribution.

An estimate is a number. It is the value taken by an estimator, given a particular set of sample data.

Solution 8.13

By Greenwood's formula:

$$\text{var}[\hat{F}(16)] \approx (1 - \hat{F}(16))^2 \sum_{t_j \leq 16} \frac{d_j}{n_j(n_j - d_j)}$$

We have:

j	t_j	d_j	n_j	$\frac{d_j}{n_j(n_j - d_j)}$
1	4	1	20	0.00263
2	5	1	19	0.00292
3	10	2	15	0.01026
4	11	1	13	0.00641
5	13	1	12	0.00758
6	15	1	10	0.01111

So:

$$\begin{aligned} \text{var}[\hat{F}(16)] &\approx (1 - 0.4060)^2 (0.00263 + 0.00292 + 0.01026 + 0.00641 \\ &\quad + 0.00758 + 0.01111) \\ &= (0.594)^2 \times 0.04091 \\ &= 0.01443 \end{aligned}$$

With a suitable choice of confidence coefficient k , we can construct a confidence interval for $\hat{F}(16)$ of the form $(0.4060 - k \times \sqrt{0.01443}, 0.4060 + k \times \sqrt{0.01443})$, assuming that a normal approximation is appropriate.

In fact, a normal approximation would be suspect here, since we are dealing with a relatively small sample. However, in real life studies involving hundreds of lives, this would be a valid approximation. (You should remember from Subject CT3 that maximum likelihood estimators are asymptotically normally distributed.)

Solution 8.14

We have:

j	t_j	d_j	n_j	$\frac{d_j}{n_j}$	$\sum_{k=1}^j \frac{d_k}{n_k}$	$1 - \exp\left(-\sum_{k=1}^j \frac{d_k}{n_k}\right)$
1	4	1	20	0.05000	0.05000	0.0488
2	5	1	19	0.05263	0.10263	0.0975
3	10	2	15	0.13333	0.23596	0.2102
4	11	1	13	0.07692	0.31288	0.2687
5	13	1	12	0.08333	0.39621	0.3271
6	15	1	10	0.10000	0.49621	0.3912
7	17	2	8	0.25000	0.74621	0.5258
8	18	2	6	0.33333	1.07954	0.6602
9	21	1	2	0.50000	1.57954	0.7939
10	22	1	1	1.00000	2.57954	0.9242

So the Nelson-Aalen estimate of $F(t)$ is:

$$\hat{F}_{NA}(t) = \begin{cases} 0 & \text{for } 0 \leq t < 4 \\ 0.0488 & \text{for } 4 \leq t < 5 \\ 0.0975 & \text{for } 5 \leq t < 10 \\ 0.2102 & \text{for } 10 \leq t < 11 \\ 0.2687 & \text{for } 11 \leq t < 13 \\ 0.3271 & \text{for } 13 \leq t < 15 \\ 0.3912 & \text{for } 15 \leq t < 17 \\ 0.5258 & \text{for } 17 \leq t < 18 \\ 0.6602 & \text{for } 18 \leq t < 21 \\ 0.7939 & \text{for } 21 \leq t < 22 \\ 0.9242 & \text{for } t \geq 22 \end{cases}$$

Comparing this with Solution 8.11, we see that the Kaplan-Meier and Nelson-Aalen estimates are quite close.

Solution 8.15

We have:

j	t_j	d_j	n_j	$\frac{d_j(n_j - d_j)}{n_j^3}$
1	4	1	20	0.00238
2	5	1	19	0.00262
3	10	2	15	0.00770
4	11	1	13	0.00546
5	13	1	12	0.00637
6	15	1	10	0.00900

So:

$$\text{var} [\tilde{\Lambda}_{16}] \approx \sum_{t_j \leq 16} \frac{d_j(n_j - d_j)}{n_j^3} = 0.03353$$

Solution 8.16

For the Weibull model:

$$S(t_i) = e^{-\alpha t_i^\beta}$$

and:

$$f(t_i) = \alpha \beta t_i^{\beta-1} e^{-\alpha t_i^\beta}$$

So, in this case the likelihood function is:

$$\begin{aligned} L &= \prod_{\text{censored lives}} S(t_i) \prod_{\text{deaths}} f(t_i) \\ &= \prod_{\text{censored lives}} e^{-\alpha t_i^\beta} \prod_{\text{deaths}} \alpha \beta t_i^{\beta-1} e^{-\alpha t_i^\beta} \\ &= \alpha^k \beta^k \left(\prod_{\text{deaths}} t_i^{\beta-1} \right) \exp \left(-\alpha \sum_{\text{all lives}} t_i^\beta \right) \end{aligned}$$

where k is the observed number of deaths.

Solution 8.17

The estimates of the survival probabilities are:

$$\hat{p}_{60} = e^{-\hat{\mu}_{60}} = e^{-0.01} = 0.99005$$

$${}_2 \hat{p}_{60} = e^{-(\hat{\mu}_{60} + \hat{\mu}_{61})} = e^{-(0.01 + 0.02)} = e^{-0.03} = 0.97045$$

$${}_3 \hat{p}_{60} = e^{-(\hat{\mu}_{60} + \hat{\mu}_{61} + \hat{\mu}_{62})} = e^{-(0.01 + 0.02 + 0.03)} = e^{-0.06} = 0.94176$$

Solution 8.18(i) ***Computing the Nelson-Aalen estimate***

Here we need to compute the estimates of the discrete hazard rates at each time of death. These are defined as:

$$\lambda_j = \frac{d_j}{n_j}$$

where d_j = number of deaths occurring at the j^{th} time of death and n_j = number of people at risk of death at the j^{th} time of death.

The first thing to do is to rewrite the data in terms of the duration since having the operation (call this t), as follows (D = died; C = censored):

Patient number	Duration t	Event
6	30	D
12	30	D
5	30	C
10	35	D
3	40	D
9	40	D
11	50	D
2	68	D
8	71	D
7	100	C
4	116	C
1	120	C

Assuming that lives who were censored at any time t were at risk of death at time t , then we can calculate the required statistics as follows:

j	j th time of death t_j	Number available to die at time t_j n_j	Number of deaths at time t_j d_j	$\hat{\lambda}_j = \frac{d_j}{n_j}$	Estimate of cumulative hazard function $\hat{\Lambda}(t)$	Value of t to which $\hat{\Lambda}(t)$ applies
0	0	12			0	$0 \leq t < 30$
1	30	12	2	0.1667	0.1667	$30 \leq t < 35$
2	35	9	1	0.1111	0.2778	$35 \leq t < 40$
3	40	8	2	0.2500	0.5278	$40 \leq t < 50$
4	50	6	1	0.1667	0.6944	$50 \leq t < 68$
5	68	5	1	0.2000	0.8944	$68 \leq t < 71$
6	71	4	1	0.2500	1.1444	$71 \leq t < 120$

The Nelson-Aalen estimate, and the range of t to which it applies, is shown in the last two columns of the above table. The estimate of the cumulative hazard function is calculated using:

$$\hat{\Lambda}(t) = \sum \hat{\lambda}_j$$

where the summation is over all j for which $t_j \leq t$.

(ii) *Estimating the survival function*

This is calculated using: $\hat{S}(t) = \exp[-\hat{\Lambda}(t)]$

The results are shown in the following table:

$\hat{S}(t)$	Value of t to which $\hat{S}(t)$ applies
1	$0 \leq t < 30$
0.8465	$30 \leq t < 35$
0.7574	$35 \leq t < 40$
0.5899	$40 \leq t < 50$
0.4994	$50 \leq t < 68$
0.4088	$68 \leq t < 71$
0.3184	$71 \leq t < 120$

(iii) ***Survival probability***

The probability of surviving for at least 70 weeks from the operation is $\hat{S}(70)$.

From the table, we can see that $\hat{S}(70) = 0.4088$.

Solution 8.19(i)(a) ***Type I censoring***

Type I censoring occurs when the censoring times are known in advance.

It is present in this investigation since we knew in advance that all lives still in the investigation on 1 January 2006 were going to be censored on that date.

(i)(b) ***Interval censoring***

Interval censoring occurs when the observational plan only allows us to say that the deaths fell within some interval of time.

Here we know the exact duration at the time of death for Patients A to E. So there is no interval censoring in respect of these patients. However, if Patient M, N or O had died between his last check-up and his first missed check-up, this would be an example of interval censoring. In this case, the only information we would have about the duration at death would be that it fell within a particular 3-month period.

Right censoring is a special case of interval censoring. It occurs when the censoring mechanism cuts short observations in progress. If contact had been lost with Patients M, N and O for a reason other than death, then this would be an example of right censoring. Right censoring also occurs at the end of the investigation since there are patients who are still alive at that time and all we know about the lifetimes of these lives is that they are greater than some known value.

(i)(c) ***Informative censoring***

Informative censoring occurs when the censoring mechanism provides some information about the future lifetimes. It is not likely to be present here.

(ii) ***Kaplan-Meier estimate of the survival function***

We assume that:

- the lives in the investigation are independent with respect to mortality and all follow the same model of mortality
- the censoring is non-informative
- the patients with whom contact is lost are censored half-way through the 3-month period in which contact with them was lost
- at duration 4 years, 4 months, the death of Patient A occurred before Patients J and K were censored.

For each life, we have to calculate the duration at exit. Duration is the length of time since a life had surgery. The values are shown in the table below:

Patient	Duration at exit	Reason for exit
A	4 years, 4 months	Death
B	6 months	Death
C	10 months	Death
D	1 year, 11 months	Death
E	10 months	Death
F	4 years, 11 months	Censored
G	4 years, 10 months	Censored
H	4 years, 9 months	Censored
I	4 years, 7 months	Censored
J	4 years, 4 months	Censored
K	4 years, 4 months	Censored
L	4 years, 2 months	Censored
M	2 years, 7½ months	Censored
N	10½ months	Censored
O	4 years 1½ months	Censored

The Kaplan-Meier estimate of the survival function is:

$$\hat{S}_{KM}(t) = \begin{cases} 1 & \text{for } 0 \leq t < \frac{6}{12} \\ \frac{14}{15} & \text{for } \frac{6}{12} \leq t < \frac{10}{12} \\ \frac{14}{15} \times \frac{12}{14} & \text{for } \frac{10}{12} \leq t < \frac{11}{12} \\ \frac{14}{15} \times \frac{12}{14} \times \frac{10}{11} & \text{for } \frac{11}{12} \leq t < \frac{4}{12} \\ \frac{14}{15} \times \frac{12}{14} \times \frac{10}{11} \times \frac{6}{7} & \text{for } \frac{4}{12} \leq t \leq 5 \end{cases}$$

$$= \begin{cases} 1 & \text{for } 0 \leq t < \frac{6}{12} \\ \frac{14}{15} & \text{for } \frac{6}{12} \leq t < \frac{10}{12} \\ \frac{4}{5} & \text{for } \frac{10}{12} \leq t < \frac{11}{12} \\ \frac{8}{11} & \text{for } \frac{11}{12} \leq t < \frac{4}{12} \\ \frac{48}{77} & \text{for } \frac{4}{12} \leq t \leq 5 \end{cases}$$

where t is measured in years since surgery.

The Kaplan-Meier estimate of the survival function is a step function that steps down at the observed death times. It always starts at 1. Since our first observed death is at time $t = \frac{6}{12}$, the first part of the estimate of the survival function is:

$$\hat{S}_{KM}(t) = 1 \quad \text{for } 0 \leq t < \frac{6}{12}$$

At time $t = \frac{6}{12}$ there are 15 patients in the at-risk group and 1 of them dies at this time. So we estimate the survival probability to be $1 - \frac{1}{15} = \frac{14}{15}$, and this stays constant until the next observed death time, ie until time $t = \frac{10}{12}$. So the second part of the estimate of the survival function is:

$$\hat{S}_{KM}(t) = \frac{14}{15} \quad \text{for } \frac{6}{12} \leq t < \frac{10}{12}$$

At time $t = \frac{10}{12}$ there are 14 patients in the at-risk group. (We started with 15 but 1 died at time $t = \frac{6}{12}$.) Out of these 14, 2 die. So we estimate the probability of not dying at time $t = \frac{10}{12}$ to be $1 - \frac{2}{14} = \frac{12}{14}$, and the probability of still being alive after time $t = \frac{10}{12}$ to be $\frac{14}{15} \times \frac{12}{14}$. (To be alive after time $t = \frac{10}{12}$, you must have not died at time $t = \frac{6}{12}$ and not died at time $t = \frac{10}{12}$.) Our estimate of the survival probability stays constant until the next observed death time, ie time $t = \frac{11}{12}$.

So the third part of the estimate of the survival function is:

$$\hat{S}_{KM}(t) = \frac{14}{15} \times \frac{12}{14} = \frac{12}{15} = \frac{4}{5} \quad \text{for } \frac{10}{12} \leq t < \frac{11}{12}$$

The rest of the function follows in a similar way.

Alternatively, you could have assumed that Patients M, N and O were censored on the dates of their last check-ups. This would give durations at censoring of 2 years 6 months, 9 months and 4 years, respectively. With this assumption, the Kaplan-Meier estimate of the survival function would be:

$$\hat{S}_{KM}(t) = \begin{cases} 1 & \text{for } 0 \leq t < \frac{6}{12} \\ \frac{14}{15} & \text{for } \frac{6}{12} \leq t < \frac{10}{12} \\ \frac{14}{15} \times \frac{11}{13} & \text{for } \frac{10}{12} \leq t < \frac{11}{12} \\ \frac{14}{15} \times \frac{11}{13} \times \frac{10}{11} & \text{for } \frac{11}{12} \leq t < \frac{4}{12} \\ \frac{14}{15} \times \frac{11}{13} \times \frac{10}{11} \times \frac{6}{7} & \text{for } \frac{4}{12} \leq t \leq 5 \end{cases}$$

$$= \begin{cases} 1 & \text{for } 0 \leq t < \frac{6}{12} \\ \frac{14}{15} & \text{for } \frac{6}{12} \leq t < \frac{10}{12} \\ \frac{154}{195} & \text{for } \frac{10}{12} \leq t < \frac{11}{12} \\ \frac{28}{39} & \text{for } \frac{11}{12} \leq t < \frac{4}{12} \\ \frac{24}{39} & \text{for } \frac{4}{12} \leq t \leq 5 \end{cases}$$

(iii) Probability that a patient will die within 4 years of surgery

From (ii), our estimate of this death probability is:

$$\hat{F}_{KM}(4) = 1 - \hat{S}_{KM}(4) = 1 - \frac{8}{11} = \frac{3}{11} = 0.27273$$

Chapter 9

Proportional hazards models



Syllabus objective

- (vi) *Describe estimation procedures for lifetime distributions.*
5. *Describe the Cox model for proportional hazards, derive the partial likelihood estimate in the absence of ties, and state its asymptotic distribution.*

0 *Introduction*

The true level of mortality for an individual is unknown in practice. In order to estimate it, we can carry out an investigation and make statistical inferences based on the observed data.

One of the main problems is *heterogeneity*. The population may include lives with very different characteristics, *eg* males and females, smokers and non-smokers, couch potatoes and fitness freaks. In such circumstances we will observe an average mortality rate over the population as a whole. It would be more informative to split the population into homogeneous subgroups of individuals with similar characteristics (*eg* male smokers, female non-smokers) and identify the level of mortality experienced by members of each subgroup.

In this chapter we will consider:

- how to incorporate in a model the different factors (called *covariates*) that are used to split the population into subgroups
- *proportional hazards models*, where the formula incorporates an adjustment to reflect the characteristics of each particular individual
- *fully parametric models*, where the hazard rate is a simple function of some time period t , and their limitations
- the *Cox model*, which is a particular type of proportional hazards model.

1 Covariates and proportional hazards models

1.1 Covariates

Estimates of the lifetime distribution, whether parametric or non-parametric, are limited in their ability to deal with some important questions in survival analysis, such as the effect of covariates on survival.

A covariate is any quantity recorded in respect of each life, such as age, sex, type of treatment, level of medication, severity of symptoms and so on. If the covariates partition the population into a small number of homogeneous groups, it is possible to compare Kaplan-Meier or other estimates of the survivor function in respect of each population, but a more direct and transparent method is to construct a model in which the effects of the covariates on survival are modelled directly: a regression model. In this section, we will assume that the values of the covariates in respect of the i th life are represented by a $1 \times p$ vector, \mathbf{z}_i .



Comment

The vector notation in this chapter requires some care. The Core Reading uses the same notation for both vectors and scalars. The notation in the ActEd material is consistent with the Core Reading. When you're trying questions on this topic, you might want to use the notation \underline{z} to denote the vector of covariates.

The covariates can be:

- direct measurements (eg age or weight)
- indicator or dummy variables (eg 0 for a male and 1 for a female or 0 for new treatment and 1 for placebo)
- a quantitative interpretation of a qualitative measurement (eg severity of symptoms from 0 to 5 with 0 representing no symptoms and 5 representing extreme severity).

For example, the vector \mathbf{z}_i might be (sex, age, weight, symptoms). If the third life is a 68-year-old male, weighing 74kg, with mild symptoms of the condition under investigation (graded as 1 on a scale from 0 to 5), then we would have $\mathbf{z}_3 = (0, 68, 74, 1)$.

1.2 **Proportional hazards models**

The most widely used regression model in recent years has been the *proportional hazards* model. Proportional hazards (PH) models can be constructed using both parametric and non-parametric approaches to estimating the effect of duration on the hazard function.

In PH models the hazard function for the i th life, $\lambda_i(t, z_i)$, may be written:

$$\lambda_i(t, z_i) = \lambda_0(t)g(z_i)$$

where $\lambda_0(t)$ is a function *only* of duration t , and $g(z_i)$ is a function *only* of the covariate vector. (In keeping with statistical habit, we denote hazards by λ rather than μ .) Here, $\lambda_0(t)$ is the hazard for an individual with a covariate vector equal to zero. It is called the *baseline hazard*.

We will see later that in this type of model, when the covariates all have value zero, the function $g(z_i)$ will equal 1.

Models can be specified in which the effect of covariates changes with duration:

$$\lambda_i(t, z_i) = \lambda_0(t)g(z_i, t)$$

but because the hazard no longer factorises into two terms, one depending only on duration and the other depending only on the covariates, these are not PH models.

We will have a look in more detail in Section 2.3 at how the proportional element works.

They are also both more complex to interpret and more computer-intensive to estimate.

2 Fully parametric models

2.1 Parametric models for the hazard function

In a fully parametric PH model, the strong assumption is made that the lifetime distribution, and hence the hazard, belongs to a given family of parametric distributions, and the regression problem is reduced to estimating the parameters from the data.

Recall from Chapter 7 that the pdf of the future lifetime random variable T_x is:

$$f_{T_x}(t) = {}_t p_x \mu_{x+t} = \mu_{x+t} \exp\left(-\int_0^t \mu_{x+s} ds\right)$$

The “hazard” referred to in the Core Reading is just the force of mortality. You may see the hazard function (from age x) written as $h(t)$ or as $h_x(t)$.

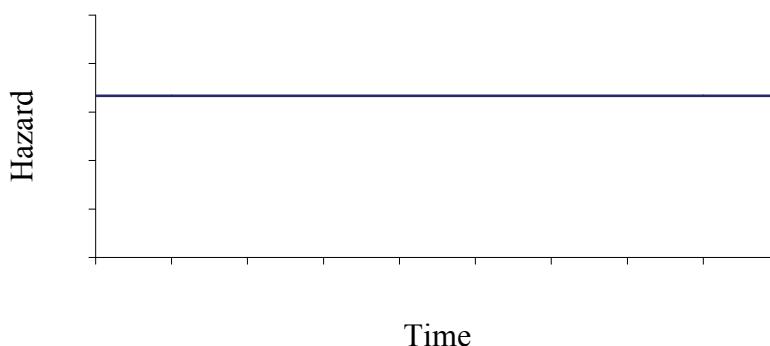
Note also that since $F_{T_x}(t) = P(T_x \leq t) = {}_t q_x = 1 - {}_t p_x$, we can write:

$$\mu_{x+t} = h_x(t) = \frac{f_{T_x}(t)}{1 - F_{T_x}(t)}$$

Distributions commonly used are the exponential (constant hazard), Weibull (monotonic hazard), Gompertz-Makeham (exponential hazard) and log-logistic (“humped” hazard).

The general shapes of the more commonly used distributions are illustrated below.

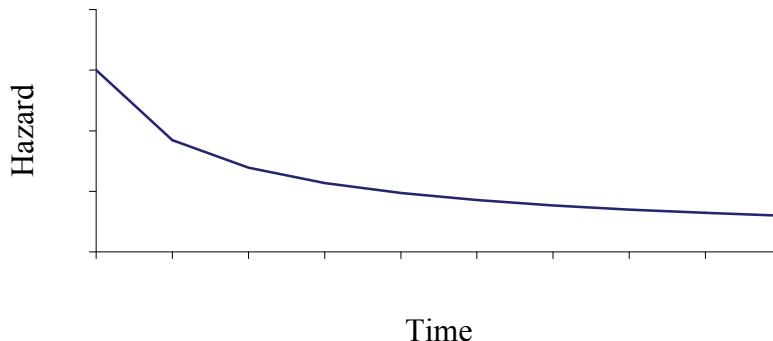
(a) *Exponential (constant hazard)*



**Question 9.1**

Explain why the constant hazard model is described as “exponential”.

- (b) **Weibull (monotonically decreasing hazard)**



The pdf of the Weibull distribution is:

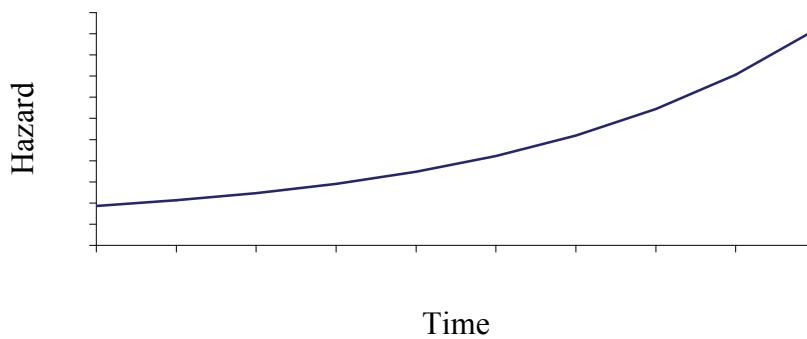
$$f(t) = c\gamma t^{\gamma-1} \exp\{-ct^\gamma\} \quad (c > 0, \gamma > 0, t > 0)$$

Note that the Weibull model can also be used for a monotonically increasing hazard.

**Question 9.2**

Write down the hazard function for the Weibull distribution. For which values of γ is this a monotonically decreasing function?

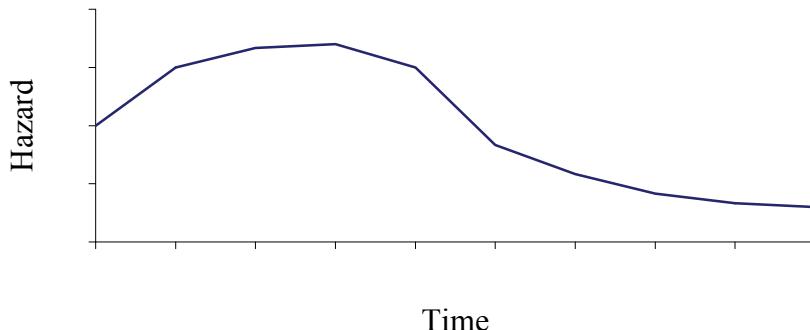
- (c) **Gompertz-Makeham (exponential hazard)**



**Question 9.3**

What is Makeham's law for the force of mortality?

- (d) *Log-logistic hazard (“humped” hazard)*



The log-logistic hazard function is:

$$h(t) = \frac{\gamma(t/\theta)^\gamma}{t[1+(t/\theta)^\gamma]}$$

**Question 9.4**

Give an example of a situation in which the hazard function may be expected to follow each of the following distributions:

- (i) exponential
- (ii) decreasing Weibull
- (iii) Gompertz-Makeham
- (iv) log-logistic.

2.2 Other applications of these models

The same distributions are often used as loss distributions with insurance claims data, but censored observations complicate the likelihoods considerably and numerical methods are usually required. For the distributions above, the likelihoods can be written down (though not always solved) explicitly.

We now look at an example containing censored data where it is possible to determine the maximum likelihood estimate analytically.



Example

Losses arising from a certain group of policies are assumed to follow an $\text{Exp}(\lambda)$ distribution. You are given the following data:

- the exact amounts x_1, x_2, \dots, x_n paid by the insurer in respect of n losses
- data from a further m losses, in respect of which the insurer paid an amount M . The actual loss amounts exceeded M , but you don't know by how much.

Calculate the maximum likelihood estimate of λ .

Solution

The contribution made by each x_i ($i = 1, 2, \dots, n$) is:

$$f(x_i) = \lambda e^{-\lambda x_i}$$

The contribution made by each of the censored loss amounts is:

$$P(X > M) = e^{-\lambda M}$$

So the likelihood function is:

$$L = \left(\prod_{i=1}^n \lambda e^{-\lambda x_i} \right) \left(e^{-\lambda M} \right)^m = \lambda^n e^{-\lambda(\sum x_i + mM)}$$

Taking logs gives:

$$\ln L = n \ln \lambda - \lambda \left(\sum x_i + mM \right)$$

Then differentiating with respect to λ , we get:

$$\frac{d \ln L}{d \lambda} = \frac{n}{\lambda} - \left(\sum x_i + mM \right)$$

Setting this equal to zero and solving:

$$\lambda = \frac{n}{\sum x_i + mM}$$

Finally, we differentiate the log-likelihood again to check the above solution is a maximum:

$$\frac{d^2 \ln L}{d\lambda^2} = -\frac{n}{\lambda^2} < 0 \Rightarrow \text{max}$$

So the maximum likelihood estimate of λ is $\hat{\lambda} = \frac{n}{\sum x_i + mM}$.

2.3 Use of parametric models

Parametric models can be used with a homogeneous population (the one-sample case) as described in Section 6 of Chapter 8, or can be fitted to a moderate number of homogeneous groups, in which case confidence intervals for the fitted parameters give a test of differences between the groups which should be better than non-parametric procedures.

The idea suggested here is that we could fit one of the parametric formulae described above to each of the separate groups (eg one for males and one for females). However, each formula might turn out to be of a different form with no natural link between the formulae used for the different groups. The proportional hazards (PH) approach attempts to get round this problem.

A parametric PH model using the Gompertz distribution might be specified as follows. The Gompertz hazard is:

$$\lambda(t) = Bc^t$$

with two parameters B and c . If we let the value of the parameter B depend on the covariate vector z_i :

$$B = \exp(\beta z_i^T)$$

where β is a $1 \times p$ vector of regression coefficients, then through the scalar product βz_i^T the influence of each factor in z_i enters the hazard multiplicatively. (Note that the "T" denotes the transpose of the vector z_i , not a lifetime.)

We then have the PH model:

$$\lambda_i(t, z_i) = c^t \exp(\beta z_i^T)$$

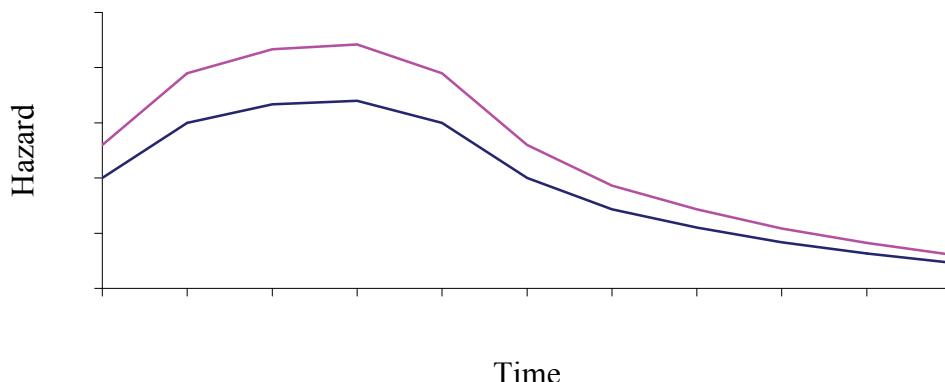
Actuaries are frequently interested in both the baseline hazard and the effect of the covariates. As long as numerical methods are available to maximise the full likelihood (and find the information matrix), which nowadays should not be a problem, it is not difficult to specify any baseline hazard required and to estimate all the parameters simultaneously, ie those in the baseline hazard and the regression coefficients.

Under PH models, the hazards of different lives with covariate vectors z_1 and z_2 are in the same proportion at all times:

$$\frac{\lambda(t, z_1)}{\lambda(t, z_2)} = \frac{\exp(\beta z_1^T)}{\exp(\beta z_2^T)}$$

Hence the name *proportional hazards* model.

The following graph shows the hazard for two lives under the proportional hazards model. The hazard functions are the same shape. The ratio of the hazard rates is constant at all times.



Moreover, the specification above ensures that the hazard is always positive and gives a linear model for the log-hazard:

$$\log \lambda_i(t, z_i) = t \log c + \beta z_i^T$$

which is very convenient in theory and practice.

However, fully parametric models are difficult to apply without foreknowledge of the form of the hazard function. Moreover, in many medical applications answers to questions depend mainly on estimating the regression coefficients. The baseline hazard is relatively unimportant. For these reasons, an alternative semi-parametric approach, originally proposed by D R Cox in 1972, has become popular.

**Question 9.5**

What are the main problems of using a parametric approach to analyse observed survival times?

3 The Cox proportional hazards model

3.1 Introduction



Definition

The Cox PH model proposes the following form of hazard function for the i th life:

$$\lambda(t; z_i) = \lambda_0(t) \exp(\beta z_i^T)$$

$\lambda_0(t)$ is the **baseline hazard**.

So the hazard for a life with covariates z_i is proportional to the baseline hazard, the proportionality factor being the exponential term $\exp(\beta z_i^T)$.

If the covariates for the i th life are $z_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ and the vector of regression parameters is $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ then $\exp(\beta z_i^T) = \exp\left(\sum_{j=1}^p \beta_j X_{ij}\right)$.



Question 9.6

If the j th covariate can take positive values only (eg age), what is the significance of:

- (i) the sign of the j th regression parameter?
- (ii) the magnitude of the j th regression parameter?



Question 9.7

The covariates for the i th observed life are (56, 183, 40) representing (age last birthday at the start of the study, height in cm, daily dose of drug A in mg).

Using the regression parameters $\beta = (0.0172, 0.0028, -0.0306)$, calculate $\lambda(t; z_i)$ in terms of $\lambda_0(t)$.

**Question 9.8**

If the covariates for the i th life are $z_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ and the vector of regression parameters is $\beta = (\beta_1, \beta_2, \dots, \beta_p)$, give an expression for the constant ratio of the hazards of lives with covariate vectors z_1 and z_2 .

3.2 *The utility of the Cox model*

The utility of this model arises from the fact that the general “shape” of the hazard function for all individuals is determined by the baseline hazard, while the exponential term accounts for differences between individuals. So, if we are not primarily concerned with the precise form of the hazard, but with the effects of the covariates, we can ignore $\lambda_0(t)$ and estimate β from the data irrespective of the shape of the baseline hazard. This is termed a **semi-parametric approach**.

In other words, by estimating the vector of parameters β , we can use the Cox model to compare the relative forces of mortality of two lives (or two homogeneous groups of lives). However, we cannot estimate the absolute force of mortality for an individual without first estimating the baseline hazard.

So useful and flexible has this proved, that the Cox model now dominates the literature on survival analysis, and it is probably the tool to which a statistician would turn first for the analysis of survival data.



Example

You are investigating the survival times of patients who have just undergone heart surgery at one of 3 city hospitals – A, B or C. You have recorded the following data for each patient:

$$Z_1 = \begin{cases} 0 & \text{for females} \\ 1 & \text{for males} \end{cases}$$

$$Z_2 = \begin{cases} 1 & \text{if patient attended Hospital B} \\ 0 & \text{otherwise} \end{cases}$$

$$Z_3 = \begin{cases} 1 & \text{if patient attended Hospital C} \\ 0 & \text{otherwise} \end{cases}$$

You have decided to model the force of mortality at time t (measured in days since the operation was performed) by an equation of the form $\lambda(t) = \lambda_0(t)e^{\beta Z^T}$, and you have estimated the parameter values to be:

$$\hat{\beta}_1 = 0.031 \quad \hat{\beta}_2 = -0.025 \quad \hat{\beta}_3 = 0.011$$

Compare the force of mortality for a female patient who attended Hospital A with that of:

- (i) a female patient who attended Hospital B
- (ii) a male patient who attended Hospital C.

Solution

- (i) According to the model, the force of mortality at time t for a female who attended Hospital A is:

$$\lambda_{female, A}(t) = \lambda_0(t)$$

and the force of mortality at time t for a female who attended Hospital B is:

$$\lambda_{female, B}(t) = \lambda_0(t)e^{-0.025}$$

The ratio of these two quantities is:

$$\frac{\lambda_{female,A}(t)}{\lambda_{female,B}(t)} = e^{0.025} = 1.0253$$

So we estimate that the force of mortality for a female who attended Hospital A is 2.53% higher than that of a female who attended Hospital B.

- (ii) Similarly, the force of mortality for a male who attended Hospital C is:

$$\lambda_{male,C}(t) = \lambda_0(t)e^{0.031+0.011} = \lambda_0(t)e^{0.042}$$

So: $\frac{\lambda_{female,A}(t)}{\lambda_{male,C}(t)} = e^{-0.042} = 0.9589$

ie we estimate that the force of mortality for a female who attended Hospital A is 4.11% lower than that of a male who attended Hospital C.

In this example we needed 3 dummy variables (*ie* 3 Z's): one for gender (which has 2 categories), and two for hospital (which has 3 categories). In general, for a covariate that has n categories, we will need $n - 1$ dummy variables.

The group of lives for whom all the dummy variables are 0 is called the baseline group. In the example above, the baseline group is females who attended Hospital A.

3.3 Summary

Before we look at the mathematics underlying the Cox model, it is useful to summarise the material we have covered so far. Understanding the “big picture” will help you to understand the mathematics without getting bogged down in the detail.

The Cox model is a popular mathematical model for the analysis of survival data. Although the model cannot help us to identify the *absolute* level of mortality of a population, it can help us to identify the factors that influence the *relative* levels of mortality between members of the population.

Under the Cox model, we assume that each individual's mortality is proportional to some general mortality function, called the baseline hazard. (This is why it is also a *proportional hazards* model.) What makes the Cox model so flexible is that we do not have to make any assumptions about the shape of this baseline hazard before looking at the data. This helps us to avoid the potential pitfall of trying to fit data to an incompatible model ("a square peg in a round hole").

The constant of proportionality for each individual depends on certain measurable quantities called *covariates*. These may be quantitative (*eg* age) or qualitative (*eg* severity of symptoms of a certain illness).

What we don't yet know is to what extent an individual's covariates affect that individual's mortality. This is the subject of the next section.

4 Estimating the regression parameters

The unknown regression parameters provide the link between an individual's (measurable) covariates and the (unknown) level of the individual's mortality. We will now consider how to estimate the regression parameters, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$.

4.1 The partial likelihood

To estimate β in the Cox model it is usual to maximise the *partial likelihood*. The partial likelihood estimates the regression coefficients but avoids the need to estimate the baseline hazard. Moreover, since (remarkably) it behaves essentially like an ordinary likelihood, it furnishes all the statistical information needed for standard inference on the regression coefficients.

Let $R(t_j)$ denote the set of lives which are at risk just before the j th observed lifetime, and for the moment assume that there is only one death at each observed lifetime, that is $d_j = 1$ ($1 \leq j \leq k$).

The partial likelihood is:

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta z_j^T)}{\sum_{i \in R(t_j)} \exp(\beta z_i^T)}$$

Intuitively, each observed lifetime contributes the probability that the life observed to die should have been the one out of the $R(t_j)$ lives at risk to die, conditional on one death being observed at time t_j .

So the contribution to the partial likelihood from the first death is the force of mortality for the first life to die divided by the total force of mortality for the lives in the at-risk group just prior to this event.

We can also describe this within the framework of Markov jump processes, which we studied in Chapters 5 and 6. Suppose that the lives are labelled Life 1, Life 2, ..., Life N . Before the first death is observed, the process is in the state where all the lives are still alive and in the "at-risk" group. Let's call this State 0. Suppose that the first person to die is Life i . When this death occurs, the process jumps into the state where everyone except the Life i is still alive. Let's call this State i .

Then the probability that the process jumps into State i when it leaves State 0 is:

$$\frac{\text{the force of transition from State 0 to State } i}{\text{the total force of transition out of State 0}}$$

Now the force of transition from State 0 to State i is the force of mortality for Life i , and the total force of transition out of State 0 is the sum of the forces of mortality for everyone in the at-risk group. So if Life i were observed to die at age x , the contribution made to the partial likelihood in respect of this death would be:

$$\frac{\mu_i(x)}{\mu_1(x) + \mu_2(x) + \dots + \mu_N(x)}$$

where $\mu_j(x)$ is the force of mortality for Life j at age x .

A contribution is made to the partial likelihood every time a death is observed, and the partial likelihood is obtained by multiplying all these contributions together.

Note that the baseline hazard cancels out and the partial likelihood depends only on the order in which deaths are observed. (The name “partial” likelihood arises because those parts of the full likelihood involving the times at which deaths were observed and what was observed between the observed deaths are thrown away.)

Unlike the Kaplan-Meier method, this partial likelihood considers observed deaths only, not the times at which the deaths occurred, nor any censoring observed between deaths. The form of the partial likelihood gives the *comparative* risk of a *particular* individual dying, given that a death occurs.

For example, if the first life to die was the tallest individual in the population and the i th covariate is height, then we may infer that height has a significant influence on mortality. In terms of the Cox model, we may infer that the value of β_i is positive.

Of course, our inferences should be based on *all* the observed deaths. By maximising this partial likelihood, our estimates of the regression parameters will be based on the *order* in which the deaths occurred. After all, the model seeks to identify the factors that influence mortality rates and hence increase or reduce the chance of an untimely death.



Example

A group of six lives was observed over a period of time as part of a mortality investigation. Each of the lives was under observation at all ages from age 55 until they died or were censored. The table below shows the sex, age at exit and reason for exit from the investigation.

Life	Sex	Age at exit	Reason for exit
1	M	56	death
2	F	62	censored
3	F	63	death
4	M	66	death
5	M	67	censored
6	M	67	censored

The following model has been suggested for the force of mortality:

$$\mu(x | Z = z) = \mu_0(x) e^{\beta z}$$

where:

- x denotes age
- $\mu_0(x)$ is the baseline hazard
- $z = 0$ for males and $z = 1$ for females.

Write down the partial likelihood for these observations using the model above.

Solution

Since there are three ages at which deaths occur, the partial likelihood will be the product of three terms – one in respect of each death.

The contribution to the partial likelihood from the first death is:

$$\frac{\mu_1(56)}{\mu_1(56) + \mu_2(56) + \dots + \mu_6(56)}$$

where $\mu_i(y)$ is the force of mortality of the i th life at age y .

In other words, we take the force of mortality for the life that dies at the youngest age and divide it by the total force of mortality for those alive at that age.

Using the model suggested, this is equivalent to:

$$\frac{\mu_0(56)}{\mu_0(56) + \mu_0(56)e^\beta + \mu_0(56)e^\beta + \mu_0(56) + \mu_0(56) + \mu_0(56)} = \frac{1}{4 + 2e^\beta}$$

Similarly, the contribution of the second death to the partial likelihood is:

$$\frac{\mu_3(63)}{\mu_3(63) + \mu_4(63) + \mu_5(63) + \mu_6(63)} = \frac{e^\beta}{e^\beta + 3}$$

Finally, the contribution of the third death to the partial likelihood is:

$$\frac{\mu_4(66)}{\mu_4(66) + \mu_5(66) + \mu_6(66)} = \frac{1}{3}$$

Multiplying these three terms together, we obtain the partial likelihood:

$$L = \frac{1}{4 + 2e^\beta} \times \frac{e^\beta}{e^\beta + 3} \times \frac{1}{3} = \frac{Ce^\beta}{(e^\beta + 2)(e^\beta + 3)}$$

where C is a constant.

4.2 Maximising the partial likelihood

Maximisation of this expression has to proceed numerically, and most statistics packages have procedures for fitting a Cox model.

Maximisation of this partial likelihood will yield our maximum likelihood estimate of the regression parameters and hence provide a link between measurable covariates and mortality (or hazard) rates. The maximisation process is complicated and often cannot be achieved directly. It may be carried out by an iterative numerical technique such as the Newton-Raphson method, which uses repeated calculations to refine the choice of regression parameters until the maximum is found to a sufficient degree of accuracy.

In the example above, there was only one covariate. In this case it is straightforward to work out the maximum likelihood estimate of the parameter β .



Question 9.9

Calculate the maximum partial likelihood estimate of β , the model parameter in the previous example.

Usually estimates of the parameters will not be so easy to calculate.

In practice there might be ties in the data, that is:

- (a) some $d_j > 1$; or
- (b) some observations are censored at an observed lifetime.

It is usual to deal with (b) by including the lives on whom observation was censored at time t_j in the risk set $R(t_j)$, effectively assuming that censoring occurs just after the deaths were observed.

Also, in a more realistic model, there will be many β parameters rather than just one.



Question 9.10

Suppose that from the investigation in the previous example we now have the following additional data:

Life	Sex	Age at exit	Reason for exit
7	M	56	censored
8	F	62	censored

How do the extra data affect the contribution to the partial likelihood from the first death?

Breslow's approximation

Accurate calculation of the partial likelihood in case (a) is messy, since all possible combinations of d_j deaths out of the $R(t_j)$ at risk at time t_j ought to contribute, and an approximation due to Breslow is often used, namely:

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta s_j^T)}{\left(\sum_{i \in R(t_j)} \exp(\beta z_i^T) \right)^{d_j}}$$

where s_j is the sum of the covariate vectors z of the d_j lives observed to die at time t_j .

So, if two lives – A and B, say – are observed to die at time t_j , we assume that the contribution to the partial likelihood from A's death is $\frac{\mu_A(t_j)}{\sum_{i \in R(t_j)} \mu_i(t_j)}$ and the

contribution to the partial likelihood from B's death is $\frac{\mu_B(t_j)}{\sum_{i \in R(t_j)} \mu_i(t_j)}$. Lives A and B are both included in the at-risk group $R(t_j)$ in each denominator. This should become clearer when you look at the example below.

Example

An investigation was carried out into the survival times (measured in months) of patients in hospital following liver transplants. The covariates are $z_{1i} = 0$ for placebo, 1 for treatment X, and $z_{2i} = \text{weight of patient}$ (measured in kg).

The observed lifetimes (with weights in brackets) were as follows:

Placebo	Treatment X
3 (83)	6*(58)
9 (68)	11(73)
14 (75)	14(68)
16 (86)	14* (49)

Observations with an asterisk represent censored observations.

Using Breslow's assumption, what contribution to the partial likelihood is made by the deaths at time 14?

Solution

Just before time 14, there were four lives at risk. The total force of mortality for these four lives at time 14 is:

$$\mu_0(14)e^{75\beta_2} + \mu_0(14)e^{\beta_1+68\beta_2} + \mu_0(14)e^{\beta_1+49\beta_2} + \mu_0(14)e^{86\beta_2}$$

where $\mu_0(t)$ denotes the baseline hazard at time t , measured in months since the transplant operation.

The individual forces of mortality for the two lives that die at time 14 are:

$$\mu_0(14)e^{75\beta_2} \quad \text{and} \quad \mu_0(14)e^{\beta_1+68\beta_2}$$

So the contribution to the partial likelihood from the deaths that occur at time 14 is:

$$\begin{aligned} & \frac{\mu_0(14)e^{75\beta_2}}{\mu_0(14)e^{75\beta_2} + \mu_0(14)e^{\beta_1+68\beta_2} + \mu_0(14)e^{\beta_1+49\beta_2} + \mu_0(14)e^{86\beta_2}} \\ & \times \frac{\mu_0(14)e^{\beta_1+68\beta_2}}{\mu_0(14)e^{75\beta_2} + \mu_0(14)e^{\beta_1+68\beta_2} + \mu_0(14)e^{\beta_1+49\beta_2} + \mu_0(14)e^{86\beta_2}} \\ = & \frac{\mu_0(14)e^{75\beta_2} \times \mu_0(14)e^{\beta_1+68\beta_2}}{\left[\mu_0(14)e^{75\beta_2} + \mu_0(14)e^{\beta_1+68\beta_2} + \mu_0(14)e^{\beta_1+49\beta_2} + \mu_0(14)e^{86\beta_2} \right]^2} \\ = & \frac{e^{\beta_1+143\beta_2}}{\left[e^{75\beta_2} + e^{\beta_1+68\beta_2} + e^{\beta_1+49\beta_2} + e^{86\beta_2} \right]^2} \end{aligned}$$

since all the baseline hazard terms cancel.

4.3 Properties of the partial likelihood

As mentioned earlier, the partial likelihood behaves much like a full likelihood; it yields an estimator for β which is asymptotically (multivariate) normal and unbiased, and whose asymptotic variance matrix can be estimated by the inverse of the observed information matrix.



Question 9.11

Explain the meaning of the statement:

“The maximum partial likelihood estimator of β is asymptotically unbiased.”

The **efficient score function, namely the vector function:**

$$u(\beta) = \left(\frac{\partial \log L(\beta)}{\partial \beta_1}, \dots, \frac{\partial \log L(\beta)}{\partial \beta_p} \right)$$

plays an important part; in particular solving $u(\hat{\beta})=0$ furnishes the maximum likelihood estimate $\hat{\beta}$.

The observed information matrix $I(\hat{\beta})$ is then the negative of the $p \times p$ matrix of second partial derivatives:

$$I(\beta)_{ij} = -\frac{\partial^2 \log L(\beta)}{\partial \beta_i \partial \beta_j} \quad (1 \leq i, j \leq p)$$

evaluated at $\hat{\beta}$.

The variance matrix is the symmetric matrix C , whose i, j th entry is equal to $\text{cov}(\tilde{\beta}_i, \tilde{\beta}_j)$. The above Core Reading is saying that, asymptotically:

$$C = [I(\hat{\beta})]^{-1} = - \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \beta_1^2} \Big|_{\beta=\hat{\beta}} & \frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_2} \Big|_{\beta=\hat{\beta}} & \dots & \frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_p} \Big|_{\beta=\hat{\beta}} \\ \frac{\partial^2 \ln L}{\partial \beta_2 \partial \beta_1} \Big|_{\beta=\hat{\beta}} & \frac{\partial^2 \ln L}{\partial \beta_2^2} \Big|_{\beta=\hat{\beta}} & \dots & \frac{\partial^2 \ln L}{\partial \beta_2 \partial \beta_p} \Big|_{\beta=\hat{\beta}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln L}{\partial \beta_p \partial \beta_1} \Big|_{\beta=\hat{\beta}} & \frac{\partial^2 \ln L}{\partial \beta_p \partial \beta_2} \Big|_{\beta=\hat{\beta}} & \dots & \frac{\partial^2 \ln L}{\partial \beta_p^2} \Big|_{\beta=\hat{\beta}} \end{bmatrix}^{-1}$$

The algebra simplifies considerably when we consider the one-parameter case.

One-parameter case

For a model with only one covariate β , say, we calculate the maximum partial likelihood estimate of β by solving the equation:

$$\frac{d \ln L}{d \beta} = 0$$

We can also estimate the variance of the maximum partial likelihood estimator $\tilde{\beta}$ using the approximation:

$$\text{var}(\tilde{\beta}) \approx \left(-\frac{d^2 \ln L}{d \beta^2} \right)^{-1} \Big|_{\beta=\hat{\beta}}$$

You may recognise the expression given here. This is just the Cramér-Rao lower bound (studied in Subject CT3).



Question 9.12

Calculate the asymptotic standard error of the estimator in Question 9.9.

**Question 9.13**

Use your results from Question 9.9 and Question 9.12 to construct an approximate 95% confidence interval for the model parameter β . What can you infer from this?

A useful feature of most computer packages for fitting a Cox model is that the information matrix evaluated at $\hat{\beta}$ is usually produced as a by-product of the fitting process (it is used in the Newton-Raphson algorithm) so standard errors of the components of $\hat{\beta}$ are available. These are helpful in evaluating the fit of a particular model.

5 Model fitting

5.1 Assessing the effect of the covariates

In a practical problem, several possible explanatory variables might present themselves, and part of the modelling process is the selection of those that have significant effects. Therefore criteria are needed for assessing the effects of covariates, alone or in combination.

A common criterion is the *likelihood ratio statistic*. Suppose we need to assess the effect of adding further covariates to the model. In general, suppose we fit a model with p covariates, and another model with $p+q$ covariates, which include the p covariates of the first model.

Each is fitted by maximising a likelihood; let L_p and L_{p+q} be the maximised log-likelihoods of the first and second models respectively.

Note that L is usually used to denote a likelihood function, with $\ln L$, $\log L$ or even just l being used for the log-likelihood. However, in this instance, the Core Reading is using L to denote a log-likelihood.



Definition

The likelihood ratio statistic is then:

$$-2(L_p - L_{p+q})$$

and it has an asymptotic χ^2 distribution on q degrees of freedom, under the hypothesis that the extra q covariates have no effect in the presence of the original p covariates.

In other words, the null hypothesis for this test is:

$$H_0 : \beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+q} = 0$$

i.e the extra covariates are not significant.

The test statistic is:

$$-2(\ln L_p - \ln L_{p+q})$$

where the log-likelihoods are calculated using the maximum partial likelihood estimates. (Note that the Core Reading is using L to denote a log-likelihood here.)

The null hypothesis will be rejected at the 5% significance level if the value of test statistic is greater than the upper 5% point of χ_q^2 .

Strictly this statistic is based upon full likelihoods, but when fitting a Cox model it is used with partial likelihoods.

In practice, the likelihood ratio statistic would be calculated numerically using a statistical computer package.



Example

Suppose we have considered a model for the effect of hypertension on survival, in which z_i has two components, with the level of $z_i^{(1)}$ representing sex and the level of $z_i^{(2)}$ representing blood pressure.

Suppose we want to test the hypothesis that cigarette smoking has no effect, allowing for sex and blood pressure.

Then we could define an augmented covariate vector $z'_i = (z_i^{(1)}, z_i^{(2)}, z_i^{(3)})$ in which $z_i^{(3)}$ is a factor (say, 0 for non-smoker and 1 for smoker) and refit the model.

The likelihood ratio statistic $-2(L_2 - L_3)$ then has an asymptotic χ^2 distribution on 1 degree of freedom, under the null hypothesis (which is that the new parameter $\beta_3 = 0$).

5.2 Building models

The likelihood ratio statistic is the basis of various model-building strategies, in which:

- (a) we start with the *null model* (one with no covariates) and add possible covariates one at a time; or
- (b) we start with a *full model* which includes all possible covariates, and then try to eliminate those of no significant effect.

In addition, it is necessary to test for *interactions* between covariates, in case their effects should depend on the presence or absence of each other. Although not required study for this subject, some examples of model building strategies, and the interpretation of likelihood ratio statistics, are given by Collett (1994).

The likelihood ratio statistic is a standard tool in model selection; for example it was used in the UK to choose members of a Gompertz-Makeham family of functions for parametric graduations (see Chapter 13).



Example (continued)

We could also extend the model developed in the example in Section 5.1 to test for *interactions* between covariates.

For example, we may wish to investigate whether mortality is influenced by a combination of smoking and high blood pressure.

In this case, we could extend the covariate vector to include the product $z^{(4)} = z^{(2)} \times z^{(3)}$, then refit the model and test the null hypothesis (which is that parameter $\beta_4 = 0$).



Example

A study is carried out to ascertain the link between the mortality of pensioners and socio-economic group. The survival times are to be modelled using a Cox regression model, which is to include allowance for two other influences on mortality – sex and smoking status. The model is to test for two-way interaction between socio-economic group and the other factors.

Suggest how the initial model might be formulated.

Solution

The initial model might be specified as:

$$\lambda(x; z_i) = \lambda_0(x) \cdot \exp(\beta z_i^\top) = \lambda_0(x) \cdot \exp\left(\sum_{j=1}^p \beta_j z_{ij}\right)$$

where x denotes age, and the covariates of the model for the i th life are:

z_{i1} = socio-economic group from 0 (low) to 4 (high)

z_{i2} = sex (0 for male, 1 for female)

z_{i3} = smoking status (0 for smoker, 1 for non-smoker)

$z_{i4} = z_{i1} \times z_{i2}$

$z_{i5} = z_{i1} \times z_{i3}$



Question 9.14

Using the model in the above example, calculate the ratio of hazards for two lives A and B of the same age, where A is a male smoker in socio-economic group 3 and B is a male non-smoker in socio-economic group 1.

**Question 9.15**

Using the model in the above example, you wish to test for interaction between socio-economic group and the other factors. State the null hypothesis, state the test statistic and describe how you would carry out the test.

5.3 Using the results

After fitting the model and analysing the likelihood ratio statistics, we can make inferences about how each covariate affects mortality. This information can be used in many different ways:

- The model may be used to assess the efficacy of a new medical treatment for patients. The treatment would be represented by a covariate, which may be a quantitative measure of dose or an indicator, *eg* 0 for placebo, 1 for treatment.
- A life insurance company may wish to know how certain covariates affect mortality, so that it can charge premiums that accurately reflect the risk for an individual, *eg* higher premiums for smokers. However, an insurance company will be restricted to covariates that can be collected easily and reliably from potential policyholders. (We return to this idea in Chapter 11, when we discuss heterogeneity within a population.)

The Cox model can provide an estimate of the relative level of an individual's mortality in comparison to the baseline hazard. By making certain assumptions about the shape and level of the baseline hazard, we can then estimate the absolute level of an individual's mortality.

6 Exam-style questions

You should now be able to attempt the following past exam questions on Cox regression.



Question 9.16

Subject 104, September 2004, Question 3

A study has been undertaken into the effect of a new treatment on the survival times of patients suffering from a tropical disease. The following model has been fitted:

$$h_i(t) = h_0(t) \exp(\underline{\beta}^T \underline{z})$$

where $h_i(t)$ is the hazard at time t , where t is the time since treatment

$h_0(t)$ is the baseline hazard at time t

\underline{z} is a vector of covariates, where

z_1 = period from diagnosis to treatment in years

z_2 = 0 if existing treatment given, 1 if new treatment given

z_3 = 0 if female, 1 if male

$\underline{\beta}$ is a vector of parameters, where

β_1 = 0.5

β_2 = 0.01

β_3 = -0.05

- (i) State the group of lives to which the baseline hazard applies. [1]
- (ii) For a male who was given the new treatment 6 months after diagnosis:
 - (a) Write down the hazard function, in terms of $h_0(t)$ only.
 - (b) Express the survival function, in terms of $h_0(t)$ only. [3]
- (iii) For a female given the new treatment at the time of diagnosis, the probability of survival for 5 years is 0.75. Calculate the probability that the male in (ii) will survive 5 years. [3]

[Total 7]

The next question examines material from Chapters 8 and 9.

**Question 9.17****Subject CT4, September 2007, Question 10**

- (i) Compare the advantages and disadvantages of fully parametric models and the Cox regression model for assessing the impact of covariates on survival. [3]

You have been asked to investigate the impact of a set of covariates, including age, sex, smoking, region of residence, educational attainment and amount of exercise undertaken, on the risk of heart attack. Data are available from a prospective study which followed a set of several thousand persons from an initial interview until their first heart attack, or until their death from a cause other than a heart attack, or until 10 years had elapsed since the initial interview (whichever of these occurred first).

- (ii) State the types of censoring present in this study, and explain how each arises.[2]
- (iii) Describe a criterion which would allow you to select those covariates which have a statistically significant effect on the risk of heart attack, when controlling the other covariates of the model. [4]

Suppose your final model is a Cox model which has three covariates: age (measured in age last birthday minus 50 at the initial interview), sex (male = 0, female = 1) and smoking (non-smoker = 0, smoker = 1), and that the estimated parameters are:

Age	0.01
Sex	-0.4
Smoking	0.5
Sex × smoking	-0.25

where “sex × smoking” is an additional covariate formed by multiplying the two covariates “sex” and “smoking”.

- (iv) Describe the final model’s estimate of the effect of sex and of smoking behaviour on the risk of heart attack. [3]
- (v) Use the results of the model to determine how old a female smoker must be at the initial interview to have the same risk of heart attack as a male non-smoker aged 50 years at the initial interview. [3]

[Total 15]

7 ***End of Part 3***

You have now completed Part 3 of the Subject CT4 Notes.

Review

Before looking at the Question and Answer Bank we recommend that you briefly review the key areas of Part 3, or maybe re-read the summaries at the end of Chapters 7 to 9.

Question and Answer Bank

You should now be able to answer the questions in Part 3 of the Question and Answer Bank. We recommend that you work through several of these questions now and save the remainder for use as part of your revision.

Assignments

On completing this part, you should be able to attempt the questions in Assignment X3.

Reminder

If you have not yet booked a tutorial, then maybe now is the time to do so.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.



Chapter 9 Summary

Covariates

A *covariate* is any quantity recorded in respect of each life, such as age, sex, type of treatment, level of medication, severity of symptoms and so on.

Proportional hazards (PH) models

In a *proportional hazards* model the hazard function for the i th life, $\lambda_i(t; z_i)$, may be written as:

$$\lambda_i(t; z_i) = \lambda_0(t)g(z_i)$$

The *baseline hazard* $\lambda_0(t)$ is a function *only* of the duration t and $g(z_i)$ is a function *only* of the covariate vector z_i .

The hazards of different lives are independent of the baseline hazard and are in the same proportion at all times. This proportion depends on the values of the covariates recorded for each life.

Fully parametric models

Fully parametric models assume a lifetime distribution based on a statistical distribution whose parameters must then be determined.

Commonly used distributions include:

- the *exponential* distribution (constant hazard)
- the *Weibull* distribution (monotonic hazard)
- the *Gompertz-Makeham* formula (exponential hazard)
- the *log-logistic* distribution (“humped” hazard).

The Cox PH model

The *Cox model* is a semi-parametric proportional hazards model under which the force of mortality (or hazard function) for an individual life is given by:

$$\lambda(t; z_i) = \lambda_0(t) \exp(\beta z_i^\top)$$

The force of mortality is proportional to the baseline hazard $\lambda_0(t)$.

The Cox model is a proportional hazards model because the hazards of different lives are independent of the baseline hazard and are in the same proportion at all times. This proportion depends on the values of the covariates recorded for each life and the values of the *regression parameters* β :

$$\frac{\lambda(t; z_1)}{\lambda(t; z_2)} = \frac{\exp(\beta z_1^\top)}{\exp(\beta z_2^\top)} = \text{constant}$$

It can be used to investigate the effect of different factors on mortality. The data collected for each life in the investigation must include information about the covariates, which may be qualitative or quantitative.

Fitting the regression parameters

The regression parameters are estimated by maximising the *partial likelihood*:

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta z_j^\top)}{\sum_{i \in R(t_j)} \exp(\beta z_i^\top)}$$

Solving the equation:

$$u(\beta) = \left(\frac{\partial \log L(\beta)}{\partial \beta_1}, \dots, \frac{\partial \log L(\beta)}{\partial \beta_p} \right) = 0$$

gives the maximum partial likelihood estimates of $\beta_1, \beta_2, \dots, \beta_p$. We denote these estimates by $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$. The maximisation procedure is usually carried out using a computer.

Breslow's approximation to the partial likelihood

If there are ties in the data, *ie* the death times are not distinct, then Breslow's approximation to the partial likelihood can be used:

$$L(\beta) = \prod_{j=1}^k \frac{\exp(\beta s_j^\top)}{\left(\sum_{i \in R(t_j)} \exp(\beta z_i^\top) \right)^{d_j}}$$

Distribution of the maximum partial likelihood estimators of the regression parameters

The maximum partial likelihood estimator of the vector of parameters β , which we denote by $\tilde{\beta}$ has the following asymptotic properties:

- It has an asymptotic multivariate normal distribution.
- It is asymptotically unbiased.
- Its variance matrix is equal to the negative of the inverse of the observed information matrix, *ie* the negative of the inverse of the matrix of second derivatives of the log-likelihood, evaluated at the point $\hat{\beta}$.

So an approximate 95% confidence interval for β_j (the j th parameter) is:

$$\hat{\beta}_j \pm 1.96 \sqrt{\text{var}(\tilde{\beta}_j)}$$

Model testing

We can compare two models using a likelihood ratio test. Suppose we want to compare a model with p covariates against an extended model with an extra q covariates.

The null hypothesis for this test is:

$$H_0 : \beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+q} = 0$$

The test statistic is:

$$-2(\ln L_p - \ln L_{p+q})$$

where $\ln L_p$ ($\ln L_{p+q}$) denotes the maximised log-likelihood of the model with p ($p + q$) covariates.

If the null hypothesis is true, then the test statistic should be a realisation of a χ_q^2 random variable. So we reject the null hypothesis at the 5% significance level if the value of the test statistic is greater than the upper 5% point of χ_q^2 .

Chapter 9 Solutions

Solution 9.1

Under the constant hazard model with hazard rate λ , the distribution function of the future lifetime of a life aged x is:

$$F_{T_x}(t) = {}_t q_x = 1 - \exp\left(-\int_0^t \lambda_{x+s} ds\right) = 1 - \exp\left(-\int_0^t \lambda ds\right) = 1 - e^{-\lambda t} \quad (t \geq 0)$$

This is the distribution function of an $Exp(\lambda)$ random variable.

Solution 9.2

The hazard function for the Weibull distribution is:

$$h(t) = c\gamma t^{\gamma-1} \quad (t > 0)$$

Differentiating this gives:

$$h'(t) = c\gamma(\gamma-1)t^{\gamma-2}$$

This is:

- negative for $\gamma < 1$
- 0 for $\gamma = 1$
- positive for $\gamma > 1$.

So the Weibull hazard is decreasing for $0 < \gamma < 1$.

Solution 9.3

Makeham's law for the force of mortality is:

$$\mu_x = A + Bc^x$$

for some parameters A , B and c . This is an exponential hazard since the variable x appears as the power.

Solution 9.4

- (i) The constant hazard model (exponential) could reflect the hazard for an individual who remains in good health. The level of hazard would reflect the risk of death from unnatural causes *eg* accident or murder.
- (ii) The decreasing hazard model (decreasing Weibull) could reflect the hazard for patients recovering from major heart surgery. The level of hazard is expected to fall as the time since the operation increases.
- (iii) The exponentially increasing hazard model (Gompertz-Makeham) could reflect the hazard for leukaemia sufferers who are not responding to treatment. The severity of the condition and the level of hazard increase with the survival time. Over longer time periods, the Gompertz-Makeham model could be suitable for describing the increasing chance of death from natural causes as age increases – we saw this in Chapter 7.
- (iv) The humped hazard (log-logistic) could reflect a hazard for patients with a disease that is most likely to cause death during the early stages *eg* TB. As the initial condition becomes more severe, the level of hazard increases. But once patients have survived the period of highest risk, the level of hazard decreases.

Solution 9.5

If an inappropriate family of parametric distributions is chosen, the hazard function will be the wrong shape. Whilst regression parameters can be chosen to maximise the likelihood for the observed data, the model will not be suitable for estimation. Typically, we do not know the form of the distribution before analysing the data.

The hazard function could also be the wrong shape if the population comprises several heterogeneous subgroups.

This type of model does not guarantee any consistency between the formulae used for the different subgroups.

Solution 9.6

- (i) If the j th regression parameter is positive, the hazard rate (*e.g.* the force of mortality) increases with the j th covariate, *i.e.* there is a positive correlation between hazard rate and covariate. For example, if obese individuals are more likely to suffer from major heart disease, we would expect to find the regression parameter associated with the covariate representing weight to be positive.

If the j th regression parameter is negative, the hazard rate (*e.g.* the force of mortality) decreases with the j th covariate, *i.e.* there is a negative correlation between hazard rate and covariate. For example, if individuals who drink a high volume of non-alcoholic liquids are less likely to suffer from liver disease, we would expect to find the regression parameter associated with the covariate representing liquid intake to be negative.

- (ii) If the magnitude of the j th regression parameter is large, the hazard rate (*e.g.* the force of mortality) is significantly affected by the j th covariate, *i.e.* there is a strong correlation (positive or negative) between hazard rate and covariate. If the magnitude of the j th regression parameter is small, the hazard rate is not significantly affected by the j th covariate, *i.e.* there is a weak correlation between hazard rate and covariate.

The significance of each covariate can be tested statistically.

Solution 9.7

$$\begin{aligned}\lambda(t; z_i) &= \lambda_0(t) \times \exp(56 \times 0.0172 + 183 \times 0.0028 + 40 \times -0.0306) \\ &= \lambda_0(t) \times \exp(0.2516) \\ &= \lambda_0(t) \times 1.286\end{aligned}$$

Solution 9.8

$$\frac{\lambda(t; z_1)}{\lambda(t; z_2)} = \frac{\exp(\beta z_1^\top)}{\exp(\beta z_2^\top)} = \frac{\exp\left(\sum_{j=1}^p \beta_j X_{1j}\right)}{\exp\left(\sum_{j=1}^p \beta_j X_{2j}\right)} = \exp\left(\sum_{j=1}^p \beta_j (X_{1j} - X_{2j})\right)$$

Solution 9.9

From the example in Section 4.1, we have:

$$L = \frac{Ce^\beta}{(e^\beta + 2)(e^\beta + 3)}$$

Taking logs gives:

$$\log L = \log C + \beta - \log(e^\beta + 2) - \log(e^\beta + 3)$$

Differentiating with respect to β :

$$\frac{d \log L}{d\beta} = 1 - \frac{e^\beta}{e^\beta + 2} - \frac{e^\beta}{e^\beta + 3}$$

Setting this equal to 0:

$$\begin{aligned} \frac{(e^\beta + 2)(e^\beta + 3) - (e^\beta + 3)e^\beta - (e^\beta + 2)e^\beta}{(e^\beta + 2)(e^\beta + 3)} &= 0 \\ \Rightarrow e^{2\beta} + 5e^\beta + 6 - e^{2\beta} - 3e^\beta - e^{2\beta} - 2e^\beta &= 0 \\ \Rightarrow -e^{2\beta} + 6 &= 0 \\ \Rightarrow 2\beta &= \log 6 \\ \Rightarrow \beta &= \frac{1}{2} \log 6 \end{aligned}$$

Differentiating the partial log-likelihood a second time gives:

$$\begin{aligned} \frac{d^2 \log L}{d\beta^2} &= -\frac{e^\beta(e^\beta + 2) - e^{2\beta}}{(e^\beta + 2)^2} - \frac{e^\beta(e^\beta + 3) - e^{2\beta}}{(e^\beta + 3)^2} \\ &= -\frac{2e^\beta}{(e^\beta + 2)^2} - \frac{3e^\beta}{(e^\beta + 3)^2} \\ &< 0 \end{aligned}$$

So the maximum likelihood estimate of β is $\hat{\beta} = \frac{1}{2} \log 6 = 0.896$.

Solution 9.10

We include the life that was censored at age 56 in the at-risk group at age 56. So we now have 5 males and 3 females at risk at age 56, and the contribution to the partial likelihood from the first death is $\frac{1}{5+3e^\beta}$.

Solution 9.11

Suppose that $\tilde{\beta}$ is the maximum partial likelihood estimator of β . Then asymptotically (*i.e.* as the sample size tends to ∞), $E(\tilde{\beta}) = \beta$.

Solution 9.12

From Question 9.9, we know that:

$$\frac{d^2 \log L}{d\beta^2} = -\frac{2e^\beta}{(e^\beta + 2)^2} - \frac{3e^\beta}{(e^\beta + 3)^2}$$

If $\hat{\beta}$ is the maximum partial likelihood estimator of β , then the asymptotic variance of $\hat{\beta}$ is given by:

$$\begin{aligned} \text{var}(\hat{\beta}) &= \left[I^{-1}(\beta) \right]_{\beta=\hat{\beta}} \\ &= \left[\frac{2e^{\hat{\beta}}}{(e^{\hat{\beta}} + 2)^2} + \frac{3e^{\hat{\beta}}}{(e^{\hat{\beta}} + 3)^2} \right]^{-1} \\ &= \left[\frac{2\sqrt{6}}{(\sqrt{6} + 2)^2} + \frac{3\sqrt{6}}{(\sqrt{6} + 3)^2} \right]^{-1} \\ &= 2.02062 \end{aligned}$$

So the asymptotic standard error is $\sqrt{2.02062} = 1.4215$.

Solution 9.13

As $\tilde{\beta}$ is asymptotically normally distributed, a 95% confidence interval for β is:

$$\hat{\beta} \pm 1.96 \sqrt{\text{var}(\tilde{\beta})} = \frac{1}{2} \ln 6 \pm (1.96 \times 1.4215) = (-1.890, 3.682)$$

Since this interval contains the value 0, we conclude on the basis of these data that sex is not a significant covariate.

Solution 9.14

The covariates for lives A and B are:

$$\begin{array}{lllll} z_{A1} = 3 & z_{A2} = 0 & z_{A3} = 0 & z_{A4} = 0 & z_{A5} = 0 \\ z_{B1} = 1 & z_{B2} = 0 & z_{B3} = 1 & z_{B4} = 0 & z_{B5} = 1 \end{array}$$

So the ratio of the hazards is:

$$\frac{\lambda_A}{\lambda_B} = \frac{\lambda_0(x)e^{3\beta_1}}{\lambda_0(x)e^{\beta_1+\beta_3+\beta_5}} = e^{2\beta_1-\beta_3-\beta_5}$$

Solution 9.15

The null hypothesis is that there is no interaction between socio-economic group and the other factors, ie:

$$H_0 : \beta_4 = \beta_5 = 0$$

We would fit a model with the first 3 covariates (z_{i1}, z_{i2}, z_{i3}) and another model with all 5 covariates (z_{i1}, \dots, z_{i5}). Each is fitted by maximising the partial likelihood, using an appropriate statistics package. Let L_1 and L_2 be the maximised log-likelihoods of the first and second models respectively.

The likelihood ratio statistic is then $-2(L_1 - L_2)$, which has an asymptotic χ^2_2 distribution if the null hypothesis is true. If the likelihood ratio statistic exceeds the upper 5% point of χ^2_2 , then the null hypothesis should be rejected.

Solution 9.16(i) ***Group of lives to which baseline hazard applies***

Lives who are:

- treated immediately following diagnosis, $z_1 = 0$
- who receive the existing treatment, $z_2 = 0$
- who are female, $z_3 = 0$.

(ii)(a) ***Hazard function for male life who received the new treatment six months after diagnosis***

We use the model parameters we are given, together with the values of the regression variables for this life:

$$z_1 = \frac{1}{2} \text{ year}, \quad z_2 = 1 \text{ for the new treatment} \quad z_3 = 1 \text{ for a male life}$$

Then:

$$h(t) = h_0(t) \exp \left\{ 0.5 \times \frac{1}{2} + 0.01 \times 1 - 0.05 \times 1 \right\} = h_0(t) e^{0.21}$$

(ii)(b) ***Survival function for male life who received the new treatment six months after diagnosis***

The survival function is:

$$\begin{aligned} S(t) &= \exp \left\{ - \int_{s=0}^t h(s) ds \right\} \\ &= \exp \left\{ - \int_{s=0}^t h_0(s) e^{0.21} ds \right\} \\ &= \exp \left\{ - e^{0.21} \int_{s=0}^t h_0(s) ds \right\} \end{aligned}$$

(iii) **Probability that the life in (ii) will survive for five years**

We use the information given about the female life to determine an expression for the baseline hazard. We can then use this expression to evaluate the probability for the male life.

For a female life given the new treatment at the time of diagnosis we can write:

$$h_f(t) = h_0(t) \exp\{0.5 \times 0 + 0.01 \times 1 - 0.05 \times 0\} = h_0(t) e^{0.01}$$

Then:

$$\begin{aligned} S_f(t) &= \exp\left\{-\int_{s=0}^5 h(s) ds\right\} = \exp\left\{-\int_{s=0}^5 h_0(s) e^{0.01} ds\right\} \\ &= \exp\left\{-e^{0.01} \int_{s=0}^5 h_0(s) ds\right\} = \left(\exp\left\{-\int_{s=0}^5 h_0(s) ds\right\}\right)^{e^{0.01}} \\ &= 0.75 \end{aligned}$$

Rearranging this result gives:

$$\exp\left\{-\int_{s=0}^5 h_0(s) ds\right\} = (0.75)^{e^{-0.01}}$$

Then using the result from (ii)(b) for $t = 5$ we can write:

$$S_m(t) = \exp\left\{-e^{0.21} \int_{s=0}^5 h_0(s) ds\right\} = \left(\exp\left\{-\int_{s=0}^5 h_0(s) ds\right\}\right)^{e^{0.21}}$$

Finally substitution gives:

$$S_m(t) = \left(\exp\left\{-\int_{s=0}^5 h_0(s) ds\right\}\right)^{e^{0.21}} = \left((0.75)^{e^{-0.01}}\right)^{e^{0.21}} = (0.75)^{e^{0.20}} = 0.7037$$

Solution 9.17

(i) **Fully parametric models versus Cox regression model**

The Cox regression model is an example of a semi-parametric approach, in which we do not pre-constrain the precise form of the hazard function. It has been the most widely used regression model in recent years and is also known as the *proportional hazards* model.

Parametric models can be used with a homogeneous population or can be fitted to a moderate number of homogeneous groups, in which case confidence intervals for the fitted parameters give a test of differences between the groups which should be better than non-parametric procedures.

However, fully parametric models are difficult to apply without foreknowledge of the form of the hazard function, which might be the very object of the study. For this reason a semi-parametric approach can be more popular.

Non-parametric approaches are limited in their ability to deal with some important questions in survival analysis, such as the effect of *covariates* on survival. A covariate is any quantity recorded in respect of each life, such as age, sex, type of treatment, level of medication, severity of symptoms and so on. If the covariates partition the population into a small number of homogeneous groups, it is possible to compare Kaplan-Meier or other non-parametric estimates in respect of each population. But a more direct and transparent method is to construct a model in which the effects of the covariates on survival are modelled directly, *i.e.* a regression model.

(ii) **Censoring present in this study**

Right censoring and Type I censoring are present at the end of the investigation.

Random censoring is present since death from a cause other than heart attack can occur at any time.

(iii) **Criterion – likelihood ratio test**

A common criterion is the *likelihood ratio test*. Suppose we need to assess the effect of adding further covariates to the model. For example, suppose we fit a model with p covariates, and another model with $p+q$ covariates (which include the p covariates of the first model).

Each model is fitted by maximising a likelihood. Let L_p and L_{p+q} be the maximised log-likelihoods of the first and second models respectively.

The null hypothesis for this test is:

$$H_0 : \beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+q} = 0$$

i.e the extra covariates are not significant.

The likelihood ratio statistic is:

$$-2(\ln L_p - \ln L_{p+q})$$

where the log-likelihoods are calculated using the maximum partial likelihood estimates. This has an asymptotic χ^2 distribution, with q degrees of freedom, under the null hypothesis.

The null hypothesis will be rejected at the 5% significance level if the value of the test statistic is greater than the upper 5% point of χ_q^2 .

(iv) ***Estimate of the effect of sex and smoking behaviour on the risk of heart attack***

The final model is:

$$\begin{aligned}\lambda(t; z_i) &= \lambda_0(t) \cdot \exp(\beta z_i^\top) \\ &= \lambda_0(t) \cdot \exp\left(\sum_{j=1}^4 \beta_j z_{ij}\right) \\ &= \lambda_0(t) \cdot \exp(0.01z_{i1} - 0.4z_{i2} + 0.5z_{i3} - 0.25z_{i4})\end{aligned}$$

where $z_{i4} = z_{i3} \times z_{i2}$.

The value of $\beta_2 = -0.4$ will decrease the hazard function for the i th life if the sex is female, $z_{i2} = 1$. This implies that, according to the model, females have a lower risk of heart attack.

The value of $\beta_3 = 0.5$ will increase the hazard function for the i th life if the smoker status is “smoker”, $z_{i3} = 1$. This implies that, according to the model, smokers have a higher risk of heart attack.

The value of $\beta_4 = -0.25$ will decrease the hazard function for the i th life if the life is both female and a smoker, ie if $z_{i4} = 1$. This implies that, according to the model, whilst female smokers have a higher risk of heart attack than female non-smokers, smoking has a much more detrimental effect on males than it does on females.

(v) ***How old a female smoker must be***

A male 50-year old non-smoker has the baseline hazard function:

$$\lambda(t) = \lambda_0(t) \exp(0) = \lambda_0(t)$$

A female smoker has the hazard function:

$$\lambda(t) = \lambda_0(t) \exp(0.01z_{i1} - 0.4 + 0.5 - 0.25) = \lambda_0(t) \exp(0.01z_{i1} - 0.15)$$

For these two hazard functions to be the same, we require:

$$0.01z_{i1} - 0.15 = 0$$

$$ie: z_{i1} = 15$$

So, according to the model, a female smoker must be 65 to have the same risk of heart attack as a male non-smoker.

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Chapter 10

Binomial and Poisson models



Syllabus objectives

- (vii) *Derive maximum likelihood estimators for the transition intensities in models of transfers between states with piecewise constant transition intensities.*
4. *Describe the Poisson approximation to the estimator (for a constant transition intensity in a Markov model) in the case of a single decrement and its advantages and disadvantages.*
- (viii) *Describe the binomial model of mortality, derive a maximum likelihood estimator for the probability of death and compare the binomial model with the multiple-state models.*
1. *Describe the binomial model of the mortality of a group of identical individuals subject to no other decrements between two ages.*
 2. *Derive the maximum likelihood estimator for the rate of mortality in the binomial model and its mean and variance.*
 3. *Describe the advantages and disadvantages of the multiple-state model and the binomial model, including consistency, efficiency, simplicity of the estimators and their distributions, application to practical observations and generality.*

0 Introduction

In this chapter we introduce two more statistical models, which are based on two discrete probability distributions – the binomial and the Poisson models. The distributions can be used to model the number of deaths observed in a mortality investigation.

The aim of the models is to derive estimates of the true underlying values of q_x (using the binomial model) or μ_x (using the Poisson model) from our observations. We will identify the appropriate estimators and their statistical properties.

We end this chapter with a thorough comparison of the advantages and disadvantages of the binomial and Poisson models, along with the two-state model that is discussed in Chapter 4.

1 Binomial-type models

1.1 The binomial model

Much of the motivation for the analysis of mortality data is provided by the following thought experiment:

Observe N identical, independent lives aged x exactly for one year, and record the number d who die. Then d is a sample value of a random variable D .

If we suppose that each life dies with probability q_x and survives with probability $1-q_x$, then D has a binomial distribution with parameters N and q_x .

In other words, the death or survival of each life can be represented by an independent Bernoulli trial with associated probabilities of q_x and $(1-q_x)$ respectively. Recall from Chapter 7 that q_x is referred to as the *initial* rate of mortality.

The number of deaths is a random variable $D \sim \text{Binomial}(N, q_x)$. The observed number of deaths d is a sample value of this random variable. You will also see the Greek letter θ (“theta”), which is commonly used to denote the observed number of deaths.

For the thought experiment outlined above, the probability that exactly d deaths will occur during the year is:

$$P[D = d] = \binom{N}{d} q^d (1-q)^{N-d} \quad d = 0, 1, 2, \dots, N$$

Proof

Since we have assumed that deaths operate independently, the probability that a specified d individuals will die during the year, and the remaining $N-d$ will not, is $q^d (1-q)^{N-d}$.

However, we need to multiply this probability by the combinatorial factor $\binom{N}{d} = \frac{N!}{d!(N-d)!}$, which is the number of ways the d deaths could be “chosen”.

**Example**

10,000 school children have been selected to take part in a one year medical study. If the initial annual rate of mortality is 0.00025 for each child and deaths are expected to occur independently, calculate the probability that 2 or more of the participants will die before the end of the study.

Solution

We have an initial population of $N = 10,000$ individuals with a constant initial rate of mortality $q = 0.00025$ and deaths are assumed to be independent. So, the binomial model applies and the number of deaths D has a $\text{Binomial}(10,000, 0.00025)$ distribution.

Using the formula for the binomial distribution:

$$P(D = 0) = (1 - 0.00025)^{10,000} = 0.0821$$

$$P(D = 1) = 10,000 \times 0.00025(1 - 0.00025)^{9,999} = 0.2052$$

So, the probability of 2 or more deaths is: $1 - 0.0821 - 0.2052 = 0.7127$

**Question 10.1**

Calculate the probability that there will be exactly 3 deaths during the study.

1.2 Estimating q_x from the data

We have seen how to calculate the probability of observing a certain number of deaths when we know the probability of an individual life surviving or dying during the year. In reality we do not know the true value of q_x and we will use the observations from an investigation to estimate this unknown quantity.

One method of deriving this estimate is to use the method of maximum likelihood. This involves selecting the value of q_x that maximises the likelihood of obtaining the observed number of deaths.

The intuitive estimate of q_x is $\hat{q}_x = d/N$, and this is also the maximum likelihood estimate.

This is certainly a very intuitive result. For example, if we were to observe 100 people for one year and 3 of the individuals were to die during the investigation, it seems natural to infer that the probability of death is 3 in 100 (*i.e.* that $q_x = 0.03$).

We will now show that this intuitive result is theoretically correct.

Proof

Under the binomial model, the likelihood of recording exactly d deaths if the rate of mortality is q is:

$$L(q) = \binom{N}{d} q^d (1-q)^{N-d}$$

which can be maximised by maximising its log:

$$\log L(q) = \log \binom{N}{d} + d \log q + (N-d) \log(1-q)$$

Differentiating with respect to q (using the function-of-a-function rule):

$$\frac{\partial}{\partial q} \log L(q) = \frac{d}{q} - \frac{N-d}{1-q}$$

This is zero at the value \hat{q} such that

$$\begin{aligned} d(1-\hat{q}) &= (N-d)\hat{q} \\ \Rightarrow d - d\hat{q} &= N\hat{q} - d\hat{q} \\ \Rightarrow d &= N\hat{q} \\ \Rightarrow \hat{q} &= d/N \end{aligned}$$

This is a *maximum* since $\frac{\partial^2}{\partial q^2} \log L(q) = -\frac{d}{q^2} - \frac{N-d}{(1-q)^2} < 0$.

The maximum likelihood estimate is the observed value of the corresponding maximum likelihood estimator $\tilde{q}_x = D/N$.

The corresponding estimator \tilde{q}_x has:

- **mean = q_x** , ie it is unbiased.
- **variance = $q_x(1-q_x)/N$** , which is in fact the smallest possible variance for an unbiased estimator of q_x . This follows from the Cramér-Rao inequality.

Proof of (i)

The observed number of deaths D has a *Binomial*(N, q_x) distribution, which means that it has mean Nq_x and variance $Nq_x(1-q_x)$. So:

$$E(\tilde{q}_x) = E(D/N) = \frac{Nq_x}{N} = q_x$$



Question 10.2

Prove result (ii):

$$\text{var}(\tilde{q}_x) = q_x(1-q_x)/N$$

Furthermore, it is a property of all maximum likelihood estimators that they are asymptotically normally distributed. In other words,

$$\tilde{q}_x \sim \text{Normal} \left(q_x, \frac{q_x(1-q_x)}{N} \right) \text{ asymptotically}$$

These properties show that \tilde{q}_x is a sensible estimator to use. Its average value equals the true value of q_x and it varies as little as possible from the true value. The normal approximation allows us to calculate approximate probabilities and confidence intervals.

This is the binomial model of mortality.

Remember that the results obtained so far have been derived from a simple, idealised thought experiment. In the next section we look at some of the problems we may encounter in reality.

1.3 Generalisation of the model

Of course, real life is rather more complicated than our idealised thought experiment.

The binomial model leads to problems if the observations are more realistic:

- **we might not observe all lives over the same interval of age**

Whilst we may limit our investigation to lives aged between x and $x+1$, we may not observe all lives for the complete year.

We assume that the i th life in our investigation will be observed between ages $x+a_i$ and $x+b_i$ ($0 \leq a_i < b_i \leq 1$).

- **there will usually be decrements other than death, and sometimes increments as well.**

For example, consider an investigation into the mortality of unemployed men aged 30. Some lives will leave the investigation because they return to employment – the data will be censored. Similarly, lives can rejoin the investigation if they become unemployed again.

In general the values of $\{a_i\}$ and $\{b_i\}$ are not the same for all the lives in the investigation. Considering the i^{th} life, we have:

$$P[D_i = 0] = P[\text{The } i\text{th life survives from age } x+a_i \text{ to } x+b_i] = 1 - {}_{b_i-a_i} q_{x+a_i}$$

$$P[D_i = 1] = P[\text{The } i\text{th life dies before age } x+b_i] = {}_{b_i-a_i} q_{x+a_i}$$

Hence:

$$P[D_i = d_i] = {}_{b_i-a_i} q_{x+a_i}^{d_i} (1 - {}_{b_i-a_i} q_{x+a_i})^{1-d_i} \quad (d_i = 0, 1)$$

This function represents the contribution to the total likelihood of the i th life in a sample of N independent lives. The likelihood is a function of the parameter $b_i-a_i q_{x+a_i}$ (which is related to q_x , the unknown quantity that we are trying to estimate from the data) and the observed statistic d_i (the information that we will use to estimate the unknown rate of mortality).

For each life under observation, we will collect information about the values of d_i , a_i and b_i .

Define the vector quantities:

$$\vec{q} = (b_1 - a_1 q_{x+a_1}, b_2 - a_2 q_{x+a_2}, \dots, b_N - a_N q_{x+a_N})$$

$$\vec{d} = (d_1, d_2, \dots, d_N)$$

\vec{q} is a $1 \times N$ vector of unknown parameters. Whilst we do not know the true values of the mortality rates, we use the $\{a_i\}$ and $\{b_i\}$ to identify what it is that we have observed, eg ${}_1 q_x$ or ${}_{1/2} q_{x+1/2}$.

\vec{d} is a $1 \times N$ vector of 0's (representing survival) and 1's (representing deaths) that is compiled from our observations, eg $(0, 0, 1, 0, \dots, 1, 0)$.

Since each life is assumed to be independent of the others, the overall likelihood is the product of the individual likelihoods. **We can write the overall likelihood as:**

$$L(\vec{q}; \vec{d}) = \prod_{i=1}^N b_i - a_i q_{x+a_i}^{d_i} (1 - b_i - a_i q_{x+a_i})^{1-d_i}.$$

1.4 Maximising the likelihood

We have to find the value of the vector \vec{q} – in general, N numbers – that maximises the likelihood.

If the intervals $\{x + a_i, x + b_i\}$ didn't overlap and the q 's were unrelated numbers, we could maximise the likelihood by setting $b_i - a_i q_{x+a_i} = 1$ wherever $d_i = 1$ and $b_i - a_i q_{x+a_i} = 0$ wherever $d_i = 0$. However, we need consistent estimates of the mortality rates. For example, it would be nonsensical to have ${}_{1/2} q_x > q_x$.

The dimension of the problem might be reduced if some of the $\{a_i\}$ and the $\{b_i\}$ are equal, ie several lives may give information about the mortality rate over the same age range.

But the usual approach is to make an assumption about the distribution of T_x in the age range $[x, x+1]$ which allows us to express any $b_i - a_i q_{x+a_i}$ in terms of q_x , making the likelihood a function of one parameter again.

Possible assumptions are:

(a) **uniform distribution of deaths:** $tq_x = t \cdot q_x$ $(0 \leq t \leq 1)$

(b) **the Balducci assumption:** $_{1-t}q_{x+t} = (1-t)q_x$ $(0 \leq t \leq 1)$

(c) **constant force of mortality:** $tq_x = 1 - e^{-\mu t}$ $(0 \leq t \leq 1)$



Question 10.3

Using the Balducci assumption, derive an expression for $_{t}q_x$ in terms of q_x .

By definition, assumption (c) implies a constant force of mortality over the year of age. The probability of a life dying in the last month of the year of age is therefore lower than the probability of death in the first month of the year of age. This is because the probability of a life dying in the last month of the year of age includes the probability of the life having survived up to that point, which is less than 1.

By contrast, assumption (a) implies a uniform distribution of deaths over the year of age. This in turn implies an *increasing* force of mortality to cancel the effect of the survival probability $_{t}p_x$ (less than 1) and maintain a constant number of deaths. (Another way to think about this is to consider a group of lives who die at a uniform rate over a given year. To maintain a constant number of deaths, the force of mortality must increase to counter the fact that the number of survivors is decreasing.)

Note that the Balducci assumption implies a decreasing force of mortality between integer ages.



Question 10.4

Show that the Balducci assumption implies a decreasing force of mortality between integer ages.



Question 10.5

Calculate (to 7 decimal places) the value of $_{1/2}q_x$ under each of the three assumptions if $q_x = 0.002$. Comment on the results.

2 The actuarial estimate

2.1 Finding a simple estimate for q_x

We would like to find a simple relationship between D , the random variable representing the number of deaths, and q_x , the underlying mortality rate. Once we know the actual waiting time and the number of observed deaths, we can then use this information to derive an estimate of q_x .

Define:

$$D_i = \begin{cases} 0 & \text{if the } i\text{th life survives} \\ 1 & \text{if the } i\text{th life dies} \end{cases}$$

Then:

$$\begin{aligned} E(D) &= \sum_{i=1}^N E(D_i) \\ &= \sum_{i=1}^N (0 \times P(i^{\text{th}} \text{ life survives}) + 1 \times P(i^{\text{th}} \text{ life dies})) \\ &= \sum_{i=1}^N b_i - a_i q_{x+a_i} \end{aligned}$$

To simplify this expression in terms of q_x , we need two intermediate steps.

First, we can see that:

$$\begin{aligned} P[\text{i}^{\text{th}} \text{ life dies between ages } x+a_i \text{ and } x+b_i] \\ = P[\text{i}^{\text{th}} \text{ life dies between } x+a_i \text{ and } x+1] \\ - P[\text{i}^{\text{th}} \text{ life dies between } x+b_i \text{ and } x+1] \end{aligned}$$

ie:

$$b_i - a_i q_{x+a_i} = 1 - a_i q_{x+a_i} - b_i - a_i p_{x+a_i} \times 1 - b_i q_{x+b_i}$$

Secondly, we can express ${}_{1-a_i}q_{x+a_i}$ and ${}_{1-b_i}q_{x+b_i}$ in terms of q_x using the Balducci assumption:

$${}_{1-t}q_{x+t} = (1-t)q_x \quad (0 \leq t \leq 1)$$

So, under the Balducci assumption:

$$\begin{aligned} E[D] &= \sum_{i=1}^N b_i - a_i q_{x+a_i} \\ &= \sum_{i=1}^N {}_{1-a_i}q_{x+a_i} - \sum_{i=1}^N b_i - a_i p_{x+a_i} \times {}_{1-b_i}q_{x+b_i} \\ &= \sum_{i=1}^N (1 - a_i)q_x - \sum_{i=1}^N (1 - E[D_i])(1 - b_i)q_x \end{aligned}$$

For simplicity we are assuming that the $\{a_i\}$ and $\{b_i\}$ are known, and that death is the only decrement.

Substituting the observed number of deaths d on the left side would usually give the moment estimate of q_x . However, the right side also involves expected deaths in such a way that it is impossible to extract all the terms in $E[D]$ and the $\{E[D_i]\}$ on one side and all the terms in q_x on the other.

Summing the last term over the observed rather than the expected survivors (for whom $d_i = 0$), we obtain:

$$E[D] \approx \sum_{i=1}^N (1 - a_i)q_x - \sum_{i=1}^N (1 - d_i)(1 - b_i)q_x$$

leading to the estimate:

$$\hat{q}_x = \frac{d}{\sum_{i=1}^N (1 - a_i) - \sum_{i:D_i=0} (1 - b_i)}$$

in which the denominator is called the *initial exposed to risk*, counting the deaths as exposed to risk until the end of the year of age.

**Question 10.6**

Can you follow the algebra on this page?

**Initial exposed to risk**

The initial exposed to risk, E_x , is defined as:

$$E_x = \sum_{i=1}^N (1-a_i) - \sum_{i:D_i=0} (1-b_i) = \sum_{\text{deaths}} (1-a_i) + \sum_{\text{censored lives}} (b_i - a_i)$$

The deaths contribute the period of length $(1-a_i)$ from $x+a_i$ to $x+1$.

The survivors contribute the period of length $(b_i - a_i)$ from $x+a_i$ to $x+b_i$.

**Question 10.7**

Can you see the logic for counting the deaths as exposed to risk until the end of the year of age?

2.2 Central exposed to risk

We have already met the central exposed to risk in Chapter 4, but in that chapter we called it the total waiting time.

**Central exposed to risk**

$E_x^c = v$ is also called the observed total waiting time at age x .

Under the crude assumption that deaths occur, on average, at age $x + \frac{1}{2}$, and ignoring the awkward possibility that $a_i > \frac{1}{2}$, we obtain the formula:

$$E_x = E_x^c + \frac{1}{2} d_x$$

Lives no longer contribute to E_x^c once they have died. By contrast, E_x includes the full potential time under observation for all deaths. Under the assumption that deaths occur, on average, at age $x + \frac{1}{2}$, the difference between E_x^c and E_x is therefore half a year for each life that has died, ie $d/2$. With this assumption, we obtain the following formula.



Actuarial estimate

$$\hat{q}_x = \frac{d}{E_x^c + d/2}$$

This is known to statisticians as the **actuarial estimate**.

Note that it is only an approximate method of moments estimate of q_x . It is based on the observed number of deaths and the observed total waiting time rather than their expected values. It is also only as reliable as the assumption we have made, namely that deaths occur, on average, at age $x + \frac{1}{2}$.



Important note on the actuarial estimate

When the exact times of all events (start of observation, censoring and death) are available for each life, the actuarial estimate is usually calculated without using the assumption about the distribution of the dates of death, ie it is calculated using the formula:

$$\hat{q}_x = \frac{d}{E_x}$$

where:

$$E_x = \sum_{deaths} (1 - a_i) + \sum_{censored lives} (b_i - a_i)$$

If the exact dates are not available, we usually use approximate methods to calculate the central exposed to risk E_x^c , and then approximate the initial exposed to risk E_x by $E_x^c + d/2$.

Approximate methods for calculating E_x^c are discussed in Chapter 11.

2.3 Strengths and weakness of the binomial model

The binomial model, and the actuarial estimate, are not without strengths. The actuarial estimate avoids numerical solution of equations and it might be used if there is a compelling reason to estimate q_x instead of something else.

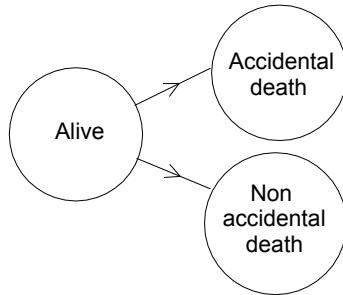
However, it cannot be said that the actuarial estimate is any simpler than the estimates based on multiple state models. Indeed, if the exposure data are of the census type (see Chapter 11), the need to compute an initial exposed to risk is a pointless complication.

As we'll see in Chapter 11 *Exposed to risk*, the data may enable us to calculate the central exposed to risk E_x^c quite easily. Calculating E_x from E_x^c would represent additional work without any real gain.

Historically, the data available to life companies were in a form that made the initial exposed to risk simpler to calculate than the central exposed to risk. With modern computer systems this is no longer the case.

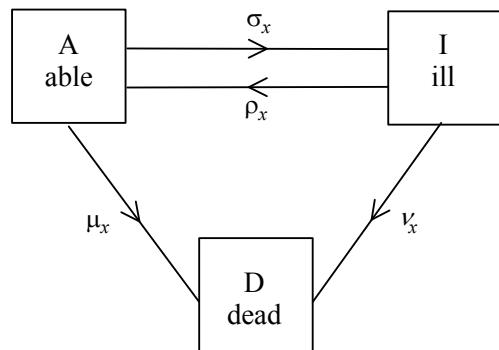
Crucially, the binomial model is not so easily generalised to settings with more than one decrement.

Here is an example of a multiple decrement model:



Even the simplest case of two decrements gives rise to difficult problems; the introduction of repeated transitions such as sickness and recovery is more difficult still.

For example:



An illness-death model

Extension of models in these directions is much simpler within the multiple-state framework.

We shall compare the relative strengths of the various models in detail at the end of this chapter.

3 Poisson models

3.1 The Poisson distribution

The Poisson distribution is used to model the number of ‘rare’ events occurring during some period of time, for example the number of particles emitted by a radioactive source in a minute. Such analogies suggest the Poisson distribution as a model for the number of deaths among a group of lives, given the time spent exposed-to-risk.

The Poisson distribution is a discrete probability distribution in which the random variable can only take non-negative integer values.

A random variable X is said to have a Poisson distribution with mean λ ($\lambda > 0$) if the probability function of X is:

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \dots$$

Remember that $E(X) = \lambda$ and $\text{var}(X) = \lambda$.

We can use the Poisson distribution to model the number of deaths observed in an investigation. We can then use our knowledge of the exposed-to-risk to derive information about the underlying force of mortality.

3.2 The Poisson model

In this section we will let E_x^c denote the total observed waiting time (or central exposed to risk). In terms of our previous notation $E_x^c = v$, the realised value of the total random waiting time V .

If we assume that we observe N individuals as before, and that the force of mortality is a constant μ , then a Poisson model is given by the assumption that

D has a Poisson distribution with parameter μE_x^c . That is:

$$P[D = d] = \frac{e^{-\mu E_x^c} (\mu E_x^c)^d}{d!}$$



Example

A small country involved in a war conscripted a cohort of healthy young men to serve in the country's army for a 3-year period starting on 1 January 1999. During this period a number of the men were killed. Given that the total period of service for the group as a whole was 10 million man-days and that the annual force of mortality for death in active service is 0.02, use a normal approximation to calculate the probability that at least 500 men were killed while in active service.

Solution

We are given that $\mu = 0.02$ and $E^c = 10,000,000/365 = 27,397$ years. So the number of deaths has a Poisson distribution with mean $\lambda = 0.02 \times 27,397 = 547.95$ (approximately).

We can approximate this distribution using a normal distribution with the same mean and variance, which will both be 547.95. So, applying a continuity correction, we find that the probability of at least 500 deaths is approximately:

$$\begin{aligned} P(\geq 500 \text{ deaths}) &= P[\text{Normal}(547.95, 547.95) > 499.5] \\ &= P\left[Z > (499.5 - 547.95)/\sqrt{547.95}\right] \\ &= 1 - \Phi(-2.070) = 0.981 \end{aligned}$$

So there is a 98.1% probability that at least 500 men were killed.

Under the observational plan described above, the Poisson model is not an exact model, since (amongst other reasons) it allows a non-zero probability of more than N deaths. But it is often a good approximation, since the probability of more than N deaths is usually negligible.

Alternatively, we might adjust the observational plan so that the Poisson model is exact. Examples of suitable (but not necessarily practicable) observational plans are:

- **to continue observation until the waiting time reaches a pre-determined value**

This will ensure that we've had exactly v person-years of exposure.

- **to replace each life who dies with an identical and independent life at the moment of death.**

This will ensure that the population size is kept constant throughout, again allowing us to determine the total exposure exactly.

Under this observational plan the total observed waiting time E_x^c is simply the constant population size multiplied by the period of the investigation.



Example

A large computer company always maintains a workforce of exactly 5,000 young workers, immediately replacing any worker who leaves. Calculate the probability that there will be fewer than 3 deaths during any 6 month period, assuming that all workers experience a constant force of mortality of 0.0008 per annum.

Solution

We have a constant population of $P = 5000$ individuals with a constant force of mortality $\mu = 0.0008$. If we assume that deaths are independent, the Poisson model applies and the number of deaths during any 6 month period has a Poisson distribution with mean $0.0008 \times 5,000 \times 6/12 = 2$.

$$\text{So: } P(\text{No deaths}) = e^{-2} = 0.1353$$

$$P(\text{Exactly 1 death}) = 2e^{-2} = 0.2707$$

$$P(\text{Exactly 2 deaths}) = \frac{2^2}{2!} e^{-2} = 0.2707$$

and the probability of fewer than 3 deaths is: $0.1353 + 0.2707 + 0.2707 = 0.6767$.


Question 10.8

The population of a small town is expected to remain constant at 50,000 over the next few years. Assuming that all inhabitants experience a force of mortality of 0.001 per annum, use a normal approximation to estimate the probability that there will be more than 225 deaths in the town during the next 4 years.

3.3 ***Estimating the underlying force of mortality***

We would like to use our knowledge about the number of deaths observed and the total exposed-to-risk (waiting time) to estimate the unknown true force of mortality.

The Poisson likelihood leads to the following estimator of (constant) μ :

$$\tilde{\mu} = \frac{D}{E_x^c}$$

Proof

The likelihood of observing d deaths if the true value of the hazard rate is μ is:

$$L(\mu) = \frac{(\mu E_x^c)^d e^{-\mu E_x^c}}{d!}$$

which can be maximised by maximising its log:

$$\log L(\mu) = d(\log \mu + \log E_x^c) - \mu E_x^c - \log d!$$

Differentiating with respect to μ :

$$\frac{\partial}{\partial \mu} \log L(\mu) = \frac{d}{\mu} - E_x^c$$

which is zero when:

$$\hat{\mu} = \frac{d}{E_x^c}$$

This is a *maximum* since $\frac{\partial^2}{\partial \mu^2} \log L(\mu) = -\frac{d}{\mu^2} < 0$.

$\hat{\mu}$ is the maximum likelihood estimate of μ . It is the realised value of the maximum likelihood estimator $\tilde{\mu} = \frac{D}{E_x^c}$.



Question 10.9

In a mortality investigation covering a 5-year period, where the force of mortality can be assumed to be constant, there were 46 deaths and the population remained approximately constant at 7,500. Estimate the force of mortality.

The estimator $\tilde{\mu}$ has the following properties:

- (i) $E[\tilde{\mu}] = \mu$, ie it is unbiased
- (ii) $\text{var}[\tilde{\mu}] = \frac{\mu}{E_x^c}$, which is in fact the smallest possible variance for an unbiased estimator of μ .

In practice, we will substitute $\hat{\mu}$ for μ to estimate these from the data.

Proof of (i)

The observed number of deaths D has a Poisson (μE_x^c) distribution, which means that it has mean μE_x^c and variance μE_x^c .

So:

$$E(\tilde{\mu}) = E\left(\frac{D}{E_x^c}\right) = \frac{\mu E_x^c}{E_x^c} = \mu$$



Question 10.10

Prove that $\text{var}(\tilde{\mu}) = \frac{\mu}{E_x^c}$.

As for the binomial model, it is a property of all maximum likelihood estimators that they are asymptotically normally distributed. In other words,

$$\tilde{\mu} \sim \text{Normal} \left(\mu, \frac{\mu}{E_x^c} \right) \text{ asymptotically}$$

These properties show that $\tilde{\mu}_x$ is a sensible estimator to use. Its mean value equals the true value of μ and it varies as little as possible from the true value. The normal approximation allows us to calculate approximate probabilities and confidence intervals.



Question 10.11

Find a 95% confidence interval for the force of mortality in Question 10.9.

3.4 Links to the two-state Markov model

Under the two-state model, $E[\tilde{\mu}] = \mu$ and $\text{Var}[\tilde{\mu}] = \mu/E[V]$, but the true values of μ and $E[V]$ are unknown and must be estimated from the data as $\hat{\mu}$ and E_x^c respectively.

So although the estimators are different, we obtain the same numerical estimates of the parameter and of the moments of the estimator, in either case.

4 Comparison of multiple-state, binomial and Poisson models

We conclude this chapter with a comparison of the three models: multiple-state, binomial and Poisson. This section provides a useful overview of the course so far.

When we compare models, we distinguish three aspects:

- (a) how well each model represents the process we are trying to model;
- (b) how easy it is to find, characterise and use the model parameters, given the data with which we must work; and
- (c) how easily each model is extended to problems other than the study of human mortality.

4.1 Modelling the underlying process

The underlying process we take to be the time(s) of death of one or more lives, considered to be indistinguishable, except in respect of their deaths. If death is the only decrement, this leads to the two specifications:

- (a) representing the time of death by the random variable T_x ; or
- (b) the two-state model parameterised by μ_{x+t} .

They can usually be taken to be equivalent, since under reasonable conditions we can derive the force of mortality starting with (a), while we can obtain the distribution of the time to death starting with (b).

In Chapter 7 we defined μ_x in terms of T_x . We then obtained $f_x(t)$, the probability density function of lifetimes, in terms of μ_x .

It is evident that the two-state model represents the process closely (in fact, almost by definition), while the binomial model represents a restricted view of the process, since it represents only the year of death, and not the time of death.

The two-state model considers the moment at which the transition took place. The binomial model simply categorises each life as *dead* or *alive* at the end of the period of investigation.

This suggests that if sufficient data are available to use the two-state model, (ie full information about the total waiting time) then using the binomial model instead will not make the fullest use of the information. This turns out to be the case.

Unless the waiting times E_x^c are fixed in advance, which would be unusual in actuarial work, the Poisson model is an approximation to the multiple-state model, in which E_x^c is regarded as non-random. This is acceptable if μ is small.

As we saw in Section 3 of this chapter, we use the *observed* waiting time E_x^c as an estimate of $E[V]$. If μ is small, E_x^c should be close to $E[V]$ and the variance of our estimator $\tilde{\mu}$ should be low.

Note that both formulations (a) and (b) above are non-parametric in the usual sense, although both can be regarded as parameterised by a function: μ_{x+t} in (b) and $F_x(t)$ or $f_x(t)$ in (a). Direct estimation of these functions (integrated in the case of μ_{x+t}) leads to the Kaplan-Meier or Nelson-Aalen estimates.

We looked at Kaplan-Meier and Nelson-Aalen estimates in Chapter 8.

4.2 Estimating the model parameters

Parametric models are obtained if we restrict attention to single years of age, and in the two-state model also assume a constant transition intensity.

The form and statistical properties of the parameter estimates (and how easy they are to find and to use) depend on the form of the likelihoods, which in turn depend on the available data.

If the exact dates of birth, entry to and exit from observation, and death (if observed) are all known, then:

- (a) we can calculate exactly the MLE of μ in the two-state model (since we can calculate the total waiting time exactly from the data);
- (b) the binomial model (or more accurately, Bernoulli model) based on individual lives is complicated, and further assumptions (such as the Balducci assumption) are needed to get results. We looked at the simplifying assumptions in Section 1.4 of this chapter.

The consequence of (b) above is that the binomial estimate of q_x has a higher variance than the estimate $\hat{q}_x = 1 - \exp(-\hat{\mu})$ obtained from the two-state model.

However, the difference is tiny unless μ is extremely high. (A rule of thumb, due to Sverdrup, is that if μ is very small, most of the information is in the number of deaths, while if μ is very large, most of the information is in the times of death.)

Likewise, when μ is very small, the actuarial estimate $\hat{q}_x = d_x / (E_x^c + d_x / 2)$ provides acceptable results.

For example, we may study the mortality of 1,000 worms over a one-year period.

- If only 4 worms died during the investigation, we would lose little information from not knowing the exact time of their deaths. The shortest possible total waiting time is 996 years – this assumes that the 4 unlucky worms died shortly after the start of the investigation. The longest possible total waiting time is 1,000 years – this assumes that the 4 unlucky worms died shortly before the end of the investigation. Whether the estimate of total waiting time is taken as 996 or 1,000 or somewhere between these figures, our estimate of the mortality rate will not vary greatly.
- If, on the other hand, all the worms died during the investigation, the actual times of their deaths is a critically important piece of information if we wish to estimate the underlying mortality rate. Our estimate would be significantly higher if all the worms had died during the first 3 months, say, than if their deaths had been spread evenly over the year.

Often, not all the dates of the relevant events are known, and then the MLE of μ in the two-state model must also be approximated (by approximating the total waiting time from the data available). **In at least one important case (the Continuous Mortality Investigation studies) the data allow easy approximation of the waiting times E_x^c (see Chapter 11).**

In terms of computation, therefore, the two-state model is preferred if complete life histories are available; otherwise both the two-state and binomial models require some degree of numerical approximation at the estimation stage.

There is no difference in practice between the two-state and Poisson models, because the maximum likelihood estimates are the same (though the estimators are not), as we saw in Section 3.4 of this chapter.

4.3 Statistical properties of the maximum likelihood estimates

The statistical properties of the MLEs in the various models differ slightly.



Question 10.12

What is meant by the terms “unbiased” and “consistent” when describing the properties of a maximum likelihood estimator?

- (a) In the multiple-state model, the MLE is consistent and asymptotically unbiased; the variance of the estimator is also only available asymptotically. Simulation experiments suggest that the results are reasonable if $d_x \geq 10$.

We considered the MLEs for the multiple-state model at the end of Chapter 5.



Question 10.13

What is the general form of the maximum likelihood estimators for the multiple-state model? What is the mean and variance of this estimator?

- (b) In the Poisson model, the MLE is consistent and unbiased. Its mean and variance are available exactly in terms of the true μ , but are estimated from the data by the same expressions as estimate the asymptotic mean and variance in the two-state model.

We looked at the MLE for the Poisson model in Section 3.3.

- (c) In the “naive” binomial model, in which N identical lives are observed for exactly one year, the MLE is consistent and unbiased, and the exact mean and variance can be obtained in terms of the true q_x . In practice, the data rarely conform to the “naive” model, so only approximate results are available.

We looked at the MLE for the binomial model in Section 1.2.

When μ is very small, there are few reasons to prefer any one of these models on the basis of the statistical properties of the MLEs alone.

4.4 Extending the models

Finally we consider the generality of the models, chiefly how easily they can be extended to more complicated processes than one decrement, and how effective they are when forces of transition are high compared with typical human mortality.

One of the most obvious extensions is the three-state health, sickness, death model, which we introduced in Chapter 5. The transition rates from healthy to sick and from sick to healthy will be much higher than the rates from healthy to dead and sick to dead – we are likely to be ill several times in our lives but we only die once.

- (a) The Markov multiple-state model is extended very simply as we have seen. No matter how complex the model, the estimators have the same simple form and statistical properties, depend only on data that will often be available exactly or approximately, and the apparatus needed in applications (such as the Kolmogorov equations) carries over without difficulty. Further extensions are possible, which complicate the calculation of probabilities but not the estimation of parameters, for example semi-Markov models.

(For the more curious amongst you, a semi-Markov model is a model in which transition probabilities depend on the past history in the current state (*eg* the length of time spent in the sick state), but not on anything occurring prior to that. Semi-Markov models are outside the scope of Subject CT4.)

- (b) The Poisson model extends just as easily to multiple decrements, but not to processes with increments.

If the population under consideration is the workforce of a company, we could use the Poisson model to count the number of decrements *eg* deaths, withdrawals, normal retirements and ill-health retirements. However, it could not deal with new recruits.

- (c) There are considerable difficulties in extending the binomial model even to multiple decrements. After all, a binomial model is based on a series of independent Bernoulli trials, which have two possible outcomes. It is relatively simple to extend the ordinary life table to multiple decrement tables, and these have long been used by actuaries. However, extending the life table (essentially a computational tool) is very far from extending any underlying probabilistic model, and, when the matter was investigated, it was found not to be a simple task (we omit details). Extension of the life table to increments is also not too hard, but extension of the binomial model is harder still.

If transition intensities are high, the loss of information (times of transitions) under the binomial model becomes more serious, while the Poisson model becomes a poorer approximation to the multiple-state model (because there is more randomness in the waiting times).

This is simply an extension of Sverdrup's rule of thumb for hazard rates that may be much higher than a force of human mortality.

In conclusion, when studying ordinary human mortality, transition intensities are so low that none of the models considered stands out on statistical grounds alone. This is why actuaries have used life tables so successfully for so long.

However, when we must model more complicated processes or higher transition intensities, which is increasingly the case as new insurance products are developed, the multiple state approach appears to offer significant advantages.

It may still be the case that a simplified approach is ultimately adopted, for example for calculations to be made by office staff, but it is best to begin with a specification which most nearly represents the process being modelled, and then make approximations as required for estimation and in applications.

5 Exam-style question

You should now be able to attempt the following past exam question.



Question 10.14

Subject CT4, April 2005, Question B7

An investigation took place into the mortality of pensioners. The investigation began on 1 January 2003 and ended on 1 January 2004. The table below gives the data collected in this investigation for 8 lives.

<i>Date of birth</i>	<i>Date of entry into observation</i>	<i>Date of exit from observation</i>	<i>Whether or not exit was due to death (1) or other reason (0)</i>
1 April 1932	1 January 2003	1 January 2004	0
1 October 1932	1 January 2003	1 January 2004	0
1 November 1932	1 March 2003	1 September 2003	1
1 January 1933	1 March 2003	1 June 2003	1
1 January 1933	1 June 2003	1 September 2003	0
1 March 1933	1 September 2003	1 January 2004	0
1 June 1933	1 January 2003	1 January 2004	0
1 October 1933	1 June 2003	1 January 2004	0

The force of mortality, μ_{70} , between exact ages 70 and 71 is assumed to be constant.

- (i) (a) Estimate the constant force of mortality, μ_{70} , using a two-state model and the data for the 8 lives in the table.
 - (b) Hence or otherwise estimate q_{70} . [7]
 - (ii) Show that the maximum likelihood estimate of the constant force, μ_{70} , using a Poisson model of mortality is the same as the estimate using the two-state model. [5]
 - (iii) Outline the differences between the two-state model and the Poisson model when used to estimate transition rates. [3]
- [Total 15]



Chapter 10 Summary

Binomial model

Under the binomial model of mortality, we consider each life as an independent Bernoulli trial with probability q_x of dying during the year of age. We record our observations only as deaths and survivals and lose information about the actual timing of the deaths.

If D denotes the number of deaths out of the original N lives, then:

$$D \sim \text{Binomial}(N, q_x)$$

and:

$$P[D = d] = \binom{N}{d} q^d (1-q)^{N-d}$$

The maximum likelihood estimator of q_x is:

$$\tilde{q}_x = D / N$$

This is asymptotically normally distributed with mean and variance:

$$E(\tilde{q}_x) = q_x \quad \text{var}(\tilde{q}_x) = q_x(1-q_x)/N$$

Actuarial estimate

In order to estimate the unknown quantity q_x , we may need to make a simplifying assumption about the probability of death over periods of less than one year. Use of the Balducci assumption leads us to the actuarial estimate:

$$\hat{q}_x = \frac{d}{E_x^c + \frac{1}{2}d}$$

Poisson model

Under the Poisson model, we assume that the force of mortality is constant between integer ages and the number of deaths has a Poisson distribution with mean μE_x^c , ie $D \sim \text{Poisson}(\mu E_x^c)$.

Then:

$$P[D = d] = \frac{e^{-\mu E_x^c} (\mu E_x^c)^d}{d!}$$

The maximum likelihood estimator of μ is:

$$\tilde{\mu} = \frac{D}{E_x^c}$$

This is asymptotically normally distributed with mean and variance:

$$E(\tilde{\mu}) = \mu \quad \text{var}(\tilde{\mu}) = \frac{\mu}{E_x^c}$$

The model is an approximation to the multiple-state model and provides the same numerical estimates of μ .

Comparison of the models

The multiple-state, binomial and Poisson models can be compared against the following criteria:

- appropriate representation of the underlying process
- properties of the model parameters
- extension of the model to other problems, eg sickness.

The maximum likelihood estimators of the underlying rate or force of mortality under each model are consistent and unbiased.

All models perform well when transition intensities are low, eg human mortality. The multiple-state model is better suited to problems with multiple decrements or higher transition intensities, eg sickness.

Chapter 10 Solutions

Solution 10.1

The probability of exactly 3 deaths is:

$$\begin{aligned}
 P(D = 3) &= \binom{10,000}{3} (0.00025)^3 (1 - 0.00025)^{10,000-3} \\
 &= \frac{10,000 \times 9,999 \times 9,998}{3 \times 2 \times 1} \times (0.00025)^3 (1 - 0.00025)^{9,997} = 0.2138
 \end{aligned}$$

Solution 10.2

$$\text{var}(\tilde{q}) = \text{var}(D/N) = \frac{Nq_x(1-q_x)}{N^2} = \frac{q_x(1-q_x)}{N}$$

Remember that constants “square up” when you take them outside a variance.

Solution 10.3

We can use the Principle of Consistency over the intervals $(x, x+t)$ and $(x+t, x+1)$ to write:

$${}_t p_x \times {}_{1-t} p_{x+t} = p_x \Rightarrow {}_t p_x = \frac{p_x}{{}_{1-t} p_{x+t}}$$

Converting the p 's to q 's, we get:

$$1 - {}_t q_x = \frac{1 - q_x}{1 - {}_{1-t} q_{x+t}}$$

Using Balducci's assumption in the denominator gives:

$$1 - {}_t q_x = \frac{1 - q_x}{1 - (1-t) q_x}$$

Rearranging gives:

$${}_t q_x = 1 - \frac{1 - q_x}{1 - (1-t) q_x} = \frac{1 - (1-t) q_x - (1 - q_x)}{1 - (1-t) q_x} = \frac{t q_x}{1 - (1-t) q_x}$$

Solution 10.4

From Chapter 4, we know that:

$$\mu_{x+t} = -\frac{\frac{\partial}{\partial t} {}_t p_x}{{}_t p_x} = -\frac{\partial}{\partial t} \ln {}_t p_x$$

Suppose that x is an integer and $0 \leq t \leq 1$. Then:

$${}_t p_x = \frac{p_x}{1-t} = \frac{p_x}{1 - {}_{1-t} q_{x+t}}$$

and under the Balducci assumption:

$${}_t p_x = \frac{p_x}{1 - (1-t)q_x} = \frac{p_x}{p_x + tq_x}$$

Taking logs gives:

$$\ln {}_t p_x = \ln p_x - \ln(p_x + tq_x)$$

and differentiating with respect to t :

$$\frac{\partial}{\partial t} \ln {}_t p_x = -\frac{q_x}{p_x + tq_x}$$

So:

$$\mu_{x+t} = -\frac{\partial}{\partial t} \ln {}_t p_x = \frac{q_x}{p_x + tq_x}$$

which is a decreasing function of t since p_x and q_x are positive and t appears in the denominator.

Solution 10.5

$$(a) \quad {}_{1/2}q_x = {}_{1/2}q_x = 0.0010000$$

$$(b) \quad \text{Using Solution 10.3, } {}_{1/2}q_x = \frac{{}_{1/2}q_x}{1 - (1 - {}_{1/2}q_x)} = 0.0010010$$

$$(c) \quad q_x = 0.002 \Rightarrow \mu = -\log_e(0.998) \Rightarrow {}_{1/2}q_x = 1 - e^{-\sqrt{2}\mu} = 0.0010005$$

As expected, we find that (a) < (c) < (b).

Solution 10.6

On Page 11 we have the equation:

$$E[D] = \sum_{i=1}^N (1 - a_i)q_x - \sum_{i=1}^N (1 - E[D_i])(1 - b_i)q_x$$

This shows that if we use the Balducci assumption, then $E[D]$, the expected number of deaths in the population, can be expressed as a sum over the entire population, and this expression involves the true mortality rate q_x . What we want to be able to do is turn this round so that we can estimate the value of q_x from the information we have about the deaths in the population.

The approach used is based on the method of moments. We replace $E[D]$, the expected number of deaths in the population, with d , the observed number. This gives:

$$d = \sum_{i=1}^N (1 - a_i)q_x - \sum_{i=1}^N (1 - E[D_i])(1 - b_i)q_x$$

The problem now lies with the $E[D_i]$'s. These are the probabilities of death (during the year) for each individual. Again, using a method of moments approach, we replace these by the observed values d_i , which will be 0 or 1 in each case. This gives:

$$d = \sum_{i=1}^N (1 - a_i)q_x - \sum_{i=1}^N (1 - d_i)(1 - b_i)q_x$$

We can now split this second sum into two separate sums, one for people who died during the year and one for the survivors:

$$d = \sum_{i=1}^N (1-a_i)q_x - \sum_{\text{Deaths}} (1-d_i)(1-b_i)q_x - \sum_{\text{Survivors}} (1-d_i)(1-b_i)q_x$$

But we know that, for the deaths $d_i = 1$, and for the survivors $d_i = 0$. So we get:

$$\begin{aligned} d &= \sum_{i=1}^N (1-a_i)q_x - \sum_{\text{Deaths}} (1-1)(1-b_i)q_x - \sum_{\text{Survivors}} (1-0)(1-b_i)q_x \\ &= \sum_{i=1}^N (1-a_i)q_x - 0 - \sum_{\text{Survivors}} (1-b_i)q_x \end{aligned}$$

Taking q_x out as a factor, we get:

$$d = q_x \left\{ \sum_{i=1}^N (1-a_i) - \sum_{\text{Survivors}} (1-b_i) \right\}$$

and we can rearrange this to get the estimate of q_x that we're after:

$$\hat{q}_x = \frac{d}{\sum_{i=1}^N (1-a_i) - \sum_{\text{Survivors}} (1-b_i)}$$

Solution 10.7

Under the binomial model we are trying to estimate q_x , the probability of death over a year of age. We count the deaths as exposed to risk until the end of the year of age so that the denominator is consistent with the quantity we are trying to estimate.

For example, if we observed just one life, which died halfway through the year, we would correctly conclude that the probability of death over the entire year was $1/1 = 1$, not $1/0.5 = 2$.

Of course, the algebra in this section justifies mathematically why the initial exposed to risk should be calculated in this way.

Solution 10.8

Since the population is stable, the total waiting time E_x^c will be $50,000 \times 4 = 200,000$. The number of deaths will have a Poisson distribution with mean $\mu E_x^c = 0.001 \times 200,000 = 200$. Since the variance of the Poisson distribution is the same as the mean, the number of deaths will have a $N(200, 200)$ distribution approximately.

Using a continuity correction, the probability of more than 225 deaths is:

$$\begin{aligned} P(\theta > 225) &\approx P[N(200, 200) > 225.5] \\ &= P[N(0,1) > \frac{225.5 - 200}{\sqrt{200}}] = P[N(0,1) > 1.803] = 0.036 \end{aligned}$$

i.e. approximately 3.6%.

Solution 10.9

The MLE of the force of mortality is $\hat{\mu} = \frac{d}{E_x^c} = \frac{46}{7,500 \times 5} = 0.00123$

Solution 10.10

$$\text{var}(\tilde{\mu}) = \text{var}(D / E_x^c) = \frac{\mu E_x^c}{(E_x^c)^2} = \frac{\mu}{E_x^c}$$

Solution 10.11

Our estimate has an approximate normal distribution: $\tilde{\mu} \sim N(\mu, \frac{\mu}{E_x^c})$

A 95% confidence interval can be obtained by taking a range covering 1.96 standard deviations either side of the mean. Since we don't know the true value of μ , we must estimate it using $\hat{\mu}$. This leads to the approximate confidence interval:

$$\mu = \hat{\mu} \pm 1.96 \sqrt{\frac{\hat{\mu}}{E_x^c}}$$

$$ie \quad \mu = 0.001227 \pm 1.96 \sqrt{\frac{0.001227}{7,500 \times 5}} = 0.001227 \pm 0.000354$$

So, an approximate 95% confidence interval is: $0.00087 < \mu < 0.00158$.

Solution 10.12

A maximum likelihood estimator $\tilde{\mu}$ is *unbiased* if $E[\tilde{\mu}] = \mu$, whatever the true value of μ is.

An estimator $\tilde{\mu}$ is consistent if the MLE converges (in probability) to the unknown parameter μ as the sample size $n \rightarrow \infty$.

Solution 10.13

If transition from state I to state II has transition intensity μ (which wouldn't necessarily represent death), then the maximum likelihood estimator is $\tilde{\mu} = D/V$, where D is the number of transitions from state I to state II and V is the total waiting time in state I.

$\tilde{\mu}$ has an asymptotic normal distribution with mean μ and variance $\frac{\mu}{E(V)}$.

Solution 10.14(i)(a) *Estimate of μ_{70}*

In each table we identify in **bold** the latest of the starting dates and the earliest of the ending dates for the year of age (70, 71) and the period of investigation beginning on 1 January 2003 and ending on 1 January 2004. The difference between these two dates gives the exposed to risk.

Life with date of birth 1 April 1932

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-04-02	01-04-03
Date of entry/exit	01-01-03	01-01-04
Exposed to risk	3 months	

Life with date of birth 1 October 1932

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-10-02	01-10-03
Date of entry/exit	01-01-03	01-01-04
Exposed to risk	9 months	

Life with date of birth 1 November 1932

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-11-02	01-11-03
Date of entry/exit	01-03-03	01-09-03
Exposed to risk	6 months	

This life contributes 1 death to θ_{70} .

Life with date of birth 1 January 1933

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-01-03	01-01-04
Date of entry/exit	01-03-03	01-06-03
Exposed to risk	3 months	

This life contributes 1 death to θ_{70} .

Life with date of birth 1 January 1933

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-01-03	01-01-04
Date of entry/exit	01-06-03	01-09-03
Exposed to risk	3 months	

Life with date of birth 1 March 1933

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-03-03	01-03-04
Date of entry/exit	01-09-03	01-01-04
Exposed to risk	4 months	

Life with date of birth 1 June 1933

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-06-03	01-06-04
Date of entry/exit	01-01-03	01-01-04
Exposed to risk	7 months	

Life with date of birth 1 October 1933

Age 70		
Start/end of period of the investigation	01-01-03	01-01-04
Birthday	01-10-03	01-10-04
Date of entry/exit	01-06-03	01-01-04
Exposed to risk	3 months	

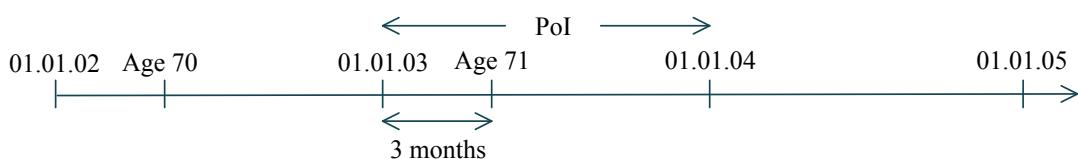
In summary, we have:

$$\theta_{70} = 2 \text{ and } E_{70}^c = 3 + 9 + 6 + 3 + 3 + 4 + 7 + 3 = 38 \text{ months}$$

So:

$$\hat{\mu}_{70} = \frac{2}{\cancel{38} / 12} = 0.63158$$

Alternatively, we can draw time lines and identify the exposure. For example for the life with date of birth 1 April 1932 we have:



(i)(b) ***Estimate of q_{70}***

We have the maximum likelihood estimate of μ_{70} , so we can use the principle of the invariance of maximum likelihood estimates to obtain the maximum likelihood estimate of q_{70} . Now:

$$q_{70} = 1 - e^{-\mu_{70}}$$

because we are told that μ_{70} is the constant force of mortality over (70, 71). Then:

$$\hat{q}_{70} = 1 - e^{-\hat{\mu}_{70}} = 1 - e^{-0.63158} = 0.46825$$

(ii) ***Maximum likelihood estimate of μ_{70} using the Poisson model***

Under the Poisson model the observed number of deaths will follow a Poisson distribution with a parameter $\frac{38}{12}\mu_{70}$. The likelihood function is:

$$L(\mu_{70}) = \frac{\left(\frac{38}{12}\mu_{70}\right)^2 \exp\left(-\frac{38}{12}\mu_{70}\right)}{2!}$$

The log-likelihood function is:

$$\log_e L(\mu_{70}) = 2 \log_e \left(\frac{38}{12} \mu_{70} \right) - \frac{38}{12} \mu_{70} - \log_e 2!$$

Then the maximum likelihood estimator of μ_{70} is the solution to:

$$\frac{\partial \log_e L(\mu_{70})}{\partial \mu_{70}} = \frac{2}{\frac{38}{12}\hat{\mu}_{70}} \times \frac{38}{12} - \frac{38}{12} = \frac{2}{\hat{\mu}_{70}} - \frac{38}{12} = 0$$

giving:

$$\hat{\mu}_{70} = \frac{2}{\frac{38}{12}} = 0.63158$$

Now:

$$\frac{\partial^2 \log_e L(\mu_{70})}{\partial \mu_{70}^2} = -\frac{2}{\mu_{70}^2} < 0$$

so this solution maximises the log-likelihood.

The estimate is the same as that for the two-state model.

(iii) ***Differences between the two-state and the Poisson model***

The Poisson model is an approximation to the two-state model.

While the two-state model can be specified so as to allow for increments (*ie* lives entering the population), this is not possible for the Poisson model.

The estimation of the transition rates in the two-state model involves the measurement of two random variables – the observed number of decrements and the exposed to risk that gave rise to these decrements.

The Poisson model assumes that the exposed to risk remains constant and estimation of the transition rates in the model only involves the measurement of the observed number of decrements.

The maximum likelihood estimators in both models are asymptotically unbiased and consistent. As we have shown, the formulae for the estimates are the same in both models.

Chapter 11

Exposed to risk



Syllabus objectives

- (ix) *Describe how to estimate transition intensities depending on age, exactly or using the census approximation.*
1. *Explain the importance of dividing the data into homogeneous classes, including subdivision by age and sex.*
 2. *Describe the principle of correspondence and explain its fundamental importance in the estimation procedure.*
 3. *Specify the data needed for the exact calculation of a central exposed to risk (waiting time) depending on age and sex.*
 4. *Calculate a central exposed to risk given the data in 3.*
 5. *Explain how to obtain estimates of transition probabilities, including in the single decrement model the actuarial estimate based on the simple adjustment to the central exposed to risk.*
 6. *Explain the assumptions underlying the census approximation of waiting times.*
 7. *Explain the concept of rate interval.*
 8. *Develop census formulae given age at birthday where the age may be classified as next, last, or nearest relative to the birthday as appropriate.*
The deaths and census data may use different definitions of age.
 9. *Specify the age to which estimates of transition intensities or probabilities in 8 apply.*

0 **Introduction**

In this chapter we will take a closer look at how to calculate mortality rates from our observed data. This might at first sight appear to be a very simple task. All we need to do is to count the number of deaths at each age occurring during a specified *observation period* and use the estimators derived in Chapter 10 of the course to obtain a set of $\{q_x\}$ or $\{\mu_x\}$ for the relevant ages.

Basically, that is all that's involved. However, there are a couple of complications that we need to overcome.

First, the multiple-state and Poisson models were based on the assumption that the force of mortality μ_x is constant over a year of age, whereas we know intuitively that it is not.

The second problem relates to data. It may be that the data that a life insurance company can provide are not classified according to age in precisely the way we would like. If this is the case, we will need to group the data according to an age “label” appropriate to the form of the available data. In order to estimate mortality rates at different ages, we will need to decide what age is implied by our arbitrary age label. Additionally, the data may be incomplete for the task ahead.

1 Calculating the exposed to risk

We have seen how exposed to risk arises in a probabilistic model of mortality. In particular, we have seen:

- (a) *Central exposed to risk* E_x^c , namely the observed waiting time in a multiple-state or Poisson model (Chapters 8, 9 and 10); and
- (b) *Initial exposed to risk* E_x , which we have derived in approximate fashion as $E_x \approx E_x^c + \frac{1}{2}d_x$, via the actuarial estimate in the binomial framework (Chapter 10).

In this chapter we consider some problems of a computational nature, concerning the approximation of exposed to risk from incomplete exposure data.

First, we comment on the difficulty, or otherwise, of the subject of exposed to risk.

- (a) Central exposed to risk (or waiting times) are very natural quantities, intrinsically observable even if observation may be incomplete in practice – that is, just record the time spent under observation by each life. Note that this is so even if lives are observed for only part of the year of age $[x, x + 1]$, for whatever reason.
- (b) Initial exposed to risk requires adjustments to be made in respect of those lives who die (and strictly in other cases too, which we have glossed over). It is thus a more complicated object than a central exposed to risk, and, unless we can use the idealised binomial model in which N lives are observed for a whole year without censoring, its interpretation is less simple.
- (c) The central exposed to risk carries through unchanged to arbitrarily complicated multiple-decrement or multiple-state models; the initial exposed to risk does not.
- (d) The central exposed to risk (as we shall see) can easily be approximated in terms of the kind of incomplete observations that are typically available in insured lives investigations. To obtain an initial exposed to risk requires further adjustments, for which it is hard to find good reasons.

You may recognise some of the arguments being put forward from the comparison of the binomial, Poisson and multiple-state models in Chapter 10. In short, the central exposed to risk is more versatile and is simpler to calculate from the data typically available.

Central exposed to risk versus initial exposed to risk

In the past, actuaries have tended to pay great attention to initial exposed to risk because the binomial model was the “prototype” of the life table. The binomial approach was also carried through to multiple decrements, which multiplied the difficulties. Not least of the drawbacks is that the initial exposed to risk becomes increasingly tricky to interpret in terms of the underlying process being modelled as we approach more elaborate situations. Compared with the simplicity of central exposed to risk, in a multiple-state or Poisson setting, much of the difficulty can now be seen to be avoidable.

In this chapter, we will confine the discussion to central exposed to risk.

We will not discuss initial exposed to risk beyond making the observation that $E_x \approx E_x^c + \frac{1}{2}d_x$ is usually a reasonable approximation, for reasons already given. Given the difficulties of initial exposed to risk, and the statistical considerations, it is strongly suggested that a multiple-state or Poisson approach be adopted. A binomial approach should be used only where the data makes the alternative approaches impossible.

Sometimes, of course, the data might be so limited that estimation in the binomial model is easiest. This is very unusual in actuarial investigations, and, in any case, the circumstances will always be different, so no prescriptive approach can be of much use. What matters is to understand how the data are related to (or generated by) the probabilistic model, and to be able to work from these first principles in any unusual case.

The approximation $E_x \approx E_x^c + \frac{1}{2}d_x$ will be reasonable when the period of investigation is long (*i.e.* several years), when we are only analysing one decrement and when that decrement rate is very low. Under these conditions the initial exposed to risk can provide reasonable results. As the Core Reading says, the central exposed to risk is generally preferable, especially if these conditions do not hold.

However, we often *want* rates of mortality (q_x) because they are easy to use and communicate. If our investigation has given us forces of mortality (μ_x) we will need to derive rates from these, and this derivation will itself involve imperfect assumptions and approximations.

It is worth repeating that for the rest of this chapter we will confine the discussion to central exposed to risk, E_x^c .

2 Homogeneity

2.1 The problem of heterogeneity

All our models and analyses are based on the assumption that we can observe groups of *identical* lives (or at least lives whose mortality characteristics are the same). Such a group is said to be *homogeneous*. In practice, this is never possible.

Even if we were to limit the scope of a mortality investigation to people of a specified age and a specified sex (eg females aged 25), there would still be a wide variety of lives – smokers and non-smokers, healthy people and ill people, rocket scientists and actuarial students. A group of lives with different characteristics is said to be *heterogeneous*.

As a result of this heterogeneity, our estimate of the mortality rate would be the estimate of the *average* rate over the whole group of lives. We could use the estimate to predict the rate of mortality for a similar group of lives but it would not provide an accurate estimate of the probability of death for any single individual. This could be a particular problem for an insurance company that wishes to set premiums that accurately reflect the risk of each individual policyholder.

For example, consider a country in which 50% of the population are smokers. If $\mu_{40} = 0.001$ for non-smokers and $\mu_{40} = 0.002$ for smokers, then a mortality investigation based on the entire population may lead us to the estimate $\hat{\mu}_{40} = 0.0015$. An insurance company that calculates its premiums using this average figure would overcharge non-smokers and undercharge smokers.



Question 11.1

A student has suggested that although the situation above is inherently unfair, it is of no real consequence to the insurance company since the average premiums will be sufficient to cover the claims. Comment on the student's reasoning.

Throughout the course we have acknowledged that mortality varies with age. This is an example of heterogeneity within a population. In this section, we extend the argument by looking briefly at the other factors affecting individual lives that can cause their underlying mortality to differ.

2.2 **The solution**

We can subdivide our data according to characteristics known, from experience, to have a significant effect on mortality. This ought to reduce the heterogeneity of each class so formed, although much will probably remain.

Among the factors in respect of which life insurance mortality statistics are often sub-divided are:

- (a) **Sex**
- (b) **Age (as we have assumed throughout)**
- (c) **Type of policy (which often reflects the reason for insuring)**
- (d) **Smoker/non-smoker status**
- (e) **Level of underwriting (eg have they undergone a medical examination?)**
- (f) **Duration in force.**

Others that might be used are:

- (g) **Sales channel**
- (h) **Policy size**
- (i) **Occupation of policyholder**
- (j) **Known impairments.**

This information will be available from the *proposal form*, which the individual must complete when applying for insurance.

“Known impairments” simply refers to any existing medical conditions that the individual has.

If sufficient data were available, we could use the Cox regression model (Chapter 9) to identify the relevant factors.

**Question 11.2**

Explain how the following factors may influence mortality rates:

- (i) sales channel (consider a mailshot to selected existing policyholders and an advert in a popular tabloid national newspaper)
- (ii) occupation of policyholder (consider a deep-sea diver, a high-street newspaper vendor and an actuary).

Two key points are:

Sub-division cannot be carried out unless the relevant information is collected, generally on the proposal form. Sometimes factors for which there is strong external evidence of an effect on mortality cannot be used because (for example) proposal forms have been kept short for marketing or administrative reasons.

Some insurance products are marketed on the strength of the simplicity and brevity of the application process, since some people may be put off by having to provide information relating to their lifestyle and medical history *etc.*

Even in quite large investigations, sub-division using many factors results in much smaller populations in each class, making the statistics more difficult. A balance must be struck between obtaining more and more homogeneity, and retaining large enough populations to make analysis possible, ie the finer the subdivision of the data, the less credible the results of the analysis. This is an example of the law of diminishing returns.

You will study heterogeneity factors in more detail in Subject CT5.

3 ***The principle of correspondence***

Mortality investigations based on estimation of q_x or $\mu_{x+\frac{1}{2}}$ at individual ages must bring together two different items of data: deaths and exposures. It is self-evident that these should be *defined consistently*, or their ratios are meaningless. Care is sometimes needed, however, because these data are often obtained from different sources in the life office. For example, death data might be obtained from the claims file, while exposure data might be obtained from the premium collection file. There is no guarantee that these use the same definition of the policyholders' ages.

In a large insurance company the payment of claims and the collection of premiums will be handled by different departments who may use different databases or computer systems.

A precise statement of what we mean by “defined consistently” is given by the *principle of correspondence*.



Principle of correspondence

A life *alive at time t* should be included in the exposure at age x at time t if and only if, were that life to die immediately, he or she would be counted in the death data d_x at age x .

This seems almost a triviality, but it is very important and useful.

This means that, when we are calculating crude estimates of mortality rates, we should try to ensure that the age definition used in the numerator (the number of deaths) is the same as the age definition used in the denominator (the exposed to risk).

Although this may seem obvious at first glance, we will see that the principle of correspondence is particularly important when we specify the ages of policyholders by definitions other than “age last birthday”. Other definitions that may be used include:

- age next birthday
- age nearest birthday.

We will consider some examples of different age definitions later in this chapter.

4 ***Exact calculation of E_x^c***

4.1 ***Working with complete data***

The procedure for the exact calculation of E_x^c is obvious:

- (a) record all dates of birth
- (b) record all dates of entry into observation
- (c) record all dates of exit from observation
- (d) compute E_x^c .

If we add to the data above the cause of the cessation of observation, we have d_x as well, and we have finished.

The central exposed to risk E_x^c for a life with age label x is the time from Date A to Date B where:

Date A is the latest of: the date of reaching age label x

the start of the investigation and

the date of entry

Date B is the earliest of: the date of reaching age label $x+1$

the end of the investigation and

the date of exit (for whatever reason)



Question 11.3

If the age label is “age nearest birthday”, when does a life reach age label x ?

Note that:

- The calculation takes account of *all* movements into and out of the population (not just deaths).
- *All* decrements contribute a fraction of a year in the year of exit and increments contribute a fraction of a year in the year of entry.
- The central exposed to risk is independent of the cause of exit under consideration.
- It is usual to assume an average of 365½ days in a year in order to convert days of exposure to years.

Although exact exposed to risk calculations are messy to do by hand, they can be done very easily on a computer (*eg* using the date functions on a spreadsheet) if we have the required information for all lives.

Conventions are often needed to define whether the day of entry or day of exit contributes to the total exposed to risk. We do not count both days.



Question 11.4

A mortality investigation covers the period 1 January 2001 to 31 December 2003. In this investigation, the age label used is “age last birthday”.

Give the range of dates for which the lives in the following table contribute to E_x^c at each age where they make a contribution. Assume that the day of entry counts in the exposed to risk but the day of exit does not.

	Date of birth	Date of joining	Date of exit	Reason for exit
A	25.04.69	07.08.99	30.10.02	Death
B	01.07.69	12.09.02	—	—
C	04.09.68	22.07.03	4.12.03	Withdrawal

**Question 11.5**

Now suppose that we are using the age label “age next birthday”. Give the range of dates for which the lives in the table above contribute to E_{34}^c .

4.2 Working with incomplete data

All of the remainder of this chapter is about approximate procedures when the data above have *not* been recorded. We will deal with two questions:

- (a) What happens when the dates of entry to and exit from observation have not been recorded? (Section 5)
- (b) What happens if the definition of age does not correspond exactly to the age interval x to $x + 1$ (for integer x)? (Section 6)

5 **Census approximations to E_x^c**

In this section we will consider how to calculate E_x^c approximately when the exact dates of entry to and exit from observation have not been recorded.

5.1 **The available data**

Suppose that we have death data of the form:

$$d_x = \text{total number of deaths age } x \text{ last birthday during calendar years } K, K+1, \dots, K+N$$

That is, we have observations over $N+1$ calendar years of all deaths between ages x and $x+1$.

However, instead of the times of entry to and exit from observation of each life being known, we have instead only the following census data:

$$P_{x,t} = \text{Number of lives under observation aged } x \text{ last birthday at time } t \\ \text{where } t = 1 \text{ January in calendar years } K, K+1, \dots, K+N, K+N+1$$

This is in fact similar to the way in which data are submitted to the CMI.

The CMI is the Continuous Mortality Investigation.

It is often quite convenient for companies to submit a total of policies in force on a date such as 1 January.

Companies may not take the time to calculate the number of policies in force every day because this information would be of limited use. However, each insurance company is likely to perform an annual actuarial valuation to assess its financial position. The number of policies in force on the annual valuation date (usually 1 January in the UK) would be calculated and recorded as part of the valuation process.

5.2 The census approximation to E_x^c

Now define $P_{x,t}$ to be the number of lives under observation, aged x last birthday, at any time t . Note that:

$$E_x^c = \int_K^{K+N+1} P_{x,t} dt$$

During any short time interval $(t, t+dt)$ there will be $P_{x,t}$ lives each contributing a fraction of a year dt to the exposure. So, integrating $P_{x,t} \times dt$ over the observation period gives the total central exposed to risk for this age. In other words, E_x^c is the area under the $P_{x,t}$ "curve" between $t = K$ and $t = K + N + 1$.

The problem is that we do not know the value of $P_{x,t}$ for all t , so we cannot work out the exact value of the integral.

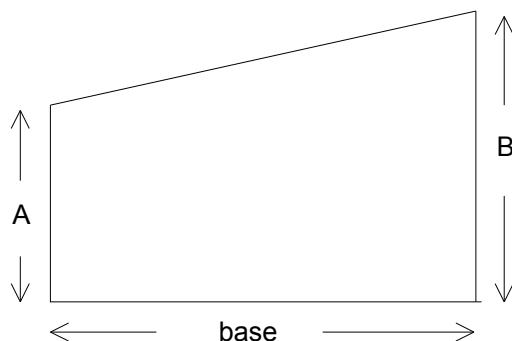
We have the values of $P_{x,t}$ only if t is a 1 January (a census date), so we must estimate E_x^c from the given census data.

The problem reduces to estimating an integral, given the integrand at just a few points (in this case, integer spaced calendar times). This is a routine problem in numerical analysis.

The simplest approximation, and the one most often used, is that $P_{x,t}$ is linear between census dates, leading to the *trapezium* approximation.

The area of a trapezium is:

$$\text{base} \times \frac{1}{2} (\text{length of side A} + \text{length of side B})$$



In this case:

- the base of the trapezium is equal to 1, ie the period between census dates
- the length of side A is $P_{x,t}$, the number of policies in force at the *start* of the year (at time t)
- the length of side B is $P_{x,t+1}$, the number of policies in force at the *end* of the year (at time $t+1$).

Using the trapezium approximation:

$$E_x^c = \int_K^{K+N+1} P_{x,t} dt \cong \sum_{t=K}^{K+N} \frac{1}{2}(P_{x,t} + P_{x,t+1})$$

This is the method used by the CMI. It is easily adapted to census data available at more or less frequent intervals, or at irregular intervals.



Example

Estimate E_{55}^c based on the following data:

Calendar year	Population aged 55 last birthday on 1 January
2005	46,233
2006	42,399
2007	42,618
2008	42,020

Solution

Using the census approximation, the central exposed to risk for the 3-year period 1 January 2005 to 1 January 2008 is:

$$\begin{aligned}
 E_{55}^c &\approx \frac{1}{2}[P_{55}(1.1.05) + P_{55}(1.1.06)] + \frac{1}{2}[P_{55}(1.1.06) + P_{55}(1.1.07)] \\
 &\quad + \frac{1}{2}[P_{55}(1.1.07) + P_{55}(1.1.08)] \\
 &= \frac{1}{2}P_{55}(1.1.05) + P_{55}(1.1.06) + P_{55}(1.1.07) + \frac{1}{2}P_{55}(1.1.08) \\
 &= \frac{1}{2}(46,233) + 42,399 + 42,618 + \frac{1}{2}(42,020) \\
 &= 129,143.5
 \end{aligned}$$

**Question 11.6**

The disreputable insurance company *Honest Sid's Mutual* had mixed fortunes in the year 2006. At both the start and the end of the year 547 policies were in force in respect of policyholders aged 40, but these figures do not tell the whole story.

There was adverse publicity early in the year linking the company's investment managers with a gambling syndicate. As a result, many policyholders "took their money elsewhere". Following a successful marketing campaign offering a free toaster to all applicants, the number of policyholders aged 40 rose from 325 at 1 June 2006 to 613 at 1 September 2006.

Calculate the value of E_{40}^c approximately.

6 Deaths classified using different definitions of age

In Section 5, we used a definition of age “ x last birthday”, which identifies the year of age $[x, x + 1]$.

All this is saying is that if someone is aged x last birthday, then their actual age is somewhere between x and $x + 1$.

Other definitions could be used, for example:

$d_x^{(2)} = \text{total number of deaths age } x \text{ nearest birthday during calendar years } K, K+1, \dots, K+N$

$d_x^{(3)} = \text{total number of deaths age } x \text{ next birthday during calendar years } K, K+1, \dots, K+N$

Each of these identifies a different year of age, called the *rate interval*.



Rate interval

A *rate interval* is a period of one year during which a life's recorded age remains the same, eg the period during which an individual is “aged 36 last birthday” or “aged 42 nearest birthday”.

The key concept is that lives carry the same age label throughout a rate interval. Given this, it follows that the rate interval starts on the day when a life's age label changes.

For example, if the age label is “age nearest birthday”, a life will go from “age 42 nearest birthday” to “age 43 nearest birthday” 6 months before the life's 43rd birthday, ie at exact age 42½. In this case, the age label changes halfway between birthdays.

Consequently, estimates \hat{q} or $\hat{\mu}$ obtained from these data ($d_x^{(2)}$ and $d_x^{(3)}$) will not be estimates of q_x or $\mu_{x+\frac{1}{2}}$, but will be estimates of q and μ at other ages.

The rate of mortality q measures the probability of death over the next year of age or, more generally, over the next rate interval. So the q -type rate applies to the age at the start of the rate interval. In contrast, μ measures the average instantaneous rate of mortality that we observe over the rate interval and the μ -type rate applies to the age in the middle of the rate interval.

We summarise the possibilities as follows:

Definition of x	Rate interval	\hat{q} estimates	$\hat{\mu}$ estimates
Age last birthday	$[x, x + 1]$	q_x	$\mu_{x+1/2}$
Age nearest birthday	$[x - 1/2, x + 1/2]$	$q_{x-1/2}$	μ_x
Age next birthday	$[x - 1, x]$	q_{x-1}	$\mu_{x-1/2}$



Important note

Once the rate interval has been identified (from the age definition used in d_x) the rule is that:

- the crude \hat{q} estimates q at the start of the rate interval; and
- the crude $\hat{\mu}$ estimates μ in the middle of the rate interval.



Example

You have details of the number of deaths aged 40 nearest birthday in a recent investigation. What initial rate and force of mortality can be estimated from this total?

Solution

In order for a death to contribute to this total, it must have occurred between ages $39\frac{1}{2}$ and $40\frac{1}{2}$, ie in the rate interval $[39\frac{1}{2}, 40\frac{1}{2}]$.

We can use the total to estimate q for the age at the start of the rate interval, ie we are estimating $q_{39\frac{1}{2}}$.

The force of mortality may vary over the rate interval. We assume that the number of deaths gives us information about the average force of mortality over the rate interval. So we can use the total to estimate μ for the age at the midpoint of the rate interval, ie we are estimating μ_{40} .

6.1 Consistency between census data and death data

Whatever the definition of age used to classify the lives, we can calculate the exact exposure to risk if we have full information about the dates of entry to and exit from observation. In practice, the information will not be complete but will take the form of census data.

We must ensure that the census data are consistent with the death data. We invoke the principle of correspondence; we must check the following:

The census data $P_{x,t}$ are consistent with the death data d_x if and only if, were any of the lives counted in $P_{x,t}$ to die *on the census date*, he or she would be included in d_x .

The definition of census data corresponding to the rate interval $[x - \frac{1}{2}, x + \frac{1}{2}]$ (ie corresponding to an age definition of age at *nearest birthday*) is:

$$P_{x,t}^{(2)} = \text{No. of lives under observation, age } x \text{ nearest birthday at time } t, \\ \text{where } t = 1 \text{ January in calendar year } K, K+1, \dots, K+N, K+N+1$$

and the definition of census data corresponding to the rate interval $[x - 1, x]$ (ie corresponding to an age definition of age at *next birthday*) is:

$$P_{x,t}^{(3)} = \text{Number of lives under observation, age } x \text{ next birthday at time } t, \\ \text{where } t = 1 \text{ January in calendar year } K, K+1, \dots, K+N, K+N+1$$

In the event that the death data and the census data use different definitions of age, we must adjust the census data. Unless it is unavoidable, we never adjust the death data, since that “carries most information” when rates of mortality are small. Hence ...



Important note

... it is always the death data that determine what rate interval to use.

For example, the CMI uses the definition “age *nearest birthday*” in its work; that is, death data as in $d_x^{(2)}$. However, some life offices contribute census data classified by “age *last birthday*”, because that is what is available from their records. The latter must be adjusted in some way.

For example, if we define:

$$P'_{x,t} = \frac{1}{2}(P_{x-1,t} + P_{x,t})$$

we can see that $P'_{x,t}$ approximates $P_{x,t}^{(2)}$.



Question 11.7

Explain why $P'_{x,t}$ approximates $P_{x,t}^{(2)}$.

6.2 Worked example

In this section we give an example where we derive a formula for the central exposed to risk for a given mortality investigation.



Example

An investigation into mortality covered the period 1 January 2007 to 1 January 2008. Time is measured in years from 1 January 2007 and $P_x(t)$ denotes the number of lives at time t aged x last birthday. The following data were recorded for each x :

d_x = number of deaths aged x last birthday

$P_x(0)$ and $P_x(1)$

- (i) Obtain an expression for the central exposed to risk in terms of the available census data that may be used to estimate the force of mortality μ_{x+f} , stating your assumptions.
- (ii) Determine the value of f , stating any assumptions you make.

Solution(i) ***Central exposed to risk***

Since the death data and the census data match (both take x to be the age last birthday), the central exposed to risk for age label x is:

$$E_x^c = \int_0^1 P_x(t) dt = \frac{1}{2} [P_x(0) + P_x(1)]$$

This assumes that $P_x(t)$ varies linearly over calendar year 2007.

(ii) ***Value off***

Since x is defined to be the age last birthday, we have a rate interval that starts on the x th birthday. So the (exact) age in the middle of the rate interval is $x+\frac{1}{2}$ (no assumptions required) and:

$$\hat{\mu}_x = \frac{d_x}{E_x^c} \text{ estimates } \mu_{x+\frac{1}{2}}$$

So $f = \frac{1}{2}$.

Now try this one yourself.

**Question 11.8**

A mortality investigation covered the period 1 January 2007 to 1 January 2008. Time is measured in years from 1 January 2007 and $P_x(t)$ denotes the number of lives at time t aged x last birthday. The following data were recorded for each x :

d_x = number of deaths aged x next birthday

$P_x(0)$ and $P_x(1)$

- (i) Obtain an expression for the initial exposed to risk in terms of the available census data that may be used to estimate the initial rate of mortality q_{x+f} , stating your assumptions.
- (ii) Determine the value of f , stating any assumptions you make.

7 Exam-style question

To finish off this chapter, you should now try the following exam-style question.

**Question 11.9****Subject 104, September 2003, Question 3 (updated)**

- (i) List the data required for the exact calculation of the central exposed to risk of lives aged x last birthday in a mortality investigation over the two-year period from 1 January 2006 to 1 January 2008. [2]
- (ii) In an investigation of mortality during the period 1 January 2006 to 1 January 2008, data are available on the number of lives under observation, aged x last birthday, on 1 January 2006, 1 July 2006 and 1 January 2008.

Derive an approximation for the central exposed to risk at age x last birthday over the period in terms of the populations recorded on each of these three dates.

[3]

[Total 5]



Chapter 11 Summary

The central exposed to risk is more versatile than the initial exposed to risk and is simpler to calculate from the data typically available.

In order to reduce *heterogeneity* amongst the lives observed in a mortality investigation, we should divide our data into homogeneous subgroups according to characteristics known to have a significant effect on mortality. This approach will only be possible if the appropriate information is available and we have sufficient data to make such detailed analysis possible.

The *principle of correspondence* states that the death data and the exposed to risk must be defined consistently, *ie* the numerator (d_x) and denominator (E_x^c) must correspond.

The exposed to risk can be calculated exactly if we have complete information for every life. In practice we may have only limited information relating to the size of the population at certain dates known as *census dates*. We can use this information to approximate the exposed to risk.

The data will be classified in terms of a *rate interval*.

Central exposed to risk

Suppose that $P_x(t)$ is the number of lives in the investigation at time t with age label x , and the same age classification has been used in both the census data (*ie* the $P_x(t)$ function) and the death data. Then:

$$E_x^c = \int_K^{K+N+1} P_x(t) dt = \sum_{t=K}^{K+N} \frac{1}{2} (P_x(t) + P_x(t+1))$$

assuming $P_x(t)$ varies linearly between the census dates.

If the census data and the death data do not match, an adjustment has to be made to the formula above to reflect the difference.

This page has been left blank so that you can keep the chapter summaries together for revision purposes.

Chapter 11 Solutions

Solution 11.1

A company that charges the same premium rate to lives that present different risks (*i.e.* to smokers and non-smokers) is in an unstable position. Its premium rate will be based on the aggregate expected risk of its applicants for insurance, assuming a certain mix of high risk and low risk lives. The office will tend to lose low risk business to its competitors if they are charging different premium rates, and will itself attract high risk business, so that its aggregate premium rate will be inadequate to meet the actual claim cost. This is called *anti-selection*. The office will then make losses, which will ultimately threaten solvency.

The company can avoid this anti-selection only by charging different premium rates appropriate to the different levels of risk presented by the applicants. This is the process of risk classification. The avoidance of anti-selection is therefore one of its key advantages, leading to improved financial stability for the insurer and a reduced risk of insolvency.

Solution 11.2

- (i) The sales channel will determine the section of the population targeted by the insurance company. For example, an advert in a popular tabloid national newspaper will typically be read by the lower socio-economic groups within the population. Different sections of society experience very different rates of mortality. The mortality experienced by the lower socio-economic groups within the population is likely to be significantly heavier than existing policyholders who have been selected according to favourable lifestyle and medical history criteria.
- (ii) Occupation can influence mortality rates directly (*e.g.* deep-sea divers suffer a high rate of accidental death whilst performing their job) and indirectly (*e.g.* actuaries may have access to company medical schemes, which will help to identify and cure medical problems before they become life threatening). Other occupations may only be carried out by a specific subsection of the population (*e.g.* high-street newspaper vendors may typically be old people whose health prevents them from doing a more active job).

Solution 11.3

Under this definition, a life attains age label x at exact age $x - \frac{1}{2}$.

Solution 11.4

- A The first life is a “starter” aged between 31 and 32. The periods of contribution to the central exposed to risk are as follows:

E_{31}^c 01.01.01 to 24.04.01

E_{32}^c 25.04.01 to 24.04.02

E_{33}^c 25.04.02 to 29.10.02

- B The second life joins at age 33 last birthday, so the contributions are:

E_{33}^c 12.09.02 to 30.06.03

E_{34}^c 01.07.03 to 31.12.03

- C The third life joins at age 34 last birthday, the contributions are:

E_{34}^c 22.07.03 to 03.09.03

E_{35}^c 04.09.03 to 03.12.03

Solution 11.5

- A The first life has age label “34 next birthday” from 25.04.02 to 24.04.03. But he dies on 30.10.02, so his contribution to E_{34}^c is from 25.04.02 to 29.10.02. (Note that his contribution to E_{34}^c is the same as his contribution to E_{33}^c in the previous question.)

- B The second life contributed to E_{34}^c from 12.09.02 to 30.06.03

- C The third life makes no contribution to E_{34}^c .

Solution 11.6

We know the values of the integrand at 1 January, 1 June, 1 September and 31 December, ie at times 0, 5/12, 8/12, 1.

Using the trapezium rule:

$$E_{40}^c \approx \left[\frac{5}{12} \times \frac{(547+325)}{2} \right] + \left[\frac{3}{12} \times \frac{(325+613)}{2} \right] + \left[\frac{4}{12} \times \frac{(613+547)}{2} \right] = 492.25$$

Solution 11.7

$P_{x,t}^{(2)}$ represents the number of lives under observation, aged x nearest birthday at time t . This group comprises all lives between ages $x - \frac{1}{2}$ and $x + \frac{1}{2}$. Those between the ages of $x - \frac{1}{2}$ and x are aged $x - 1$ last birthday. Those between ages x and $x + \frac{1}{2}$ are aged x last birthday. Assuming that birthdays are uniformly distributed over the calendar year, then $\frac{1}{2}$ of the $P_{x,t}^{(2)}$ lives will be aged $x - 1$ last birthday and $\frac{1}{2}$ of the $P_{x,t}^{(2)}$ lives will be aged x last birthday.

So $P_{x,t}^{(2)}$ can be approximated by taking the average of:

- the number of lives aged between $x - 1$ and x , ie the number of policyholders aged $x - 1$ last birthday at time t given by $P_{x-1,t}$, and
- the number of lives aged between x and $x + 1$, ie policyholders aged x last birthday at time t given by $P_{x,t}$.

This approximation assumes that the birthdays of the individuals involved are spread uniformly over the calendar year, which is usually approximately true.

Solution 11.8(i) ***Initial exposed to risk***

Since the death data and the census data don't match, we define a new census function $P'_x(t)$ that does match with the death data. So let:

$$P'_x(t) = \text{number of lives at time } t \text{ aged } x \text{ next birthday}$$

The central exposed to risk for age label x is:

$$E_x^c = \int_0^1 P'_x(t) dt = \frac{1}{2} [P'_x(0) + P'_x(1)]$$

This assumes that $P'_x(t)$ varies linearly between time 0 and time 1.

Now:

$$\begin{aligned} P'_x(0) &= \text{number of lives at time 0 (1 January 2007) aged } x \text{ next birthday} \\ &= \text{number of lives at time 0 aged } x-1 \text{ last birthday} \\ &= P_{x-1}(0) \end{aligned}$$

Similarly:

$$P'_x(1) = P_{x-1}(1)$$

So:

$$E_x^c = \frac{1}{2} [P_{x-1}(0) + P_{x-1}(1)]$$

and the initial exposed to risk at age label x can be calculated from the central exposed to risk using the formula:

$$E_x = E_x^c + \frac{1}{2} d_x$$

Here we are making the assumption that deaths occur uniformly over the rate interval, which in this case is the same as saying that deaths occur uniformly over each year of age (since this is a rate interval starting at age $x-1$ and ending at age x).

(ii) ***Value of f***

Since x is defined to be the age next birthday, we have a rate interval that ends on the x th birthday. So the exact age at the end of the rate interval is x , and the exact age at the start of the rate interval is $x-1$. Hence:

$$\hat{q}_x = \frac{d_x}{E_x} \text{ estimates } q_{x-1}$$

So $f = -1$.

Solution 11.9(i) ***Data required for exact calculation of exposed to risk***

For each life observed during the investigation period we need:

- date of birth
- date of joining the investigation (if after 1 January 2006)
- date of leaving the investigation (if before 1 January 2008)
- reason for leaving the investigation (*ie* death or an alternative cause).

(ii) ***Deriving a census formula***

This will be much easier to do with some symbols. So first define the data:

$P_x(t)$ = number of lives under observation at time t years (where $t = 0$ is 1 January 2006), aged x last birthday at time t .

We actually have $P_x(0)$, $P_x(\frac{1}{2})$ and $P_x(2)$.

Note that we do not have $P_x(1)$!

Now the exact exposed to risk for lives aged x last birthday over the two-year investigation period is:

$$E_x^c = \int_0^2 P_x(t) dt$$

We can use the trapezium rule between successive census dates to calculate the approximate value of the integral, assuming $P_x(t)$ varies linearly between the census dates, which gives:

$$\begin{aligned} E_x^c &= \int_0^{\frac{1}{2}} P_x(t) dt + \int_{\frac{1}{2}}^2 P_x(t) dt \\ &= \frac{1}{2} \times \frac{[P_x(0) + P_x(\frac{1}{2})]}{2} + \frac{1}{2} \times \frac{[P_x(\frac{1}{2}) + P_x(2)]}{2} \\ &= \frac{1}{4} P_x(0) + P_x(\frac{1}{2}) + \frac{3}{4} P_x(2) \end{aligned}$$

Chapter 12

Graduation and statistical tests



Syllabus objectives

- (x) *Describe how to test crude estimates for consistency with a standard table or a set of graduated estimates, and describe the process of graduation.*
1. *Describe the following statistical tests of crude estimates, for comparison with a standard table:*
 - *chi-square test*
 - *standardised deviations test*
 - *sign test*
 - *cumulative deviation test*
 - *grouping of signs test*
 - *serial correlations test*

For each test describe:

 - *the formulation of the hypothesis*
 - *the test statistic*
 - *the distribution of the test statistic using approximations where appropriate*
 - *the application of the test statistic*
 2. *Describe the reasons for graduating crude estimates of transition intensities or probabilities, and state the desirable properties of a set of graduated estimates.*
 3. *Describe a test for smoothness of a set of graduated estimates.*

0 Introduction

0.1 Graduation of observed mortality rates

In previous chapters we have introduced models for mortality over a single year of age, x to $x+1$. In practice, an investigation will include a considerable range of ages. For example a national life table will include all ages from 0 to over 100.

The crude mortality rates derived from a mortality investigation will not be the final rates that are published for use in actuarial calculations. The rates will have to pass through a further process called *graduation*.

Graduation refers to the process of using statistical techniques to improve the estimates provided by the crude rates. The aims of graduation are to produce a smooth set of rates that are suitable for a particular purpose, to remove random sampling errors (as far as possible) and to use the information available from adjacent ages to improve the reliability of the estimates. Graduation results in a “smoothing” of the crude rates.

The graduation process itself is covered in Chapter 13, *Methods of graduation*. In this chapter we look at the aims of graduation and the statistical tests that are used to check the reasonableness of the graduated rates.

0.2 Hats, squiggles and circles

The Subject CT4 Core Reading uses various “accents” to indicate different versions of the mortality rates q_x and μ_x . Here is a summary of the notation that we use:

q_x is the true mortality rate

\hat{q}_x is the numerical value of the crude (observed) estimate

\tilde{q}_x is the random variable representing the crude (observed) estimator

$\overset{\circ}{q}_x$ is the graduated rate

Here is a paragraph that uses all of these:

If we divide the deaths recorded at age 50 next birthday by the corresponding initial exposed to risk, we obtain the crude initial rate $\hat{q}_{50} = 0.00527$. If the assumptions of the binomial model for mortality hold, \tilde{q}_{50} should have a standard normal distribution (approximately). After graduation (a process that uses information from nearby ages to reduce random errors), we obtain an improved estimate $\tilde{q}_{50}^o = 0.00540$. Because of the age definition used for recording deaths (“age next birthday”), this is actually an estimate of q_{49} .

Note that, although this is the notation adopted by the Core Reading, it is not universally recognised, and you may well encounter alternative notation – even in the exams. If there is any scope for doubt, you should define the symbols you are using.

0.3 *The underlying assumptions*

We now suppose that we have data for all ages from the lowest, denoted x_1 , to the highest, x_m , depending on the investigation.

(a) **If we are using the Poisson or multiple-state model, we have:**

$$\left. \begin{array}{ll} \text{Deaths} & d_x \\ \text{Exposed-to-risk} & E_x^c \\ \text{Crude estimates} & \hat{\mu}_{x+\frac{1}{2}} \end{array} \right\} x = x_1, x_2, \dots, x_m$$

and we will use the approximate asymptotic distribution:

$$D_x \sim \text{Normal} (E_x^c \mu_{x+\frac{1}{2}}, E_x^c \mu_{x+\frac{1}{2}})$$

$$\text{or } \tilde{\mu}_{x+\frac{1}{2}} \sim \text{Normal} \left(\mu_{x+\frac{1}{2}}, \frac{\mu_{x+\frac{1}{2}}}{E_x^c} \right)$$

(b) If we are using the binomial model, we have:

$$\left. \begin{array}{ll} \text{Deaths} & d_x \\ \text{Exposed-to-risk} & E_x \equiv E_x^c + \frac{1}{2}d_x \\ \text{Crude estimates} & \hat{q}_x \end{array} \right\} \quad x = x_1, x_2, \dots, x_m$$

and we will use the approximation:

$$D_x \sim \text{Binomial}(E_x, q_x)$$

$$\text{or } \tilde{q}_x \sim \frac{1}{E_x} \text{Binomial}(E_x, q_x)$$

(The notation used here means that \tilde{q}_x has the same statistical distribution as the values from a $\text{Binomial}(E_x, q_x)$ distribution after they have been divided by E_x .)

We also use the further approximation

$$D_x \sim \text{Normal}(E_x q_x, E_x q_x(1 - q_x))$$

$$\text{or } \tilde{q}_x \sim \text{Normal}\left(q_x, \frac{q_x(1 - q_x)}{E_x}\right)$$

You should recognise these underlying assumptions as some of the important results from Chapter 10 of the course.



Question 12.1

If we are using the Poisson or multiple-state model, why do our crude estimates represent $\hat{\mu}_{x+\frac{1}{2}}$ rather than $\hat{\mu}_x$?

1 Comparison with another experience

1.1 Introduction

Given the data above (the observed numbers of deaths and the exposed-to-risks for each age, and our crude estimates), **we often want to know if it is consistent with another, known experience.** For example, if it is the recent experience of the policyholders of a life insurance company, we might ask:

- (a) Is it consistent with the company's own past experience, or is the experience changing? This could be important for pricing life insurance contracts.
- (b) Is it consistent with the published life tables? This is important if the company plans to use published tables for any financial calculations.

It is important for an insurance company to be aware of the extent to which the mortality experienced by its policyholders differs from that of its past experience and published life tables. The difference will be reflected in the premiums charged for life assurance contracts.



Question 12.2

What would be the major problem of charging premiums that are:

- (a) too low?
- (b) too high?

1.2 Standard tables

Published life tables based on large amounts of data are called standard tables. The main examples are:

- (a) National life tables, based on the census data and death registration data of a whole country. In the UK, these are published every 10 years; the largest are the English Life Tables (actually based on the population of England and Wales).

Most countries have a similar approach. For instance, in Italy national life tables are also prepared every ten years based on the national population. The most recently published sets are called SIM90 and SIF90.

- (b) **Tables based on data from life insurance companies.** In the UK, most life insurance companies contribute data to the Continuous Mortality Investigation (CMI), which publishes extensive tables for different types of business from time to time.

The latest were based on 1991-94 data, and are known as the “92 series” tables. Most (UK) life insurance companies use these standard tables very extensively, so it is important that they check whether or not their own mortality experience is consistent with that of the tables.

The term “consistent” covers two concepts: the *shape* of the mortality curve over the range of ages and the *level* of mortality rates.

1.3 Comparison with standard tables

We introduce the following notation. The superscript “s” will denote a quantity from a published standard table, eg q_x^s or $\mu_{x+\frac{1}{2}}^s$.

In rough terms, the question is whether our estimates \hat{q}_x or $\hat{\mu}_{x+\frac{1}{2}}$ are consistent with the given q_x^s or $\mu_{x+\frac{1}{2}}^s$. We will formulate this more precisely, in a way that allows us to derive statistical tests.

We have:

- (a) the probabilistic model (multiple-state, Poisson or binomial);
- (b) the data (the observed numbers of deaths, the exposed-to-risks and our crude estimates \hat{q}_x or $\hat{\mu}_{x+\frac{1}{2}}$); and
- (c) a standard table.

The hypothesis that we wish to test is that the standard table quantities $\{q_x^s\}$ or $\{\mu_{x+\frac{1}{2}}^s\}$ are the “true” parameters of the model at each age x .

In other words, our null hypothesis is:

H_0 : the true underlying mortality rates at each age x for the experience are the rates in the standard table.

We can derive tests of this hypothesis using the distributional assumptions of Section 0.3:

- (a) If we are using a multiple-state or Poisson model under the hypothesis:

$$D_x \sim \text{Normal} \left(E_x^c \mu_{x+\frac{1}{2}}^s, E_x^c \mu_{x+\frac{1}{2}}^s \right) \text{ (approximately)}$$

- (b) If we are using a binomial model under the hypothesis:

$$D_x \sim \text{Normal} \left(E_x q_x^s, E_x q_x^s (1 - q_x^s) \right) \text{ (approximately).}$$

Hence we can find test statistics comparing the *actual* deaths d_x with the *expected* deaths given by these distributions. We will describe suitable statistical tests later in this chapter.

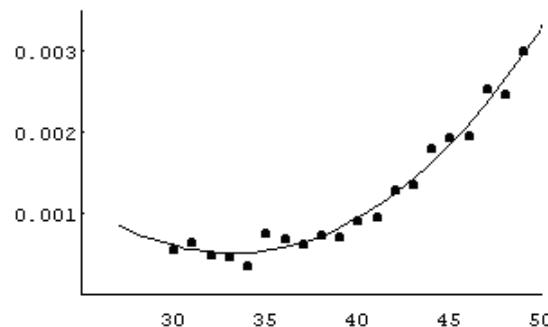
First, however, we must discuss some general features of mortality experiences, and the extent to which we might want to adjust (*ie* graduate) the crude estimates so that they reflect these features.

2 The need for graduation

The crude estimates $\{\hat{q}_x\}$ or $\{\hat{\mu}_{x+\frac{1}{2}}\}$ will progress more or less roughly, ie it is unlikely that the crude estimates will progress smoothly. In large part, this is because they have each been estimated independently and hence suffer independent sampling errors. The smaller the sample size (ie the population studied), the less smoothly the crude estimates are likely to progress.

For several reasons (discussed below) we would prefer to work with q_x or μ_x , which are smooth functions of age. Therefore, we graduate or smooth the crude estimates, to produce a set of graduated estimates that do progress smoothly with age. We denote these $\{\check{q}_x\}$ or $\{\check{\mu}_{x+\frac{1}{2}}\}$.

The graph below illustrates the type of relationship we would expect to see between the crude rates (■) and the graduated rates (—). The true underlying rates are likely to be very close to the graduated rates.



Three questions that we must answer are:

- (a) Why do we want smoothed estimates? We discuss this in Section 3.
- (b) How do we carry out the graduation? (ie produce the \check{q}_x from the \hat{q}_x or the $\check{\mu}_{x+\frac{1}{2}}$ from the $\hat{\mu}_{x+\frac{1}{2}}$). This is the subject of Chapter 13.
- (c) How do we decide that a given attempt to graduate the crude estimates is satisfactory? We discuss this in Section 4, before resuming our discussion of statistical tests of a mortality experience, because statistical tests form part of the answer.

3 Reasons for graduation

3.1 The theoretical argument

At the heart of our desire to graduate is the intuitive idea that quantities such as q_x or μ_x should be smooth functions of age. There is some evidence from large investigations to support this, but it is nevertheless an *assumption*.

It follows that a crude estimate at any age x (\hat{q}_x or $\hat{\mu}_{x+\frac{1}{2}}$) also carries information about the values of q_{x-1} , q_{x+1} or $\mu_{x-\frac{1}{2}}$, $\mu_{x+\frac{1}{2}}$ etc.

For example, if q_x is smooth and not changing too rapidly, then \hat{q}_x should not be too far away from estimating q_{x-1} and q_{x+1} , as well as being the “best” estimate, in some sense, of q_x . By smoothing, we can make use of the data at adjacent ages to improve the estimate at each age.

Another way of looking at this is that smoothing reduces the sampling errors at each age.



Question 12.3

It is intuitively sensible to think that mortality is a smooth function of age. But can you think of any reasons for mortality not to be smooth at certain ages?

3.2 The practical argument

A purely practical reason for smoothing mortality data is that we will use the life table to compute financial quantities, such as premiums for life insurance contracts. It is very desirable that such quantities progress smoothly with age, since irregularities (jumps or other anomalies) are hard to justify in practice.

We could calculate these quantities using our crude mortality rates, and then smooth the premium rates etc directly, but it is much more convenient to have smoothed mortality rates to begin with.

We would never, in any case, apply the results of a mortality experience directly to some financial problem without considering carefully its suitability. This means comparing it with other relevant experiences and tables, not just in aggregate but over age ranges of particular financial significance. It is often the case that a mortality experience must be adjusted in some way before use, in which case there is little point in maintaining the roughness of the crude estimates.

3.3 Limitations

What graduation *cannot* do is remove any bias in the data arising from faulty data collection or otherwise.

Graduation can only produce results as reliable as the original data. This principle is known as “garbage in, garbage out”.

3.4 Summary

The crude estimates of mortality (\hat{q}_x or $\hat{\mu}_{x+\frac{1}{2}}$) provide an estimate of the true underlying mortality for a particular age. However, since we believe that the underlying rates of mortality will follow a smooth curve as the age varies, we can use the additional information provided by the numbers of deaths at nearby ages to improve our estimate. This process of applying statistical techniques to improve the estimates provided by crude rates over a range of ages is called *graduation*.

The aims of graduation are:

- to produce a smooth set of rates that are suitable for a particular purpose
- to remove random sampling errors
- to use the information available from adjacent ages



Question 12.4

A student has said “If the data includes the whole population, there is no need to graduate the crude rates because there will be no sampling errors”. Discuss briefly.

4 Desirable features of a graduation

We list three desirable features of a graduation:

- (a) **smoothness;**
- (b) **adherence to data; and**
- (c) **suitability for the purpose to hand.**

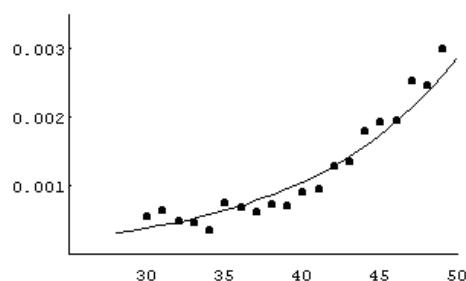
4.1 Smoothness versus adherence to data

The reasons for desiring smoothness were discussed above. At one extreme, we could easily smooth the crude estimates by ignoring the data altogether; we want to avoid such extremes since we want the graduation to be representative of the experience. We say that we require **adherence to data or goodness of fit**.

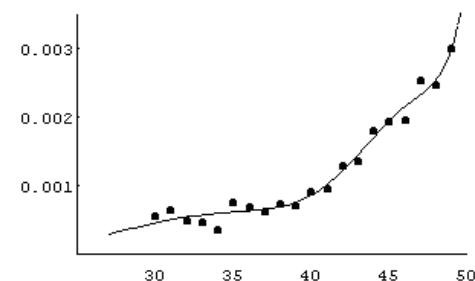
Smoothness and adherence to data are usually conflicting requirements. Perfect smoothness (extreme example: a straight line) pays no attention to the data, while perfect adherence to the data means no smoothing at all.

If the graduation process results in rates that are smooth but show little adherence to the data, then we say that the data may be *overgraduated*. The graph in Graduation A (see below) is very smooth, but it consistently overestimates the crude rates at the younger ages and underestimates them at the older ages.

Overgraduation has an opposite, referred to as *undergraduation*. This refers to the case where insufficient smoothing has been carried out. This will tend to produce a curve of inadequate smoothness, but better adherence to data. In this case, the graduated rates will follow the crude rates very closely, but will show an irregular progression over the range of ages. The graph in Graduation B (which uses the same data as Graduation A) adheres very closely to the crude rates, but it twists and turns erratically.



Graduation A - overgraduated



Graduation B - undergraduated

4.2 Testing smoothness and adherence to data

The “art” of graduation lies in finding a satisfactory compromise.

The compromise is between smoothness and adherence to data. The balance between the two will be first addressed when we carry out the graduation itself (see Chapter 13).

To assist in this task, we have a battery of tests of smoothness and of adherence to data. We describe the usual test of smoothness in Section 5.

The tests of adherence to data have much in common with the statistical tests of an experience against a standard table, and we will consider these together later in this chapter. They rely on the assumption that the “true” parameters of the underlying probability model are the graduated estimates $\{\hat{q}_x\}$ or $\{\hat{\mu}_{x+\frac{1}{2}}\}$.

When comparing observed data against a standard table, our null hypothesis is:

H_0 : the true underlying mortality rates at each age x for the experience are the rates in the standard table (ie the observed data come from a population that experiences mortality at the graduated rates).

When testing the adherence of a graduation to the observed data, our null hypothesis becomes:

H_0 : the true underlying mortality rates at each age x for the experience are the graduated rates.

That is, for example, we replace the assumption:

$$D_x \sim \text{Normal} \left(E_x^c \mu_{x+\frac{1}{2}}^s, E_x^c \mu_{x+\frac{1}{2}}^s \right)$$

with the assumption:

$$D_x \sim \text{Normal} \left(E_x^c \hat{\mu}_{x+\frac{1}{2}}, E_x^c \hat{\mu}_{x+\frac{1}{2}} \right)$$

and then proceed to test the statistical hypothesis (almost) as before.



Question 12.5

How does the assumption change for the binomial model?

4.3 Suitability for the purpose in hand

The suitability of a graduation for practical work depends very much on what that work is, and can only be assessed in particular cases. However, two very important observations are:

- (a) In life insurance work, losses result from premature deaths (benefits are paid sooner than expected) so we must not *underestimate* mortality.

In the case of term assurance policies, the insurance company will pay a benefit in respect of policyholders who die within the specified term. If we were to underestimate the mortality rates when calculating the premiums to charge, the insurance company would make a loss – the premiums would be insufficient to cover the benefits paid.

- (b) In pensions or annuity work, losses result from delayed deaths (benefits are paid for longer than expected) so we must not *overestimate* mortality.

When an individual buys an annuity, the insurance company agrees to provide a regular income for the remaining lifetime of that individual. The company will provide a higher income for an individual with a lower expected lifetime, *i.e.* a higher rate of mortality. To limit potential losses, the company should err on the low side when determining the appropriate rates of mortality to use.

4.4 Two examples of graduation

The examples in this chapter are based on Graduation A and Graduation B, which were shown graphically in Section 4.1. The data for these graduations are shown in the tables below. The underlying model for the data is the binomial model.

The left hand part of each table shows the crude data (d_x , E_x and \hat{q}_x). The centre column shows the graduated rates (\mathring{q}_x). The right hand columns show some quantities we will use in the graduation tests, which we will cover in detail later.

The details of each graduation will be explained in Chapter 13, *Methods of graduation*.

Graduation A

x	E_x	d_x (A)	\hat{q}_x	$\overset{\circ}{q}_x$	$E_x \overset{\circ}{q}_x$ (E)	$\frac{(A-E)^2}{E}$	$E_x \overset{\circ}{q}_x (1 - \overset{\circ}{q}_x)$	z_x
30	70,000	39	0.000557	0.000387	27.09	5.24	27.08	2.29
31	66,672	43	0.000645	0.000428	28.54	7.33	28.53	2.71
32	68,375	34	0.000497	0.000473	32.34	0.09	32.32	0.29
33	65,420	31	0.000474	0.000523	34.21	0.30	34.19	-0.55
34	61,779	23	0.000372	0.000579	35.77	4.56	35.75	-2.14
35	66,091	50	0.000757	0.000640	42.30	1.40	42.27	1.18
36	68,514	48	0.000701	0.000707	48.44	0.00	48.41	-0.06
37	69,560	43	0.000618	0.000782	54.40	2.39	54.36	-1.55
38	65,000	48	0.000738	0.000864	56.16	1.19	56.11	-1.09
39	66,279	47	0.000709	0.000955	63.30	4.20	63.24	-2.05
40	67,300	62	0.000921	0.001056	71.07	1.16	70.99	-1.08
41	65,368	63	0.000964	0.001167	76.28	2.31	76.19	-1.52
42	65,391	84	0.001285	0.001290	84.35	0.00	84.24	-0.04
43	62,917	86	0.001367	0.001426	89.72	0.15	89.59	-0.39
44	66,537	120	0.001804	0.001576	104.86	2.19	104.69	1.48
45	62,302	121	0.001942	0.001742	108.53	1.43	108.34	1.20
46	62,145	122	0.001963	0.001926	119.69	0.04	119.46	0.21
47	63,856	162	0.002537	0.002129	135.95	4.99	135.66	2.24
48	61,097	151	0.002471	0.002353	143.76	0.36	143.42	0.60
49	61,110	184	0.003011	0.002600	158.89	3.97	158.48	1.99
Total		1,561			1,515.65	43.30	1,513.32	

Graduation A assumed that $\log(q_x/p_x)$ satisfied a linear equation $\alpha + \beta x$ (2 parameters), which was fitted using the method of least squares.

Graduation B

x	E_x	d_x (A)	\hat{q}_x	$\overset{\circ}{q}_x$	$E_x \overset{\circ}{q}_x$ (E)	$\frac{(A-E)^2}{E}$	$E_x \overset{\circ}{q}_x (1 - \overset{\circ}{q}_x)$	z_x
30	70,000	39	0.000557	0.000460	32.20	1.44	32.19	1.20
31	66,672	43	0.000645	0.000508	33.87	2.46	33.85	1.57
32	68,375	34	0.000497	0.000548	37.47	0.32	37.45	-0.57
33	65,420	31	0.000474	0.000578	37.81	1.23	37.79	-1.11
34	61,779	23	0.000372	0.000600	37.07	5.34	37.05	-2.31
35	66,091	50	0.000757	0.000616	40.71	2.12	40.68	1.46
36	68,514	48	0.000701	0.000632	43.30	0.51	43.27	0.71
37	69,560	43	0.000618	0.000654	45.49	0.14	45.46	-0.37
38	65,000	48	0.000738	0.000693	45.04	0.19	45.01	0.44
39	66,279	47	0.000709	0.000761	50.44	0.23	50.40	-0.48
40	67,300	62	0.000921	0.000867	58.35	0.23	58.30	0.48
41	65,368	63	0.000964	0.001018	66.54	0.19	66.47	-0.43
42	65,391	84	0.001285	0.001215	79.45	0.26	79.35	0.51
43	62,917	86	0.001367	0.001494	94.00	0.68	93.86	-0.83
44	66,537	120	0.001804	0.001701	113.18	0.41	112.99	0.64
45	62,302	121	0.001942	0.001945	121.18	0.00	120.94	-0.02
46	62,145	122	0.001963	0.002155	133.92	1.06	133.63	-1.03
47	63,856	162	0.002537	0.002332	148.91	1.15	148.56	1.07
48	61,097	151	0.002471	0.002545	155.49	0.13	155.09	-0.36
49	61,110	184	0.003011	0.003002	183.45	0.00	182.90	0.04
Total		1,561			1,557.87	18.09	1,555.24	

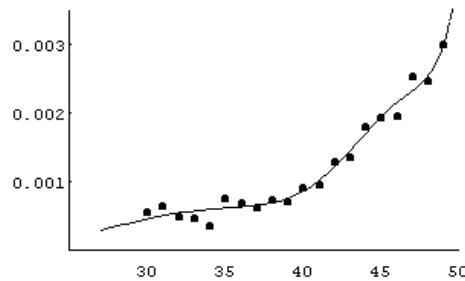
Graduation B assumed that q_x satisfied a polynomial of degree 15 (16 parameters), which was fitted using the method of least squares.

5 Testing the smoothness of a graduation

5.1 What is a smooth graduation?

Mathematical smoothness is usually defined in terms of differentiability, but this is of little use in graduation work because many functions that misbehave wildly between integer ages are nevertheless differentiable many times.

The test for smoothness will be used as a check for undergraduation. It is possible to fit a high-order polynomial to any set of observed data, as in the graph below. The fitted polynomial is smooth in the mathematical sense, *ie* it is differentiable many times, but it does not progress smoothly from age to age.



Instead, we seek a more rough-and-ready measure of smoothness having regard to the scale with which we work (usually the year of age).

To test smoothness, we need to calculate the *third differences* of the graduated quantities $\{\overset{\circ}{q}_x\}$ or $\{\overset{\circ}{\mu}_{x+\frac{1}{2}}\}$. For example:

The first difference $\Delta \overset{\circ}{q}_x = \overset{\circ}{q}_{x+1} - \overset{\circ}{q}_x$.

The second difference $\Delta^2 \overset{\circ}{q}_x = \Delta \overset{\circ}{q}_{x+1} - \Delta \overset{\circ}{q}_x$.

The third difference $\Delta^3 \overset{\circ}{q}_x = \Delta^2 \overset{\circ}{q}_{x+1} - \Delta^2 \overset{\circ}{q}_x$.

The third differences measure the change in curvature.

The criterion of smoothness usually used is that the third differences of the graduated quantities $\{\overset{\circ}{q}_x\}$ or $\{\overset{\circ}{\mu}_{x+\frac{1}{2}}\}$ should:

- (a) **be small in magnitude compared with the quantities themselves; and**
- (b) **progress regularly.**

How to judge if this criterion is met takes some practice. However, since most methods of graduation now in use automatically give smooth results, this is not of great importance. Of the three methods discussed in Chapter 13, only the graphical method presents difficulties in achieving smoothness.

You may like to review this section after reading Chapter 13, *Methods of graduation*.

5.2 A numerical example

In this section we show how to carry out the smoothness test.



Example

Compare the smoothness of the rates in Graduations A and B over the age range 40 to 45.

Solution

Finding the third differences by calculating successive differences (*i.e.* $\Delta \overset{\circ}{q}_x = \overset{\circ}{q}_{x+1} - \overset{\circ}{q}_x$, $\Delta^2 \overset{\circ}{q}_x = \Delta \overset{\circ}{q}_{x+1} - \Delta \overset{\circ}{q}_x$ and $\Delta^3 \overset{\circ}{q}_x = \Delta^2 \overset{\circ}{q}_{x+1} - \Delta^2 \overset{\circ}{q}_x$) gives:

Graduation A

x	$\overset{\circ}{q}_x$	$\Delta \overset{\circ}{q}_x$	$\Delta^2 \overset{\circ}{q}_x$	$\Delta^3 \overset{\circ}{q}_x$
40	0.001056	0.000111	0.000012	0.000001
41	0.001167	0.000123	0.000013	0.000001
42	0.001290	0.000136	0.000014	0.000002
43	0.001426	0.000150	0.000016	
44	0.001576	0.000166		
45	0.001742			

For this graduation, the third differences are very small (in fact, as small as rounding permits), which indicates that the graduated rates are very smooth.

Graduation B

x	$\overset{\circ}{q}_x$	$\Delta \overset{\circ}{q}_x$	$\Delta^2 \overset{\circ}{q}_x$	$\Delta^3 \overset{\circ}{q}_x$
40	0.000867	0.000151	0.000046	0.000036
41	0.001018	0.000197	0.000082	-0.000154
42	0.001215	0.000279	-0.000072	0.000109
43	0.001494	0.000207	0.000037	
44	0.001701	0.000244		
45	0.001945			

For Graduation B, the third differences are relatively large (with a total magnitude of 0.000299), which indicates that the rates are irregular.

6 Statistics refresher



Note

In Section 7 we will look at the statistical tests to assess the reasonableness of a graduation. This section gives a brief review of the background to statistical tests. Students who are familiar with this material (which is all covered in Subject CT3) may wish to skip this section.

6.1 Statistical tests

Hypotheses

Statistical tests assess the plausibility of a particular *null hypothesis* in relation to an *alternative hypothesis*. The null hypothesis (denoted by H_0) corresponds to a neutral conclusion. In graduation tests, the null hypothesis will correspond to a statement that some aspect of a proposed graduation is “OK”. The alternative hypothesis (denoted by H_1) corresponds to a definite conclusion. In graduation tests, the alternative hypothesis will correspond to a statement that some aspect of a proposed graduation is “no good”.

Test process

In statistical tests, we start by first assuming that the null hypothesis is correct. A *test statistic* is then calculated from the data, on that assumption. Using statistical theory, the distribution of the values that might be obtained from the test statistic, assuming that the null hypothesis is correct, can be determined. If it turns out that the value actually obtained for the test statistic is one that would be very unlikely if the null hypothesis were correct, then we conclude that the null hypothesis is not plausible, and we reject it in favour of the alternative hypothesis.

Probability value

In order to decide what can be considered as “very unlikely”, a *significance level* must be selected at the beginning of the test. The significance level usually used is 5%, which means that if H_0 was true the value of the test statistic would only be this extreme by chance 1 time in 20. Using a significance level of 1% would give a stricter test in the sense that we would require a more extreme result to indicate that the null hypothesis is not valid.

The *probability value* (or *p-value*) is the probability, calculated assuming H_0 is true, of obtaining a value of the test statistic as extreme as the actual value obtained. If the probability value is smaller than the significance level chosen, then we reject the null hypothesis. A smaller probability value indicates a more definite (“significant”) result.

A statistical test may be *one-tailed* or *two-tailed*, depending on the nature of the test and the feature we are interested in:

One-tailed tests

In a one-tailed (or one-sided) test, we will be suspicious about a test value that is unusually extreme in one direction only. For instance, a very high test value might worry us whereas a very low value would not.

For example, if our null hypothesis was $H_0 : P(\text{heads on a coin}) < 0.6$, then we would use a one-tailed test because a high number of heads would cast doubt on the null hypothesis but a low number would not.

Two-tailed tests

In a two-tailed (or two-sided) test, we will be suspicious about a test value that is unusually extreme in either direction, *ie* either very high or very low.

For example, if our null hypothesis was $H_0 : P(\text{heads on a coin}) = 0.5$, then we would use a two-tailed test because a low or high number of heads would cast doubt on the validity of the null hypothesis.

Conclusions

The test will result in one of two outcomes:

1. The probability value (*eg* 2%) is lower than the significance level (*eg* 5%). In this case, we conclude that the test provides sufficient evidence for us to reject the null hypothesis.
2. The probability value (*eg* 12%) is not lower than the significance level (*eg* 5%). In this case, we conclude that the test did not provide sufficient evidence for us to reject the null hypothesis.


Question 12.6

Consider the test:

H_0 : Smoking has no effect on mortality, *versus*

H_1 : Smoking increases mortality

- (a) Is this a one-sided or two-sided test?
- (b) What are the possible conclusions of this test?

6.2 Continuity correction

We often use a continuous distribution as an approximation to a discrete distribution, *eg* the normal distribution as an approximation to the binomial. When we do so, we must be careful to take into account the fact that the discrete distribution can only take integer values, whilst the continuous distribution can take any value.

To ensure that the approximation is acceptable, we estimate the probability of observing a particular integer value under the discrete distribution (*ie* $X = x$) by calculating the probability of the continuous distribution being in the range $[x - \frac{1}{2}, x + \frac{1}{2}]$. This is the *continuity correction*.

For example, if we toss a fair coin 20 times, the number of heads has a $\text{Binomial}(20, \frac{1}{2})$ distribution. Under the central limit theorem, the number of heads (X) will have an approximate normal distribution with mean $20 \times \frac{1}{2} = 10$ and variance $20 \times \frac{1}{2} \times \frac{1}{2} = 5$. Using the continuity correction:

$$\begin{aligned} P(10 \text{ heads}) &= P(9.5 \leq X \leq 10.5) \\ &= P\left(-\frac{0.5}{\sqrt{5}} \leq Z \leq \frac{0.5}{\sqrt{5}}\right) \approx \Phi(0.2236) - \Phi(-0.2236) = 0.177 \end{aligned}$$

and:

$$\begin{aligned} P(16 \text{ heads or more}) &= P(15.5 \leq X) \\ &= P\left(\frac{5.5}{\sqrt{5}} \leq Z\right) \approx 1 - \Phi(2.4597) = 1 - 0.993 = 0.007 \end{aligned}$$

6.3 Chi square tests

Purpose

A chi square test can be used to decide whether the observed numbers of individuals who fall into specified categories are consistent with a model that predicts the expected numbers in each category. It is a test for overall *goodness of fit*.

For example, the following categories might be used:

Dead/alive at each age: This would enable us to test whether the observed numbers of deaths and survivors at each age are consistent with the numbers predicted by the probabilities q_x for a particular graduation.

Cause of death: If the deaths within a population have been classified by cause of death, this would enable us to test whether the numbers dying from each cause are consistent with the numbers predicted from an assumed set of proportions.

Rationale

The chi square test is based on the statistic $\chi^2 = \sum \frac{(A-E)^2}{E}$, where A is the *actual number* observed in a particular category and E is the *expected number* predicted by the assumed probabilities and the sum is over *all possible categories*. Each term in the sum represents the square of the discrepancy between the actual and expected values for one group (with an appropriate weighting factor applied). A high value for the total indicates that the overall discrepancy is quite large and would lead us to reject the model. A low value indicates that the observed data fit the model well.

In some cases, the model assumed in the null hypothesis doesn't specify the precise probabilities, but just gives a general formula or a family of distributions. In such cases, it will be necessary to estimate any unknown parameters to calculate the E 's.

Chi square distribution

The theory of multinomial distributions tells us that, in situations where a large number of individuals can be allocated to different categories based on fixed (but unknown) probabilities, this statistic has a *chi square distribution* (very nearly), which is tabulated in the statistics section of the *Formulae and Tables*.

Degrees of freedom

The chi square distribution has one parameter, called the number of *degrees of freedom* (DF), which can take the values 1,2,3,...,. This parameter reflects the amount of “play” present in the system. The correct number of degrees of freedom to use in a chi square test depends on the number of constraints that restrict the way individuals can be allocated to the different categories.

To determine the number of degrees of freedom:

1. Start with the number of groups. (Each combined group counts as one group. See below.)
2. If the groups form a set of mutually exclusive and exhaustive categories (so that their probabilities must add up to 1) or the expected numbers for each category were determined based on the total number for all groups, then subtract 1.
3. Subtract a further 1 for each parameter that has been estimated.

Small groups

The chi square distribution provides a good approximation provided the numbers in each group are not too small. If the expected number in a group is small (less than 5 say), a difference of just one person can make a big difference to the value of $\Delta^3 q_x$ and the approximation becomes unreliable. This problem can be overcome by combining the expected and actual numbers in small groups.



Example

A study of causes of death in elderly men in the 1970s showed the proportions given in the table below. Carry out a chi square test to determine whether these percentages can still be considered to provide an accurate description of causes of death in 2003.

Cause of death	Proportion of deaths in 1975	Number of deaths in 2003
Cancer	8%	286
Heart disease	22%	805
Other circulatory disease	40%	1,548
Respiratory diseases	19%	755
Other causes	11%	464

Solution

The total number of deaths is $286 + 805 + 1,548 + 755 + 464 = 3,858$.

We can calculate the expected numbers of deaths from each cause by applying the proportions to this total. For example, the expected number of cancer deaths is $0.08 \times 3,858 = 308.64$. The figures are given in the table below.

Cause of death	Actual <i>A</i>	Expected <i>E</i>	$\frac{(A - E)^2}{E}$
Cancer	286	308.64	1.661
Heart disease	805	848.76	2.256
Other circulatory disease	1,548	1,543.20	0.015
Respiratory diseases	755	733.02	0.659
Other causes	464	424.38	3.699
Total	3,858	3,858.00	$\chi^2 = 8.290$

Here, we have 5 categories. We haven't estimated any parameters, but we have calculated the expected numbers by assuming that the total is the same as for the actual numbers. So the number of degrees of freedom to use is $5 - 1 = 4$.

From the *Formulae and Tables* (p169), the upper 5% point of the χ^2_4 distribution is 9.488. Our observed value is less than this. So we don't have sufficient evidence to conclude that there has been a change in the pattern of causes of death.



Question 12.7

The mortality rates for a population for the age range 30-34 were estimated by fitting a straight line $\alpha + \beta x$ to the crude values of $\log_e(q_x/p_x)$. Test whether this model (with estimated parameter values of $\alpha = -10.9446$ and $\beta = 0.110404$) can be considered to give a good fit to the data shown in the table below for 2003.

Age x	30	31	32	33	34
Number of deaths in 2003	335	391	428	436	458

The initial exposed to risk in 2003 was approximately 700,000 at each age.

7 Statistical tests of a mortality experience

Here we describe some statistical tests based on the hypothesis that:

- (a) the numbers of deaths at different ages are independent
- (b) In the case where we are comparing the experience with a standard table:

$$D_x \sim \text{Normal} \left(E_x^c \mu_{x+\frac{1}{2}}^s, E_x^c \mu_{x+\frac{1}{2}}^s \right) \text{ (Poisson)}$$

$$D_x \sim \text{Binomial} \left(E_x, q_x^s \right) \quad \text{(Binomial)}$$

- (c) In the case where we are testing the adherence to data of a graduation:

$$D_x \sim \text{Normal} \left(E_x^c \hat{\mu}_{x+\frac{1}{2}}, E_x^c \hat{\mu}_{x+\frac{1}{2}} \right) \text{ (Poisson)}$$

$$D_x \sim \text{Binomial} \left(E_x, \hat{q}_x \right) \quad \text{(Binomial)}$$

In parts (b) and (c) of the hypothesis, the Normal distribution is used as an approximation to the Poisson distributions with small intensity and large exposure, and as an approximation to the Binomial distribution (which is acceptable if $E_x q_x^s$ or $E_x \hat{q}_x > 5$).

Many of the tests that we will describe can be based on the standardised deviations, which we now define.



Definition

- (a) In the Poisson or multiple-state models, the **deviation** at age x is defined to be:

Actual deaths – Expected deaths

$$= d_x - E_x^c \mu_{x+\frac{1}{2}}^s \quad \text{or} \quad d_x - E_x^c \overset{\circ}{\mu}_{x+\frac{1}{2}}$$

and the **standardised deviation**, denoted z_x is:

$$z_x = \frac{d_x - E_x^c \mu_{x+\frac{1}{2}}^s}{\sqrt{E_x^c \mu_{x+\frac{1}{2}}^s}} \quad \text{or} \quad \frac{d_x - E_x^c \overset{\circ}{\mu}_{x+\frac{1}{2}}}{\sqrt{E_x^c \overset{\circ}{\mu}_{x+\frac{1}{2}}}}$$

- (b) In the Binomial model, the **deviation** at age x is likewise defined to be:

Actual deaths – Expected deaths

$$= d_x - E_x q_x^s \quad \text{or} \quad d_x - E_x \overset{\circ}{q}_x$$

and the **standardised deviation**, denoted z_x is:

$$z_x = \frac{d_x - E_x q_x^s}{\sqrt{E_x q_x^s (1 - q_x^s)}} \quad \text{or} \quad \frac{d_x - E_x \overset{\circ}{q}_x}{\sqrt{E_x \overset{\circ}{q}_x (1 - \overset{\circ}{q}_x)}}$$

The z_x 's are often referred to as *individual* standardised deviations to distinguish them from *cumulative* deviations, which we will meet shortly.

Note that if $q \approx 0$ then $(1 - q) \approx 1$ and we can use the approximations:

$$z_x \approx \frac{d_x - E_x q_x^s}{\sqrt{E_x q_x^s}} \quad \text{or} \quad \frac{d_x - E_x \overset{\circ}{q}_x}{\sqrt{E_x \overset{\circ}{q}_x}}$$

Under the assumption that there is a sufficient number of (independent) lives at each age x , we can replace all our hypotheses, under all our models, with the following, by virtue of the Central Limit Theorem:

- (a) $z_x \sim \text{Normal}(0, 1)$ $x = x_1, x_2, \dots, x_m$
- (b) The z_x 's at different ages are mutually independent.



Question 12.8

Verify the figure shown in the table for the standardised deviation at age 30 in Graduation B.

7.1 Chi square (χ^2) test

The first test we describe is the χ^2 -test. Unfortunately, this is the one test where we must pay attention to whether we are comparing an experience with a standard table, or testing the adherence to data of a graduation.

In either case, the test statistic is the same.

As we shall see, the difference relates to the appropriate number of degrees of freedom to use.

Purpose

To test whether the observed numbers of deaths at each age are consistent with a particular set of graduated mortality rates or a particular graduation formula. The chi square test will indicate overall goodness of fit.

Rationale

The chi square test can be applied to the numbers of deaths/survivors at each age (or in age groups).

A high value of the chi square statistic indicates that the discrepancies between the observed numbers and those predicted by the graduated table are large, *i.e.* the fit is not very good. This may be because of overgraduation.

Assumptions

1. There is no heterogeneity of mortality (*ie* no variation in the mortality rates) within each age group and lives are independent.
2. The expected numbers of deaths are high enough (usually at least 5 in each cell) for the chi square approximation to be valid.

Method

Step 1

Combine any small groups by pooling the actual and expected deaths, so that the expected number of deaths is never less than 5.

Step 2

Calculate the test statistic:

$$X = \sum_{\substack{\text{all ages} \\ x}} z_x^2$$

X is called the χ^2 -statistic.

Step 3

Determine the appropriate number of degrees of freedom and compare the observed value of the test statistic with the appropriate percentage point of the chi square distribution given on Page 169 of the *Tables*.

- (a) If we are comparing an experience with a standard table, then X can be assumed to have a χ^2 distribution with m degrees of freedom. (m is just the number of age groups in our notation.) Large values of X indicate excessive deviations, so we will test X against the upper 5% percentage point of the χ_m^2 distribution, and say that the test fails if $X > \chi_m^2; 0.95$.

For example, suppose we are comparing the mortality of a population that is divided into 30 age groups with a standard table. Then the critical value for the chi-squared test is the upper 5% point of χ_{30}^2 , *ie* 43.77.

- (b) If we are testing the adherence to data of a graduation, X can be assumed to have a χ^2 distribution, but with fewer than m degrees of freedom. How many fewer depends on the method of graduation, so we defer further comment to Chapter 13.

Once you have read Chapter 13, you may wish to read this section again.

Conclusion

If the chi square statistic exceeds the upper 95% point, this indicates a poor fit or overgraduation. The contributions to the chi square statistic from each term in the sum can be used to identify the ages where the fit is worst.

Note that if the chi square statistic is very low, this may indicate undergraduation. However, if we suspect undergraduation is a problem we will usually test for it in other ways (eg by considering the standardised deviations or applying the smoothness test described in Section 5).

Strengths and weaknesses

The chi square test is a good test for overall goodness of fit.

The χ^2 -test will fail to detect several defects that could be of considerable financial importance. (These comments apply particularly when we are testing a graduation, and for ease of exposition we will write as if that were the case.)

- (a) There could be a few large deviations offset by a lot of very small deviations. In other words, the χ^2 -test could be satisfied although the data do not satisfy the distributional assumptions that underlie it. This is, in essence, because the χ^2 -statistic summarises a lot of information in a single figure.
- (b) The graduation might be biased above or below the data by a small amount. The χ^2 -statistic can often fail to detect consistent bias if it is small, but we should still wish to avoid it for the reasons given in Section 4.3.
- (c) Even if the graduation is not biased as a whole, there could be significant groups of consecutive ages (called *runs* or *clumps*) over which it is biased up or down. This is still to be avoided.

It should be noted that because the χ^2 -test is based on squared deviations, it tells us nothing about the direction of any bias or the nature of any lack of adherence to data of a graduation, even if the bias is large or the lack of adherence manifest. To ascertain this there is no substitute for an inspection of the experience.

Accordingly, we devise tests that will detect these defects (at least, will do so better than does the χ^2 -test).

These tests are described in the following sections.



Example

Apply the chi square test to Graduation A.

Solution

The actual numbers of deaths A correspond to the values of d_x . The expected numbers of deaths E predicted by the graduation are calculated as $E_x \overset{\circ}{q}_x$.

The chi square statistic can then be calculated:

$$\chi^2 = \sum \frac{(A - E)^2}{E} = 5.24 + 7.33 + \dots + 3.97 = 43.30$$

In this example, it is not difficult to work out how many degrees of freedom to use. There are 20 ages. We have not constrained the totals. The graduated rates have been calculated by estimating 2 parameters. So, the number of degrees of freedom is $20 - 2 = 18$.

From the *Tables*, the upper 95% point for the χ^2_{18} distribution is 28.87. The observed value of the test statistic exceeds this, so we reject the null hypothesis. (In fact, the test statistic also exceeds 37.16, the upper 99.5% point.)

So, we conclude that the mortality experience does not conform to a formula of the type assumed in the graduation.



Question 12.9

Apply the chi square test to Graduation B.

7.2 Standardised deviations test

Purpose

We can use the **standardised deviations test** to look for the first defect of the chi-square test.

The defect to which the Core Reading is referring, is the failure to detect a number of excessively large deviations. This test can detect overall goodness of fit. Where the test reveals a problem this may be due to under/overgraduation, heterogeneity or duplicates. (The problem of duplicates is covered in Chapter 13.)

Rationale

The test looks at the distribution of the values of the standardised deviations.

Under the hypothesis, the z_x 's comprise m independent sample values from a $N(0, 1)$ distribution. This test just tests for that normality.

If the graduated rates are not a good fit, the distribution will not be “centred correctly”. If there is heterogeneity within the age groups or deaths are not independent, the variance will be (respectively) smaller or greater than we would expect if our underlying model were correct. If we have undergraduation, then we would expect the standardised deviations to be tightly bunched. Conversely, if we have overgraduation, we would expect the standardised deviations to be too spread out.

Assumptions

The normal approximation provides a good approximation at all ages.

Method

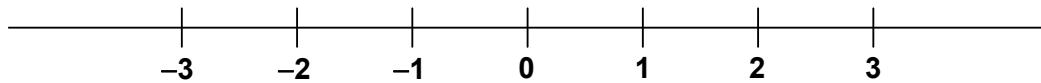
Step 1

Calculate the standardised deviation z_x for each age or age group.

Step 2

Divide the real (number) line into any convenient intervals (the more age groups, the more intervals it might be reasonable to use).

For example:



where the intervals at either end are $(-\infty, -3]$ and $[+3, +\infty)$.

Plot or count the number of standardised deviations falling into each of the ranges.

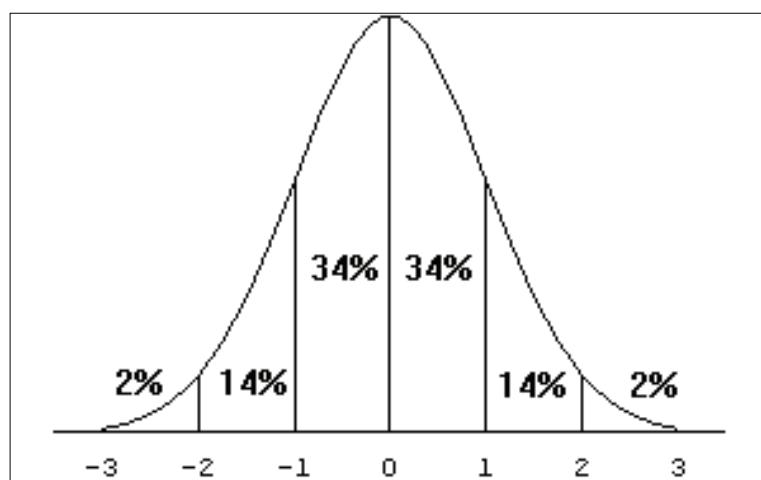
Step 3

We can then compare

- (a) the observed number of the z_x that fall in each interval; and
- (b) the expected number of the z_x that should fall in each interval, under the hypothesis, ie $z_x \sim \text{Normal}(0, 1)$

In this example, the expected numbers are:

Interval	$(-\infty, -3)$	$(-3, -2)$	$(-2, -1)$	$(-1, 0)$	$(0, 1)$	$(1, 2)$	$(2, 3)$	$(3, \infty)$
Expected number	0	0.02m	0.14m	0.34m	0.34m	0.14m	0.02m	0



To formalise the comparison, we can form a χ^2 -statistic (nothing to do with the use of the χ^2 -test mentioned previously):

$$X = \sum_{\text{all intervals}} \frac{(\text{Actual} - \text{Expected})^2}{\text{Expected}}$$

which here should have a χ^2 -distribution with 7 degrees of freedom (since we have used 8 intervals).

Note how this differs from the way we previously applied the χ^2 -test, which was to test whether the observed numbers of deaths were consistent with a given set of graduated rates. Here, we are testing whether the observed pattern of the individual standardised deviations (*i.e.* the numbers falling in each interval) is consistent with a standard normal distribution.

If the number of age groups is small, we should use a smaller number of intervals, ensuring that the expected number of standardised deviations in each interval is not less than five (as a rule of thumb), and we then reduce the number of degrees of freedom in the χ^2 -test appropriately.

If there are only a few age groups, a test must be carried out “by eye”, by considering the following features of the normal distribution:

Overall shape

The number of values in each of the ranges should conform broadly with the percentages for the normal distribution.

Absolute deviations

A related test derives from the fact that, if the z_x 's comprise independent samples from a $N(0,1)$ distribution, half of them should lie in the interval $(-2/3, 2/3)$ – that is, half of them should have an absolute value greater than $2/3$. Thus if we have m ages, the number of standardised deviations exceeding $2/3$ in absolute value is a binomial random variable with parameters m and 0.5.

If there are a lot of values in the tails (*i.e.* the absolute deviations are too big), this indicates overgraduation or the existence of duplicates. (See later.)

In this case a one-tailed test is appropriate, as we usually wish only to identify instances where the number of absolute deviations exceeding $2/3$ is large. We reject the null hypothesis (of no difference between the standard table and the mortality underlying the experience, or of no difference between the graduated rates and the mortality underlying the experience) if this number falls in the upper 5% tail of the $\text{Binomial}(m, 0.5)$ distribution. If $m > 20$ a Normal approximation to the Binomial can be used.

Outliers

In addition to these two tests, we should also look at the values of the individual standardised deviations. If the z_x 's are $N(0,1)$, individual z_x 's with absolute values greater than 1.96 should form at most 1 in 20 of the whole set, and there should be only 1 in 100 with an absolute value greater than 2.57. In practice the number of ages we work with is often quite small, so these rules should be applied with some flexibility. As a guideline, we can say that with fewer than 20 ages we should be suspicious about any individual standardised deviation with a value greater than about 2.0, and regardless of the number of ages we should be concerned about any z_x with an absolute value greater than about 2.5.

Symmetry

There should be roughly equal numbers of positive and negative standardised deviations (since the normal distribution is symmetrical). An excess of positive values indicates that the graduation has introduced a positive bias (*ie* the graduated rates are too low). An excess of negative values indicates that the graduation has introduced a negative bias (*ie* the graduated rates are too high).

Conclusion

If the standardised deviations do not appear to conform to a standard normal distribution, this indicates that the observed mortality rates do not conform to the model with the rates assumed in the graduation. The features considered above will indicate the nature of the discrepancy.

Strengths and weaknesses

Looking at the distribution of the standardised deviations is a good all round test that detects most of the problems that might be present in a graduation.

**Example**

Analyse the distribution of the standardised deviations for Graduation A.

Solution

The actual and expected numbers in each range are shown in the table below.

Interval	$(-\infty, -3)$	$(-3, -2)$	$(-2, -1)$	$(-1, 0)$	$(0, 1)$	$(1, 2)$	$(2, 3)$	$(3, \infty)$
Actual	0	2	4	4	3	4	3	0
Expected	0.0	0.4	2.8	6.8	6.8	2.8	0.4	0

There are only 7 values in the range $(-2/3, 2/3)$. So, there appear to be too few values in the centre of the distribution and too many in the tails *ie* the variance is greater than predicted by the binomial model. This might indicate overgraduation (an inappropriate graduation formula) or the presence of duplicates.

The values are symmetrical (10 positive and 10 negative). So, this shows no evidence of bias in the graduated rates.

If we combine the small groups by pooling the values in the ranges $(-\infty, -1)$ and $(1, \infty)$, we can apply a chi square test to the resulting 4 groups:

$$\chi^2 = \frac{(6-3.2)^2}{3.2} + \frac{(4-6.8)^2}{6.8} + \frac{(3-6.8)^2}{6.8} + \frac{(7-3.2)^2}{3.2} = 10.24$$

This exceeds 7.815, the upper 95% point of the chi square distribution with 3 degrees of freedom, which confirms that the deviations do not conform to a standard normal distribution.

(Note that, strictly speaking, the χ^2 -test should not be used even with this broad grouping since we have $E = 3.2 < 5$.)

**Question 12.10**

Analyse the distribution of the standardised deviations for Graduation B.

7.3 Signs test

Purpose

The signs test is a simple test for overall bias, ie whether the graduated rates are too high or too low.

This test will identify the second deficiency of the chi square test, ie failure to detect where there is an imbalance between positive and negative deviations.

Rationale

If the graduated rates do not tend to be higher or lower than the crude rates on average, we would expect roughly half the graduated values to be above the crude rates and half below. So, if there are m age groups, the number above (or below) should have a $B(m, \frac{1}{2})$ distribution. An excessively high number of positive or negative deviations will indicate that the rates are biased.

This will normally be a two-tailed test, ie we are looking for both positive and negative bias.

Assumptions

None.

Method

Step 1

Count how many of the graduated rates lie above/below the crude rates. We will do this by looking at the signs of the individual standardised deviations. (This can also be done by comparing the crude rates with a graduated mortality curve plotted on a graph or by comparing the numerical values of \hat{q}_x and $\overset{\circ}{q}_x$.)

Step 2

Calculate the probability value for the test by finding the probability of obtaining a split of positive/negative values as extreme as observed.

Define the test statistic:

P = Number of z_x that are positive.

Alternatively, you could base the test on the number of negative z_x 's.

Under the hypothesis, $P \sim \text{binomial}(m, \frac{1}{2})$

So the probability function of P is:

$$P(P = x) = \binom{m}{x} \left(\frac{1}{2}\right)^m, \quad x = 0, 1, \dots, m.$$

An excess of either negative or positive deviations is a defect, so we apply a two-tailed test. We cannot do so exactly, since the binomial distribution is discrete. We could find k^* , defined as the smallest value of k for which

$$\sum_{j=0}^{k^*} \binom{m}{j} \left(\frac{1}{2}\right)^m \geq 0.025$$

i.e. the probability that $P < k^*$ is less than 2.5%. Similarly, the probability that $P > m - k^*$ is less than 2.5%.

The test would be satisfied (at the 5% level) if $k^* \leq P \leq m - k^*$. Or, (perhaps more satisfactorily) we could just find the p -value corresponding to P .

Another way to carry out the test is to calculate the p -value of the test using the probabilities for the binomial distribution given on Pages 186-188 of the *Tables*. However, we can only use the *Tables* if that the sample size is one of those listed ($n = 2, 3, \dots, 12$ or 20).



Example

A graduation covers 20 age groups and has resulted in 6 positive and 14 negative deviations. Carry out a signs test on these data.

Solution

Under the null hypothesis, $P \sim \text{Bin}(20, \frac{1}{2})$. The p -value of the test is:

$$p = 2P(P \leq 6) = 2 \times 0.0577 = 0.1154$$

The probability of 0.0577 can be found on Page 188 of the Tables by looking up $n = 20$, $p = 0.5$, $x = 6$. We multiply this probability by 2 because this is a two-tailed test.

Since the p -value is greater than 5%, there is insufficient evidence to reject the null hypothesis at the 5% significance level.

If the number of age groups is large, we can use the approximation

$$P \sim \text{Normal}(\frac{1}{2}m, \frac{1}{4}m)$$

If we use a normal approximation, then we should use a continuity correction since we would be approximating a discrete (binomial) distribution with a continuous (normal) distribution.

For a two-tailed test, the probability value must be based on the total probability for both tails of the distribution. For example, we can only reject H_0 at the 5% level if the observed number of positives or negatives is greater than the upper 97.5% point or less than the lower 2.5% point of the binomial distribution.

Conclusion

If the test shows that the number of positive values is very high or very low, this indicates that the rates are on average too low or too high (respectively). An examination of the pattern of the signs will indicate the range of ages where the bias is worst.

Strengths and weaknesses

Just looking at the signs of the deviations provides no indication of the extent of the discrepancy. This test is qualitative rather than quantitative.

**Example**

What conclusion can be drawn from an examination of the signs of the deviations for Graduation A and Graduation B?

Solution

When we looked at the distribution of the standardised deviations, there were 10 positive and 10 negative values for both graduations. So, there is no evidence of any overall bias on the basis of this test.

**Question 12.11**

What would your conclusions be from the sign test for the following graduations:

- (a) 13 positive and 27 negative standardised deviations (over the age range 25 to 64)
- (b) 5 positive and 2 negative standardised deviations (5 year age groups over the age range 65 to 99)?

7.4 Cumulative deviations

Purpose

To test whether the overall number of deaths conforms to the model with the mortality rates assumed in the graduation. This test can detect overall goodness of fit. Where the fit is not good this may be due to heterogeneity or duplicates. It addresses the problem of the inability of the chi square test to detect a large positive or negative cumulative deviation over part (or the whole) of the age range.

The cumulative deviations test detects overall bias or long runs of deviations of the same sign.

Rationale

Consider the hypothesis,

$$d_x \sim \text{Normal}(E_x^c \hat{\mu}_{x+\frac{1}{2}}, E_x^c \hat{\mu}_{x+\frac{1}{2}})$$

(taking the multiple-state model as an example).

Here, the deviation has (approximate) distribution:

$$d_x - E_x^c \hat{\mu}_{x+\frac{1}{2}} \sim \text{Normal}(0, E_x^c \hat{\mu}_{x+\frac{1}{2}})$$

So the accumulated deviation, over the whole age range, has distribution:

$$\sum_{\text{all ages}} (d_x - E_x^c \hat{\mu}_{x+\frac{1}{2}}) \sim \text{Normal}\left(0, \sum_{\text{all ages}} E_x^c \hat{\mu}_{x+\frac{1}{2}}\right)$$

and, upon standardising,

$$\frac{\sum_{\text{all ages}} (d_x - E_x^c \hat{\mu}_{x+\frac{1}{2}})}{\sqrt{\sum_{\text{all ages}} E_x^c \hat{\mu}_{x+\frac{1}{2}}}} \sim \text{Normal}(0, 1)$$

This can be tested in the usual way, using a two-tailed test, since either positive or negative deviations are of concern. As well as applying this test to the whole age range, we can apply it to parts of the age range of possible financial significance, provided we choose which sub-ranges to test without reference to the data.

Assumptions

The normal approximation is reasonable at all ages.

Method (illustrated using the Poisson model)

Step 1

Decide which range of ages to test. This might be the whole table or just the range of ages that has the most financial importance. Note that the test is not valid if the age range is selected *after* looking at the pattern of the data. For example, if you spot a blip in one part of the table and decide to apply the cumulative deviation test to that part of the table only, the results would be meaningless.

Step 2

Calculate $\sum(d_x - E_x^c \hat{\mu}_{x+\frac{1}{2}})$ and $\sum E_x^c \hat{\mu}_{x+\frac{1}{2}}$ summed over the selected age range.
(Remember that $E_x^c \hat{\mu}_{x+\frac{1}{2}}$ is just the expected number of deaths for age x .)

Step 3

Calculate the test statistic $\frac{\sum(d_x - E_x^c \hat{\mu}_{x+\frac{1}{2}})}{\sqrt{\sum E_x^c \hat{\mu}_{x+\frac{1}{2}}}}$ and use this to find the probability value

using the tables for the standard normal distribution.



Question 12.12

How would the method differ if the underlying model had been the binomial model?

Conclusion

If the magnitude (*i.e.* the absolute value) of the calculated test statistic is high, this indicates that either:

- the graduated rates are biased (too low if the test statistic is positive, too high if the test statistic is negative), or
- the variance is higher than predicted by the binomial or Poisson model for the range of ages considered. This could be as a result of duplicate policies, which we will discuss in Section 6 of Chapter 13.

The cumulative deviation test can only detect features that are present over the whole age range considered. An excess of positive deviations over one age range may “cancel out” an excess of negatives over another range.

A word of warning: many methods of graduation result in a cumulative deviation of zero as part of the fitting process, in which case this test cannot be applied.



Question 12.13

What conclusion would you draw from applying the cumulative deviation test to the whole age range of Graduation A?



Question 12.14

Why is the cumulative deviation test invalid if the range of ages tested is selected *after* examining the data?

7.5 Grouping of signs test

Purpose

To test for overgraduation.

The grouping of signs test (also called Stevens' test) detects “clumping” of deviations of the same sign. It relies on some simple combinatorics.

Rationale

The test looks at the number of groups (or *runs*) of deviations of the same sign and compares this with the number that would be expected if the positive and negative signs were arranged in random order.

If the graduated rates are overgraduated, the standardised deviations will not swap from positive to negative very often and there will be fewer runs than expected. If the rates are undergraduated, the standardised deviations will swap from positive to negative very often and there will be more runs than expected. However, we do not usually use this test to look for undergraduation. So it is a one-sided test as we are worried about a *low* number of groups.



Question 12.15

Why do you think we don't usually use this test to detect undergraduation?

Define the test statistic:

$$G = \text{Number of groups of positive } z_x \text{'s.}$$

Also, suppose that of the m deviations, n_1 are positive and n_2 are negative.

The hypothesis is that the given n_1 positive deviations and n_2 negative deviations are in random order. We, therefore, compute the probability that the number of positive groups will be at least G given n_1 and n_2 .

Let $t \leq G$.

- (a) There are $\binom{n_2 + 1}{t}$ ways to arrange t positive groups among n_2 negative signs.

There are $(n_2 + 1)$ places in which the t positive groups can be located: before the first negative sign, after the last negative sign or in any of the $(n_2 - 1)$ gaps between the signs.

- (b) There are $\binom{n_1 - 1}{t - 1}$ ways to arrange n_1 positive signs into t positive groups.



Question 12.16

Show that there are $\binom{n_1 - 1}{t - 1}$ ways to arrange n_1 positive signs into t positive groups, assuming that the positive signs occur in a given order.

- (c) There are $\binom{m}{n_1}$ ways to arrange n_1 positive and n_2 negative signs,

since, by definition, $m = n_1 + n_2$.

Hence, the probability of exactly t positive groups is $\frac{\binom{n_1 - 1}{t - 1} \binom{n_2 + 1}{t}}{\binom{m}{n_1}}$.

This formula is given on Page 34 of the *Tables*.

Assumptions

None.

Method

Step 1

Determine the sign of the deviation at each age.

Step 2

Count the number of groups of positive signs ($= G$).

Step 3

Calculate the probability value for the test by finding the probability of obtaining a number of groups as extreme as observed.

Since every pair of positive groups must be separated by a negative group, the numbers of positive and negative groups will be small or large alike, so a one-tailed test is appropriate. We should find the smallest k such that

$$\sum_{t=1}^k \frac{\binom{n_1 - 1}{t - 1} \binom{n_2 + 1}{t}}{\binom{m}{n_1}} \geq 0.05$$

and say that the test has been failed (at the 5% level) if $G < k$.

(Note that the contribution from the $t = 0$ term in the above summation is zero.)

Alternatively, you could look up the critical value of the test on Page 189 of the *Tables*. If the number of groups of positive deviations is less than or equal to the critical value given in the *Tables*, we reject the null hypothesis.



Example

A graduation covers 20 age groups. The number of positive deviations is 6, and the number of groups of positive deviations is 2. Carry out a grouping of signs test using these data.

Solution

From Page 189 of the *Tables*, we see that the critical value is 2 when $n_1 = 6$ and $n_2 = 14$. Since we have observed 2 groups of positive deviations, we reject the null hypothesis at the 5% significance level and conclude that there is evidence of grouping of deviations of the same sign.

However, if m is large enough ($m \geq 20$ or so), we can use a Normal approximation as follows:

$$G \sim \text{Normal} \left(\frac{n_1(n_2 + 1)}{n_1 + n_2}, \frac{(n_1 n_2)^2}{(n_1 + n_2)^3} \right)$$

This result is also given on Page 34 of the *Tables*. Because the test statistic can only take integer values, a continuity correction should be applied when using a normal approximation.

Conclusion

If there are too few runs, this indicates that the rates are overgraduated. The rates do not adhere closely enough to the crude data and may be consistently too high or too low over certain parts of the table.

Strengths and weaknesses

When applying this test, we arbitrarily chose to count the *positive* groups, rather than the *negative* groups. This test can, in some cases, lead to different conclusions depending on whether positive or negative groups are considered.

**Example**

Test Graduation A for overgraduation using the grouping of signs test.

Solution

From the table, we see that there are 3 positive runs and 2 negative runs in the deviations. Assuming the number of positive runs is approximately normally distributed with mean

$$\frac{10(10+1)}{(10+10)} = 5.5 \text{ and variance } \frac{(10 \times 10)^2}{(10+10)^3} = 1.25.$$

There are fewer positive runs than expected. We will therefore find the probability (under the null hypothesis) of having as few as 3 positive runs.

Applying a continuity correction, this corresponds to a value of $(3.5 - 5.5)/\sqrt{1.25} = -1.79$ on the standard normal distribution.

This gives a probability value of 3.7%. So, we can reject the null hypothesis at the 5% level and conclude that the rates are overgraduated.

**Question 12.17**

Apply the grouping of signs test to Graduation B.

7.6 Serial correlations test

Purpose

To test for overgraduation.

The serial correlations test detects grouping of signs of deviations.

It does this by analysing the relationship between the deviations at nearby ages, taking into account the magnitude of the values. The test will address the problem of the inability of the chi square test to detect excessive clumping of deviations of the same sign.

Rationale

If the graduated rates are neither overgraduated nor undergraduated, we would expect the individual standardised deviations at consecutive ages to behave as if they were independent.

However, if the graduated rates are overgraduated, the graduated mortality curve will tend to stay the same side of the crude rates for relatively long periods and, although there will be random variations in the numbers of deaths, we would expect the values of consecutive deviations to have similar values, *ie* they will be positively correlated.

(Conversely, if the rates are undergraduated, the graduated curve will cross the crude rates quite frequently and the values of consecutive deviations will tend to oscillate, *ie* they will be negatively correlated. However we will use this test as a one sided test to test for overgraduation since undergraduation would be tested by means of the smoothness test.)

If correlations are present, we would expect the effect to be strongest at adjacent ages or at ages separated by 2 or 3 years.

Under the null hypothesis, the two sequences (of length $m - 1$):

$$z_1, z_2, \dots, z_{m-2}, z_{m-1}$$

and $z_2, z_3, \dots, z_{m-1}, z_m$

should be uncorrelated.

So should the sequences (of length $m - 2$):

$$z_1, z_2, \dots, z_{m-3}, z_{m-2}$$

and $z_3, z_4, \dots, z_{m-1}, z_m$

We call these the lagged sequences, with lag 1 and lag 2 respectively, and we define sequences with longer lags in the obvious way.

The correlation coefficient of the j th lagged sequences is:

$$r_j = \frac{\sum_{i=1}^{m-j} (z_i - \bar{z}^{(1)})(z_{i+j} - \bar{z}^{(2)})}{\sqrt{\sum_{i=1}^{m-j} (z_i - \bar{z}^{(1)})^2 \sum_{i=1}^{m-j} (z_{i+j} - \bar{z}^{(2)})^2}}$$

$$\text{where } \bar{z}^{(1)} = \frac{1}{m-j} \sum_{i=1}^{m-j} z_i \text{ and } \bar{z}^{(2)} = \frac{1}{m-j} \sum_{i=1}^{m-j} z_{i+j}.$$

This ratio gives the *serial correlation coefficients* r_j , which can take values in the range $-1 \leq r_j \leq 1$. A positive value indicates that nearby values of z_x tend to have similar values, whereas a negative value indicates that they tend to have opposite values.

If m is large enough, we can approximate $\bar{z}^{(1)}$ and $\bar{z}^{(2)}$ by $\bar{z} = \frac{1}{m} \sum_{i=1}^m z_i$, and simplify the above, to obtain:

$$r_j \equiv \frac{\sum_{i=1}^{m-j} (z_i - \bar{z})(z_{i+j} - \bar{z})}{\frac{m-j}{m} \sum_{i=1}^m (z_i - \bar{z})^2} \quad \text{or} \quad r_j \equiv \frac{\frac{1}{m-j} \sum_{i=1}^{m-j} (z_i - \bar{z})(z_{i+j} - \bar{z})}{\frac{1}{m} \sum_{i=1}^m (z_i - \bar{z})^2}$$

The second form shows that this is just the ratio of two averages. This formula for the approximation is listed on Page 34 of the *Tables*.

The difference between this approximation and the exact formula above is negligible for large m . However, the approximate formula may be inappropriate for small values of m . It is generally acceptable in practice, but its limitations should be borne in mind and, in particular, if the test is to be carried out on *real* data where m is small (eg less than 20) then the exact formula should be used.

It is known that $r_j \sim \text{Normal}(0, 1/m)$, under the null hypothesis.

So multiplying r_j by \sqrt{m} should give a value that comes from a standard normal distribution. This value is called the *T ratio*.

Hence, $r_j\sqrt{m}$ (the *T ratio*) can be tested against the **Normal (0,1) distribution. Too high a value indicates a tendency for deviations of the same sign to cluster.**

Assumptions

None.

Method

Step 1

Calculate the standardised deviations z_x for each age or age group.

Step 2

Calculate the serial correlation coefficients using the formula:

$$r_j = \frac{\sum_{i=1}^{m-j} (z_i - \bar{z})(z_{i+j} - \bar{z})}{\frac{m-j}{m} \sum_{i=1}^m (z_i - \bar{z})^2}$$

where $\bar{z} = \frac{1}{m} \sum_{i=1}^m z_i$ is the overall average of z_x for the m ages (or age groups).

Note that you can use the fact that r_j must take values in the range $-1 \leq r_j \leq 1$ to check your calculations for reasonableness.

Step 3

Multiply by \sqrt{m} to find the *T ratio* t_j and compare this with the percentage points of the standard normal distribution.

Conclusion

If the T ratio is “too positive”, this indicates that the rates are overgraduated. The rates do not adhere closely enough to the crude data and may be consistently too high or too low over certain parts of the table.

Strengths and weaknesses

The serial correlation test is a *parametric* test, *ie* it takes into account actual numerical values, whereas the sign test and grouping of signs test are *nonparametric*, since they only look at “how many”.

Because the serial correlation test takes into account the numerical values of the deviations, it is possible for correlations in one part of the age range to be cancelled out by opposite correlations in another part. This means that the signs test and grouping of signs test are usually more powerful, *ie* they are more likely to detect overgraduation if this is present.



Example

Calculate the serial correlation coefficients and the T ratio for Graduation A with lag 1 and interpret your result.

Solution

The mean of the individual standardised deviations is:

$$\bar{z} = \frac{1}{20} \sum_{x=30}^{49} z_x = \frac{1}{20} (2.29 + 2.71 + \dots + 1.99) = 0.19$$

The average of the squared deviation is:

$$\frac{1}{20} \sum_{x=30}^{49} (z_x - \bar{z})^2 = 2.13$$

The average of the cross products is:

$$\frac{1}{19} \sum_{x=30}^{48} (z_x - \bar{z})(z_{x+1} - \bar{z}) = 0.94$$

So:

$$r_1 = 0.94 / 2.13 = 0.44$$

The T ratio is found by multiplying by the square root of the number of ages (*i.e.* 20):

$$t_1 = \sqrt{20} \times 0.44 = 1.97$$

The T ratio is positive, which might suggest the rates are overgraduated. The value of t_1 is just above the upper 2½% point of the normal distribution.



Question 12.18

Calculate the T ratio for Graduation B with lag 1. What does the value indicate about the graduation?

Note: For Graduation B you are given that:

$$\sum(z_x - \bar{z})(z_{x+1} - \bar{z}) = -1.87$$

and:

$$\sum(z_x - \bar{z})^2 = 18.09$$

7.7 Testing actual versus expected rates

We have mainly looked at the tests in the context of testing graduations. However, the tests above can also be used where we wish to test a set of observed rates against an existing table to which we think the rates conform.

The tests will be carried out as previously. In this case though there is no equivalent to undergradaution. We are looking only for goodness of fit.

The statistical tests that assess the significance of differences between observed and expected operate on the basis that the true underlying mortality rates at each age are those specified by the expected mortality basis.

The null hypothesis in this case is:

H_0 : The true underlying mortality rates at each age x for the lives in the investigation are the expected mortality rates (*i.e.* the observed mortality data come from a population that suffers mortality in line with the standard table).

The purpose of each of the various tests in this case is summarised below.

Chi square test

This will be a one sided test for goodness of fit. The null hypothesis will be rejected if the test statistic exceeds the upper 95% level. Note that in this case the number of degrees of freedom will normally just be the number of age groups being considered.

Distribution of ISDs

This test will again be used to examine goodness of fit. In particular it will identify any excessively large deviations.

Signs

This test is used to identify any imbalance between positive and negative deviations, *ie* to ensure that the observed rates are not consistently above or below the expected rates.

Serial correlation

This is a one sided test. A large positive value of the test statistic indicates that the shape of the true rates underlying the observed rates is significantly different from the expected mortality rates at least over part of the range.

Cumulative deviations

This test will detect a large positive or negative cumulative deviation, as previously.

Grouping of signs test

This test detects excessive clumping of deviations of the same sign.

8 Exam-style question

You should now be able to attempt the following past exam question.



Question 12.19

Subject 104, September 2001, Question 8

- (i) Explain why graduated rates, rather than crude estimates of mortality rates are used in the construction of standard mortality tables. [3]
- (ii) A graduation of the mortality experience of the male population of a region of the United Kingdom has been carried out using a graphical method. The following is an extract from the results.

Age	Actual number of deaths	Graduated mortality rate	Initial exposed to risk	
x	θ_x	$\overset{o}{q}_x$	E_x	$\overset{o}{E_x q}_x$
14	3	0.00038	12,800	4.86
15	8	0.00043	15,300	6.58
16	5	0.00048	12,500	6.00
17	14	0.00053	15,000	7.95
18	17	0.00059	16,500	9.74
19	9	0.00066	10,100	6.67
20	15	0.00074	12,800	9.47
21	10	0.00083	13,700	11.37
22	10	0.00093	11,900	11.07
Total	91		73.71	

Use the Chi-squared test to test the adherence of the graduated rates to the data. State clearly the null hypothesis you are testing and comment on the result. [4]

- (iii) Perform two other tests which detect different aspects of the adherence of the graduation to the data. For each test state clearly the features of the graduation which the test is able to detect, and comment on your results. [8]

[Total 15]



Chapter 12 Summary

Purpose of graduation

The crude mortality rates derived from a mortality investigation are graduated to make them acceptable for use in actuarial calculations.

Graduation refers to the process of using statistical techniques to improve the estimates provided by the crude rates. The aims of graduation are:

- to produce a smooth set of rates that are suitable for a particular purpose
- to remove random sampling errors (as far as possible)
- to use the information available from adjacent ages to improve the reliability of the estimates.

Graduation results in a “smoothing” of the crude rates.

Graduated rates should move smoothly between adjacent years of age. This is based on the theoretical assumption that underlying mortality progresses smoothly from year to year and on the practical desire to perform financial calculations (*eg* premiums) that are consistent.

The process of graduation involves a trade-off between smoothness and goodness of fit. The suitability of a graduation can be assessed using statistical tests.

Null hypothesis

The null hypothesis for each of these tests is that the graduated rates are representative of the true underlying mortality rates.

Testing smoothness

Smoothness (undergraduation) can be judged by examining the third differences of the graduated rates.

$$\Delta \ddot{q}_x = \ddot{q}_{x+1} - \ddot{q}_x \quad \Delta^2 \ddot{q}_x = \Delta \ddot{q}_{x+1} - \Delta \ddot{q}_x \quad \Delta^3 \ddot{q}_x = \Delta^2 \ddot{q}_{x+1} - \Delta^2 \ddot{q}_x$$

Testing goodness of fit

Many of the tests of goodness of fit are based on the values of the individual standardised deviations, which provide information about the individual ages or age groups.

Standardised deviations

The individual standardised deviations are given by:

$$z_x = \frac{d_x - E_x^c \overset{\circ}{\mu}_{x+\frac{1}{2}}}{\sqrt{E_x^c \overset{\circ}{\mu}_{x+\frac{1}{2}}}} \quad \text{or} \quad z_x = \frac{d_x - E_x \overset{\circ}{q}_x}{\sqrt{E_x \overset{\circ}{q}_x (1 - \overset{\circ}{q}_x)}}$$

Under the null hypothesis, these approximately normally distributed.

Chi square test

The test statistic is:

$$X = \sum_{\substack{\text{all ages} \\ x}} z_x^2$$

Under the null hypothesis, X has a chi-squared distribution. The number of degrees of freedom depends on the number of ages and the method of graduation.

The chi square test gives an overall assessment of the goodness of fit, but can miss features such as:

- an excess of positive (or negative) deviations over all or part of the table
- clumping of the signs of deviations (which may indicate overgraduation)
- an excessive cumulative deviation over all or part of the table
- outliers balanced by small deviations.

Cumulative deviations test

The test statistic is:

$$\frac{\sum_{\text{all ages}} (d_x - E_x^c \bar{\mu}_{x+\frac{1}{2}})}{\sqrt{\sum_{\text{all ages}} E_x^c \bar{\mu}_{x+\frac{1}{2}}}} \quad (\text{similarly for binomial model})$$

Under the null hypothesis, this should come from a standard normal distribution.

The cumulative deviation test looks at the overall deviation over a range of ages.

Signs test

Under the null hypothesis:

$$\text{Number of +ve signs} \sim \text{Binomial } (n, \frac{1}{2})$$

The signs test looks at the distribution of positive and negative deviations, but it ignores the magnitude of the deviations. The test can be carried out by calculating the p -value. Cumulative probabilities are listed on Pages 186-188 of the *Tables* for certain values of n . Alternatively, if n is large, we can use a normal approximation with a continuity correction.

Grouping of signs test

This is a one-tailed test for which the critical values are given in the *Tables*. It checks for excessive clumping of deviations of the same sign. If n_1 = number of positive signs and n_2 = number of negative signs, then:

$$P(t \text{ positive groups}) = \frac{\binom{n_1-1}{t-1} \binom{n_2+1}{t}}{\binom{n_1+n_2}{n_1}}, \quad n_1 \geq 1, t \geq 1$$

$$\text{Number of positive groups} \sim N\left(\frac{n_1(n_2+1)}{n_1+n_2}, \frac{(n_1n_2)^2}{(n_1+n_2)^3}\right) \text{ approximately}$$

These formulae are given on Page 34 of the *Tables*.

The test can also be carried out by comparing the number of groups of positive deviations with the critical value. Critical values are given on Page 189 of the *Tables*.

Serial correlations test

The approximation formulae are:

$$r_j = \frac{\sum_{i=1}^{m-j} (z_i - \bar{z})(z_{i+j} - \bar{z})}{\frac{m-j}{m} \sum_{i=1}^m (z_i - \bar{z})^2} \quad \text{where } \bar{z} = \frac{1}{m} \sum_{i=1}^m z_i$$

$$t_j \sim \text{Normal}(0,1) \quad \text{where } t_j = r_j \sqrt{m}$$

These are given on Page 34 of the *Tables*.

The serial correlation test looks at the relationship between the deviations at consecutive ages.

Each of these tests concentrates on different features of the graduation. Several tests must be applied before deciding that a set of graduated rates is acceptable. However, an unfavourable result on just one test may be sufficient to reject a set of rates.

We can also use these tests to check whether a set of observed mortality rates conforms to an existing standard table.

Chapter 12 Solutions

Solution 12.1

We covered this point in Chapter 4, *The two-state Markov model*. Here is the relevant section of Core Reading from that chapter:

Having estimated piecewise constant intensities over single years of age, we can use these (if required) to estimate the function μ_x as a smooth function of age (the process of smoothing is called graduation). For this purpose we usually assume that $\hat{\mu}$ estimates $\mu_{x+1/2}$.

Having assumed that μ_x is a constant (μ) over the year of age, our estimate $\hat{\mu}$ will represent the average of μ_x at all ages over the period from x to $x+1$, ie at average age $x + \frac{1}{2}$.

Solution 12.2

- (a) If premiums are too low, the business will be unprofitable. The insurance company may pay out more in claims and maturities than the invested premiums can provide.
- (b) If premiums are too high, the insurance company is likely to be uncompetitive and may lose business. The office may not write enough business to cover its fixed costs and may ultimately need to cease trading.

Solution 12.3

Mortality rates may show some significant changes at certain ages. For example, there is often a marked increase in mortality amongst young males around the age when individuals start to drive cars or ride motorbikes, or start drinking alcohol. Nevertheless, this change is spread over a period of several years.

Solution 12.4

There will still be sampling errors (ie the actual numbers will not be the same as the expected numbers) because the study involves a finite population and a finite time period. Also there are other reasons for graduating crude mortality rates, other than to remove sampling errors.

Solution 12.5

We replace the assumption

$$D_x \sim \text{Normal} \left(E_x q_x^s, E_x q_x^s (1 - q_x^s) \right)$$

with the assumption

$$D_x \sim \text{Normal} \left(E_x \overset{\circ}{q}_x, E_x \overset{\circ}{q}_x (1 - \overset{\circ}{q}_x) \right)$$

Solution 12.6

- (a) This is a one-sided test, since we are only concerned about an *increase* in mortality. The corresponding two-sided test would be:

$$H_0: \text{Smoking has no effect on mortality} \text{ vs } H_1: \text{Smoking affects mortality}$$

Here H_1 includes the possibility that smoking could also *reduce* mortality.

- (b) If the probability value of the calculated test statistic is smaller than the significance level, the conclusion is: “The test provides sufficient evidence to reject the null hypothesis and conclude that smoking increases mortality.”

(Note that this conclusion may not actually be correct. If we were using a significance level of 5%, we would arrive at this conclusion 5% of the time, even if smoking didn’t affect mortality. However, if the probability value is very small *eg* 0.1%, this is so unlikely that nobody would doubt the result.)

If the probability value of the calculated test statistic is greater than the significance level, the conclusion is: “The test does not provide sufficient evidence to reject the null hypothesis that smoking has no effect on mortality.”

(Again, this conclusion may not actually be correct. It may be that our study wasn’t big enough or the test we used wasn’t powerful enough to give a convincing result.)

Solution 12.7

Starting with the equation $\log_e(q_x/p_x) = \alpha + \beta x$, we can take exponentials of both sides to get:

$$\frac{q_x}{p_x} = e^{\alpha + \beta x}$$

Then cross-multiplying:

$$q_x = e^{\alpha + \beta x} (1 - q_x)$$

Collecting the q_x terms on the LHS:

$$q_x (1 + e^{\alpha + \beta x}) = e^{\alpha + \beta x}$$

So:

$$q_x = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} = \frac{1}{e^{-(\alpha + \beta x)} + 1}$$

We can therefore calculate the graduated rates using the formula: $\overset{\circ}{q}_x = \frac{1}{1 + e^{-(\alpha + \beta x)}}$, and using the estimated values for α and β , we can work out the expected numbers of deaths at each age.

Age	$\overset{\circ}{q}_x$	Actual A_x	Expected E_x	$\frac{(A_x - E_x)^2}{E_x}$
30	0.0004842	335	338.94	0.046
31	0.0005407	391	378.49	0.413
32	0.0006038	428	422.66	0.067
33	0.0006742	436	471.94	2.737
34	0.0007529	458	527.03	9.041
Total		2,048	2,139.06	$\chi^2 = 12.304$

Here, we have 5 ages and we haven't imposed any constraints on the totals ($\sum E_x \neq \sum A_x$). However, we have used 2 estimated parameters. So the number of degrees of freedom to use is $5 - 2 = 3$. From the *Tables* (p107), the upper 5% point of the χ^2_3 distribution is 7.815. Our observed value is greater than this. (In fact, it is greater than the upper 1% point.) So we can conclude that this is not a satisfactory model of mortality over this age range.

Solution 12.8

$$z_{30} = \frac{d_{30} - E_{30} \overset{\circ}{q}_{30}}{\sqrt{E_{30} \overset{\circ}{q}_{30} (1 - \overset{\circ}{q}_{30})}} = \frac{39 - 70,000(0.000460)}{\sqrt{70,000 \times 0.000460(1 - 0.000460)}} = \frac{6.80}{5.673} = 1.20$$

Solution 12.9

Referring to the table, the actual numbers of deaths A_x correspond to the values of d_x . The expected numbers of deaths E_x predicted by the graduation are calculated as $E_x \overset{\circ}{q}_x$.

The chi square statistic can then be calculated:

$$\chi^2 = \sum_i \frac{(A_i - E_i)^2}{E_i} = 1.44 + 2.46 + \dots + 0.00 = 18.09$$

There are 20 ages. We have not constrained the totals. The graduated rates have been calculated by estimating 16 parameters. So, the number of degrees of freedom is $20 - 16 = 4$.

From the *Tables*, the upper 95% point for the χ^2_4 distribution is 9.488. The observed value of the test statistic exceeds this, so we reject the null hypothesis. (In fact, the test statistic also exceeds 14.86, the upper 99.5% point.)

So, we conclude that the mortality experience does not conform to a formula of the type assumed in the graduation.

Solution 12.10

Interval	$(-\infty, -3)$	$(-3, -2)$	$(-2, -1)$	$(-1, 0)$	$(0, 1)$	$(1, 2)$	$(2, 3)$	$(3, \infty)$
Actual	0	1	2	7	6	4	0	0
Expected	0.0	0.4	2.8	6.8	6.8	2.8	0.4	0

The actual numbers appear to be very close to the expected numbers. There are 11 values in the range $(-2/3, 2/3)$. So, this provides no evidence that the binomial model is not appropriate. The values are symmetrical (10 positive and 10 negative). So, this shows no evidence of bias in the graduated rates.

If we combine the small groups by pooling the values in the ranges $(-\infty, -1)$ and $(1, \infty)$, we can apply a chi square test to the resulting 4 groups:

$$\chi^2 = \frac{(3-3.2)^2}{3.2} + \frac{(7-6.8)^2}{6.8} + \frac{(6-6.8)^2}{6.8} + \frac{(4-3.2)^2}{3.2} = 0.31$$

This is nowhere near 7.815, the upper 95% point of the chi square distribution with 3 degrees of freedom. So, again, there is no evidence to reject the binomial model or the graduated rates.

Solution 12.11

- (a) There are enough values to use a normal approximation:

$$\begin{aligned} P(\geq 27 \text{ negatives}) &= P[B(40, \frac{1}{2}) \geq 27] \\ &\approx P[N(20, 10) > 26.5] \\ &= P[N(0, 1) > (26.5 - 20)/\sqrt{10}] = 1 - \Phi(2.055) = 0.020 \end{aligned}$$

So, the probability value is 2.0% for a one-sided test and 4.0% for a two-sided test. In either case, we conclude that the sign test provides sufficient evidence to conclude that the graduation is producing rates that are generally higher than the true rates.

- (b) Since there are only 7 values, an exact calculation must be used:

$$\begin{aligned} P(\geq 5 \text{ positives}) &= P[B(7, \frac{1}{2}) \geq 5] \\ &= P[B(7, \frac{1}{2}) = 5] + P[B(7, \frac{1}{2}) = 6] + P[B(7, \frac{1}{2}) = 7] \\ &= \binom{7}{5} \left(\frac{1}{2}\right)^7 + \binom{7}{6} \left(\frac{1}{2}\right)^7 + \binom{7}{7} \left(\frac{1}{2}\right)^7 \\ &= \frac{21 + 7 + 1}{2^7} \\ &= 0.227 \end{aligned}$$

So, the probability value is 0.227 for a one-sided test and 0.454 for a two-sided test. In either case, we conclude that the sign test doesn't provide enough evidence to conclude that the graduation is biased.

Solution 12.12

Step 1 is unchanged.

In Step 2, we will calculate $\sum(d_x - E_x \hat{q}_x)$ and $\sum E_x \hat{q}_x (1 - \hat{q}_x)$ summed over the selected age range.

In Step 3, we will calculate the test statistic $\frac{\sum(d_x - E_x \hat{q}_x)}{\sqrt{\sum E_x \hat{q}_x (1 - \hat{q}_x)}}$ and use this to find the probability value using the tables for the standard normal distribution.

Solution 12.13

From the table:

$$\text{The test statistic} = \frac{\sum_{x=30}^{49} (d_x - E_x \hat{q}_x)}{\sqrt{\sum E_x \hat{q}_x (1 - \hat{q}_x)}} = \frac{1,561 - 1,515.65}{\sqrt{1,513.32}} = 1.166$$

This is a two-tailed test, so we compare the value of the test statistic with the upper and lower 2.5% points of $N(0,1)$, ie ± 1.96 . As $-1.96 < 1.166 < 1.96$, there is insufficient evidence to reject the null hypothesis.

So, the cumulative deviations test does not provide evidence that the graduated rates are biased or the variance is different from the value predicted by the binomial model.

Solution 12.14

In any set of data there will be some “unusual” features eg a number of consecutive positive values. If the age range is selected on the basis that it looked unusual, then the calculated probability value is not valid.

Solution 12.15

Any problems resulting from undergraduation will usually already have been picked up by either the smoothness (third differences) test (if adhering too closely to the crude rates has led to an erratic pattern of rates) or by the chi square test (if adhering too closely to the crude rates over one part of the age range has led to large discrepancies in another part).

We will see in the next chapter that, with some methods of graduation, the nature of the method makes it impossible for the graduated rates to be undergraduated.

Solution 12.16

We can define a “separator” as a boundary marking the end of one group and the start of another group. The problem of splitting the n_1 signs into t groups is equivalent to the problem of placing $t-1$ separators in the n_1-1 gaps between the n_1 signs. (No separator can come before the first or after the last sign.)

There are $\binom{n_1-1}{t-1}$ ways of placing the $t-1$ separators in the n_1-1 gaps, and so this is the number of ways of splitting the n_1 signs into t groups.

Solution 12.17

Since the grouping of signs test is a one-tailed test and we are only testing to see if there are too few positive groups. From the *Tables*, we find that the critical value of the test is 3. The observed value of 8 is greater than this and therefore is not significant.

Solution 12.18

The serial correlation coefficient is:

$$r_1 = \frac{\frac{1}{m-1} \sum (z_x - \bar{z})(z_{x+1} - \bar{z})}{\frac{1}{m} \sum (z_x - \bar{z})^2} = \frac{-1.87/19}{18.09/20} = -0.109$$

The T ratio is found by multiplying by the square root of the number of ages (*i.e* 20):

$$t_1 = \sqrt{20} \times -0.109 = -0.49$$

The T ratio is negative, which might suggest the rates are undergraduated. However, it is not sufficiently extreme when compared with the standard normal distribution for us to be confident of this. The smoothness test could be carried out to check for undergradation.

Solution 12.19(i) ***Reasons for graduation***

We expect the true rates to progress smoothly, with no irregularities.

Graduation reduces sampling errors at each age by using information from adjacent ages.

Standard tables are used for premium and reserve calculations, where it is important to have unbiased estimates of the true underlying rates.

Premiums should vary smoothly with age (as policyholders would expect).

(ii) ***Chi-squared test***

The null hypothesis is:

H_0 : the graduated rates are the true underlying mortality rates for the population

We first calculate the (approximate) individual standardised deviations at each age using the formula:

$$z_x = \frac{\theta_x - E_x \ddot{q}_x}{\sqrt{E_x \ddot{q}_x (1 - \ddot{q}_x)}} \approx \frac{\theta_x - E_x \ddot{q}_x}{\sqrt{E_x \ddot{q}_x}}$$

(The approximation holds because $1 - \ddot{q}_x \approx 1$ for all x since the \ddot{q}_x terms are small.)

The ISDs are:

-0.8437, 0.5536, -0.4082, 2.1457, 2.3263, 0.9022, 1.7970, -0.4063, -0.3216

The test statistic for the chi-squared test is:

$$\sum z_x^2 = 15.51$$

We now compare this with a χ^2 distribution. Since the graduation was carried out graphically, we lose 2 or 3 degrees of freedom for every 10 age groups included in the graduation. We were given data from 9 ages, so we are left with 6 or 7 degrees of freedom.

From the *Tables*, we find that:

- the upper 5% point of χ_6^2 is 12.59
- the upper 5% point of χ_7^2 is 14.07.

As the value of the test statistic exceeds both of these, we reject the null hypothesis and conclude that the graduated rates do not provide a good fit to the data. In particular, it looks like the graduated rates are too low for ages 17 to 20.

(iii) ***Two other tests***

You can take your pick here from the individual standardised deviations test, the signs test, the cumulative deviations test, the grouping of signs test and the serial correlation test.

ISD Test

This is a good all round test that detects most of the problems that might be present in a graduation.

For this test we compare the z_x 's with a standard normal distribution:

	$(-\infty, -3)$	$(-3, -2)$	$(-2, -1)$	$(-1, 0)$	$(0, 1)$	$(1, 2)$	$(2, 3)$	$(3, \infty)$
Obs	0	0	0	4	2	1	2	0
Exp	0.01	0.19	1.22	3.07	3.07	1.22	0.19	0.01

There are 4 things to check for here:

- Outliers – there are no ISDs greater than 3 in absolute value, which is good; however there are more in the (2,3) range than expected.
- The balance of positive and negative deviations, which is OK.
- Symmetry – the distribution is a bit positively skewed, which is not so good.
- Proportion of ISDs lying in the range $(-\frac{2}{3}, \frac{2}{3})$ should be $\frac{1}{2}$ – it is $\frac{4}{9}$ here, which is OK.

The graduated rates fail this test as the ISDs are not normally distributed. In particular, the graduated rates appear to be too low at ages 17 and 18.

Signs test

This is a simple two-tailed test for overall bias.

There should be roughly equal number of positive and negative ISDs. We have 5 positives and 4 negatives, which is OK.

Cumulative deviations test

This is a two-tailed test for overall bias.

The test statistic is:

$$\frac{\sum \theta_x - \sum E_x \hat{q}_x}{\sqrt{\sum E_x \hat{q}_x}} = \frac{90.5 - 73.71}{\sqrt{73.71}} = 1.96$$

For a test at the 5% significance level, we compare the value of the test statistic with the lower and upper 2.5% points of $N(0,1)$, ie with ± 1.96 . So the result of this test is borderline.

If you don't incorporate the continuity correction, you would get a test statistic of 2.014. This would lead you to reject the null hypothesis and conclude that the graduated rates are too low overall.

Grouping of signs test

This is a one-tailed test that detects clumping of deviations of the same sign.

The observed number of positive runs is 2, the observed number of positive deviations is 5, and the observed number of negative deviations is 4.

From the *Tables*, we find that the critical value of the test (ie the smallest number of positive runs that would lead us to reject the null hypothesis) is 1.

So we do not reject the null hypothesis in this case, and we conclude that there is no evidence of grouping of signs.

Serial correlation test

This is an alternative test for grouping of signs, but it takes much longer to carry out this test so we don't recommend that you do it unless you absolutely have to. Make sure that you don't carry out both the grouping of signs test and the lag-1 serial correlation test since they test for the same thing!

The serial correlation coefficient at lag 1 is:

$$r_1 \approx \frac{\frac{1}{8} \sum_{x=1}^8 (z_x - \bar{z})(z_{x+1} - \bar{z})}{\frac{1}{9} \sum_{x=1}^9 (z_x - \bar{z})^2} = \frac{0.2156}{1.3161} = 0.1638$$

and the test statistic is:

$$r_1 \sqrt{m} = 0.1638 \times 3 = 0.491$$

As we are only testing for positive correlation, we compare the value of the test statistic with 1.645, the upper 5% point of $N(0,1)$. We find that there is insufficient evidence to reject the null hypothesis or, in other words, there is no evidence of grouping of signs.

Comment

The graduation has not fully taken into account the accident hump, *i.e* the increase in mortality around the late teens and early twenties.

Chapter 13

Methods of graduation



Syllabus objectives

- (x) *Describe how to test crude estimates for consistency with a standard table or a set of graduated estimates, and describe the process of graduation.*
4. *Describe the process of graduation by the following methods, and state the advantages and disadvantages of each:*
- *parametric formula*
 - *standard table*
 - *graphical.*
- (The student will not be required to carry out a graduation.)*
5. *Describe how the statistical tests (for comparison of crude estimates with a standard table) should be amended to compare crude and graduated sets of estimates.*
6. *Describe how the statistical tests (for comparison of crude estimates with a standard table) should be amended to allow for the presence of duplicate policies.*
7. *Carry out a comparison of a set of crude estimates and a standard table, or a set of crude estimates and a set of graduated estimates.*

0 Introduction

In this chapter, we will look at three methods of carrying out a graduation:

- graduation by parametric formula
- graduation by reference to a standard table
- graphical graduation.

The most appropriate method of graduation to use will depend on the quality of the data available and the purpose for which the graduated rates will be used.

The general methodology of graduation is essentially the same under each method. Once we have decided on the appropriate method, we will choose a model to represent the underlying force (or rates) of mortality, fit the model to the crude observed rates and test the graduation for adherence to data and (if necessary) smoothness. Each method can produce many possible graduations. The graduation chosen will be the one whose adherence and smoothness best meet the requirements for which the rates are intended.

Graduation is a compromise between adherence to data (goodness of fit) and smoothness. The balance that we want between these two conflicting objectives is a subjective choice and will depend on how the graduated rates will be used. For example:

- If we are constructing a standard table of national population mortality, we will be interested in maximising the accuracy. We will put more emphasis on adherence and less emphasis on smoothness.
- If the rates are to be used to calculate premiums and reserves (which we'll discuss in CT5) for a life insurance company, we will want to ensure that the rates (and hence the premiums and reserves) progress smoothly from age to age to avoid sudden changes and inconsistencies. We will put more emphasis on smoothness and less emphasis on adherence. The mortality rates at ages around the accident hump will be less important in this situation as few policyholders are likely to be in the age range 18-22.

You may remember that the precise form of some of the statistical tests that we described in Chapter 12 depends on the method of graduation used. We strongly recommend that you re-read the relevant sections of Chapter 12 after you have worked through this chapter.

1 ***Graduation by parametric formula***

1.1 ***Overview***

The method of graduation most often used for reasonably large experiences is to fit a parametric formula to the crude estimates.

The underlying assumption is that μ_x (or q_x) can be modelled using an appropriate mathematical formula with unknown parameters. The parameters are typically calculated automatically by a computer using numerical methods.

If the formula used does not include enough parameters, it will not be flexible enough to follow the crude rates closely, which may result in overgraduation. If too many parameters are included, sudden bends may appear in the graduated curve, which may result in undergraduation.

For different values of the parameters, we can assess the smoothness and adherence to data of the fitted model. (In practice we will not need to check smoothness if the number of parameters is sufficiently small.)

We will choose the values of the parameters that provide the most appropriate model, according to some pre-defined criterion in respect of goodness of fit.

1.2 ***Choosing and fitting parametric formulae***

Two simple (but useful) formulae are:

$$\text{Gompertz (1825)} \quad \mu_x = Bc^x$$

$$\text{Makeham (1860)} \quad \mu_x = A + Bc^x$$

We described these simple laws of mortality in Chapter 7.

In practice, it is usually found that μ_x follows an exponential curve quite closely over middle and older ages (in human populations) so most successful formulae include a Gompertz term. Makeham's formula is interpreted as the addition of accidental deaths, independent of age, to a Gompertz term representing senescent deaths.

The most recent standard tables produced for use by UK life insurance companies used formulae of the form

$$\mu_x = \text{polynomial}_1 + \exp(\text{polynomial}_2)$$

which includes Gompertz and Makeham as special cases.



Question 13.1

Define polynomial_1 and polynomial_2 for the special case of Makeham's formula.

A wide range of techniques is available to choose and to fit a curve to a set of crude estimates. Here we just describe how the fitting was carried out in the case of the most recent UK life insurance standard tables.

1.3 A practical example – the “92 series” tables

The available data were deaths and central exposed to risk (ie d_x and E_x^c), and the Poisson model was used. The data were collected by life insurance companies during 1991-94, and were analysed by the Continuous Mortality Investigation (CMI). Different tables were prepared for males and females, and for different classes of insurance and pension business. Collectively they are known as the “92 series” tables.

“Classes” of insurance refers to the different types of policy, eg whole life annuity or term assurance, which we'll meet in Subject CT5.

The formulae were of the form:

$$\overset{\circ}{\mu}_x = f(\alpha_1, \alpha_2, \dots, \alpha_r, \alpha_{r+1}, \alpha_{r+2}, \dots, \alpha_{r+s}, x)$$

where

$$\text{polynomial}(1) = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \dots + \alpha_r x^{r-1}$$

$$\text{polynomial}(2) = \alpha_{r+1} + \alpha_{r+2} x + \alpha_{r+3} x^2 + \dots + \alpha_{r+s} x^{s-1}$$

In other words, a formula with $(r+s)$ parameters was fitted for each table of the form $\mu_x = \text{polynomial}_1 + \exp(\text{polynomial}_2)$. You should note that this is not the formula given in the *Tables*. The differences are discussed in Section 4 of Chapter 7.

Under the Poisson model (covered in Chapter 10), the probability of observing d_x deaths from a central exposed-to-risk E_x^c over the age interval $[x, x+1]$ is given by:

$$P[D_x = d_x] = \frac{(\mu_{x+\frac{1}{2}} E_x^c)^{d_x} \exp(-\mu_{x+\frac{1}{2}} E_x^c)}{d_x!}$$

(Note that we're using $\mu_{x+\frac{1}{2}}$ here rather than μ_x because the average age over the age interval is $x + \frac{1}{2}$. We discussed this in Chapter 11.)

Therefore, in respect of the age interval $[x, x+1]$, the likelihood in the Poisson model is:

$$\begin{aligned} & (\mu_{x+\frac{1}{2}})^{d_x} \cdot \exp(-\mu_{x+\frac{1}{2}} E_x^c) \times \text{constants} \\ &= f(\alpha_1, \dots, \alpha_{r+s}, x + \frac{1}{2})^{d_x} \cdot \exp(-f(\alpha_1, \dots, \alpha_{r+s}, x + \frac{1}{2}) E_x^c) \times \text{constants} \end{aligned}$$

So the total likelihood, ignoring constants, is:

$$\prod_{\substack{\text{all ages} \\ x}} f(\alpha_1, \dots, \alpha_{r+s}, x + \frac{1}{2})^{d_x} \cdot \exp(-f(\alpha_1, \dots, \alpha_{r+s}, x + \frac{1}{2}) E_x^c).$$

This likelihood was maximised numerically to obtain maximum likelihood estimates of the parameters $\alpha_1, \alpha_2, \dots, \alpha_{r+s}$, and hence $\hat{\mu}_x$.

Other ways to fit a parametric formula include:

- minimising the χ^2 -statistic $\sum \frac{(A - E)^2}{E}$, where the expected number of deaths is calculated using the fitted model; and
- minimising the value of the weighted least squares, ie $\sum w_x (\hat{\mu}_x - \hat{\mu}_x)^2$, the sum of the squares of the differences between the crude and fitted values of μ_x (or q_x) with a weighting w_x based on the exposed-to-risk at each age.

1.4 Other considerations

Using additional information from other investigations

For practical use, it is not sufficient to choose and fit a formula using statistical methods alone. It is always necessary to inspect the results in the light of previous knowledge of mortality experiences, especially at very young and very old ages where the data may be scarce.



Question 13.2

What experiences may be available to check our results?

The graduated estimates should also be compared with other experiences to see if they behave as we would expect. Examples of the checks that would be applied are:

- (a) The mortality of males is higher than the mortality of females.
- (b) The mortality of persons with life insurance policies is lower than that of the population as a whole.
- (c) The mortality of persons who have recently taken out life insurance is lower than that of persons who took out life insurance a long time ago (because they have to be in good health to obtain life insurance).

It might be necessary to adjust the graduation to obtain a satisfactory final result.

Our observations may suggest a different pattern of mortality to that experienced in previous investigations. In these circumstances, we must decide on the relative levels of reliance that we can place on each source of information. If our study is small, we may be more confident that the previous investigations reflect the true position. If our study is large, we may be more confident that the pattern of underlying mortality is genuinely different from that of previous investigations.



Question 13.3

Why may the mortality of people with life insurance policies be lower than that of the population as a whole?

Financial risks

We should always consider where the financial risks lie.

- (a) **A life insurance contract pays out on death, so the insurance company will charge inadequate premiums if it underestimates mortality.**
- (b) **A pension or annuity contract pays out on continued survival, so the insurance company will charge inadequate premiums if it overestimates mortality.**

So, if an insurance company wishes to protect itself against the risk of charging inadequate premiums, it will try to overestimate mortality for life insurance contracts and underestimate mortality for pension or annuity contracts. At the same time, the insurance company will need to ensure that its margins are not so large as to make the premiums uncompetitive.

Changes in mortality

Since insurance companies will use graduated mortality tables to estimate *future* mortality (under insurance contracts yet to be sold) but investigations must be of *past* mortality, the trend of mortality is important. In most countries mortality has been falling for a long time, which means that past mortality is likely to be on the safe side for insurance business but not adequate for pension or annuity business. In respect of the latter it is necessary to make some projection of future improvements in mortality.

When the CMI published the “80 series” tables (which preceded the “92 series” tables) in the UK, it also provided a formula with which to model future improvements in mortality. The projections can be used in two ways:

- to model the level of mortality that may be experienced by a population as a whole in some future year, *eg* 2020
- to model the mortality that may be experienced by an individual over his or her lifetime by projecting the improvements in mortality at each future year of age.

1.5 ***The graduation process***

The curve-fitting process described above is only one of several stages that must be carried out, often repeatedly, before a satisfactory result is obtained.

Step 1 – select a graduation formula

A particular parametric family of curves must be chosen. For example, the first few (useful) families of the general type used by the CMI are:

$$\alpha_1 \exp(\alpha_2 x) \quad (\text{Gompertz})$$

$$\alpha_1 + \alpha_2 \exp(\alpha_3 x) \quad (\text{Makeham})$$

$$\alpha_1 + \alpha_2 \exp(\alpha_3 x + \alpha_4 x^2)$$

and so on.

Step 2 – determine parameter values

Given a family of curves, the best-fitting values of the parameters must be found. The CMI used maximum likelihood, but there are many other suitable procedures, eg minimising either the χ^2 -statistic or the value of the weighted least squares. This is usually performed on a computer using a statistics package.

Step 3 – calculate graduated rates

Calculate the graduated rates at each age using the fitted parametric formula. This can be done on a computer using a spreadsheet program.

Step 4 – test

Given the best-fitting curve of a given family, the graduated rates must be compared with the original data to see if they are acceptably close, according to some test procedures (see Chapter 12).

Usually this process will be carried out for several families of curves, and the final choice will be influenced by the “goodness of fit”. However, many other factors influence the outcome, and it is not always the best-fitting graduation (in the statistical sense) that gives the most suitable result for practical use.

2 **Graduation by reference to a standard table**

2.1 Overview

The underlying rationale of the method is that, if the class of lives involved in the graduation is sufficiently similar to the class of lives whose experience formed the basis of a particular standard table, then the true underlying mortality of our class of lives should be quite similar to that of the standard table. Even if overall levels of mortality differ between the two, it would still be expected that the overall progression of rates from age to age would be similar.

A “standard table” means a published life table based upon sufficient data to be regarded as reliable (for appropriate applications). Examples include national life tables based on a country’s entire population (eg the English Life Tables) and insured lives tables based on large numbers of insured lives (eg the “92 series” tables).

A standard table will always be based on a well-defined class of lives, although this does not mean that that class of lives will be perfectly homogeneous. If we are given the mortality experience of a similar group of lives, we might reasonably suppose that it should share some of the characteristics of the experience underlying the standard table, such as its overall shape. This is useful if we do not have much data from the experience in which we are interested.

So, we’ll tend to use this method of graduation when we do not have a large amount of data *and* there is a standard table that we think appropriate to the population. We make use of the valuable information provided by the standard table relating to the general shape of mortality. An appropriate simple equation, involving unknown parameters, is selected to reflect the relationship between the mortality rates for the new experience and the graduated rates for the standard table. For example, if we think that the true underlying mortality rates are a linear function of the standard table rates, we may try an equation of the form $\mathring{q}_x = aq_x^S + b$. The graduated rates are then a combination of the shape provided by the standard table and the level of mortality observed in our investigation.

Only if the standard table is sufficiently similar to the underlying experience will a satisfactory fit to the crude estimates be possible.

2.2 The graduation process

Step 1 – select standard table

We select an existing standard mortality table that is believed to have a similar pattern of mortality rates over the age range of interest. The appropriateness of a particular standard table will be assessed by comparing the characteristics of the lives on which it was based and those in the current investigation, *eg* sex, geographical area, period of investigation.

Step 2 – find simple link to a standard table

Let q_x^s or μ_x^s be the rates or forces of mortality of the standard table. Then we try to exploit the assumed similarity of the experiences by seeking a reasonably simple function $f()$ such that $\hat{q}_x = f(q_x^s)$ or $\hat{\mu}_x = f(\mu_x^s)$.

Examples include:

$$\hat{q}_x = a + bq_x^s$$

$$\hat{q}_x = (a + bx)q_x^s$$

$$\hat{\mu}_x = \mu_x^s + k$$

$$\hat{\mu}_x = \mu_{x+k}^s$$

where a , b and k are suitable constants.

The search for a suitable function $f()$ can be aided by making some simple plots:

- (a) a plot of \hat{q}_x against q_x^s might indicate a linear relationship in q_x
- (b) a plot of $-\log(1 - \hat{q}_x)$ against $-\log(1 - q_x^s)$ might indicate a linear relationship in μ_x

and so on.

If it is not possible to find a simple relationship, then the supposition that the experiences have similar characteristics should perhaps be reconsidered. It should be remembered that if data are scarce, too close a fit to any suggested relationship is not to be expected, especially at extreme ages.

Step 3 – determine parameter values

Once a possible relationship has been identified, the best-fitting parameters must be found. Any suitable method might be used, for example:

- (a) maximum likelihood: the underlying model is that

$$q_x = f(\alpha_1, \dots, \alpha_n, q_x^s) \text{ or } \mu_x = f(\alpha_1, \dots, \alpha_n, \mu_x^s)$$

where the $\alpha_1, \dots, \alpha_n$ are unknown parameters. The MLEs are then found by maximising the likelihoods as in Section 1.3.

- (b) least squares: the parameter values are found that minimise

$$\sum_{\substack{\text{all ages} \\ x}} w_x (\hat{q}_x - \hat{q}_x)^2 \text{ or } \sum_{\substack{\text{all ages} \\ x}} w_x (\hat{\mu}_x - \hat{\mu}_x)^2$$

where the $\{w_x\}$ are suitable weights. Natural weights would be the exposures to risk (E_x or E_x^c) at each age, or the inverse of the estimated variance of \hat{q}_x or $\hat{\mu}_x$.



Question 13.4

Why are the weights based on the *inverse* of the estimated variance?

Step 4 – calculate graduated rates

The assumed relationship with the standard table is then used to calculate the graduated rates.

Step 5 – test

The resulting graduation would be subjected to tests of goodness-of-fit to the data (see Chapter 12) and, if there is more than one candidate function $f()$ (or more than one suitable standard table), the goodness-of-fit may be used to assist in the final choice.

The remarks in Section 1.4 apply also to experiences graduated by this method.



Example

A set of crude forces of mortality ($\hat{\mu}_x$) is to be graduated by reference to a standard table using the relationship $\hat{\mu}_x = a\mu_x^s + b$. The parameters are to be determined using the method of weighted least squares. Assuming that all the crude rates are strictly positive, write down a formula for the expression to be minimised in this case and suggest an appropriate set of weights.

Solution

We need to minimise:

$$S = \sum_x w_x (\hat{\mu}_x - \hat{\mu}_x^s)^2 = \sum_x w_x [\hat{\mu}_x - (a\mu_x^s + b)]^2$$

The weights should be proportional to the reciprocal of the variance of the crude rates. So here we would use:

$$w_x = \frac{E_x^c}{\hat{\mu}_x}$$

So the quantity to be minimised is:

$$S = \sum_x \frac{E_x^c}{\hat{\mu}_x} [\hat{\mu}_x - (a\mu_x^s + b)]^2$$

In this case the calculation could be done using calculus.

3 Graphical graduation

3.1 Overview

It is always possible to employ the simple technique of plotting the crude estimates \hat{q}_x or $\hat{\mu}_x$ on graph paper and drawing a curve through them. This might reasonably be done if all that is needed is a “quick and dirty” result, or a visual impression, though quite obviously it has severe limitations.

Graphical graduation can also be useful in other contexts, for example when looking at the rates of withdrawal or retirement of members of a pension scheme at different ages, particularly if small numbers of people are involved and only a rough idea of the level and pattern of the rates is required.

When performing a graphical graduation various techniques may be used to help in producing a smooth curve. The data may be grouped before attempting the graduation or confidence intervals may be used to help determine the range of likely values for the true underlying rates. Once a suitable graph has been drawn, the graduated rates at each age can be read off. If necessary, these graduated rates can be subjected to a final smoothing procedure called *handpolishing*.

3.2 The graphical graduation process

Here we describe the process of graphical graduation, along with a few useful aids to obtaining a satisfactory result.

Step 1 – plot crude estimates

The crude estimates (\hat{q}_x or $\hat{\mu}_x$) or an appropriate function (eg $\log \hat{q}_x$ or $\log \hat{\mu}_x$) are plotted on graph paper.

Linear or log scale?

Human rates of mortality range from the very small at low ages (< 0.001) to the very large at high ages (> 0.1). It is difficult, if not impossible, to plot values of such different orders of magnitude on the same, linear, scale.

Unless dealing with a very short age range, it is best to plot the crude data on a logarithmic scale. Since a Gompertz term is often a good approximation over some age ranges, this has the advantage that a straight-line fit might be a good approximation on the logarithmic plot.

Coping with scarce data

The problem with scarcity of data at certain ages is that the random errors in the number of deaths can be relatively large. It will then be difficult to determine where the true rates of mortality lie, *i.e.* it will be difficult to decide where exactly to draw the graph.

If data are scarce, group different ages together. There is no settled view on how to choose age groups. Some possibilities are:

- (i) choose groups so that the grouped crude rates run more smoothly;
- (ii) use evenly-spaced groups, eg quinquennial (5-year) age bands; or
- (iii) group so that there is a reasonable number of deaths (eg at least five) in each group.

The first of these can be difficult to achieve.

Having grouped, it is necessary to decide the age to which each grouped crude rate applies, in order to plot it on the graph. Clearly, this is some kind of “mean” age representing the group. The simplest approach would be to use the unweighted mean age. A slightly more sophisticated approach would be to use the exposures to risk at each individual age as weights.

For example, if we have grouped the data into five-year age bands, the mean age \bar{x} representing the age group 40-44 could be calculated as either:

$$(1) \text{ the unweighted average age, } \bar{x} = \frac{40 + 41 + 42 + 43 + 44}{5} = 42, \text{ or}$$

$$(2) \text{ the weighted average age, } \bar{x} = \frac{\sum_{x=40}^{44} x.E_x^c}{\sum_{x=40}^{44} E_x^c} \text{ or } \bar{x} = \frac{\sum_{x=40}^{44} x.E_x}{\sum_{x=40}^{44} E_x}$$



Question 13.5

Why do we weight the average age by the exposed-to-risk at each age?

Step 2 – mark confidence intervals

**Plot approximate confidence limits or error bars around the plotted crude rates.
Rough 95% error bars are given by**

$$\hat{q}_x \pm \frac{2\sqrt{d_x}}{E_x} \quad \text{or} \quad \hat{\mu}_x \pm \frac{2\sqrt{d_x}}{E_x^c}$$

It is not necessary that the graduation should pass through every confidence interval (indeed, it should not) but it should not pass outside more than about 1 in 20.

The confidence intervals give an idea of the acceptable departure of the graduated rates from the crude estimates. To simplify the task of drawing the curve through the crude estimates, we can construct a “corridor” by drawing two lines – one joining successive values of the upper confidence limits and the other joining successive values of the lower confidence limits. When we draw the curve, it should only go outside this corridor for about 1 point in 20.

We can also plot confidence intervals for grouped data as long as the groups have been defined without reference to the data, *eg* evenly spaced age bands. In such cases there has been no loss of randomness.

We cannot plot confidence intervals when the groups have been defined with reference to the data, *eg* if data have been subdivided to produce grouped crude rates that run more smoothly. In such cases $\hat{\mu}_{\bar{x}}$ or $\hat{q}_{\bar{x}}$ (the force or rate of mortality for the group) is an adjusted random variable whose distribution we do not know.

Step 3 – sketch graph

Before sketching the graph, we may wish to refer to existing tables based on similar experience to get an idea of the likely shape of the graph, particularly at the extreme ages where data may be sparse.

Having made preparations such as those above, draw a curve as smoothly as possible, trying to capture the overall shape of the crude rates.

This cannot easily be described – the difficulties involved can only be grasped by trying to do it. It is, in fact, very difficult to draw a curve that is truly smooth (see Chapter 12 for what we mean by “smooth”).

Note that we do not necessarily need to produce a “truly smooth” curve at this stage, since we can improve the smoothness later on using numerical techniques (*eg* handpolishing – see Step 7).

Step 4 – read off values

The graduated values at integer ages can now be read off the plotted curve. It will not usually be possible to discern more than three significant figures in any part of the curve.

Step 5 – test

The graduation can then be subjected to the same tests of goodness-of-fit as before, and also to tests of adequate smoothness (which are rarely necessary with the other two methods described here).

If the data are very scarce, formal statistical tests might not add very much to what we can learn simply by inspection of the graph, but they should still be carried out.

By its approximate nature, the graphical graduation process may well produce a set of rates that do not satisfy the test for smoothness. The next two steps look at ways of dealing with this potential problem.

Step 6 – redraw if necessary

If tests show severe lack of smoothness or adherence to data, then we may wish to redraw the curve, *ie* return to Step 3. We may find that better results can be obtained by plotting a different graph, *eg* $\log \hat{q}_x$.

Step 7 – ‘handpolish’ the graduated rates

Once adequate adherence and a reasonably smooth progression have been obtained, the rates are *handpolished* to iron out any irregularities introduced when the values were read off the graph. (Any unnecessary points of inflexion are also usually removed). Handpolishing is done by manually adjusting the differences between successive figures to ensure a steady progression of values from one age to the next. For example, the differences ($\times 1,000$) should go 3,3,3,4,4,5,5,5, rather than 3,5,3,5,4,3,5,4.

4 Comparison of different methods

First, we note that the three methods of graduation described above by no means cover all possible methods. We take parametric formula graduation to be an example of approaches used with reasonably large data sets, and the other two to be examples of methods used with smaller data sets.

4.1 Graduation by parametric formula

The mathematical formula method produces extremely precise results, good smoothness, and can be a fully automated process. The ability to optimise statistically eliminates any subjectivity from the fitting process. The only subjectivity remaining is in the ultimate choice of formula to use. These properties, along with the independence of such a graduation from any other experience, make this method ideally suited for use in constructing standard tables.

The main problem with the method is that a phenomenon such as mortality, which is subject to a great number of different influences and to different extents over an individual's lifetime, may be impossible to represent adequately by a single mathematical formula. As a result no single function may produce a satisfactory fit, at least over the whole range of age. Heterogeneous data (*ie* data covering a mixture of people with different patterns of mortality – males and females, for example) can also render it more difficult to find a function that produces an adequate fit.

Some specific points about parametric formula graduation are:

- (a) It is a natural extension of the simple probabilistic models for single years of age, parameterised by q_x or $\mu_{x+1/2}$. It is straightforward to extend the statistical theory of estimation from one parameter to several, including estimation of standard errors and so on, and very often computer software is available to carry out the necessary optimisations.
- (b) Provided a reasonably small number of parameters is used, the resulting graduation will be acceptably smooth.
- (c) Sometimes, when comparing several experiences, it is useful to fit the same parametric formula to all of them. Differences between the fitted parameters, given their standard errors, then give insight into the differences between the experiences. For example, the difference between parameters may help us to identify trends in mortality over time.

- (d) **The approach is very well-suited to the production of standard tables from large amounts of data.** It is not possible to use the method successfully where data are scanty over a wide age range.
- (e) **It can, however, be very difficult to find a suitable curve that fits an experience well at all ages. Partly this is because of the different features that predominate at different ages (eg infant mortality, the accident hump and exponential mortality after middle age). Partly it may be because cross-sectional studies mix up different generations at different ages.**

A very likely reason is that there is still a good deal of heterogeneity in all mortality studies, even if we classify the data by age, sex, policy type, calendar period and so on. For example, if we tried to remove heterogeneity by limiting our investigation to 40-year old female policyholders, we might still find large differences in the underlying rates because of differences in diet, smoking habits, levels of exercise *etc.*

- (f) **Care is required when extrapolating. Most methods of curve fitting will result in a good fit where there is most data, which in graduation usually means at middle ages. The form of the curve at the extreme ages is therefore sometimes determined by the best-fitting parameters at other ages, which means that the curve is, to a large degree, extrapolated from the middle ages. The results at extreme ages can, therefore, be quite poor, and might require adjustment. The same warning applies if the graduation is extrapolated beyond the ages for which there are data.**

Given this problem, we may decide to abandon the formula at extreme ages and use an adjusted standard table instead.

- (g) The optimisation procedures can make quantitative allowance for our relative confidence in each observed rate by reflecting the amount of data available at each age via the weighting factors.

4.2 **Graduation by reference to a standard table**

This method is a very simple way of obtaining a workable set of graduated rates in many practical situations. The process will often follow naturally from an experience investigation involving a comparison against a standard table, as described earlier. Because of the fact that a great deal of the form of the function $f(x)$ is provided by the standard table, the graduation formula (and hence the process itself) is greatly simplified compared with other methods. As only a few parameters may need to be fitted from the data, the amount of information that the data need to provide is also less than other methods.

This dependence on the form of the standard table often leads to a very significant difficulty. The features of the actual experience (and hence the probable progression of the true underlying rates with age) may differ significantly from the features displayed by any standard table. Where this is the case, it would never be possible to obtain satisfactory adherence to the data using this method, at least without further adjustment to the graduated rates. It also makes the method unsuitable for the purpose of *producing* standard tables, which need to be fully representative of the data. In this case it would be inappropriate for the graduation to be influenced in this way by another experience.

Some points about graduation by reference to a standard table are:

- (a) **It can be used to fit relatively small data sets where a suitable standard table exists. For example, the CMI have fitted the life office experiences for periods after 1991-94 as linear functions of the “92 series” tables, as a way of gauging the subsequent development of insured lives mortality.** The results based on small data sets will be correspondingly less reliable.
- (b) **Provided a simple function is chosen (eg a polynomial or exponential function of low order), and the standard table is smooth to begin with, a smooth graduation should result.**
- (c) **The collateral information obtained from the standard table can be particularly useful in deciding the shape of the graduation at the extreme ages, where there might be little or no data.**
- (d) **The method is not suitable for the preparation of standard tables based on large amounts of data.**
- (e) **The choice of standard table is important; choosing an inappropriate table could impart the wrong shape to the entire graduation.** Features exhibited by the standard table will also be exhibited by the graduated rates. These may not be desirable (or representative of the data) for the graduation being performed.
- (f) **It is not always easy to choose an appropriate standard table.**
- (g) The simple form of the function means that the fitting process (*ie* estimation of the parameters) is usually easy to carry out.

4.3 Graphical graduation

The method is useful in practical situations where a high degree of precision is not required, and where there is no standard table available that is sufficiently similar to the experience to enable a graduation by reference to a standard table to be carried out successfully. It would therefore be quite inappropriate for the production of standard tables, which require a high degree of smoothness and generally require a much more objective approach. The method does, however, produce adequate results for most practical purposes, and has great flexibility.

Some points about graphical graduation are:

- (a) **It can be used for scanty data sets where any more sophisticated method would be hard to justify. But if the data are too scanty, the results will be unreliable.**
- (b) **It is extremely difficult to obtain results that are both smooth and adhere satisfactorily to the data. It is usually necessary to make several attempts, and to adjust the results by hand (rather than by re-drawing the curve) to restore smoothness (a process sometimes called handpolishing).**
- (c) **It does allow an experienced person to allow for known (or likely) features of the experience, which might escape any more mechanical method.**
- (d) **The other side of (c); any bias or erroneous assumptions on the part of the graduator will be reflected in the outcome, ie the method is subjective. In fact, different practitioners may come up with significantly different graduations.**
- (e) **It is not a suitable method for the production of standard tables based on large amounts of data.**
- (f) **The shape of the curve is completely flexible.**
- (g) **We can use confidence intervals to “pin down” the rates at the ages with most data.**
- (h) **The graduated rates can only be determined to around 3 significant figures (since they have to be read from the axes of the graph).**
- (i) **Graphical graduation requires a great deal of skill and patience.**

**Question 13.6**

Indicate, with reasons, which one of the three methods of graduation described you think would be most appropriate in each of the following situations:

- (a) investigating mortality rates over the last century in a small town to display in a historical exhibition at the town's centenary celebration
- (b) investigating the proportions of policyholders surrendering a particular type of endowment policy at different durations for use in profit test calculations
- (c) investigating mortality rates in a third world country as part of an international comparison study.

5 Statistical tests of a graduation

We discussed statistical tests of graduations in Chapter 12. You may remember that the precise form of some tests depends on the method of graduation. In this section we will return to the subject of statistical tests and cover the outstanding issues, now that we have covered these graduation methods.

You may find it useful to refresh your memory of the statistical tests in Chapter 12. In this chapter we will be looking at just two of the tests – the chi square (χ^2) test and the cumulative deviations test – as these are the only ones that we need to modify according to the method of graduation employed.

5.1 Comparing one experience with another

In Chapter 12, we introduced statistical tests of the hypothesis that one experience was the same as another. Often, the question is whether or not an experience for which we have data and crude estimates is consistent with a given standard table.

The tests depended on comparison of the actual deaths observed at each age x in one experience, d_x , with the number expected on the basis of the other experience.

For example, if the second experience was represented by a standard table $\{q_x^s\}$ or $\{\mu_{x+\frac{1}{2}}^s\}$, we devised tests based on the deviations

$$d_x - E_x q_x^s \quad \text{or} \quad d_x - E_x^c \mu_{x+\frac{1}{2}}^s$$

5.2 Testing a graduation

The same tests can be used to test the hypothesis that the graduation adheres to the data, by substituting the graduated estimates for the standard table quantities above, and using the deviations

$$d_x - E_x \hat{q}_x \quad \text{or} \quad d_x - E_x^c \hat{\mu}_{x+\frac{1}{2}}$$

In effect, we are asking whether or not the observed numbers of deaths are consistent with the numbers expected if the graduated estimates are “correct”.

There are two problems. The first is with the χ^2 -test. The second relates to the cumulative deviations test.

Chi square test

Given m age groups (x_1, \dots, x_m) , the χ^2 -statistics:

$$\sum_{x=x_1}^{x_m} \frac{(d_x - E_x q_x^s)^2}{E_x q_x^s (1 - q_x^s)} \quad \text{or} \quad \sum_{x=x_1}^{x_m} \frac{(d_x - E_x^c \mu_{x+1/2}^s)^2}{E_x^c \mu_{x+1/2}^s}$$

have a χ^2 distribution with m degrees of freedom. (The superscript “s” denotes the standard table as usual.) This leads to a simple statistical test.

A crucial point in the above reasoning is that the two experiences in question should not be the same. In other words, the data upon which the *observed* deaths are based should *not* be the same as the data upon which the *expected* deaths are based.

This clearly does not hold when we are testing the adherence of a graduation to the observed data – we compute the expected deaths using the graduated quantities $\{\hat{q}_x\}$ or $\{\hat{\mu}_{x+1/2}\}$, which are themselves based on the observed deaths.

It is still legitimate to use the χ^2 -test in these circumstances. The χ^2 -statistic is unchanged (except that mortality according to the standard table has been replaced by mortality according to our graduation):

$$\sum_{x=x_1}^{x_m} \frac{(d_x - E_x \hat{q}_x)^2}{E_x \hat{q}_x (1 - \hat{q}_x)} \quad \text{or} \quad \sum_{x=x_1}^{x_m} \frac{(d_x - E_x^c \hat{\mu}_{x+1/2})^2}{E_x^c \hat{\mu}_{x+1/2}}$$

but we must reduce the number of degrees of freedom.

- (a) If we used parametric formula graduation, we lose one degree of freedom for each parameter fitted.
- (b) If we used standard table graduation, we lose one degree of freedom for each parameter fitted, and we lose some further (indeterminate) number of degrees of freedom because of the constraints imposed by the choice of standard table. Rather than suggest how many, it is more important to be aware of the problem when it comes to interpreting the result of the test.

- (c) If we used graphical graduation, it is very difficult to determine how many degrees of freedom are lost. A rule of thumb sometimes used is that two or three degrees of freedom are lost for every ten or so ages or age groups fitted, corresponding roughly to the determination of the height, slope and (perhaps) curvature over that section of the curve.



Question 13.7

For a given set of data you have calculated crude estimates $\{\hat{\mu}_{x+\frac{1}{2}}\}$ for ages $x = 30, 31, \dots, 79$ under the Poisson model. You have graduated the rates assuming that the underlying force of mortality follows Makeham's law $\mu_x = A + Bc^x$.

You now wish to assess the adherence of the graduation to the observed data. State the test statistic and the form of the test.

Cumulative deviations test

The cumulative deviations test cannot be used if the cumulative deviation is zero because of the graduation procedure.

Some methods of graduation may force the cumulative distribution to be close to zero over the age range being graduated. The test is invalidated for such graduations due to the non-random way in which the curve has been fitted to the data.

6 The effect of duplicate policies

6.1 Introduction

The investigations of life office mortality carried out in the UK by the CMI have one particular feature that affects the statistical properties of the resulting estimates: they are based on policies and not lives.

That is, instead of observing persons and recording:

$$E_x^c = \text{Number of person-years observed}$$

$$d_x = \text{Number of deaths}$$

the CMI observe policies, and record:

$$E_x^c = \text{Number of policy-years observed}$$

$$d_x = \text{Number of policies becoming claims by death.}$$

So, if a female policyholder born on 1 January 1960 were to own 3 separate life assurance policies, she would contribute a maximum of 3 years to the value of E_{45}^c in the year 2005. If she were to die in 2006, she would contribute 3 to the value of d_{46} in that year.

The reasons for observing policies rather than lives are that life office record-keeping is based on policies, and that it can be very difficult to establish when two policies are, in fact, owned by the same person, especially if many life offices pool their data (as in the CMI investigations).

It can be particularly difficult to establish if two policies are owned by the same person if the policies were bought from different insurance companies.

The outcome is that we can no longer be sure that we are observing a collection of independent claims; it is quite possible that two distinct death claims are the result of the death of the same life. This is called the problem of *duplicate policies*.



Question 13.8

Why do you think the existence of duplicate policies might be a problem?

6.2 The effect of duplicate policies

We analyse the effect of duplicates in the simplest possible setting. Suppose we observe N lives from age x to $x+1$, with no censoring, new entrants or other complications.

The lives are statistically independent in the usual way, but a proportion π_i ($i = 1, 2, 3, \dots$) of the N lives each own i insurance policies. Thus the total number of policies observed is $\sum_i i\pi_i N$. Suppose each life dies during the year with probability q_x .

The problem we face is that, in a real investigation, we only know that we observe $\sum_i i\pi_i N$ policies. The proportions π_i ($i = 1, 2, 3, \dots$) are unknown.

For example, we may observe the mortality experienced in relation to 1,000 insurance policies. These policies may relate to 1,000 individuals. Alternatively, they may relate to, say, 900 individuals with 1 policy, 35 individuals with 2 policies and 10 individuals with 3 policies, ie only 945 separate individuals. We just can't tell from the data.

Were we to assume that the number of claims, C , is distributed as a Binomial ($\sum_i i\pi_i N, q_x$) random variable, we might be in error because of the dependencies.

Let D_i be the number of *deaths* among the $\pi_i N$ lives each with i policies, and let C_i be the number of *claims* among the same lives. We can say that:

$$D_i \sim \text{Binomial}(\pi_i N, q_x)$$

because we have independence of *deaths*.

So, for the example above, we have:

$$D_1 \sim \text{Binomial}(900, q_x) \quad C_1 = D_1$$

$$D_2 \sim \text{Binomial}(35, q_x) \quad C_2 = 2D_2$$

$$D_3 \sim \text{Binomial}(10, q_x) \quad C_3 = 3D_3$$

$$C = C_1 + C_2 + C_3 = D_1 + 2D_2 + 3D_3$$

and the expected number of claims is:

$$\begin{aligned} E(C) &= E(D_1) + 2E(D_2) + 3E(D_3) \\ &= 900q_x + 2(35q_x) + 3(10q_x) \\ &= 1,000q_x \end{aligned}$$

This is just equal to the total number of policies (1,000) multiplied by q_x . So it is not affected by the presence of duplicates.

More generally, we can write:

$$\begin{aligned} \mathbf{E}[C] &= \mathbf{E}\left[\sum_i C_i\right] = \mathbf{E}\left[\sum_i i D_i\right] = \sum_i i \mathbf{E}[D_i] \\ &= \sum_i i \pi_i N q_x \end{aligned}$$

Now consider the variance of the number of claims. In our example:

$$\begin{aligned} \text{var}(C) &= \text{var}(D_1) + 4 \text{var}(D_2) + 9 \text{var}(D_3) \quad (\text{by independence}) \\ &= 900q_x(1-q_x) + 4(35q_x(1-q_x)) + 9(10q_x(1-q_x)) \\ &= 1,130q_x(1-q_x) \end{aligned}$$

However, if we observed 1,000 independent policies, then the variance of the number of claims would be $1,000q_x(1-q_x)$. So, in this case, the presence of duplicates has increased the variance of the number of claims by a factor of 1.13.

Generalising this result, we obtain:

$$\begin{aligned}\text{var}[C] &= \text{var} \left[\sum_i C_i \right] = \text{var} \left[\sum_i i D_i \right] \\ &= \sum_i i^2 \text{var}[D_i] \quad (\text{independence of deaths}) \\ &= \sum_i i^2 \pi_i N q_x (1 - q_x)\end{aligned}$$

and again we want to compare this to the result we would have for independent policies...

Were we to observe $\sum_i i \pi_i N$ independent lives (or policies) we would have:

$$E[C] = \sum_i i \pi_i N q_x$$

$$\text{var}[C] = \sum_i i \pi_i N q_x (1 - q_x)$$

So, the effect of duplicate policies is to increase the variance of the number of claims, in the ratio

$$r = \frac{\sum_i i^2 \pi_i}{\sum_i i \pi_i}$$

This is called a **variance ratio**. The proportions π_i ($i = 1, 2, 3, \dots$) depend on age, so we get a different variance ratio r_x at each age.

If the variance ratios were known, we could make allowance for the increased variances in tests of a graduation and so on. Usually they are not known for any particular investigation, but the CMI has carried out special investigations from time to time to match up duplicate policies in force and hence derive estimates of r_x suitable for use.

The values of r_x should be recalculated thoroughly from time to time to ensure that the estimates of the variance remain reasonably accurate as the underlying proportions π_i ($i = 1, 2, 3, \dots$) change. In one recent study, the values of r_x varied between 1.18 and 1.75, with an average of 1.46.

Remember that the form of the variance ratio above comes from our assumption that the number of deaths follows a binomial model of mortality.



Question 13.9

Calculate the variance ratio for an investigation based on the observation of the policies in the following table, using the binomial model of mortality.

Number of policies, n	Number of lives holding n policies
1	1,107
2	62
3	16
4	0
5	2

In addition, calculate the expectation and variance of the number of policies becoming claims by death if the policies are all held by 50 year olds with $q_{50} = 0.00534$.

7 Exam-style questions

To finish this chapter, you should now try the following past exam questions.



Question 13.10 Subject D2, September 1999, Question 13 (adapted)

The mortality experience of some whole of life assurance policyholders has been compared with a standard mortality table for assured lives. The following is an extract from the data:

Age, x	Actual deaths, θ_x	Expected deaths, $E_x q_x^s$	$\theta_x - E_x q_x^s$
60	37	42.88	-5.88
61	40	61.73	-21.73
62	28	38.06	-10.06
63	41	47.23	-6.23
64	34	40.36	-6.36
65	40	49.98	-9.98
66	27	25.13	1.87
67	15	22.25	-7.25
68	16	26.23	-10.23
69	30	27.61	2.39
70	23	25.11	-2.11
Total	331	406.56	-75.57

- (i) Carry out a comparison between the actual and expected mortality experience, using the following statistical tests:
- (a) Chi-squared test
 - (b) Cumulative deviations test
 - (c) Serial correlations test.

You should state the appropriate null hypothesis and, for each test, the conclusion reached with regard to this hypothesis. [12]

- (ii) Summarise what you can infer about the mortality experience of these policyholders from your analysis, giving your reasons. [3]

- (iii) You have now decided to produce graduated rates of mortality by fitting the following function to these data using weighted least squares techniques:

$$g(x) = a q_x^s$$

- (a) Determine the value of the parameter a that minimises the weighted least squares criterion for these data.
- (b) The graduated rates are to be tested using the same tests as in part (i). Without performing any further calculations, state how you would expect the results to differ, if at all, from the results of the tests you carried out in part (i).
- (c) Hence state whether you think this graduation is likely to be satisfactory, giving your reasons.

[6]

[Total 21]

**Question 13.11****Subject CT4, September 2006, Question B6 (part)**

- (i)
 - (a) Describe the general form of the polynomial formula used to graduate the most recent standard tables produced for use by UK life insurance companies.
 - (b) Show how the Gompertz and Makeham formulae arise as special cases of this formula.
- (ii) An investigation was undertaken of the mortality of persons aged between 40 and 75 years who are known to be suffering from a degenerative disease. It is suggested that the crude estimates be graduated using the formula:

$$\hat{\mu}_{x+\frac{1}{2}} = \exp \left[b_0 + b_1 \left(x + \frac{1}{2} \right) + b_2 \left(x + \frac{1}{2} \right)^2 \right]$$

- (a) Explain why this might be a sensible formula to choose for this class of lives.
- (b) Suggest two techniques which can be used to perform the graduation.

[3]

[Total 6]

8 **End of Part 4**

You have now completed Part 4 of the Subject CT4 Notes.

Review

Before looking at the Question and Answer Bank we recommend that you briefly review the key areas of Part 4, or maybe re-read the summaries at the end of Chapters 10 to 13.

Question and Answer Bank

You should now be able to answer the questions in Part 4 of the Question and Answer Bank. We recommend that you work through several of these questions now and save the remainder for use as part of your revision.

Assignments

On completing this part, you should be able to attempt the questions in Assignment X4.



Chapter 13 Summary

Methods of graduation

Three of the most common methods of graduation are:

- graduation by parametric formula – we assume that mortality can be modelled using a mathematical formula
- graduation by reference to a standard table – we assume that there is a simple relationship between the observed mortality and an appropriate standard table
- graphical graduation – we draw a curve by hand on a graph of the crude estimates.

The strengths and weaknesses of these methods can be assessed in terms of the following criteria:

- smoothness
- precision of calculated rates
- fidelity to crude rates (goodness of fit)
- ease of use
- amount of data required
- flexibility in allowing for special features
- validity of the method given the problems.

We must take care when using some of the statistical tests to assess the adherence of a graduation to the observed crude estimates. Since the actual number of deaths (those observed) and the expected number (based on the graduated rates) are based on the same set of data, we must reduce the number of degrees of freedom for the χ^2 -test.

Duplicate policies

Duplicate policies (*i.e.* lives with more than one policy) can distort the results of an investigation. Allowance can be made for the increase in the variance of the number of claims observed due to the existence of duplicate policies.

Distribution of claims with duplicate policies (under the binomial model)

$$E[C] = \sum_i i\pi_i Nq_x$$

$$\text{var}[C] = \sum_i i^2\pi_i Nq_x(1-q_x)$$

Variance ratio (under the binomial model)

$$r = \frac{\sum_i i^2\pi_i}{\sum_i i\pi_i}$$

Chapter 13 Solutions

Solution 13.1

$$\mu_x = A + Bc^x$$

So: polynomial₁ = A

$$\text{polynomial}_2 = \log B + x \log c$$

Solution 13.2

We may be able to check our results against:

- previous investigations of the same population
- recent investigations of a different population with similar characteristics
- changes to mortality observed in recent investigations in other countries.

Solution 13.3

The mortality of people with life insurance policies may be lower than that of the population as a whole for several reasons. These include the facts that individuals in poor health may be refused life insurance and that policyholders may generally belong to a higher socio-economic group with a lower rate of mortality.

We'll look at this topic in more detail in Subject CT5.

Solution 13.4

The weights are based on the inverse (*i.e.* the reciprocal) of the variance because this will give more weighting to the ages where the estimated variance is low (which means that we are more confident about the true rate) and less weighting to the ages where the estimated variance is high (which means that we are less confident about the true rate).

Solution 13.5

The crude estimate for the age group is the observed value of a weighted average of the underlying rates at each age within the age group. By weighting the age by the exposure to risk, the representative age for the group is influenced by the relative amount of data we have at each age, and hence the relative contribution that mortality at each age has made to the observations.

As an extreme example, suppose we had 1 person aged 40 and 99 people aged 44 in the age band. It would be more appropriate to consider this to be the rate for the weighted age $0.01 \times 40 + 0.99 \times 44 = 43.96$, which is close to 44, the age of the vast majority of the lives, than to use $\frac{1}{2} \times (40 + 44) = 42$, the midpoint of the age band.

Solution 13.6

- (a) Great accuracy is not required and data will be very limited (because of the relatively small numbers of deaths). So a graphical method is probably most appropriate.

If the pattern of mortality in the town is sufficiently similar to the national experience, graduation by reference to the national population mortality could be used (assuming such statistics are available for the country in question).

- (b) The crude rates are likely to exhibit features that are dependent on the precise terms of the policies (*eg* surrender values may only be available after the end of the third year). So a graphical method is probably most appropriate.

If the rates show a simple pattern, or surrender rates are required for different age bands, it might be possible/necessary to find a suitable parametric formula to describe the rates.

- (c) If the rates appear to follow a similar pattern to another country (or group of countries) in the study whose mortality has already been studied, graduation by reference to a standard table could be used. This would make comparisons with other countries easier.

A graphical method would probably not be accurate enough. A parametric formula might be difficult to fit if the data are unreliable, *eg* if the country does not register births, the ages recorded will only be approximate.

Solution 13.7

We have 50 age groups (ages $x = 30, 31, \dots, 79$). In fitting the model, we will have estimated three parameters (A, B and c).

The test statistic is $\sum_{x=30}^{79} \frac{(d_x - E_x^c \hat{\mu}_{x+\frac{1}{2}})^2}{E_x^c \hat{\mu}_{x+\frac{1}{2}}}$

We will use this statistic in a one-sided χ^2 -test with 47 (ie $50 - 3$) degrees of freedom. We will reject the model if the test statistic exceeds $\chi^2_{0.05, 47} \approx 64$.

Solution 13.8

The death of an individual with multiple life assurance policies would have a disproportionately large impact on the crude estimate of mortality at that age. In particular, the presence of duplicates increases the variance of the estimates of the mortality rates and invalidates many of the graduation tests, which assume that the recorded deaths are statistically independent.

Solution 13.9

The 1,187 lives own $(1 \times 1,107) + (2 \times 62) + (3 \times 16) + (5 \times 2) = 1,289$ policies between them.

To calculate the variance ratio, it is easier to work with the number of policies ($\pi_i N$) rather than the proportions (π_i). Both methods will produce the same answer.

The variance ratio is:

$$\begin{aligned} r &= \frac{\sum_i i^2 \pi_i}{\sum_i i \pi_i} = \frac{\sum_i i^2 \pi_i N}{\sum_i i \pi_i N} \\ &= \frac{(1^2 \times 1,107) + (2^2 \times 62) + (3^2 \times 16) + (5^2 \times 2)}{(1 \times 1,107) + (2 \times 62) + (3 \times 16) + (5 \times 2)} \\ &= \frac{1,549}{1,289} = 1.2017 \end{aligned}$$

The expected number of policies becoming claims by death is:

$$E[C] = \sum_i i \pi_i N q_x = 1,289 \times 0.00534 = 6.883$$

The variance is:

$$\text{var}[C] = \sum_i i^2 \pi_i N q_x (1 - q_x) = 1,549 \times 0.00534 \times (1 - 0.00534) = 8.227$$

Solution 13.10(i) **Statistical tests**

The null hypothesis is that the standard table is representative of the true underlying mortality rates for this group of policyholders.

(a) **Chi-squared test**

If the null hypothesis is true, then the chi-squared test statistic will have a χ^2_{11} distribution.

From the given data, we have:

Age	Standardised deviation, z_x	z_x^2
60	-0.898	0.8063
61	-2.766	7.6493
62	-1.631	2.6591
63	-0.907	0.8218
64	-1.001	1.0022
65	-1.412	1.9928
66	0.373	0.1392
67	-1.537	2.3624
68	-1.997	3.9898
69	0.455	0.2069
70	-0.421	0.1773
Total	-11.741	21.807

The upper 5% point of χ^2_{11} is 19.68. We therefore reject the null hypothesis at the 5% significance level.

(b) *Cumulative deviations test*

$$\sum_x (D_x - E_x q_x^s)$$

If the null hypothesis is true, then the test statistic $Z(c) = \frac{\sum_x (D_x - E_x q_x^s)}{\sqrt{\sum_x E_x q_x^s}} \sim N(0,1)$

where D_x = random number of deaths in age group x .

The observed value of the test statistic is:

$$z(c) = \frac{-75.57}{\sqrt{406.56}} = -3.748$$

The p -value of this (two-tailed) test is:

$$2P(Z(c) > 3.748) = 2(1 - \Phi(3.748)) = 2(1 - 0.99991) = 0.00018$$

Since this is very small, there is very strong evidence against the null hypothesis. .

(c) *Serial correlation test*

If the null hypothesis is true, then $r_1 \sqrt{m} \sim N(0,1)$. The formula for r_1 is given on Page 34 of the *Tables*:

$$r_1 = \frac{\frac{1}{m-1} \sum_{i=1}^{m-1} (z_i - \bar{z})(z_{i+1} - \bar{z})}{\frac{1}{m} \sum_{i=1}^m (z_i - \bar{z})^2}$$

In this case $m = 11$ and $\bar{z} = \frac{-11.741}{11} = -1.067$. The denominator of r_1 is the variance of the standardised deviations, calculated by dividing the sum of squares by m (rather than $m-1$). You can therefore calculate the denominator using the statistical functions on your calculator. This gives:

$$\frac{1}{11} \sum_{i=1}^{11} (z_i - \bar{z})^2 = 0.843$$

The numerator of r_1 has to be calculated manually. We have:

$z_i - \bar{z}$	$(z_i - \bar{z})(z_{i+1} - \bar{z})$
0.169	-0.287
-1.699	0.958
-0.564	-0.090
0.160	0.011
0.066	-0.023
-0.345	-0.497
1.440	-0.677
-0.470	0.437
-0.930	-1.416
1.522	0.983
0.646	-
Total	-0.601

So:

$$r_1 = \frac{\frac{1}{10} \times -0.601}{0.843} = -0.071$$

and the test statistic is $-0.071\sqrt{11} = -0.236$. This is a one-tailed test, and the null hypothesis is rejected if the test statistic lies in the upper tail of $N(0,1)$. Since the test statistic is negative, we do not reject the null hypothesis. (There is also no possibility of undergraduation, since we are comparing against a standard table.)

(ii) *Comments*

From the serial correlation test the shape of the mortality curve is consistent with that of the standard table.

However, by inspection of the individual standard deviations they are consistently negative. This is confirmed by the result from the cumulative deviations test and the chi-squared test. Hence we conclude that the rates from the standard table have a similar shape to those of the experience, but the rates from the standard table are too high.

(iii) **Weighted least squares method**

Value of a that minimises weighted least squares

We need to minimise:

$$\sum w_x \left(\hat{q}_x - \ddot{q}_x \right)^2$$

where $\hat{q}_x = \frac{\theta_x}{E_x}$, $\ddot{q}_x = a q_x^s$, and the weights w_x are inversely proportional to $\text{var}(\tilde{q}_x)$.

From Chapter 10:

$$\text{var}(\tilde{q}_x) = \frac{q_x(1-q_x)}{E_x} \approx \frac{q_x}{E_x}$$

However, the q_x values are unknown, but we can replace them with the standard table rates and use weights $w_x = \frac{E_x}{q_x}$. Hence, the expression to be minimised is:

$$Q = \sum \frac{E_x}{q_x^s} \left(\frac{\theta_x}{E_x} - a q_x^s \right)^2$$

Differentiating with respect to a gives:

$$\frac{dQ}{da} = -2 \sum \frac{q_x^s E_x}{q_x^s} \left(\frac{\theta_x}{E_x} - a q_x^s \right)$$

Equating this to zero and simplifying gives:

$$\sum \theta_x = a \sum E_x q_x^s$$

Hence:

$$a = \frac{\sum \theta_x}{\sum E_x q_x^s} = \frac{331}{406.56} = 0.814$$

Comment on tests

The cumulative deviations will now be zero, which is very acceptable.

There will be a large reduction in the chi-squared value. (It won't be zero since the test looks at the sum of the squares of the deviations.)

The serial correlations test statistic will still be acceptable.

(c) ***Suitable?***

Almost definitely yes!

The amendment will greatly reduce the problem of the standard table rates being too high, without changing the shape of the mortality curve.

Solution 13.11

(i)(a) ***Form of the polynomial***

The general form of the polynomial is:

$$\mu_x = \text{poly}(1) + \exp[\text{poly}(2)]$$

where $\text{poly}(1)$ and $\text{poly}(2)$ are polynomials in x .

(i)(b) ***Gompertz and Makeham***

We can generate Gompertz' formula by setting $\text{poly}(1)$ equal to zero, and setting $\text{poly}(2)$ equal to a linear function of x , say $a + bx$. We can now write:

$$\mu_x = \exp(a + bx) = e^a (e^b)^x$$

which is Gompertz' law, $\mu_x = Bc^x$ with $B = e^a$ and $c = e^b$.

We can obtain Makeham's law in exactly the same way, but by setting $\text{poly}(2)$ equal to a constant A instead of to zero.

(ii)(a) ***Explanation***

Mortality is likely to be an exponentially increasing function over the age range from 40 to 75, so this might be a sensible formula since it behaves in the correct way.

We use a formula with $x + \frac{1}{2}$ since putting in integer values of x will give the force of mortality half way through the year of life, which is what we require.

(ii)(b) ***Graduation techniques***

Two possible methods for fitting the graduation formula in this context might be:

- (a) maximum likelihood estimation, and
- (b) least squares estimation.

Part 1 – Questions

Introduction

The Question and Answer Bank is divided into five parts. The first four parts of the Question and Answer Bank are split into two sections:

- Section 1 – Development questions. The aim of these questions is to build on your understanding, test key Core Reading and bring your knowledge and skills to the level required to tackle exam-style questions.
- Section 2 – Exam-style questions. These questions are of the level of difficulty you are likely to face in the examination. It is very important that you focus on these questions as preparation for the exam.

The last part contains a set of exam-style questions covering the whole course.

For each part the questions may require knowledge from earlier parts of the course.

We strongly recommend that you use these questions to practise the *thinking* necessary to pass the exam. Do not use them as a set of material to *learn* but attempt the questions for yourself under strict exam-style conditions, before looking at the solutions provided.

This distinction represents the difference between active studying and passive studying. Given that the examiners will be aiming to set questions to make you think (and in doing so they will be devising questions you have not seen before) it is much better if you practise the skills that they will be testing.

It may also be useful to you if you group a number of the questions together to attempt under exam time conditions. Ideally three hours would be set aside, but anything from one hour (*i.e.* 35 marks) upwards will help your time management.

Note that the split between Development questions and Exam-style questions is somewhat subjective. For example, there have been past CT4 exam questions that test knowledge of the Core Reading, and so are similar to what we've included here as Development questions. The Exam-style questions involve more application and a wider range of ideas and are typically the more challenging questions in the exam.

1 ***Development questions***

Question 1.1

Consider each of the following statistics associated with a bank account:

- (i) number of times the account has been overdrawn since it was opened
- (ii) status (overdrawn, in credit) of the account on the last day of each month
- (iii) number of direct debits paid since the account was opened
- (iv) status (overdrawn, in credit) of the account at any time since the account was opened.

Each statistic is to be modelled by a stochastic process. In each case:

- (a) state whether the state space should be discrete or continuous
- (b) state whether the time set should be discrete or continuous
- (c) suggest the name of a stochastic process that might be a suitable model and give a brief justification for your suggestion. [8]

Question 1.2

Explain what is meant by “a stochastic model” and give two advantages these have over deterministic models. [3]

Question 1.3

For a stochastic process X_n with time set J and state space S , define the terms:

- (i) stationary [1]
 - (ii) weakly stationary [1]
 - (iii) increment [1]
 - (iv) Markov property. [1]
- [Total 4]

Question 1.4

A moving average (stochastic) process, X_n , has a discrete time domain and a continuous state space and is defined as:

$$X_n = Z_n + \alpha_1 Z_{n-1} + \alpha_2 Z_{n-2} + \alpha_3 Z_{n-3}$$

where $\{Z_n, n \in \mathbb{Z}\}$ are independent and identically distributed $N(0, \sigma^2)$ random variables and $\alpha_1, \alpha_2, \alpha_3$ are constants.

- (i) Prove that X_n is weakly stationary. [5]
 - (ii) Explain whether the Markov property holds. [2]
 - (iii) Deduce whether the process has independent increments. [1]
- [Total 8]

Question 1.5

- (i) (a) Define a Poisson process with rate λ .
 (b) Define a compound Poisson process. [2]
- (ii) The cumulative amount of claims reaching an insurance company is modelled using a compound Poisson process.
- (a) Explain why the compound Poisson process has the Markov property.
 (b) Comment on whether this seems reasonable for the given insurance model.
 (c) State whether the compound Poisson process is weakly stationary.
 (d) Explain whether you expect the cumulative insurance claims to follow a weakly stationary process. [5]
- [Total 7]

Question 1.6

The price of an ordinary share is modelled as a stochastic process X_n ; $n = 0, 1, 2, 3, \dots$ with initial condition $X_0 = x_0 > 0$, where:

$$X_n = x_0 \prod_{j=1}^n U_j \quad n \geq 1$$

and U_n is a white noise process.

- (i) Show that the process $\log X_n$, $n \geq 0$ has independent increments. [2]
 (ii) Explain why X_n is a Markov process. [1]
- [Total 3]

Question 1.7

A simple NCD system for motor insurance has four levels of discount – 0%, 20%, 40% and 60%. A new policyholder starts on 0% discount. At the end of each policy year, policyholders will change levels according to the following rules:

- At the end of a claim free year, a policyholder moves up one level, or remains on the maximum discount.
- At the end of a year in which exactly one claim was made, a policyholder drops back one level, or remains at 0%.
- At the end of a year in which more than one claim was made, a policyholder drops back to zero discount.

For a particular driver in any year, the probability of a claim free year is 0.9, the probability of exactly one claim is 0.075, and the probability of more than one claim is 0.025.

A policyholder takes out a policy for the first time on 1 January 2004, and by 1 January 2007 he/she had made only one claim, on 3 May 2006. Calculate the probability that he/she is on 20% discount in 2009. [3]

Question 1.8

A Markov chain is determined by the transition matrix:

$$P = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Determine which of the following are stationary distributions for the chain.

I $(\frac{9}{35}, \frac{1}{7}, \frac{3}{5}, \frac{1}{10})$

II $(0, 0, 0, 1)$

III $(\frac{9}{70}, \frac{1}{14}, \frac{3}{10}, \frac{1}{2})$

[3]

Question 1.9

A Markov chain $\{X_n\}_{n=0}^{\infty}$ has a discrete state space S . The initial probability distribution is given by $P[X_0 = i] = q_i$. The one-step transition probabilities are denoted by $P[X_{m+1} = i_{m+1} | X_m = i_m] = p_{i_m i_{m+1}}^{(m,m+1)}$.

- (i) State the Markov property for such a process. [1]
 - (ii) Write down expressions for the following in terms of p 's and q 's.
 - (a) $P[X_0 = i_0, X_1 = i_1, \dots, X_n = i_n]$
 - (b) $P[X_4 = i]$ [2]
- [Total 3]

Question 1.10

A new actuarial student is planning to sit one exam each session. He expects that his performance in any exam will only be dependent on whether he passed or failed the last exam he sat. If he passes a given exam, the probability of passing the next will be α , regardless of the nature of the exam. If he fails an exam, the probability of passing the next will be β .

- (i) Calculate the probability that:
 - (a) the first exam he fails is the seventh, given that he passes the first
 - (b) he passes the fifth exam, given that he fails the first three. [4]
- (ii) Explain the results above in terms of a Markov chain, specifying the state space and transition matrix. (For the purposes of this model, assume that we are only interested in predicting passing or failing, not in the number of exams passed so far.) [3]

[Total 7]

Question 1.11

The stochastic process $\{X_t\}$ is defined by the relationship $X_t = Z_t + Z_{t-1}$, where $\{Z_t\}$ is a sequence of independent random variables with probability function:

$$Z_t = \begin{cases} 1 & \text{with probability } p \\ 1,000 & \text{with probability } q \end{cases}$$

where $p + q = 1$ and $q < p$.

(i) Find expressions in terms of p and q for each of the following probabilities:

(a) $P(X_5 = 1,001)$

(b) $P(X_5 = 1,001 | X_4 = 1,001)$

(c) $P(X_5 = 1,001 | X_4 = 1,001, X_3 = 1,001)$. [6]

(ii) State, with reasons, whether $\{X_t\}$ has the Markov property. [2]

[Total 8]

Question 1.12

Determine all the stationary distributions for a Markov chain with transition matrix:

$$P = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & \frac{4}{5} & \frac{1}{5} & 0 & 0 \\ \frac{1}{2} & \frac{3}{10} & \frac{1}{5} & 0 & 0 \end{pmatrix}$$

[5]

Question 1.13

Show that, under the assumptions of the two-state Markov model for mortality:

$$\text{var}[D_i - \mu V_i] = E[D_i] \quad [6]$$

The notation used here is the same as in the Core Reading, and you may assume that $E[D_i - \mu V_i] = 0$ (which is proved in Chapter 4 of the Course Notes).

Question 1.14

A survival model for an elderly population has two states A and D, representing alive and dead. The force of mortality at age t years $\mu(t)$, ie the transition rate from A to D, is given by:

$$\mu(t) = 0.0001 \times (1.10)^t$$

Calculate the probability that a 60-year-old will survive to age 80. [2]

2 Exam-style questions

Question 1.15

Describe the role of simulation in sensitivity analysis.

[3]

Question 1.16

List the factors that should be considered when deciding whether a model is suitable for a particular application.

[5]

Question 1.17

The classification of stochastic processes according to discrete or continuous time variable, discrete or continuous state space gives rise to a four-way classification. Give four examples, one of each type, of stochastic models that may be used to model observed processes.

[2]

Question 1.18

- (i) X_n is a stochastic process with a discrete state space and a discrete time set. Show that if non-overlapping increments of this process are independent, then the process satisfies the Markov property.

[2]

- (ii) Show that a white noise process in discrete time with a discrete state space does not have independent increments, but is a Markov process.

[2]

[Total 4]

Question 1.19

An insurer has initial capital of u and receives premium income continuously at the rate of c per annum. Let $S(t)$ denote the total claim amount up to time t .

- (i) Describe a model that would allow the insurer to estimate its probability of ruin (*ie* the probability that its claims outgo is more than its available funds). State any assumptions that you make.

[3]

- (ii) Write down an expression for the probability of ruin in terms of u , c and $S(t)$.

[1]

[Total 4]

Question 1.20

At the end of each year an independent organisation ranks the performance of the unit trusts invested in a particular sector, and classifies them into four quartiles (with quartile 1 being the best performer). Past experience has shown that, at the end of each year, a fund will either remain in the same quartile or will move to a neighbouring quartile.

In fact, there is a probability $1-2\alpha$ that a fund will remain in the same quartile and, where upward or downward movements are both possible, these are equally likely. However, it has been found that a fund that has remained in the top or bottom quartile for two consecutive years has a probability of $1-\beta$ ($\beta < \alpha$) of remaining in the same quartile the following year.

- (i) Construct a discrete-time Markov chain with six states to model this situation, defining the states in your model and drawing a state transition diagram. [3]
 - (ii) Write down the transition matrix for your model. [2]
 - (iii) Explain whether this Markov chain is irreducible, periodic or both. [2]
 - (iv) Show that, if a stationary distribution exists with a quarter of the funds in each quartile, then $\beta = \frac{\alpha(1-2\alpha)}{1-\alpha}$. [4]
 - (v) Last year 20% of funds in the second quartile moved up to the top quartile. Assuming the fund rankings have reached a stationary state, estimate the probability that a fund that has been in the top quartile for the last two years will remain in the top quartile for a third consecutive year. [2]
- [Total 13]

Question 1.21

A simple NCD system has four levels of discount – 0%, 20%, 40% and 60%. A new policyholder starts on 0% discount. At the end of each policy year, policyholders will change levels according to the following rules:

- At the end of a claim free year, a policyholder moves up one level, or remains on the maximum discount.
- At the end of a year in which exactly one claim was made, a policyholder drops back one level, or remains at 0%.
- At the end of a year in which more than one claim was made, a policyholder drops back to zero discount.

For a particular policyholder in any year, the probability of a claim free year is $\frac{7}{10}$, the probability of exactly one claim is $\frac{1}{5}$ and the probability of more than one claim is $\frac{1}{10}$.

- (i) Write down the transition matrix for this time homogeneous Markov chain. [2]
 - (ii) Calculate the 2-step transition probabilities from state i to state j , $p_{ij}^{(2)}$. [3]
 - (iii) If the policyholder starts with no discount, calculate the probability that this policyholder is at the maximum discount level 5 years later. [5]
 - (iv) If a large number of people having the same claim probabilities take out policies at the same time, calculate the proportion would you expect to be in each discount category in the long run. [5]
- [Total 15]

Question 1.22

Consider the following two Markov chains:

- Chain I is defined on the state space $\{1, 2, 3, 4\}$ and has transition matrix:

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ \left(\begin{array}{cccc} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{array} \right) \end{array}$$

- Chain II is defined on the state space $\{1, 2, 3, 4, 5\}$ and has transition matrix:

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ \left(\begin{array}{ccccc} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{array} \right) \end{array}$$

Let X_t denote the state occupied at time t . For each chain:

- Draw a transition diagram, including on your diagram the probability of each possible transition. [2]
 - Calculate:
 - $P(X_2 = 1 | X_0 = 1)$
 - $P(X_4 = 1 | X_0 = 1)$ [4]
 - Explain whether the chain is irreducible and/or aperiodic. [3]
 - Explain whether or not the process will converge to a stationary distribution given that it is in State 1 at time 0. If it does converge, determine the stationary distribution. [3]
- [Total 12]

Question 1.23

An author is about to start writing a book that will contain 20 chapters. The author plans to write a new chapter each week. However, when he reviews his work at the end of each week, there is a probability of 0.25 (which is independent of the current state of the book) that he will not be happy with one of the chapters he has written. In this case, he will spend the following week rewriting that particular chapter instead of embarking on a new one. He may decide to rewrite any one chapter, including a new one he has just finished or one that he has previously rewritten.

Let X_k denote the number of chapters that the author is happy with at the end of week k , and define $X_0 = 0$.

- (i) Explain why X_k can be modelled as a Markov chain. [2]
 - (ii) Calculate the probability that the author will complete the book in exactly 25 weeks. [2]
 - (iii) Calculate the expected number of weeks it will take the author to complete the book. [2]
- [Total 6]

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Part 1 – Solutions

Solution 1.1

- (i) ***Number of times the account has been overdrawn since it was opened***
- (a) Discrete state space of the non-negative integers. [½]
- (b) Continuous time set. [½]
- (c) This is a counting process that counts the number of overdrawn events. It is Markov because the probability of being in a particular state at the next point in time depends only on the current state. If events occur at a rate of λ per unit time, ie the rate of events is independent of time, then a Poisson process might be appropriate. [1]
- (ii) ***Status of the account on the last day of each month***
- (a) Discrete state space with two states: “overdrawn” and “in credit”. [½]
- (b) Discrete time set of non-negative integers measuring the “number” of the month since the start of the process. [½]
- (c) If the probability of being in a particular state at the next time point depends only on the current state, then the process is Markov so a Markov chain would be appropriate. [1]
- (iii) ***Number of direct debits paid since the account was opened***
- (a) Discrete state space of the non-negative integers. [½]
- (b) Continuous time set. [½]
- (c) This is a counting process that counts the number of direct debits. It is Markov because the probability of being in a particular state at the next point in time depends only on the current state. If direct debits happen at a rate of λ per unit time, ie the rate of events is independent of time, then a Poisson process might be appropriate. [1]

- (iv) ***Status of the account at any time since the account was opened***
- (a) Discrete state space with two values: “overdrawn” and “in credit”. [½]
- (b) Continuous time set. [½]
- (c) If the probability of being in a particular state depends only on the current state, then the process is Markov so a Markov jump process would be appropriate. [1]

Solution 1.2

A stochastic model is one that recognises the random nature of the input components. [1]

Stochastic models have the following advantages over deterministic models:

1. To reflect reality as accurately as possible, the model should imitate the random nature of the variables involved.
2. A stochastic model can provide information about the distribution of the results (eg probabilities, variances etc), not just a single best estimate figure.
3. Stochastic models allow you to use Monte Carlo simulation, which is an extremely powerful technique for solving complex problems. [2]

Solution 1.3

- (i) ***Stationary***

A stochastic process X_n is stationary if the joint distributions of $X_{t_1}, X_{t_2}, \dots, X_{t_m}$ and $X_{t_1+k}, X_{t_2+k}, \dots, X_{t_m+k}$ are identical for all $t_1, t_2, \dots, t_m, k + t_1, k + t_2, \dots, k + t_m \in J$ and all integers m . [1]

- (ii) ***Weakly stationary***

The process is weakly stationary if the expectations $E[X_t]$ are constant with respect to t and the covariances $\text{cov}(X_t, X_{t+k})$ depend only on the lag k . [1]

(iii) ***Increment***

If t and $t+u$ are in J then the increment for time u will be $X_{t+u} - X_t$. Often $u=1$ is used.

[1]

(iv) ***Markov property***

The Markov property states that:

$$P[X_t \in A | X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_m} = x_m] = P[X_t \in A | X_{t_m} = x_m]$$

for all times $t_1 < t_2 < \dots < t_m < t \in J$, all states $x_1, x_2, \dots, x_m, x \in S$ and all subsets A of S .

[1]

Solution 1.4(i) ***Weak stationarity***

The Z_j are independent and identically distributed, and the α_j are constants. So:

$$E[X_n] = (1 + \alpha_1 + \alpha_2 + \alpha_3) E[Z] = (1 + \alpha_1 + \alpha_2 + \alpha_3) \times 0 = 0 \quad [1]$$

and:

$$\begin{aligned} \text{var}[X_n] &= \text{var}[Z] + \alpha_1^2 \text{var}[Z] + \alpha_2^2 \text{var}[Z] + \alpha_3^2 \text{var}[Z] \\ &= (1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2) \sigma^2 \end{aligned} \quad [1]$$

which is constant.

The covariance at lag 1 is:

$$\begin{aligned} \text{cov}(X_n, X_{n+1}) &= \text{cov}(Z_n + \alpha_1 Z_{n-1} + \alpha_2 Z_{n-2} + \alpha_3 Z_{n-3}, Z_{n+1} + \alpha_1 Z_n + \alpha_2 Z_{n-1} + \alpha_3 Z_{n-2}) \\ &= \alpha_1 \text{var}(Z) + \alpha_1 \alpha_2 \text{var}(Z) + \alpha_2 \alpha_3 \text{var}(Z) \\ &= (\alpha_1 + \alpha_1 \alpha_2 + \alpha_2 \alpha_3) \sigma^2 \end{aligned} \quad [1]$$

The covariance at lag 2 is:

$$\begin{aligned}
 & \text{cov}(X_n, X_{n+2}) \\
 &= \text{cov}(Z_n + \alpha_1 Z_{n-1} + \alpha_2 Z_{n-2} + \alpha_3 Z_{n-3}, Z_{n+2} + \alpha_1 Z_{n+1} + \alpha_2 Z_n + \alpha_3 Z_{n-1}) \\
 &= \alpha_2 \text{ var}(Z) + \alpha_1 \alpha_3 \text{ var}(Z) \\
 &= (\alpha_2 + \alpha_1 \alpha_3) \sigma^2
 \end{aligned} \quad [\frac{1}{2}]$$

The covariance at lag 3 is:

$$\begin{aligned}
 & \text{cov}(X_n, X_{n+3}) \\
 &= \text{cov}(Z_n + \alpha_1 Z_{n-1} + \alpha_2 Z_{n-2} + \alpha_3 Z_{n-3}, Z_{n+3} + \alpha_1 Z_{n+2} + \alpha_2 Z_{n+1} + \alpha_3 Z_n) \\
 &= \alpha_3 \text{ var}(Z) \\
 &= \alpha_3 \sigma^2
 \end{aligned} \quad [\frac{1}{2}]$$

The covariances at lags 4, 5, 6 ... are 0. [\frac{1}{2}]

So the covariance depends only on the lag and not on the value of n . Thus the process X_n is weakly stationary. [\frac{1}{2}]

(ii) ***Markov?***

For a Markov process, the value of X_n only depends on the most recently known value. However, X_n depends on the previous X values so it does not possess the Markov property. [2]

(iii) ***Independent increments?***

If the increments of a process are independent, then that process must have the Markov property. Since we've said that this process is not a Markov process, it cannot have independent increments. [1]

Solution 1.5(i)(a) ***Poisson process***

A Poisson process N_t , $t \geq 0$, with rate λ is a continuous time process satisfying:

- $N_0 = 0$
- N_t has Poisson distributed independent increments. [1]

(i)(b) ***Compound Poisson process***

Let $\{X_n\}_{n=0}^{\infty}$ be independent identically distributed random variables. A compound Poisson process with rate λ is defined for $t \geq 0$ to be:

$$S_t = X_1 + X_2 + \dots + X_{N_t}$$

and:

$$S_0 = 0 \quad [1]$$

(ii)(a) ***Markov property***

It is sufficient to show that the compound process has independent increments, since then the Markov property must hold. However, having independent increments is part of the definition of the compound process. [1]

(ii)(b) ***Reasonableness***

This is consistent with insurance claims, since we would only expect the cumulative insurance claims by time t to depend on the most recently known value. For example, if we know the cumulative claims after day one are £1,000, and by day ten are £15,000, you wouldn't expect the older value of £1,000 to add any useful information to the more recent value of £15,000. [2]

(ii)(c) ***Weak stationarity***

The process cannot be stationary since, for example, $E(S_t)$ changes with t . [1]

(ii)(d) ***Is cumulative claim amount weakly stationary?***

You wouldn't expect $E[S_t]$ to be constant since the cumulative claims generally increases with time. This would be a constant only in the trivial case where the individual claim amounts are £0, which is rather uninteresting. [1]

Note that to show something is not stationary, it is sufficient to show that any one of the conditions fails to hold.

Solution 1.6(i) ***Show that $\ln X_n$ has independent increments***

By definition:

$$\begin{aligned}\ln X_n &= \ln x_0 + \sum_{j=1}^n \ln U_j \\ &= \ln x_0 + \sum_{j=1}^n Z_j\end{aligned}\quad [\frac{1}{2}]$$

where $Z_j = \ln U_j$ is a white noise process, ie are a set of independent and identically distributed random variables. Then:

$$\ln X_n - \ln X_{n-1} = \ln U_n = Z_n \quad [\frac{1}{2}]$$

Because $Z_n, n = 0, 1, \dots$ are independent, $\ln X_n$ has independent increments. [1]

(ii) ***Explain why X_n is a Markov process***

$\ln X_n$ has independent increments

$\Rightarrow \ln X_n$ is a Markov process

$\Rightarrow X_n = \exp(\ln X_n)$ is a Markov process.

because exponentiation merely rescales the state space of the process. [1]

Solution 1.7

The policyholder starts with 0% discount in 2004. He makes no claims in 2004 or 2005, and so have a 40% discount in 2006. He makes exactly one claim that year so he falls back to 20% discount for 2007. So we are looking for the probability of being at the 20% discount level in 2009, given that in 2007 the discount was also 20%. [½]

The transition matrix is:

$$P = \begin{pmatrix} 0.1 & 0.9 & 0 & 0 \\ 0.1 & 0 & 0.9 & 0 \\ 0.025 & 0.075 & 0 & 0.9 \\ 0.025 & 0 & 0.075 & 0.9 \end{pmatrix} \quad [1]$$

We want the (20%,20%) entry (or (2,2)th entry) in P^2 . [½]

Multiplying the 2nd row by the 2nd column gives:

$$0.1 \times 0.9 + 0.9 \times 0.075 = 0.1575 \quad [1]$$

Solution 1.8

I $(\frac{1}{35}, \frac{1}{7}, \frac{3}{5}, \frac{1}{10})$ is not a distribution as its entries do not sum to 1. [1]

II $(0,0,0,1)P = (0,0,0,1)$, so $(0,0,0,1)$ is a stationary distribution. [1]

III $(\frac{1}{70}, \frac{1}{14}, \frac{3}{10}, \frac{1}{2})P = (\frac{1}{70}, \frac{1}{14}, \frac{3}{10}, \frac{1}{2})$, so $(\frac{1}{70}, \frac{1}{14}, \frac{3}{10}, \frac{1}{2})$ is a stationary distribution. [1]

Solution 1.9

(i) **Markov property**

The Markov property means a lack of dependence on the past of the process:

$$P[X_n = j | X_0 = i_0, X_1 = i_1, \dots, X_{m-1} = i_{m-1}, X_m = i] = P[X_n = j | X_m = i] \quad [1]$$

(ii) **Probabilities**

$$(a) P[X_0 = i_0, X_1 = i_1, \dots, X_n = i_n] = q_{i_0} p_{i_0 i_1}^{(0,1)} p_{i_1 i_2}^{(1,2)} \cdots p_{i_{n-1} i_n}^{(n-1,n)} \quad [1]$$

Note that this is the probability of the process taking a unique given path.

$$(b) P[X_4 = i] = \sum_{i_0 \in S} \sum_{i_1 \in S} \sum_{i_2 \in S} \sum_{i_3 \in S} q_{i_0} p_{i_0 i_1}^{(0,1)} p_{i_1 i_2}^{(1,2)} p_{i_2 i_3}^{(2,3)} p_{i_3 i}^{(3,4)} \quad [1]$$

Here we need to sum over all the possible starting points and then over all paths from these starting points to end up in state i at time 4.

Solution 1.10(i) **Probabilities**

$$(a) \alpha^5 (1 - \alpha) \quad [2]$$

$$(b) (1 - \beta)\beta + \beta\alpha = \beta(\alpha - \beta + 1) \quad [2]$$

(ii) **Explanation**

Because the student's performance depends only upon whether he passed or failed the last exam, we can think of the problem as a Markov chain on the state space $\{F, P\}$ representing "failed the last exam" and "passed the last exam" respectively. The transition matrix is:

$$\begin{pmatrix} 1 - \beta & \beta \\ 1 - \alpha & \alpha \end{pmatrix}$$

In (i)(a) we've calculated the probability of the unique path:

$$P \rightarrow P \rightarrow P \rightarrow P \rightarrow P \rightarrow P \rightarrow F$$

given that we start in P . [1]

In (b) we calculated the transition probability $p_{FP}^{(2)}$, which is the *FP* entry ((1,2)th entry) in the matrix :

$$\begin{pmatrix} 1-\beta & \beta \\ 1-\alpha & \alpha \end{pmatrix}^2 = \begin{pmatrix} (1-\beta)^2 + \beta(1-\alpha) & (1-\beta)\beta + \beta\alpha \\ (1-\alpha)(1-\beta) + (1-\alpha)\alpha & (1-\alpha)\beta + \alpha^2 \end{pmatrix} \quad [2]$$

Solution 1.11

(i) **Probabilities**

$$\begin{aligned} \text{(a)} \quad P(X_5 = 1,001) &= P(Z_5 + Z_4 = 1,001) \\ &= P(Z_5 = 1,000, Z_4 = 1) + P(Z_5 = 1, Z_4 = 1,000) \\ &= qp + pq = 2pq \end{aligned}$$

[2]

(b) Using the result in (i)(a) to evaluate the denominator:

$$\begin{aligned} P(X_5 = 1,001 | X_4 = 1,001) &= \frac{P(X_5 = 1,001, X_4 = 1,001)}{P(X_4 = 1,001)} \\ &= \frac{P(Z_5 + Z_4 = 1,001, Z_4 + Z_3 = 1,001)}{2pq} \\ &= \frac{P(Z_5 = 1,000, Z_4 = 1, Z_3 = 1,000)}{2pq} \\ &\quad + \frac{P(Z_5 = 1, Z_4 = 1,000, Z_3 = 1)}{2pq} \\ &= \frac{q^2 p + p^2 q}{2pq} \\ &= \frac{pq(q+p)}{2pq} \\ &= \frac{1}{2} \quad \text{because } p+q=1. \end{aligned} \quad [2]$$

(c) Using the expression for the numerator in (i)(b) to evaluate the denominator:

$$\begin{aligned}
 & P(X_5 = 1,001 | X_4 = 1,001, X_3 = 1,001) \\
 &= \frac{P(X_5 = 1,001, X_4 = 1,001, X_3 = 1,001)}{P(X_4 = 1,001, X_3 = 1,001)} \\
 &= \frac{P(Z_5 + Z_4 = 1,001, Z_4 + Z_3 = 1,001, Z_3 + Z_2 = 1,001)}{pq(p+q)} \\
 &= \frac{P(Z_5 = Z_3 = 1,000, Z_4 = Z_2 = 1)}{pq(p+q)} \\
 &\quad + \frac{P(Z_5 = Z_3 = 1, Z_4 = Z_2 = 1,000)}{pq(p+q)} \\
 &= \frac{2p^2q^2}{pq(p+q)} \\
 &= \frac{2pq}{(p+q)} \\
 &= 2pq \quad \text{because } p+q=1. \tag{2}
 \end{aligned}$$

(ii) **Markov?**

If $\{X_t\}$ had the Markov property, the probabilities in (i)(b) and (i)(c) would be the same. Since they are not, it doesn't. (Note that $2pq < \frac{1}{2}$ when $q < p$.) [2]

Solution 1.12

We are solving:

$$(\pi_1, \pi_2, \pi_3, \pi_4, \pi_5) \begin{pmatrix} 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{5} & \frac{4}{5} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & \frac{4}{5} & \frac{1}{5} & 0 & 0 \\ \frac{1}{2} & \frac{3}{10} & \frac{1}{5} & 0 & 0 \end{pmatrix} = (\pi_1, \pi_2, \pi_3, \pi_4, \pi_5)$$

This gives five equations:

$$\begin{aligned} + \frac{1}{2}\pi_5 &= \pi_1 \\ \frac{1}{5}\pi_2 &+ \frac{4}{5}\pi_4 + \frac{3}{10}\pi_5 = \pi_2 \\ \frac{4}{5}\pi_2 &+ \frac{1}{5}\pi_4 + \frac{1}{5}\pi_5 = \pi_3 \\ \frac{1}{2}\pi_1 &+ \frac{1}{3}\pi_3 = \pi_4 \\ \frac{1}{2}\pi_1 &+ \frac{2}{3}\pi_3 = \pi_5 \end{aligned} \quad [1]$$

Rearranging we get:

$$\begin{aligned} -\pi_1 &+ \frac{1}{2}\pi_5 = 0 \\ -\frac{4}{5}\pi_2 &+ \frac{4}{5}\pi_4 + \frac{3}{10}\pi_5 = 0 \\ \frac{4}{5}\pi_2 &- \pi_3 + \frac{1}{5}\pi_4 + \frac{1}{5}\pi_5 = 0 \\ \frac{1}{2}\pi_1 &+ \frac{1}{3}\pi_3 - \pi_4 = 0 \\ \frac{1}{2}\pi_1 &+ \frac{2}{3}\pi_3 - \pi_5 = 0 \end{aligned}$$

We will ignore the third equation since one equation is always redundant. So we are trying to solve:

$$\begin{aligned} -\pi_1 &+ \frac{1}{2}\pi_5 = 0 \\ -\frac{4}{5}\pi_2 &+ \frac{4}{5}\pi_4 + \frac{3}{10}\pi_5 = 0 \\ \frac{1}{2}\pi_1 &+ \frac{1}{3}\pi_3 - \pi_4 = 0 \\ \frac{1}{2}\pi_1 &+ \frac{2}{3}\pi_3 - \pi_5 = 0 \end{aligned} \quad [1]$$

We will choose π_1 as the working variable.

From the first equation we have $\pi_5 = 2\pi_1$. Substituting this in the fourth equation gives $\pi_3 = \frac{9}{4}\pi_1$. Using the third we can then obtain $\pi_4 = \frac{1}{2}\pi_1 + \frac{1}{3} \times \frac{9}{4}\pi_1 = \frac{5}{4}\pi_1$. [1]

Finally from the second equation we see that:

$$\pi_2 = \frac{5}{4} \left(\frac{4}{5} \times \frac{5}{4}\pi_1 + \frac{3}{10} \times 2\pi_1 \right) = 2\pi_1$$

Thus our solution in terms of π_1 is $(1, 2, \frac{9}{4}, \frac{5}{4}, 2)\pi_1$. [1]

Now apply the condition of summing to 1 to get:

$$\pi_1 = \frac{1}{(1+2+\frac{9}{4}+\frac{5}{4}+2)} = \frac{4}{34} = \frac{2}{17}$$

and therefore the stationary distribution is $(\frac{2}{17}, \frac{4}{17}, \frac{9}{34}, \frac{5}{34}, \frac{4}{17})$. [1]

Solution 1.13

We can start with the usual formula for variances:

$$\text{var}[D_i - \mu V_i] = E[(D_i - \mu V_i)^2] - \{E[D_i - \mu V_i]\}^2$$

Since $E[D_i - \mu V_i] = 0$, this is just:

$$\text{var}[D_i - \mu V_i] = E[(D_i - \mu V_i)^2] \quad [1]$$

We can evaluate this using the definition of the expectation, by considering that death will either occur at some time v_i in the interval $(0, b_i - a_i)$, in which case $D_i - \mu V_i = 1 - \mu v_i$, or the life will survive to the end of this interval, in which case $D_i - \mu V_i = 0 - \mu(b_i - a_i)$. So we get:

$$\text{var}[D_i - \mu V_i] = \int_0^{b_i - a_i} (1 - \mu v_i)^2 \cdot \mu e^{-\mu v_i} dv_i + [\mu(b_i - a_i)]^2 e^{-\mu(b_i - a_i)} \quad [1]$$

If we expand the integrand, we get:

$$\begin{aligned} \text{var}[D_i - \mu V_i] &= \int_0^{b_i-a_i} \mu e^{-\mu v_i} dv_i \\ &\quad - 2\mu \int_0^{b_i-a_i} v_i \cdot \mu e^{-\mu v_i} dv_i + \mu^2 \int_0^{b_i-a_i} v_i^2 \cdot \mu e^{-\mu v_i} dv_i + \mu^2 (b_i - a_i)^2 e^{-\mu(b_i-a_i)} \end{aligned} \quad [1]$$

The first of these integrals is just $E[D_i]$, as proved in the Course Notes. So we have:

$$\begin{aligned} \text{var}[D_i - \mu V_i] &= E[D_i] \\ &\quad + \mu^2 \left\{ -2 \int_0^{b_i-a_i} v_i \cdot e^{-\mu v_i} dv_i + \int_0^{b_i-a_i} v_i^2 \cdot \mu e^{-\mu v_i} dv_i + (b_i - a_i)^2 e^{-\mu(b_i-a_i)} \right\} \end{aligned} \quad [1]$$

So we just need to show that the three terms in the curly brackets sum to zero.

From equation (*) in Section 4.2 of Chapter 4 of the Course Notes, we know that:

$$\int_0^{b_i-a_i} \mu e^{-\mu v_i} dv_i + e^{-\mu(b_i-a_i)} = 1$$

If we differentiate this with respect to μ , we get:

$$\int_0^{b_i-a_i} (e^{-\mu v_i} - v_i \cdot \mu e^{-\mu v_i}) dv_i - (b_i - a_i) e^{-\mu(b_i-a_i)} = 0 \quad [1]$$

And if we differentiate this *again* with respect to μ , we get:

$$\int_0^{b_i-a_i} (-2v_i \cdot e^{-\mu v_i} + v_i^2 \cdot \mu e^{-\mu v_i}) dv_i + (b_i - a_i)^2 e^{-\mu(b_i-a_i)} = 0 \quad [1]$$

Since the LHS of this equation is the same expression as in the curly brackets, we have established the result:

$$\text{var}[D_i - \mu V_i] = E[D_i]$$

Solution 1.14

The survival probability is given by:

$${}_{20}p_{60} = \exp\left(-\int_{60}^{80} 0.0001(1.1)^t dt\right) = \exp\left(-0.0001\left[\frac{(1.1)^t}{\ln 1.1}\right]_{60}^{80}\right) = 0.16 \quad [2]$$

Solution 1.15

Sensitivity analysis involves testing the robustness of the model by making small changes to the input parameters. This should result in small changes to the output from the model that are consistent with the real-world behaviour of the situation we are modelling. [1]

The usual method of carrying out a sensitivity analysis is to run a large number of computer simulations based on the original parameter values, then to repeat these using several sets of slightly different parameter values. [1]

For consistency, the same set of pseudo-random numbers should be used for each set of simulations. This removes the effect of the additional source of randomness that would be introduced by using different pseudo-random numbers. [1]

This question was Question 1(iii) on the Subject 103, September 2004 exam paper.

Solution 1.16

The factors that should be considered when assessing the suitability of a model for a particular exercise are:

- the objectives of the modelling exercise [½]
 - the validity of the model for the purpose to which it is to be put [½]
 - the validity of the data to be used [½]
 - the possible errors associated with the model or parameters used not being a perfect representation of the real world situation being modelled [½]
 - the impact of correlations between the random variables that drive the model [½]
 - the extent of correlations between the various results produced from the model [½]
 - the current relevance of models developed and used in the past [½]
 - the credibility of the data input [½]
 - the credibility of the results output [½]
 - the dangers of spurious accuracy [½]
 - the ease with which the model and its results can be communicated. [½]
- [Maximum 5]

Solution 1.17

Using the models in the course, we could give any of the following examples:

- | | |
|---|--|
| Discrete time / discrete state space: | Markov chain, simple random walk |
| Discrete time / continuous state space: | Time series, general random walk, white noise |
| Continuous time / discrete state space: | Markov jump process, Poisson process |
| Continuous time / continuous state space: | Diffusion process, Brownian motion, compound Poisson process [2] |

This question was Question 5(i) on the Subject 103, September 2002 exam paper.

Solution 1.18(i) **Proof**

$$\begin{aligned} P[X_n = a | X_{n-m} = x, X_{n-m-1} = x_{n-m-1}, X_{n-m-2} = x_{n-m-2}, X_{n-m-3} = x_{n-m-3}, \dots] \\ = P[X_n - X_{n-m} + x = a | X_{n-m} = x, X_{n-m-1} = x_{n-m-1}, X_{n-m-2} = x_{n-m-2}, \dots] \end{aligned}$$

for all times $m > 0$ and all states $a, x_{n-m}, x_{n-m-1}, \dots$ of the state space, S . [1]

$$\begin{aligned} P[X_n - X_{n-m} + x = a | X_{n-m} = x, X_{n-m-1} = x_{n-m-1}, X_{n-m-2} = x_{n-m-2}, \dots] \\ = P[X_n - X_{n-m} + x = a | X_{n-m} = x] \\ = P[X_n = a | X_{n-m} = x] \end{aligned}$$

if non-overlapping increments are independent. [1]

(ii) **White noise process**

For a discrete time, discrete state white noise process $\{Z_n : n = 1, 2, 3, \dots\}$, where Z_n are independent and identically distributed random variables with mean μ and variance σ^2 , we have:

$$\text{cov}(Z_n - Z_{n-1}, Z_{n-1} - Z_{n-2}) = \text{cov}(-Z_{n-1}, Z_{n-1}) = -\sigma^2$$

So non-overlapping increments are not independent. [1]

However:

$$P[Z_n = z_n | Z_{n-m} = z_{n-m}, Z_{n-m-1} = z_{n-m-1}, Z_{n-m-2} = z_{n-m-2}, \dots] = P[Z_n = z_n]$$

and $P[Z_n = z_n | Z_{n-m} = z_{n-m}] = P[Z_n = z_n]$

because the random variables are independent. So the process satisfies the Markov property. [1]

Solution 1.19(i) ***Model***

Let $N(t)$ denote the number of claims received by the insurer up to time t . $N(t)$ can be modelled as a Poisson process. [1]

Let X_j denote the amount of the j th claim. Then the cumulative claim amount up to time t is given by:

$$S(t) = X_1 + X_2 + \cdots + X_{N(t)}$$

If we assume that the random variables X_j are independent and identically distributed, and they are independent of $N(t)$, then $S(t)$ is a compound Poisson process. [2]

(ii) ***Probability of ruin***

The probability of ruin for the insurer is the probability that, for some time t , its claims outgo up to time t is greater than its initial capital plus premium income up to time t . In symbols, this is:

$$P(S(t) > u + ct \text{ for some } t > 0) \quad [1]$$

Solution 1.20(i) **Markov chain**

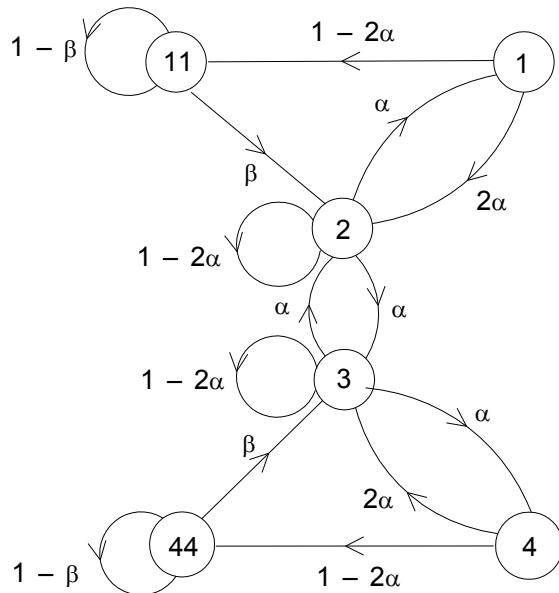
We need to subdivide the top and bottom quartiles in order to satisfy the Markov property. This results in the following 6 states:

- State 11: Funds in the 1st quartile this year and last year
- State 1: Funds in the 1st quartile this year but not last year
- State 2: Funds in the 2nd quartile this year
- State 3: Funds in the 3rd quartile this year
- State 4: Funds in the 4th quartile this year but not last year
- State 44: Funds in the 4th quartile this year and last year

[1]

The labels for the states need not match the ones given here.

The state diagram then looks like this:



[2]

(ii) ***Transition matrix***

The transition matrix is:

$$P = \begin{bmatrix} & 11 & 1 & 2 & 3 & 4 & 44 \\ 11 & 1-\beta & \beta & & & & \\ 1 & 1-2\alpha & 2\alpha & & & & \\ 2 & & \alpha & 1-2\alpha & \alpha & & \\ 3 & & & \alpha & 1-2\alpha & \alpha & \\ 4 & & & & 2\alpha & 1-2\alpha & \\ 44 & & & & \beta & 1-\beta & \end{bmatrix} \quad [2]$$

(iii) ***Irreducible and periodic?***

This chain is irreducible since it is possible to move from each state to any other, eg by following the route $\dots \rightarrow 11 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 44 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 11 \rightarrow \dots$. [1]

A periodic chain is one in which a state can only be revisited at multiples of some fixed number $d > 1$. This is not the case here since any of the four states 11, 2, 3, 44 can be repeated an arbitrary number of times. [1]

(iv) ***Proof***

If a stationary distribution exists with a quarter of the funds in each quartile, then the stationary probabilities π_i must satisfy:

$$\pi_{11} + \pi_1 = \pi_2 = \pi_3 = \pi_4 + \pi_{44} = \frac{1}{4} \quad [1]$$

The stationary probabilities also satisfy the matrix equation $\pi = \pi P$.

The first column of this matrix equation tells us that:

$$(1-\beta)\pi_{11} + (1-2\alpha)\pi_1 = \pi_{11} \Rightarrow -\beta\pi_{11} + (1-2\alpha)\pi_1 = 0 \Rightarrow \pi_{11} = \frac{(1-2\alpha)}{\beta}\pi_1$$
[1]

But we want $\pi_{11} + \pi_1 = \frac{1}{4}$.

So:

$$\frac{(1-2\alpha)}{\beta} \pi_1 + \pi_1 = \frac{1}{4} \text{ ie } \left(1 + \frac{1-2\alpha}{\beta}\right) \pi_1 = \frac{1}{4} \quad [1]$$

The second column of this matrix equation tells us that:

$$\alpha \pi_2 = \pi_1 \text{ ie } \alpha \times \frac{1}{4} = \pi_1$$

Combining these two equations gives:

$$\left(1 + \frac{1-2\alpha}{\beta}\right) \alpha \times \frac{1}{4} = \frac{1}{4} \Rightarrow \left(1 + \frac{1-2\alpha}{\beta}\right) \alpha = 1 \Rightarrow \beta = \frac{\alpha(1-2\alpha)}{1-\alpha} \quad [1]$$

(v) ***Estimated probability***

The probability of a fund in the second quartile moving up to the top quartile is α . So we estimate $\hat{\alpha} = 0.2$. Hence the probability of the fund remaining in the top quartile for a third consecutive year is estimated to be:

$$1 - \hat{\beta} = 1 - \frac{\hat{\alpha}(1-2\hat{\alpha})}{1-\hat{\alpha}} = 1 - \frac{0.2 \times 0.6}{0.8} = 0.85 \quad [2]$$

Solution 1.21(i) **Transition matrix**

The one-step transition matrix is:

$$P = \frac{1}{10} \begin{pmatrix} 3 & 7 & 0 & 0 \\ 3 & 0 & 7 & 0 \\ 1 & 2 & 0 & 7 \\ 1 & 0 & 2 & 7 \end{pmatrix} = \begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0.1 & 0.2 & 0 & 0.7 \\ 0.1 & 0 & 0.2 & 0.7 \end{pmatrix} \quad [2]$$

(ii) **Two-step transition probabilities**

We use the fact that $p_{ij}^{(2)} = (P^2)_{ij}$.

$$P^2 = \frac{1}{100} \begin{pmatrix} 3 & 7 & 0 & 0 \\ 3 & 0 & 7 & 0 \\ 1 & 2 & 0 & 7 \\ 1 & 0 & 2 & 7 \end{pmatrix} \begin{pmatrix} 3 & 7 & 0 & 0 \\ 3 & 0 & 7 & 0 \\ 1 & 2 & 0 & 7 \\ 1 & 0 & 2 & 7 \end{pmatrix} = \frac{1}{100} \begin{pmatrix} 30 & 21 & 49 & 0 \\ 16 & 35 & 0 & 49 \\ 16 & 7 & 28 & 49 \\ 12 & 11 & 14 & 63 \end{pmatrix} \quad [3]$$

(iii) **Probability of being at maximum discount in 5 years**

We shall represent the states 0%, 20%, 40% and 60% by 0,1,2 and 3 respectively. In order to calculate $p_{0,3}^{(5)}$ we can use $(P^5)_{0,3} = \sum_{k=0}^3 (P^2)_{0,k} (P^3)_{k,3}$. So we can first calculate the fourth column of P^3 :

$$P^3 = \frac{1}{1,000} \begin{pmatrix} 30 & 21 & 49 & 0 \\ 16 & 35 & 0 & 49 \\ 16 & 7 & 28 & 49 \\ 12 & 11 & 14 & 63 \end{pmatrix} \begin{pmatrix} 3 & 7 & 0 & 0 \\ 3 & 0 & 7 & 0 \\ 1 & 2 & 0 & 7 \\ 1 & 0 & 2 & 7 \end{pmatrix} = \frac{1}{1,000} \begin{pmatrix} * & * & * & 343 \\ * & * & * & 343 \\ * & * & * & 539 \\ * & * & * & 539 \end{pmatrix} \quad [2]$$

Now we have:

$$\begin{aligned} (P^5)_{0,3} &= \sum_{k=0}^3 (P^2)_{0,k} (P^3)_{k,3} = \frac{1}{100,000} (30 \times 343 + 21 \times 343 + 49 \times 539) \\ &= \frac{43904}{100,000} = 0.43904 \end{aligned} \quad [3]$$

(iv) ***Long-term proportions on each discount level***

This is equivalent to finding the stationary distribution, ie solving the matrix equation:

$$(\pi_0, \pi_1, \pi_2, \pi_3) \begin{pmatrix} 3 & 7 & 0 & 0 \\ 3 & 0 & 7 & 0 \\ 1 & 2 & 0 & 7 \\ 1 & 0 & 2 & 7 \end{pmatrix} = 10(\pi_0, \pi_1, \pi_2, \pi_3)$$

This matrix equation is equivalent to the simultaneous equations:

$$\begin{aligned} 3\pi_0 + 3\pi_1 + \pi_2 + \pi_3 &= 10\pi_0 \\ 7\pi_0 + 2\pi_2 &= 10\pi_1 \\ 7\pi_1 + 2\pi_3 &= 10\pi_2 \\ 7\pi_2 + 7\pi_3 &= 10\pi_3 \end{aligned} \quad [1]$$

We can ignore one of the equations, say the first. Rearranging we get:

$$\begin{aligned} 7\pi_0 - 10\pi_1 + 2\pi_2 &= 0 \\ 7\pi_1 - 10\pi_2 + 2\pi_3 &= 0 \\ 7\pi_2 - 3\pi_3 &= 0 \end{aligned} \quad [1]$$

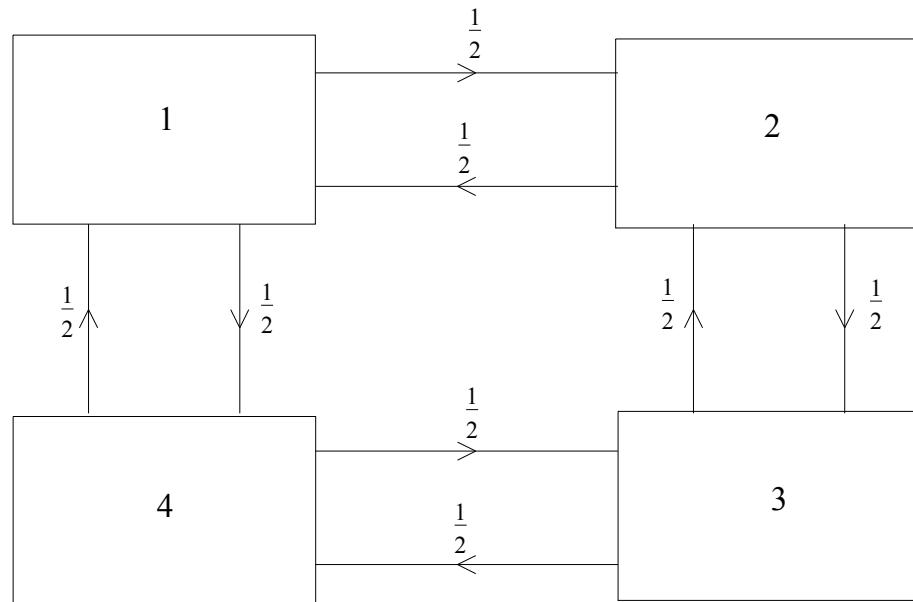
Use π_3 (say) as the working variable. From the third equation we have $\pi_2 = \frac{3}{7}\pi_3$. Substituting in the second:

$$\pi_1 = \left(\frac{10}{7} \times \frac{3}{7} - \frac{2}{7} \right) \pi_3 = \frac{16}{49} \pi_3 \quad [1]$$

Finally from the first equation $\pi_0 = \left(\frac{10}{7} \times \frac{16}{49} - \frac{2}{7} \times \frac{3}{7} \right) \pi_3 = \frac{118}{343} \pi_3$. [1]

So we have the stationary distribution $(118, 112, 147, 343) \frac{\pi_3}{343}$. Applying the summation condition this gives:

$$\frac{1}{720} (118, 112, 147, 343) = (0.1639, 0.1556, 0.2042, 0.4764) \quad [1]$$

Solution 1.22**Chain I**(i) **Transition diagram**

[1]

(ii) **Probabilities**

The initial distribution is $(1, 0, 0, 0)$. Repeated postmultiplication of this vector by the transition matrix for Chain I gives:

$$(1, 0, 0, 0) \rightarrow \left(0, \frac{1}{2}, 0, \frac{1}{2}\right) \rightarrow \left(\frac{1}{2}, 0, \frac{1}{2}, 0\right) \rightarrow \left(0, \frac{1}{2}, 0, \frac{1}{2}\right) \rightarrow \left(\frac{1}{2}, 0, \frac{1}{2}, 0\right) \rightarrow \dots$$

So:

$$(a) P(X_2 = 1 | X_0 = 1) = \frac{1}{2} \quad [1]$$

$$(b) P(X_4 = 1 | X_0 = 1) = \frac{1}{2} \quad [1]$$

Alternatively, because we are only asked about particular probabilities, we could evaluate all the possible paths corresponding to each event and add their probabilities.

$$P(X_2 = 1 | X_0 = 1)$$

Time	0	1	2	Probability
Path	1	2	1	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
Path	1	4	1	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$$P(X_4 = 1 | X_0 = 1)$$

Time	0	1	2	3	4	Probability
Path	1	2	1	2	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	2	3	2	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	4	1	4	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	4	3	4	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	2	1	4	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	4	1	2	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	2	3	4	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	4	3	2	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

(iii) ***Irreducible and/or aperiodic?***

Chain I is irreducible since every state can be reached from every other state. [½]

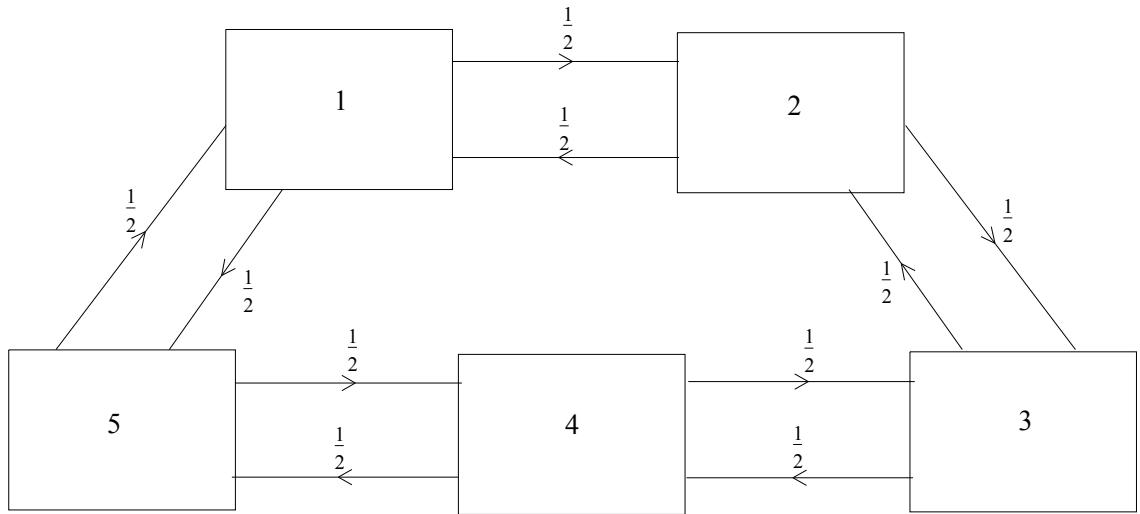
Because the chain is irreducible every state will have the same period. It is possible to return to state 1 in 2, 4, 6, 8 ... moves. State 1 has a period of 2 and so every state has period of 2. The chain is not aperiodic. [1]

(iv) ***Will the process converge to a stationary distribution?***

The process is finite and irreducible, so it has a unique stationary distribution but the process may not conform to this distribution the long term. In fact, this process will not converge to a stationary distribution. In the solution to part (ii), we saw that the distribution will alternate between $(\frac{1}{2}, 0, \frac{1}{2}, 0)$ and $(0, \frac{1}{2}, 0, \frac{1}{2})$. [1]

Chain II

- (i)
- Transition diagram***



[1]

- (ii)
- Probabilities***

The initial distribution is $(1, 0, 0, 0, 0)$. Repeated postmultiplication of this vector by the transition matrix for Chain II gives:

$$(1, 0, 0, 0, 0) \rightarrow \left(0, \frac{1}{2}, 0, 0, \frac{1}{2}\right) \rightarrow \left(\frac{1}{2}, 0, \frac{1}{4}, \frac{1}{4}, 0\right) \rightarrow \left(0, \frac{3}{8}, \frac{1}{8}, \frac{1}{8}, \frac{3}{8}\right) \rightarrow \left(\frac{3}{8}, \dots, \dots, \dots, \dots\right)$$

So:

$$(a) P(X_2 = 1 | X_0 = 1) = \frac{1}{2} \quad [1]$$

$$(b) P(X_4 = 1 | X_0 = 1) = \frac{3}{8} \quad [1]$$

Alternatively, because we are only asked about particular probabilities, we could evaluate all the possible paths corresponding to each event and add their probabilities.

$$P(X_2 = 1 | X_0 = 1)$$

Time	0	1	2	Probability
Path	1	2	1	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
Path	1	5	1	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$$P(X_4 = 1 | X_0 = 1)$$

Time	0	1	2	3	4	Probability
Path	1	2	1	2	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	2	3	2	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	5	1	5	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	2	1	5	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	5	1	2	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
Path	1	5	4	5	1	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

(iii) ***Irreducible and/or aperiodic?***

Chain II is irreducible since every state can be reached from every other state. [½]

Because the chain is irreducible every state will have the same period. It is possible to return to state 1 in 2, 4, 5, 6, 7, 8 ... moves. State 1 has a period of 1 (it is aperiodic) and so every state is aperiodic. The chain is aperiodic. [1]

(iv) ***Will the process converge to a stationary distribution?***

Yes. The chain has a finite number of states, is irreducible and is aperiodic. So there will be a unique stationary distribution that the process will conform to in the long term.

[1]

By symmetry, this stationary distribution is $\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$. [1]

Solution 1.23(i) **Markov chain**

The process has the Markov property since the probability of moving to the next state does not depend on the history of the process prior to entering the current state. [1]

In fact, we have:

$$X_k = \begin{cases} X_{k-1} + 1 & \text{with probability 0.75} \\ X_{k-1} & \text{with probability 0.25} \end{cases}$$

for $X_{k-1} \neq 20$ and $P(X_k = 20 | X_{k-1} = 20) = 1$.

X_t has a discrete state space, namely $\{0, 1, 2, \dots, 20\}$, and a discrete time set since the value of the process is recorded at the end of each week. So the process is a Markov chain. [1]

(ii) **Probability**

To calculate the probability that the book is finished in exactly 25 weeks, we need the probability that the last chapter is completed in the 25th week and, in the first 24 weeks there were 5 chapters rewritten. So the probability is:

$$\binom{24}{5} 0.25^5 \times 0.75^{19} \times 0.75 = 0.13163 \quad [2]$$

(iii) **Expected number of weeks until completion**

Let N denote the number of weeks it takes to complete the book. The possible values of N are 20, 21, 22, ... and:

$$P(N = 20) = 0.75^{20}$$

$$P(N = 21) = \binom{20}{1} 0.25 \times 0.75^{20} = \binom{20}{19} 0.25 \times 0.75^{20}$$

$$P(N = 22) = \binom{21}{2} 0.25^2 \times 0.75^{20} = \binom{21}{19} 0.25^2 \times 0.75^{20}$$

and so on. So N has a Type 1 negative binomial distribution with $k = 20$ and $p = 0.75$. [1]

Hence:

$$E(N) = \frac{k}{p} = \frac{20}{0.75} = 26.67 \text{ weeks} \quad [1]$$

Alternatively, you could define m_k to be the expected time until the book is finished, given that there are currently k chapters completed. Then, for $k = 0, 1, \dots, 19$:

$$m_k = 1 + 0.75m_{k+1} + 0.25m_k$$

ie, in one week's time, there is a 75% chance of having $k+1$ completed chapters and a 25% chance of still having k completed chapters. Rearranging this equation, we get:

$$0.75m_k = 1 + 0.75m_{k+1}$$

or:

$$m_k = \frac{1}{0.75} + m_{k+1}$$

Since $m_{20} = 0$, we have:

$$m_{19} = \frac{1}{0.75}$$

$$m_{18} = \frac{1}{0.75} + \frac{1}{0.75} = \frac{2}{0.75}$$

$$m_{17} = \frac{1}{0.75} + \frac{2}{0.75} = \frac{3}{0.75}$$

and so on. In general, we have:

$$m_k = \frac{20-k}{0.75}$$

So the expected time until the book is completed is:

$$m_0 = \frac{20}{0.75} = 26.67 \text{ weeks}$$

Part 2 – Questions

Note that the split between Development questions and Exam-style questions is somewhat subjective. For example, there have been past CT4 exam questions that test knowledge of the Core Reading, and so are similar to what we've included here as Development questions. The Exam-style questions involve more application and a wider range of ideas and are typically the more challenging questions in the exam.

1 Development questions

Question 2.1

For a Poisson process with parameter λ , state the distribution of the inter-arrival time random variable T . [2]

Question 2.2

For a Poisson process with rate λ , give an expression for the probability that exactly one event will occur during a finite time interval of length t . [2]

Question 2.3

For a particular insurance company, the number of reported claims of a particular type can be modelled as a Poisson process with rate 10 per day. The company has a 24-hour hotline. Of the reported claims, 40% are for amounts over £1,000, and the distribution of these among all claims is independent of anything else. Calculate the expected waiting time until another claim over £1,000 is reported. [2]

Question 2.4

Claims on a portfolio of policies occur according to a Poisson process with a mean rate of 5 claims per day. Claim amounts are 10, 20 or 30. 20% of claims are of amount 10, 70% are of amount 20 and 10% are of amount 30.

- (i) Calculate the expected waiting time until the first claim of amount 30. [1]
- (ii) Calculate the probability that there are at least 10 claims during the first 2 days, given that there were exactly 6 claims during the first day. [2]
- (iii) Calculate the probability that there are at least 2 claims of amount 20 during the first day and at least 3 claims of amount 20 during the first 2 days. [3]
- (iv) Calculate the conditional variance of the number of claims during the first day, given that there are 2 claims of amount 10 during the first day. [3]

[Total 9]

Question 2.5

$\{X_t\}$ is a Markov jump process with state space $S = \{0, 1, 2, \dots\}$ and $X_0 = 0$. The transition rates are given by:

$$\mu_{ij} = \begin{cases} \lambda & \text{if } j = i + 1 \\ -\lambda & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

- (i) Write down the transition probabilities $P_{ij}(t)$. [2]
- (ii) Define the term *holding time*. [1]
- (iii) Find the distribution of the first holding time T_0 . [2]
- (iv) State the value of X_{T_0} . [1]
- (v) Given that the increments are stationary and independent, state the distributions of T_0, T_1, T_2, \dots . Justify your answer. [3]

[Total 9]

Question 2.6

A Markov jump process has transition probabilities given by $p_{ij}(s, t)$.

- (i) Stating any assumptions you make, define the transition rates. [2]
- (ii) Show that the sum of the transition rates out of any state i is zero. [2]
- (iii) Write down the Chapman-Kolmogorov equations and differentiate to obtain the Kolmogorov forward and backward differential equations. [4]
- (iv) Explain why the forward and backward equations often more useful than the original transition matrix. [2]

[Total 10]

Question 2.7

In a Markov jump process model of sickness and death there are three states: healthy, sick and dead, denoted by $\{H, S, D\}$. The transition rates are given by the matrix:

$$A(t) = \begin{pmatrix} -\sigma(t) - \mu(t) & \sigma(t) & \mu(t) \\ \rho(t) & -\rho(t) - \nu(t) & \nu(t) \\ 0 & 0 & 0 \end{pmatrix}$$

State whether each of the following is a correct equation for this process.

I $p_{HS}(s, t) = p_{HH}(s, u)p_{HS}(u, t) + p_{HS}(s, u)p_{SS}(u, t) \quad s < u < t$

II $p_{HS}(s, t) = \int_0^{t-s} p_{SS}(s+w, t) \sigma(s+w) e^{-\int_s^{s+w} (\sigma(u) + \mu(u)) du} dw$

III $p_{HS}(s, t) = \int_0^{t-s} p_{HH}(s, t-w) \sigma(t-w) e^{-\int_{t-w}^t (\sigma(u) + \mu(u)) du} dw \quad [3]$

Question 2.8

In a Markov jump process model of sickness and death there are three states: healthy, sick and dead, {H,S,D}. The transition rates are given by the matrix:

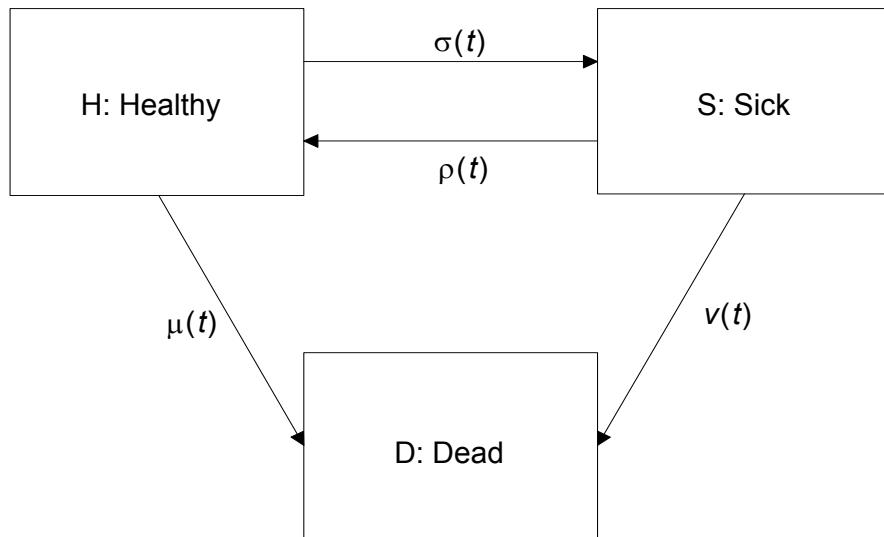
$$A(t) = \begin{pmatrix} -\sigma(t) - \mu(t) & \sigma(t) & \mu(t) \\ \rho(t) & -\rho(t) - \nu(t) & \nu(t) \\ 0 & 0 & 0 \end{pmatrix}$$

Explain the following equation by general reasoning:

$$p_{SH}(s, t) = \int_0^{t-s} p_{SS}(s, t-w) \rho(t-w) e^{-\int_{t-w}^t (\sigma(u) + \mu(u)) du} dw \quad [7]$$

Question 2.9

In a Markov jump process model of sickness and death there are three states: healthy, sick and dead, $\{H, S, D\}$. The transition graph is shown below. Denote the state of the process at time t by X_t .



- (i) Give the transition rates in the form of a matrix. [2]
 - (ii) Define the residual holding time R_s . [1]
 - (iii) If someone is sick at time s , give an integral expression for the probability they remain sick for a further period of at least w . [2]
 - (iv) State the probability density function of R_s (given $X_s = S$). [2]
 - (v) If we are told that a transition from H has just taken place, give an expression for the probability that it was to S . [2]
 - (vi) Give the integral form of the backward Kolmogorov equation for $p_{SD}(s, t)$, the probability that an individual who is sick at time s is dead at time t . [5]
 - (vii) Explain your expression by general reasoning. [4]
- [Total 18]

Question 2.10

A Markov jump process is used to model sickness and death. Four states are included $S = \{H, S_1, S_2, D\}$ representing healthy, sick, terminally sick and dead respectively. We are told that the people who are terminally sick never recover and die at a rate of $1.03(1.01)^t$ where t is their age in years. Calculate the probability that a terminally sick 50-year-old dies within a year. [2]

Question 2.11

A Poisson process is being used to model the number of “hits” that will be received by a new website.

Data from other web sites suggest that hits will be received at a rate of 2 per second.

- (i) Describe how you would simulate the process. [3]
 - (ii) Describe how you would verify the model when the new website is “live”. [2]
- [Total 5]

2 Exam-style questions

Question 2.12

Claims on an insurer’s travel insurance policies arriving in the claims department (state A) wait for an average of two days before being logged and classified by a claims administrator as requiring:

- investigation by a loss adjuster (state L),
- more details from the insured (state I),
- no further information is required and the claim should be settled immediately (state S).

Only one new claim in ten is classified as requiring investigation by a loss adjuster, and five in ten require further information from the insured.

If needed, investigation by a loss adjuster takes an average of 10 days, after which 30% of cases require further information from the insured and 70% are sent for immediate settlement.

Collecting further information from the insured takes an average of 5 days to complete, and immediate settlement takes an average of 2 days before the settlement process is complete (state C).

It is suggested that a time-homogeneous Markov process with states A, L, I, S and C could be used to model the progress of claims through the settlement system with the ultimate aim of reducing the average time to settle a claim.

- (i) Calculate the generator matrix, $\{\mu_{ij}; i, j = A, L, I, S, C\}$, of such a model. [2]
- (ii) Calculate the proportion of claims that eventually require more details from the insured. [2]
- (iii) Using a forward differential equation, derive an expression for the probability that a claim is yet to be logged and classified by a claims administrator at time t . [4]
- (iv) Write down the backward integral equation for the transition probability $P_{AI}(t)$, and hence show that:

$$P_{AI}(t) = 0.05 \int_0^t e^{-0.50w} P_{LI}(t-w) dw + 0.25e^{-0.20t} \int_0^t e^{-0.30w} dw \quad [3]$$

- (v) Using the expression in (iv), show that:

$$P_{AI}(t) = \frac{3}{80}e^{-0.10t} + \frac{47}{60}e^{-0.20t} - \frac{197}{240}e^{-0.50t} \quad [6]$$

- (vi) Calculate the probability that a claim is still awaiting information from the insured after 10 days. [1]
 - (vii) Describe briefly how you could verify that this time homogeneous rather than an alternative time inhomogeneous model was a suitable model of the claims' settlement process. [2]
- [Total 20]

Question 2.13

An n -state, time-homogeneous Markov jump process with transition probability matrix $P(t)$ over a period of length t , is said to have a stationary distribution if:

$$(1) \quad \pi P(t) = \pi$$

$$(2) \quad 0 \leq \pi_i \leq 1 \text{ for each } i = 1, 2, \dots, n$$

$$(3) \quad \sum_{i=1}^n \pi_i = 1$$

- (i) Show that condition (1) is equivalent to $\pi \Sigma = 0$ where Σ is the matrix of transition rates. [1]

In a particular company the salary scale has only two different levels. On average, an employee spends 2 years at level 1 before moving on to the higher level, or leaving the company. An employee at the maximum level spends an average of 5 years before leaving. Nobody is demoted, and promotion can occur at any time.

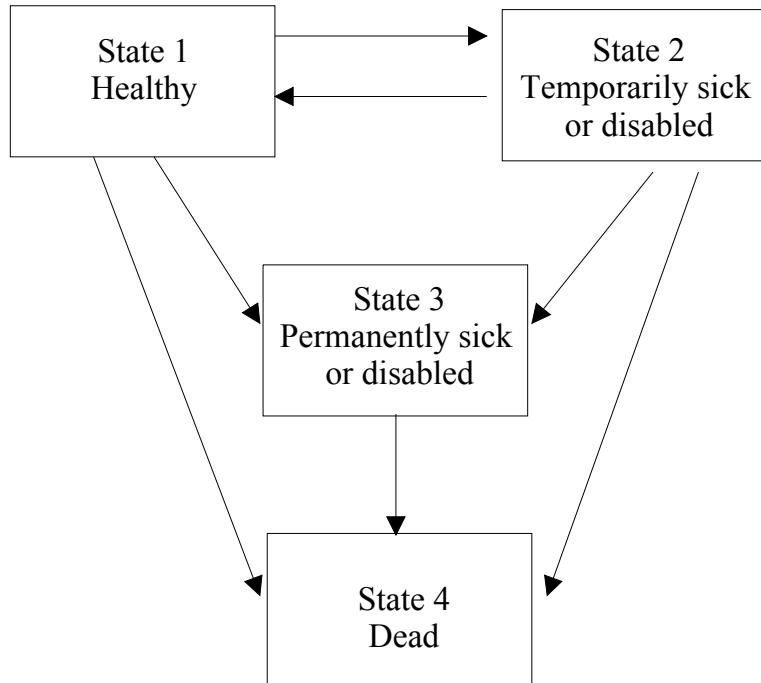
Upon leaving level 1, the probability that an employee moves to level 2 is 50%.

- (ii) Explain how you could model this as a Markov process, commenting on any assumptions that you make. [2]
- (iii) Derive the generator matrix of the Markov jump process. [2]
- (iv) The company currently has 1,000 employees. The proportions at levels 1 and 2 are 60% and 40% respectively. Assuming that nobody joins the company in the future, determine the distribution of these employees in five years' time. [6]

[Total 11]

Question 2.14

The following diagram represents a four-state Markov model.



The force of transition from state i to state j ($i \neq j$) at age x is denoted by μ_x^{ij} , and the probability that a life, who is in state i when aged x , will be in state j at age $x+t$ is ${}_t p_x^{ij}$.

- (i) Derive from first principles a differential equation for ${}_t p_x^{23}$, stating all assumptions made. [5]
- (ii) Given that, for $x = 40, 41$:

$${}_1 p_x^{12} = 0.03, {}_1 p_x^{13} = 0.002, {}_1 p_x^{14} = 0.001,$$

$${}_1 p_x^{21} = 0.4, {}_1 p_x^{23} = 0.1, {}_1 p_x^{24} = 0.01 \text{ and } {}_1 p_x^{34} = 0.3,$$

calculate ${}_2 p_{40}^{13}$. [2]

- (iii) An insurance company issues a combined sickness, disability and assurance contract that provides the following benefits:
- an income payable while the policyholder is temporarily sick or disabled; and
 - a lump sum payable either on becoming permanently sick or disabled, or on death.

The contract terminates as soon as the lump sum has been paid.

Explain how the model could be simplified for the purpose of modelling the claims process involved. State how your answer to (i) would be altered as a result of this change. (You are not required to derive this result from first principles). [2]

[Total 9]

Question 2.15

An investigator wishes to construct a multiple decrement model of the mortality of a population, subdivided by the following causes of death.

Cause 1: Cancer

Cause 2: Heart disease

Cause 3: All other causes

You are given the following definitions:

μ_x^i is the force of mortality due to cause i ($i = 1, 2, 3$) at exact age x

${}_u q_x^i$ is the probability that a life at exact age x dies due to cause i ($i = 1, 2, 3$) before reaching exact age $x + u$ ($u \geq 0$)

${}_u p_x$ is the probability that a life aged exactly x is still alive at exact age $x + u$ ($u > 0$)

You may assume that $\frac{u q_x^i}{u} \rightarrow \mu_x^i$ as $u \rightarrow 0$.

- (i) Derive an expression for ${}_t p_x$ ($t \geq 0$), in terms of the forces of mortality, using only the above functions in your derivation. [5]

- (ii) Write down a formula for q_x^i in terms of $_t p_x$ and the appropriate force(s) of mortality. (Note that $q_x^i = {}_1 q_x^i$.) [2]
- (iii) By reference to the expressions obtained in (i) and (ii) above, explain briefly why q_x^i is often referred to as a dependent rate of mortality. [2]
- (iv) Assuming that the force of mortality from each cause i is a constant μ^i between integer ages x and $x+1$, show that:

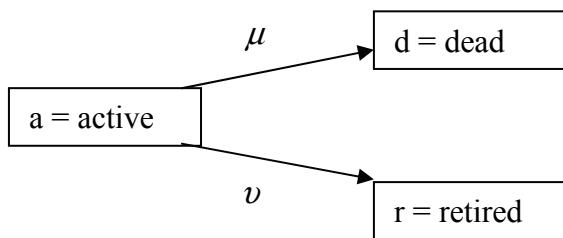
$$q_x^i = \frac{\mu^i}{\sum_{i=1}^3 \mu^i} \times q_x \text{ where } q_x = 1 - p_x \quad [5]$$

(Note that $p_x = {}_1 p_x$.)

[Total 14]

Question 2.16

- (i) The following multiple state model has been suggested as a representation of deaths and retirements between the ages of 59 and 60. There are no other decrements and the forces of decrement μ and ν are constant. Let $_t p_x^{ij}$ denote the probability that a life is in state j at age $x+t$ given that he was in state i at age x .



- (a) State the assumptions underlying the above model.
- (b) Show that $_t p_x^{aa} = e^{-(\mu+\nu)t}$ for $59 \leq x \leq x+t \leq 60$.

- (c) Suppose that you make the following observations in respect of n identical and statistically independent lives:

v = time spent in the active state

d = number of deaths

r = number of retirements

Assuming that lives are only observed to the earlier of death or retirement, show that the likelihood for μ and v given these observations is:

$$L(\mu, v) = e^{-(\mu+v)v} \mu^d v^r$$

- (d) Give formulae (without proof) for:

- the maximum likelihood estimator of the parameter v
- the asymptotic expected value of the estimator
- an estimate of the asymptotic variance of the estimator. [16]

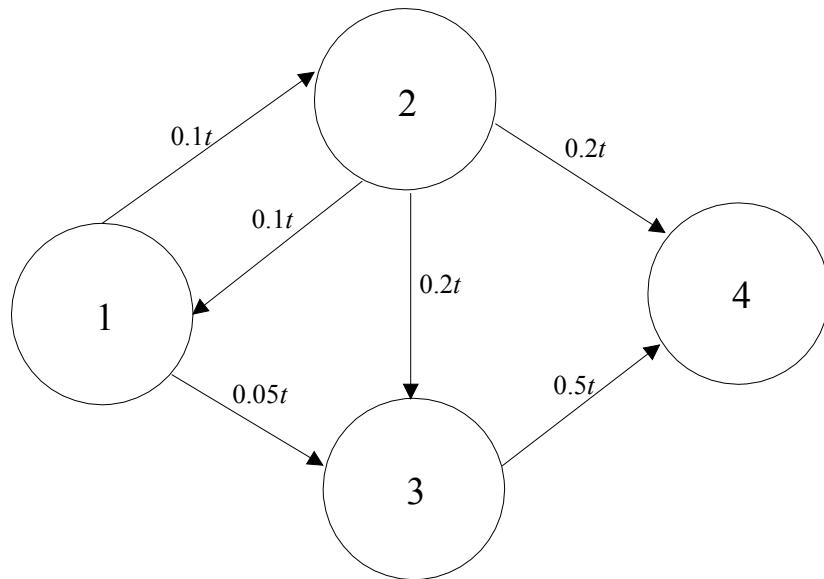
- (ii) Suppose that you learn that retirements can only take place on a birthday, so that r is the number of retirements at exact age 60. In addition to v , d and r you also observe:

m = number of lives attaining exact age 60, where $m \leq n$. Suppose that any life attaining exact age 60 will retire with probability k , where $0 < k < 1$.

- (a) State the likelihood for μ and k , given v , d , r and m .
- (b) Give a formula (without proof) for the maximum likelihood estimator of the parameter k . [4]
- [Total 20]

Question 2.17

Consider the following time-inhomogeneous Markov jump process with transition rates as shown below:



- (i) Write down the generator matrix at time t . [2]
- (ii) Write down the Kolmogorov backward differential equations for $P_{33}(s,t)$ and $P_{13}(s,t)$. [3]
- (iii) Using the technique of separation of variables, or otherwise, show that the solution of the differential equation for $P_{33}(s,t)$ is:

$$P_{33}(s,t) = e^{-0.25(t^2-s^2)} \quad [4]$$

- (iv) Show that the probability that the process visits neither state 2 nor state 4 by time t , given that it starts in state 1 at time 0, is:

$$\frac{8}{7}e^{-0.075t^2} - \frac{1}{7}e^{-0.25t^2} \quad [6]$$

- (v) State the limiting value as $t \rightarrow \infty$ of the probability in (iv). Explain why this must be the case for this particular model. [2]

[Total 17]

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Part 2 – Solutions

Solution 2.1

The distribution of the inter-arrival time random variable is $\text{Exp}(\lambda)$. [2]

Solution 2.2

The distribution of the number of occurrences in a time period t is $\text{Poisson}(\lambda t)$. So the probability of exactly one event is $\lambda t e^{-\lambda t}$. [2]

Solution 2.3

The number of claims over £1,000 is a Poisson process with parameter 4 per day. The holding time has an exponential distribution with parameter 4 so the expected wait is $\frac{1}{4}$ day or 6 hours. [2]

Solution 2.4

(i) ***Expected waiting time until the first claim of amount 30***

Claims of amount 30 occur according to a Poisson process with a mean of $0.1 \times 5 = 0.5$ per day. So the waiting time until the first claim of amount 30 has an $\text{Exp}(0.5)$ distribution and the expected waiting time is $\frac{1}{0.5} = 2$ days. [1]

(ii) ***Probability that there are at least 10 claims during the first 2 days, given that there were exactly 6 claims during the first day***

Let $N(t)$ denote the number of claims during the interval $[0, t]$. Then:

$$\begin{aligned} P(N(2) \geq 10 | N(1) = 6) &= P(N(2) - N(1) \geq 4 | N(1) - N(0) = 6) \\ &= P(N(2) - N(1) \geq 4) \end{aligned} \quad [1]$$

since $N(0) = 0$ and the numbers of claims in non-overlapping time intervals are independent.

Now $N(2) - N(1) \sim Poi(5)$, so:

$$\begin{aligned}
 P(N(2) \geq 10 | N(1) = 6) &= P(Poi(5) \geq 4) \\
 &= 1 - P(Poi(5) \leq 3) \\
 &= 1 - e^{-5} \left(1 + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} \right) \\
 &= 0.73497
 \end{aligned} \tag{1}$$

- (iii) ***Probability that there are at least 2 claims of amount 20 during the first day and at least 3 claims of amount 20 during the first 2 days***

Let $N_{20}(t)$ denote the number of claims of amount 20 in the interval $[0, t]$. We want:

$$P(N_{20}(1) \geq 2, N_{20}(2) \geq 3)$$

If we have 3 or more claims during the first day, then the second condition is automatically satisfied. If we have exactly 2 claims on the first day, then we need at least 1 claim on the second day. So the required probability is:

$$P(N_{20}(1) \geq 3) + P(N_{20}(1) = 2, N_{20}(2) - N_{20}(1) \geq 1) \tag{1}$$

Now $N_{20}(1)$ and $N_{20}(2) - N_{20}(1)$ are both Poisson with mean $0.7 \times 5 = 3.5$. Also, $N_{20}(1)$ and $N_{20}(2) - N_{20}(1)$ are independent. So:

$$P(N_{20}(1) \geq 3) = 1 - P(N_{20}(1) \leq 2) = 1 - e^{-3.5} \left(1 + \frac{3.5^1}{1!} + \frac{3.5^2}{2!} \right) = 0.679153 \quad [\frac{1}{2}]$$

and:

$$\begin{aligned}
 P(N_{20}(1) = 2, N_{20}(2) - N_{20}(1) \geq 1) &= P(N_{20}(1) = 2) P(N_{20}(2) - N_{20}(1) \geq 1) \\
 &= P(N_{20}(1) = 2) [1 - P(N_{20}(2) - N_{20}(1) = 0)] \\
 &= \frac{e^{-3.5} 3.5^2}{2!} [1 - e^{-3.5}] \\
 &= 0.179374
 \end{aligned} \tag{1}$$

The required probability is:

$$\begin{aligned} P(N_{20}(1) \geq 3) + P(N_{20}(1) = 2, N_{20}(2) - N_{20}(1) \geq 1) &= 0.679153 + 0.179374 \\ &= 0.85853 \end{aligned} \quad [\frac{1}{2}]$$

- (iv) ***Conditional variance of the number of claims during the first day, given that there are 2 claims of amount 10 during the first day***

Let $N_j(t)$, $j = 10, 20, 30$, denote the number of claims of amount j in the interval $[0, t]$. Then:

$$N(1) = N_{10}(1) + N_{20}(1) + N_{30}(1)$$

and:

$$\begin{aligned} \text{var}[N(1) | N_{10}(1) = 2] &= \text{var}[N_{10}(1) + N_{20}(1) + N_{30}(1) | N_{10}(1) = 2] \\ &= \text{var}[2 + N_{20}(1) + N_{30}(1)] \\ &= \text{var}[N_{20}(1) + N_{30}(1)] \\ &= \text{var}[N_{20}(1)] + \text{var}[N_{30}(1)] \end{aligned} \quad [2]$$

by independence.

Now, since $N_{20}(1) \sim \text{Poi}(3.5)$ and $N_{30}(1) \sim \text{Poi}(0.5)$:

$$\text{var}[N(1) | N_{10}(1) = 2] = 3.5 + 0.5 = 4 \quad [1]$$

Solution 2.5

The process defined is a Poisson process with parameter λ .

(i) **Transition probabilities**

Since the distribution of the increments is $\text{Poisson}(\lambda t)$, these are:

$$P_{ij}(t) = \frac{e^{-\lambda t} (\lambda t)^{j-i}}{(j-i)!} \quad \text{for } j \geq i \quad [2]$$

(ii) **Holding time**

The holding times are inter-event times. In other words, the time spent in a particular state between transitions. For the process given, the i th holding time T_{i-1} is the time spent in state $i-1$ before the transition to state i . [1]

(iii) **Distribution of the first holding time**

We have:

$$P[T_0 > t | X_0 = 0] = P[X_t = 0 | X_0 = 0] = P_{00}(t) = e^{-\lambda t} \quad [1]$$

and it follows that T_0 has an $\text{Exp}(\lambda)$ distribution. [1]

(iv) **Value of X_{T_0}**

$X_{T_0} = 1$ since we choose the sample paths to be right-continuous. So at time T_0 it has just jumped to 1. [1]

(v) **Distribution of i th holding time**

Consider T_i :

$$\begin{aligned} P\left[T_i > t | X_0 = 0, \sum_{j=0}^{i-1} T_j = s\right] &= P\left[X_{t+s} - X_s = 0 | X_0 = 0, \sum_{j=0}^{i-1} T_j = s\right] \\ &= P[X_t - X_0 = 0] = P_{00}(t) = e^{-\lambda t} \end{aligned} \quad [2]$$

The second equality is due to the increments being independent and stationary. Hence T_i also has an exponential distribution with parameter λ . [1]

Solution 2.6

(i) **Transition rates**

Assume that the functions $p_{ij}(s, t)$ are continuously differentiable. The transition rates are defined by differentiation with respect to t :

$$\mu_{ij}(s) = \left[\frac{\partial}{\partial t} p_{ij}(s, t) \right]_{t=s} = \lim_{h \rightarrow 0} \frac{p_{ij}(s, s+h) - p_{ij}(s, s)}{h} \quad [2]$$

(ii) **Sum of transition rates is zero**

Since the sum of probabilities from any fixed state i to all other states j must be 1 we have:

$$\sum_{j \in S} \mu_{ij}(s) = \sum_{j \in S} \left[\frac{\partial}{\partial t} p_{ij}(s, t) \right]_{t=s} = \left[\frac{\partial}{\partial t} \sum_{j \in S} p_{ij}(s, t) \right]_{t=s} = \left[\frac{\partial}{\partial t} 1 \right]_{t=s} = 0 \quad [2]$$

(iii) **Chapman-Kolmogorov equations**

The Chapman-Kolmogorov equations are:

$$p_{ij}(s, t) = \sum_{k \in S} p_{ik}(s, u) p_{kj}(u, t) \quad [\frac{1}{2}]$$

To obtain the forward equations we differentiate with respect to t and evaluate at $u = t$:

$$\frac{\partial}{\partial t} p_{ij}(s, t) = \sum_{k \in S} \left[p_{ik}(s, u) \left(\frac{\partial}{\partial t} p_{kj}(u, t) \right) \right]_{u=t} = \sum_{k \in S} p_{ik}(s, t) \mu_{kj}(t) \quad [1]$$

Similarly the backward equations are obtained by differentiating with respect to s and setting $u = s$:

$$\frac{\partial}{\partial s} p_{ij}(s, t) = \sum_{k \in S} \left[\left(\frac{\partial}{\partial s} p_{ik}(s, u) \right) p_{kj}(u, t) \right]_{u=s} = - \sum_{k \in S} \mu_{ik}(s) p_{kj}(s, t) \quad [1]$$

We now need to explain where the minus sign in the RHS comes from.

The definition of the transition rates is such that:

$$p_{ik}(s, s+h) = \delta_{ik} + h\mu_{ik}(s) + o(h)$$

or equivalently:

$$p_{ik}(s-h, s) = \delta_{ik} + h\mu_{ik}(s-h) + o(h)$$

Rearranging this gives:

$$\mu_{ik}(s-h) = \frac{p_{ik}(s-h, s) - \delta_{ik} - o(h)}{h} \quad [1/2]$$

Now taking the limit of both sides as $h \rightarrow 0$ and noting that $p_{ik}(s, s) = \delta_{ik}$, we get:

$$\mu_{ik}(s) = -\lim_{h \rightarrow 0} \frac{p_{ik}(s-h, s) - p_{ik}(s, s) - o(h)}{-h} = -\left[\frac{\partial}{\partial s} p_{ik}(s, t) \right]_{t=s} \quad [1]$$

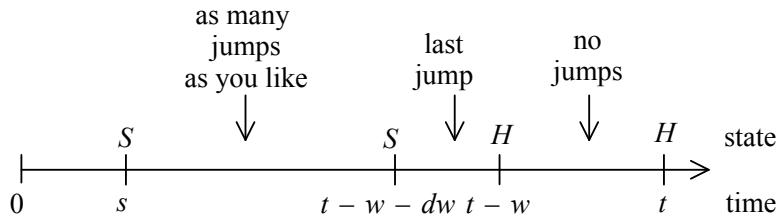
(iv) ***Usefulness of forward and backward equations***

We normally wish to calculate transition probabilities. This cannot usually be done directly from the transition matrix. However, the complete set of Chapman-Kolmogorov equations gives us a set of simultaneous differential equations, which we can solve fairly easily (usually by numerical methods run on a computer). [2]

Solution 2.7

- I Correct since $p_{DS}(u, t) = 0$ the given equation is the standard form of the Chapman-Kolmogorov equation. [1]
- II This is the correct form of the backward equation. [1]
- III This looks like the forward equation but is not valid. The exponential part should refer to the probability that we remain continuously in S from $t-w$ to t . However, the transition rate for this is $-(\rho(u) + \nu(u))$ from the matrix. [1]

When verifying equations like those in (II) and (III) it is often helpful to quickly sketch the transition graph and draw a time line diagram for the process.

Solution 2.8

The above diagram helps us to understand the transition probability $p_{SH}(s,t)$. At time s the process is in state S , *i.e.* sick, and at a later time t it is back in state H . In order for this to occur there must be at least one transition during the period (s,t) . We set up one possible path from time s to time t by conditioning on the last jump before time t occurring in the interval $(t-w-dw, t-w)$. [1]

The overall probability given the last transition is at time $t-w$ is therefore composed of:

- $p_{SS}(s, t-w-dw)$ the probability of starting at time s in state S and being in state S at time $t-w-dw$. [1]
- The transition probability at time $t-w-dw$ of going from S to H given that there is a transition. We can assume the transition takes place over a small time interval dw , then we can approximate this probability by using the transition rate from state S to state H at time $(t-w-dw)$:

$$\rho(t-w-dw) dw \quad [1]$$

- Finally, the process is continuously in state H from time $t-w$ until time t . This is the occupancy or survival probability conditional on being in state H at time $t-w$:

$$p_{HH}(t-w, t) = e^{-\int_{t-w}^t (\sigma(u) + \mu(u)) du} \quad [2]$$

Putting these three probabilities together we obtain:

$$p_{SS}(s, t-w) \rho(t-w-dw) e^{-\int_{t-w}^t (\sigma(u) + \mu(u)) du} dw \quad [1]$$

Letting dw become infinitesimally small and integrating over the possibilities for w , *i.e.* paths from state S at time s to state H at time t , we get:

$$p_{SH}(s, t) = \int_{w=0}^{w=t-s} p_{SS}(s, t-w) \rho(t-w-dw) e^{-\int_{t-w}^t (\sigma(u) + \mu(u)) du} dw \quad [1]$$

Solution 2.9(i) ***Matrix of transition rates***

The transition matrix is:

$$A(t) = \begin{pmatrix} -\sigma(t) - \mu(t) & \sigma(t) & \mu(t) \\ \rho(t) & -\rho(t) - \nu(t) & \nu(t) \\ 0 & 0 & 0 \end{pmatrix} \quad [2]$$

(ii) ***Residual holding time***

The residual holding time at time s is the random variable representing the remaining time until the next jump. [1]

(iii) ***Integral expression***

The required expression is:

$$P[R_s > w | X_s = S] = e^{-\int_s^{s+w} (\rho(u) + \nu(u)) du} \quad [2]$$

(iv) ***Probability density function***

By differentiating the probability in (iii), we obtain:

$$f_{R_s}(w) = (\rho(s+w) + \nu(s+w)) e^{-\int_s^{s+w} (\rho(u) + \nu(u)) du}, \quad w > 0 \quad [2]$$

(v) ***Conditional probability of transition to state S***

We want the ratio of the rate from H to S to the rate from H to either S or D. We don't consider H to H since we have been told that a transition has taken place. The required probability is:

$$\frac{\mu_{HS}(t)}{\mu_{HS}(t) + \mu_{HD}(t)} = \frac{\sigma(t)}{\sigma(t) + \mu(t)} \quad [2]$$

(vi) ***Backward Kolmogorov equation***

Considering the two possible states the first transition could move to:

$$\begin{aligned}
 p_{SD}(s, t) &= \int_0^{t-s} e^{-\int_s^{s+w} (\rho(u) + \nu(u)) du} \nu(s+w) p_{DD}(s+w, t) dw \\
 &\quad + \int_0^{t-s} e^{-\int_s^{s+w} (\rho(u) + \nu(u)) du} \rho(s+w) p_{HD}(s+w, t) dw \\
 &= \int_0^{t-s} e^{-\int_s^{s+w} (\rho(u) + \nu(u)) du} \nu(s+w) dw \\
 &\quad + \int_0^{t-s} e^{-\int_s^{s+w} (\rho(u) + \nu(u)) du} \rho(s+w) p_{HD}(s+w, t) dw \quad [5]
 \end{aligned}$$

(vii) ***General reasoning explanation***

The backward equation is constructed by conditioning on the first transition time. Let $s+w$ be the time of the first transition from state S. Let this transition be to state k , which can be either H or D. [1]

There are three stages to the process:

The process is in state S from time s to $s+w$. This has probability density:

$$f_{R_s}(w) = (\rho(s+w) + \nu(s+w)) e^{-\int_s^{s+w} (\rho(u) + \nu(u)) du} \quad [1]$$

There is then the assumed transition from S to k , which has probability:

$$\frac{\mu_{Sk}(s+w)}{\rho(s+w) + \nu(s+w)} \quad [1]$$

Finally we need the probability $p_{kD}(s+w, t)$ for going from state k to state D.

Putting these together and integrating over the possible times for w we obtain the given expression. We can simplify using the fact that $p_{DD}(s,t) = 1$. [1]

Solution 2.10

The probability of surviving the year is:

$$p_{50} = \exp\left(-\int_{50}^{51} 1.03(1.01)^t dt\right) \quad [\frac{1}{2}]$$

Noting that $(1.01)^t = e^{t \ln 1.01}$, this is an exponential integral:

$$\int_{50}^{51} 1.03(1.01)^t dt = 1.03 \left[\frac{(1.01)^t}{\ln 1.01} \right]_{50}^{51} = 1.70242 \quad [1]$$

So the probability of dying within the year is:

$$1 - e^{-1.70242} = 0.818 \quad [\frac{1}{2}]$$

Solution 2.11(i) ***Simulation of Poisson process***

The times (measured in seconds) between the arrival of hits will be a series of independent exponential random variables with parameter $\lambda = 2$. [1]

We can simulate these by inverting the cumulative density function of an $Exp(2)$ random variable. The steps in the process are:

(1) Generate a $U(0,1)$ random deviate. [½]

(2) Then:

$$u = F_X(x) = 1 - e^{-2x}$$

$$\Rightarrow x = -\frac{1}{2} \log_e(1-u)$$

where x is the simulated value of the $Exp(2)$ random variable. [1]

(3) The cumulative sum of a series of these random deviates will be the simulated arrival times of the hits (in seconds). [½]

(ii) ***Verifying the model***

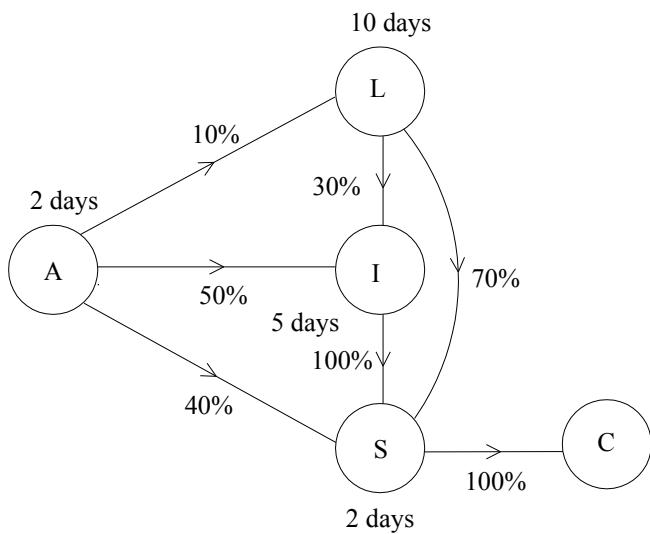
We identify some key characteristics of the process, eg average number of hits per minute. [½]

We calculate summary frequency distributions of these characteristics from the simulated data, and compare them with the frequency distributions calculated from observations on the real process. [1]

If these distributions are in close agreement for all the important characteristics of the process, then we have verified the model. [½]

Solution 2.12(i) **Generator matrix**

The information given in the question about the occupancy times in each state and the transition probabilities in the Markov jump chain can be summarised as:



The generator matrix is:

$$\begin{array}{ccccc}
 & A & L & I & S & C \\
 \begin{matrix} & -0.50 & \frac{1}{2} \times 0.10 & \frac{1}{2} \times 0.50 & \frac{1}{2} \times 0.40 & 0 \\ & 0 & -0.1 & \frac{1}{10} \times 0.30 & \frac{1}{10} \times 0.70 & 0 \\ & 0 & 0 & -0.20 & \frac{1}{5} \times 1.00 & 0 \\ & 0 & 0 & 0 & -0.50 & \frac{1}{2} \times 1.00 \\ & 0 & 0 & 0 & 0 & 0 \end{matrix} & = & \begin{bmatrix} -0.50 & 0.05 & 0.25 & 0.20 & 0 \\ 0 & -0.1 & 0.03 & 0.07 & 0 \\ 0 & 0 & -0.20 & 0.20 & 0 \\ 0 & 0 & 0 & -0.50 & 0.50 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} & [2]
 \end{array}$$

(ii) ***Proportion of claims that require further details from the insured***

We can list all the paths that correspond to the event of visiting state I if the process starts in state A .

These are $A \rightarrow L \rightarrow I \rightarrow S \rightarrow C$ and $A \rightarrow I \rightarrow S \rightarrow C$. [1]

The probabilities of these paths are $0.10 \times 0.30 \times 1 \times 1 = 0.03$ and $0.50 \times 1 \times 1 = 0.50$. The total probability is 0.53. [1]

Alternatively we can use a more general approach. This has the advantage of working in more complicated situations where the path counting approach becomes very cumbersome.

Let $p_i = P[\text{never visit state } I \mid \text{currently in state } i]$, then using the Markov jump chain transition matrix:

$$\begin{bmatrix} 0 & 0.10 & 0.50 & 0.40 & 0 \\ 0 & 0 & 0.30 & 0.70 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

we can write:

$$p_A = 0.1p_L + 0.5p_I + 0.4p_S = 0.07 + 0.40 = 0.47$$

$$p_L = 0.3p_I + 0.7p_S = 0.7$$

$$p_I = 0$$

$$p_S = 1$$

$$p_C = 1$$

The required probability is $1 - p_A = 1 - 0.47 = 0.53$.

- (iii) **Probability that a claim is yet to be logged and classified by a claims administrator at time t**

The Chapman-Kolmogorov equation is:

$$P_{AA}(t+dt) = P_{AA}(t)P_{AA}(dt)$$

assuming that the process satisfies the Markov property.

[½]

Then the law of probability allows us to write:

$$P_{AA}(dt) + P_{AL}(dt) + P_{AI}(dt) + P_{AS}(dt) + P_{AC}(dt) = 1 \quad [½]$$

The definition of the transition rates gives:

$$P_{AL}(dt) \approx 0.05dt + o(dt)$$

$$P_{AI}(dt) \approx 0.25dt + o(dt)$$

$$P_{AS}(dt) \approx 0.20dt + o(dt) \quad [½]$$

Also, $P_{AC}(dt) = o(dt)$ because it involves more than one transition.

[½]

Substituting we obtain:

$$P_{AA}(dt) = 1 - 0.05dt - 0.25dt - 0.20dt + o(dt)$$

$$\Rightarrow P_{AA}(t+dt) = P_{AA}(t)(1 - 0.05dt - 0.25dt - 0.20dt + o(dt))$$

$$\Rightarrow \frac{P_{AA}(t+dt) - P_{AA}(t)}{dt} = -0.50P_{AA}(t) + \frac{o(dt)}{dt}$$

$$\Rightarrow \frac{dP_{AA}(t)}{dt} = -0.50P_{AA}(t) \quad [1]$$

Using an integrating factor of $e^{0.5t}$ and a boundary condition $P_{AA}(0) = 1$ gives the solution:

$$P_{AA}(t) = e^{-0.50t} \quad [1]$$

(iv) ***Backward integral equation***

We add the probabilities of the mutually exclusive and exhaustive events representing the two possible paths from state A at time 0 to state I at time t with a first jump out of state A in the time interval $(w, w+dw)$. Then: [½]

$$P_{AI}(t) = \int_{w=0}^t P_{\overline{AA}}(w) 0.05 P_{LI}(t-w) dw + \int_{w=0}^t P_{\overline{AA}}(w) 0.25 P_{II}(t-w) dw \quad [1]$$

$$P_{\overline{AA}}(t) = e^{-0.50t}$$

and:

$$P_{II}(t) = e^{-0.20t}$$

So we can write: [½]

$$\begin{aligned} P_{AI}(t) &= \int_{w=0}^t e^{-0.50w} 0.05 P_{LI}(t-w) dw + \int_{w=0}^t e^{-0.50w} 0.25 e^{-0.20(t-w)} dw \\ &= 0.05 \int_{w=0}^t e^{-0.50w} P_{LI}(t-w) dw + 0.25 e^{-0.20t} \int_{w=0}^t e^{-0.30w} dw \end{aligned} \quad [1]$$

(v) ***Expression for $P_{AI}(t)$***

To evaluate this expression we need an integral expression for $P_{LI}(t)$. There is one possible path from state L at time 0 to state I at time t with a first jump out of state L in the time interval $(w, w+dw)$.

We write:

$$P_{LI}(t) = \int_{w=0}^t P_{\overline{LL}}(w) 0.03 P_{II}(t-w) dw \quad [1]$$

But $P_{LL}(t) = e^{-0.10t}$ and $P_{II}(t) = e^{-0.20t}$. So we can write:

$$\begin{aligned}
 P_{LI}(t) &= \int_{w=0}^t P_{LL}(w) 0.03 P_{II}(t-w) dw \\
 &= \int_{w=0}^t e^{-0.10w} 0.03 e^{-0.20(t-w)} dw \\
 &= 0.03 e^{-0.20t} \int_{w=0}^t e^{0.10w} dw \\
 &= \frac{3}{10} e^{-0.20t} (e^{0.10t} - 1) \\
 &= \frac{3}{10} e^{-0.10t} - \frac{3}{10} e^{-0.20t} \tag{2}
 \end{aligned}$$

Then substituting we can write:

$$\begin{aligned}
 P_{AI}(t) &= 0.05 \int_{w=0}^t e^{-0.50w} P_{LI}(t-w) dw + 0.25 e^{-0.20t} \int_{w=0}^t e^{-0.30w} dw \\
 &= \frac{0.15}{10} e^{-0.10t} \int_{w=0}^t e^{-0.40w} dw - \frac{0.15}{10} e^{-0.20t} \int_{w=0}^t e^{-0.30w} dw \\
 &\quad + 0.25 e^{-0.20t} \int_{w=0}^t e^{-0.30w} dw \\
 &= \frac{15}{1,000} e^{-0.10t} \int_{w=0}^t e^{-0.40w} dw + \frac{235}{1,000} e^{-0.20t} \int_{w=0}^t e^{-0.30w} dw \\
 &= -\frac{15}{400} e^{-0.10t} (e^{-0.40t} - 1) - \frac{235}{300} e^{-0.20t} (e^{-0.30t} - 1) \\
 &= -\left(\frac{15}{400} + \frac{235}{300}\right) e^{-0.50t} + \frac{15}{400} e^{-0.10t} + \frac{235}{300} e^{-0.20t} \\
 &= \frac{3}{80} e^{-0.10t} + \frac{47}{60} e^{-0.20t} - \frac{197}{240} e^{-0.50t} \tag{3}
 \end{aligned}$$

(vi) ***Probability that a claim is awaiting more details from the insured at time 10***

$$\begin{aligned}
 P_{AI}(10) &= \frac{3}{80}e^{-0.10 \times 10} + \frac{47}{60}e^{-0.20 \times 10} - \frac{197}{240}e^{-0.50 \times 10} \\
 &= 0.0138 + 0.1060 - 0.0055 \\
 &= 0.1143
 \end{aligned}
 \quad [1]$$

(vii) ***Verifying the time homogeneous model***

The time homogeneous model has transition rates that are constant over time. A consequence of this is that the waiting times in each state follow an exponential distribution with a parameter equal to the total transition rate out of the state. [1]

We could check this by recording data on the times spent in each state by claims during the settlement process. We could then carry out a chi-squared goodness-of-fit test to check if the observed data conformed to the distributions implied by the model, *ie* time in $A \sim \text{Exp}(0.5)$, time in $L \sim \text{Exp}(0.1)$, time in $I \sim \text{Exp}(0.2)$ and time in $S \sim \text{Exp}(0.5)$. [1]

Solution 2.13(i) ***Equivalent first condition for stationary***

We can differentiate the equation $\pi P(t) = \pi$ with respect to t , and then set $t = 0$ afterward (since $\Sigma = P'(0)$). Since the distribution is stationary, the vector π is constant and the derivative of the RHS is zero. [1]

(ii) ***Modelling as a Markov process***

This is a 3-state Markov jump process. The states are (1) level 1; (2) level 2; (3) left the company. [½]

We have made the Markov assumption, *ie* that the probability of jumping to any particular state depends only on knowing the current state that is occupied. [1]

We have assumed that transition rates between states are constant over time. [½]

(iii) ***Generator matrix***

The average waiting time in each state, i is exponentially distributed with mean $\frac{1}{\lambda_i}$.

The mean times in states 1 and 2 are 2 and 5 years respectively. The values of the exponential parameters are:

$$\lambda_1 = \frac{1}{2} \quad \lambda_2 = \frac{1}{5} \quad [\frac{1}{2}]$$

The transition matrix of the jump chain, p_{ij} is:

$$\begin{array}{ccc} \text{level 1} & \text{level 2} & \text{left} \\ \left[\begin{array}{ccc} 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array} \right] & & [\frac{1}{2}] \end{array}$$

The off-diagonal elements of the matrix of transition rates, μ_{ij} are given by:

$$\mu_{ij} = \lambda_i p_{ij}$$

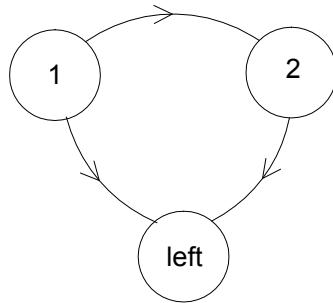
The diagonal elements are chosen to make the row sums of the matrix equal to 0.

The generator matrix (matrix of transition rates) is:

$$\begin{array}{ccc} \text{level 1} & \text{level 1} & \text{left} \\ \left[\begin{array}{ccc} -0.50 & 0.25 & 0.25 \\ 0 & -0.20 & 0.20 \\ 0 & 0 & 0 \end{array} \right] & & [1] \end{array}$$

(iv) ***Distribution of employees in five years***

The model is:



We are given the initial state as $[600 \quad 400 \quad 0]$. We can use the five-year transition probabilities to estimate the numbers in each state in five years' time.

The number in state 1 will be:

$$600P_{11}(5)$$

The number in state 2 will be:

$$600P_{12}(5) + 400P_{22}(5)$$

And the number of employees who have left the company can be obtained by deducting the numbers in states 1 and 2 from 1000.

The occupancy probabilities for states 1 and 2 are given by:

$$P_{11}(t) = P_{\overline{11}}(t) = e^{-0.5t} \quad [\frac{1}{2}]$$

$$P_{22}(t) = P_{\overline{22}}(t) = e^{-0.2t} \quad [\frac{1}{2}]$$

Using the generator matrix, we can write the forward differential equation for $P_{12}(t)$:

$$\begin{aligned} \frac{d}{dt}P_{12}(t) &= 0.25P_{11}(t) - 0.2P_{12}(t) \\ \Rightarrow \quad \frac{d}{dt}P_{12}(t) + 0.2P_{12}(t) &= 0.25e^{-0.5t} \end{aligned} \quad [1]$$

This can be solved using the integrating factor method. The integrating factor is $e^{0.2t}$. Multiplying through by the integrating factor gives:

$$e^{0.2t} \frac{d}{dt} P_{12}(t) + 0.2e^{0.2t} P_{12}(t) = 0.25e^{-0.5t} e^{0.2t} = 0.25e^{-0.3t} \quad [\frac{1}{2}]$$

Integrating both sides:

$$e^{0.2t} P_{12}(t) = -\frac{5}{6}e^{-0.3t} + C \quad [\frac{1}{2}]$$

The boundary condition is $P_{12}(0) = 0$. So:

$$0 = -\frac{5}{6} + C \Rightarrow C = \frac{5}{6} \quad [\frac{1}{2}]$$

Simplifying then gives:

$$P_{12}(t) = \frac{5}{6} \left(e^{-0.2t} - e^{-0.5t} \right) \quad [\frac{1}{2}]$$

So the number of employees on level 1 in 5 years' time is:

$$600P_{11}(5) = 600e^{-2.5} = 49 \quad [\frac{1}{2}]$$

and the number of employees on level 2 in 5 years' time is:

$$600P_{12}(5) + 400P_{22}(5) = 600 \times \frac{5}{6} \left(e^{-1} - e^{-2.5} \right) + 400e^{-1} = 290 \quad [1]$$

The number of lives who have left the company is:

$$1,000 - 49 - 290 = 661 \quad [\frac{1}{2}]$$

Solution 2.14(i) **Differential equation for ${}_t p_x^{23}$**

Using the Markov assumption, which says that the probabilities of being found in any state at any future age depend only on the ages involved and the current state occupied ... [½]

... we can write:

$${}_{t+h} p_x^{23} = {}_t p_x^{21} \cdot {}_h p_{x+t}^{13} + {}_t p_x^{22} \cdot {}_h p_{x+t}^{23} + {}_t p_x^{23} \cdot {}_h p_{x+t}^{33} + {}_t p_x^{24} \cdot {}_h p_{x+t}^{43} \quad [1]$$

According to the law of total probability: ${}_h p_{x+t}^{33} = 1 - {}_h p_{x+t}^{34}$ [½]

Assuming that, for $(i \neq j)$, ${}_h p_{x+t}^{ij} = h \cdot \mu_{x+t}^{ij} + o(h)$ [½]

... where $\lim_{h \rightarrow 0+} \frac{o(h)}{h} = 0$ [½]

... and noting that ${}_t p_x^{43} = 0$ [½]

... then ${}_{t+h} p_x^{23} = {}_t p_x^{21} \cdot h \cdot \mu_{x+t}^{13} + {}_t p_x^{22} \cdot h \cdot \mu_{x+t}^{23} + {}_t p_x^{23} \left(1 - h \cdot \mu_{x+t}^{34}\right) + o(h)$. [½]

So:

$$\frac{{}_{t+h} p_x^{23} - {}_t p_x^{23}}{h} = {}_t p_x^{21} \cdot \mu_{x+t}^{13} + {}_t p_x^{22} \cdot \mu_{x+t}^{23} - {}_t p_x^{23} \cdot \mu_{x+t}^{34} + \frac{o(h)}{h} \quad [½]$$

and:

$$\frac{\partial}{\partial t} {}_t p_x^{23} = \lim_{h \rightarrow 0+} \frac{{}_{t+h} p_x^{23} - {}_t p_x^{23}}{h} = {}_t p_x^{21} \cdot \mu_{x+t}^{13} + {}_t p_x^{22} \cdot \mu_{x+t}^{23} - {}_t p_x^{23} \cdot \mu_{x+t}^{34} \quad [½]$$

(ii) **Calculate ${}_2 p_{40}^{13}$**

$${}_2 p_{40}^{13} = \left({}_1 p_{40}^{11}\right) \left({}_1 p_{41}^{13}\right) + \left({}_1 p_{40}^{12}\right) \left({}_1 p_{41}^{23}\right) + \left({}_1 p_{40}^{13}\right) \left({}_1 p_{41}^{33}\right) \quad [1]$$

$$= (1 - 0.03 - 0.002 - 0.001) \times 0.002 + 0.03 \times 0.1 + 0.002 \times (1 - 0.3) = 0.006334 \quad [1]$$

(iii) ***Simplified model?***

The transition from permanently sick and disabled to dead is not necessary for the modelling of the claims process, because this transition has no effect on the incidence of any claim payments. (Note that a lump sum claim is payable only on transitions from state 1 to 3, 1 to 4, 2 to 3 or 2 to 4; no payment is made on transition from state 3 to 4.)

[1]

The revised differential equation would be:

$$\frac{\partial}{\partial t} {}_t p_x^{23} = {}_t p_x^{21} \cdot \mu_{x+t}^{13} + {}_t p_x^{22} \cdot \mu_{x+t}^{23} \quad [1]$$

Solution 2.15(i) ***Derivation***

The probability that an individual survives a period is one minus the probability that they die. Consider a short time interval of length h .

This means that:

$${}_h p_{x+u} = 1 - \sum_{i=1}^3 {}_h q_{x+u}^i \quad [1]$$

The question tells us that:

$${}_h q_{x+u}^i = h \mu_{x+u}^i + o(h)$$

Substituting this into the first expression gives:

$${}_h p_{x+u} = 1 - h \sum_{i=1}^3 \mu_{x+u}^i + o(h)$$

From the Markov assumption, we know that ${}_{u+h} p_x = {}_u p_x \times {}_h p_{x+u}$. [1]

Hence:

$$\frac{{}_{u+h} p_x - {}_u p_x}{h} = {}_u p_x \left(- \sum_{i=1}^3 \mu_{x+u}^i + \frac{o(h)}{h} \right) \quad [1]$$

Letting h tend to zero gives:

$$\frac{\partial}{\partial u} {}_u p_x = - {}_u p_x \sum_{i=1}^3 \mu_{x+u}^i \quad [\frac{1}{2}]$$

Dividing both sides by ${}_u p_x$ gives:

$$\frac{\partial}{\partial u} \log {}_u p_x = - \sum_{i=1}^3 \mu_{x+u}^i \quad [\frac{1}{2}]$$

Integrating gives:

$$\log {}_t p_x - \log {}_0 p_x = - \int_0^t \sum_{i=1}^3 \mu_{x+u}^i du \quad [\frac{1}{2}]$$

Since $\log {}_0 p_x = \log 1 = 0$, we find that:

$${}_t p_x = \exp \left(- \int_0^t \sum_{i=1}^3 \mu_{x+u}^i du \right) \quad [\frac{1}{2}]$$

(ii) **Formula**

Note that q_x^i is just ${}_t q_x^i$ when $t = 1$, ie it is the probability that an individual aged x leaves through cause i during the coming year. Expressed as an integral, this is:

$$q_x^i = \int_0^1 {}_t p_x \mu_{x+t}^i dt$$

since an individual who leaves through cause i during the year must survive all decrements up to a time t (in the range $0 < t < 1$), then must leave through cause i during the time interval $(t, t + dt)$. [2]

(iii) **Why called dependent rate of mortality**

Suppose that one of the rates of decrement (say decrement 2) changes. Then we see from the formula in part (i) that the value of ${}_t p_x$ will change. The formula in part (ii) then tells us that q_x^3 (say) will also change (even though μ_x^3 has not changed).

So the decrements are functionally dependent on each other.

It seems fairly logical that if the rate of heart disease increases then you are less likely to die of other causes (since you will only die once and the chance of death being caused by heart disease has increased). [2]

(iv) ***Proof***

Combining the formulae derived in parts (i) and (ii), we have:

$$q_x^i = \int_0^1 p_x \mu_{x+t}^i dt = \int_0^1 \exp\left(-\int_0^t \sum_{i=1}^3 \mu_{x+u}^i du\right) \mu_{x+t}^i dt \quad [2]$$

Since the force of mortality is constant over the year, this is:

$$q_x^i = \mu^i \int_0^1 \exp\left(-\int_0^t \sum_{i=1}^3 \mu^i du\right) dt = \mu^i \int_0^1 \exp\left(-t \sum_{i=1}^3 \mu^i\right) dt \quad [1]$$

Integrating gives:

$$q_x^i = \frac{\mu^i}{\sum_{i=1}^3 \mu^i} \left[-\exp\left(-t \sum_{i=1}^3 \mu^i\right) \right]_0^1 = \frac{\mu^i}{\sum_{i=1}^3 \mu^i} \left[1 - \exp\left(-\sum_{i=1}^3 \mu^i\right) \right] \quad [1]$$

This simplifies to:

$$q_x^i = \frac{\mu^i}{\sum_{i=1}^3 \mu^i} [1 - p_x] = \frac{\mu^i}{\sum_{i=1}^3 \mu^i} \times q_x \quad [1]$$

Solution 2.16

In this question you have to be very careful not to mix up your v 's (V for Victor) and your ν 's (the Greek letter "nu").

(i)(a) **Assumptions**

The lives are independent and identical. [½]

The transition probabilities depend only upon the individual's current state. They do not depend upon the previous transitions for the individual. [½]

These probabilities are given by:

$${}_t p_x^{ad} = \mu t + o(t)$$

and: ${}_t p_x^{ar} = \nu t + o(t)$ for $59 \leq x \leq x+t \leq 60$ [1]

(i)(b) **Proof**

A life who remains active for $t+h$ years must first remain active for t years, then remain active for a further h years (where h represents a short time interval).

Expressed in terms of probabilities, this is:

$${}_{t+h} p_x^{aa} = {}_t p_x^{aa} \times {}_h p_{x+t}^{aa} \quad [1]$$

This follows from the "Markov" property, ie that the probabilities in different time periods are independent of each other. [1]

During a short time period $(t, t+h)$, an active life must remain active, die or retire.

So:

$${}_h p_{x+t}^{aa} + {}_h p_{x+t}^{ad} + {}_h p_{x+t}^{ar} = 1 \quad [1]$$

If we assume that the probability of two or more transitions in the short time h is $o(h)$, then we can express the transition probabilities in terms of the forces of transition:

$${}_h p_{x+t}^{aa} + \mu h + \nu h + o(h) = 1 \quad [½]$$

Substituting this into the first equation, we find that:

$${}_{t+h} p_x^{aa} = {}_t p_x^{aa} \times [1 - h(\mu + v) + o(h)] \quad [\frac{1}{2}]$$

Rearranging and letting $h \rightarrow 0$ gives:

$$\frac{\partial}{\partial t} {}_t p_x^{aa} = -(\mu + v) {}_t p_x^{aa} \quad [\frac{1}{2}]$$

ie:

$$\frac{\partial}{\partial t} \log {}_t p_x^{aa} = -(\mu + v) \quad [\frac{1}{2}]$$

Integrating with respect to t with limits of 0 and s :

$$\left[\log {}_t p_x^{aa} \right]_0^s = -(\mu + v)s \quad [\frac{1}{2}]$$

Hence:

$${}_s p_x^{aa} = e^{-(\mu + v)s} \quad [\frac{1}{2}]$$

for $59 \leq x \leq x + s \leq 60$.

(i)(c) **Likelihood**

Here are two possible approaches to this part.

During the year, individual i will either survive to the end, die or retire. Using the result in (i)(b) and writing t_i for this individual's waiting time in the active state, the likelihood corresponding to each of these is:

Survival: $e^{-(\mu + v)t_i}$

Death: $e^{-(\mu + v)t_i} \times \mu$

Retirement: $e^{-(\mu + v)t_i} \times v$

Since the experiences of the individuals are assumed to be independent, the overall likelihood for all the lives will be:

$$L(\mu, v) = \prod_{survivors} e^{-(\mu + v)t_i} \times \prod_{deaths} \mu e^{-(\mu + v)t_i} \times \prod_{retirements} v e^{-(\mu + v)t_i} \quad [3]$$

This can be simplified to give:

$$\begin{aligned} L(\mu, \nu) &= \prod_{all\ lives} e^{-(\mu+\nu)t_i} \times \prod_{deaths} \mu \times \prod_{retirements} \nu \\ &= e^{-(\mu+\nu)\sum t_i} \times \mu^d \times \nu^r = e^{-(\mu+\nu)v} \times \mu^d \times \nu^r \end{aligned} \quad [2]$$

Alternatively, we can write down the probability density/mass function for life i as a single function:

$$f_i(d_i, r_i, v_i) = \begin{cases} v_i p_x & (d_i = 0, r_i = 0) \\ v_i p_x \mu & (d_i = 1) \\ v_i p_x \nu & (r_i = 1) \end{cases}$$

where d_i and r_i represent the number of deaths and retirements experienced by this individual during the year (which will be 0 or 1).

We can then express these three “combinations” in a single formula as:

$$f_i(d_i, r_i, v_i) = {}_{v_i} p_x \times \mu^{d_i} \times \nu^{r_i} = \exp[-(\mu + \nu) v_i] \times \mu^{d_i} \times \nu^{r_i}$$

So the joint likelihood for the whole group will be:

$$L = \prod_{i=1}^N \exp[-(\mu + \nu) v_i] \times \mu^{d_i} \times \nu^{r_i} = \exp[-(\mu + \nu)v] \mu^d \nu^r$$

(i)(d) **Formulae**

The MLE of ν is $\tilde{\nu} = \frac{R}{V}$. [1]

Asymptotically, this has moments:

$$\text{Mean: } E(\tilde{\nu}) = \nu \quad [1]$$

$$\text{Variance: } \text{var}(\tilde{\nu}) = \frac{\nu}{v} \quad [1]$$

(ii)(a) **Likelihood**

The likelihood function is now found by combining the likelihood of observing d deaths during the year with the likelihood of observing r retirements out of the m lives who survived to age 60. This second part is a binomial probability, and we get:

$$e^{-\mu v} \mu^d \times \binom{m}{r} k^r (1-k)^{m-r} \quad [3]$$

(ii)(b) **Formula for MLE of k**

Since we have m lives at age 60 and r are observed to retire, the MLE of k is just the binomial proportion, $\tilde{k} = \frac{r}{m}$. [1]

Solution 2.17(i) **Generator matrix at time t**

The generator matrix is the matrix of transition rates (denoted by $A(t)$ in the Course Notes).

At time t we have:

$$A(t) = \begin{pmatrix} -0.15t & 0.1t & 0.05t & 0 \\ 0.1t & -0.5t & 0.2t & 0.2t \\ 0 & 0 & -0.5t & 0.5t \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad [2]$$

Note that all the rows sum to 0.

(ii) **Backward differential equations**

The matrix form of the backward differential equations is:

$$\frac{\partial}{\partial s} P(s, t) = -A(s)P(s, t)$$

Since this model is time-inhomogeneous and we're asked for the backward differential equation, we are differentiating with respect to s .

For this model:

$$\frac{\partial}{\partial s} P_{33}(s, t) = -[-0.5s P_{33}(s, t)] = 0.5s P_{33}(s, t) \quad [1]$$

and:

$$\begin{aligned} \frac{\partial}{\partial s} P_{13}(s, t) &= -[-0.15s P_{13}(s, t) + 0.1s P_{23}(s, t) + 0.05s P_{33}(s, t)] \\ &= 0.15s P_{13}(s, t) - 0.1s P_{23}(s, t) - 0.05s P_{33}(s, t) \end{aligned} \quad [2]$$

(iii) ***Solving the differential equation***

Separating the variables gives:

$$\frac{\frac{\partial}{\partial s} P_{33}(s, t)}{P_{33}(s, t)} = 0.5s \quad [\frac{1}{2}]$$

and changing the variable from s to u :

$$\frac{\partial}{\partial u} \ln P_{33}(u, t) = 0.5u \quad [\frac{1}{2}]$$

Integrating both sides with respect to u between the limits of $u = s$ and $u = t$, we get:

$$[\ln P_{33}(u, t)]_s^t = \int_s^t 0.5u \, du = \left[0.25u^2 \right]_s^t \quad [1]$$

i.e:

$$\ln P_{33}(t, t) - \ln P_{33}(s, t) = 0.25(t^2 - s^2) \quad [1]$$

However, since $P_{33}(t, t) = 1$ and $\ln 1 = 0$, we have:

$$-\ln P_{33}(s, t) = 0.25(t^2 - s^2)$$

The expression above can be rearranged to give:

$$P_{33}(s,t) = e^{-0.25(t^2-s^2)} \quad [1]$$

(iv) ***Probability of having visited neither state 2 nor state 4 by time t***

There are two possible ways for the process, which started in state 1 at time 0, to have visited neither state 2 nor state 4 by time t. These are:

1. *the process has stayed in state 1 throughout the time interval [0,t], or*
2. *for some s, 0 < s < t, the process has stayed in state 1 throughout the time interval [0,s), jumped into state 3 at time s, and stayed in state 3 throughout the time interval (s,t].*

So the probability that we require is the sum of the probabilities of events 1 and 2 above.

Event 1

The probability that the process stays in state 1 throughout the time interval [0,t] is:

$$P_{11}^{-}(0,t) = \exp\left(-\int_0^t 0.15s \, ds\right) = \exp\left[-0.075s^2\right]_0^t = e^{-0.075t^2} \quad [1]$$

Event 2

The probability that for some s, 0 < s < t, the process stays in state 1 throughout the time interval [0,s), jumps into state 3 at time s, and stays in state 3 throughout the time interval (s,t] is:

$$\int_0^t P_{11}^{-}(0,s) \sigma_{13}(s) P_{33}^{-}(s,t) \, ds \quad [1]$$

From above:

$$P_{11}^{-}(0,s) = e^{-0.075s^2} \quad [\frac{1}{2}]$$

Also, since a return to state 3 is impossible, we know from (iii):

$$P_{\bar{3}3}(s,t) = P_{33}(s,t) = e^{-0.25(t^2-s^2)} \quad [1]$$

So the probability of event 2 is:

$$\int_0^t e^{-0.075s^2} 0.05s e^{-0.25(t^2-s^2)} ds = 0.05e^{-0.25t^2} \int_0^t s e^{0.175s^2} ds$$

Making the substitution $u = 0.175s^2$ (so that $du = 0.35s ds$), the integral on the RHS above becomes:

$$\int_0^{0.175t^2} \frac{e^u}{0.35} du = \left[\frac{e^u}{0.35} \right]_0^{0.175t^2} = \frac{1}{0.35} (e^{0.175t^2} - 1) \quad [1]$$

So the probability of event 2 is:

$$0.05e^{-0.25t^2} \times \frac{1}{0.35} (e^{0.175t^2} - 1) = \frac{1}{7} (e^{-0.075t^2} - e^{-0.25t^2}) \quad [1]$$

Hence the probability that the process has visited neither state 2 nor state 4 by time t is:

$$e^{-0.075t^2} + \frac{1}{7} (e^{-0.075t^2} - e^{-0.25t^2}) = \frac{8}{7} e^{-0.075t^2} - \frac{1}{7} e^{-0.25t^2} \quad [\frac{1}{2}]$$

(v) ***Limiting value***

As $t \rightarrow \infty$, the probability in (iv) tends to 0. [1]

This must be the case for this particular model because eventually the process will end up in state 4, which is an absorbing state. In other words the probability of visiting state 4 by time t tends to 1 as $t \rightarrow \infty$. So the probability of not having visited state 4 tends to 0. [1]

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Part 3 – Questions

Note that the split between Development questions and Exam-style questions is somewhat subjective. For example, there have been past CT4 exam questions that test knowledge of the Core Reading, and so are similar to what we've included here as Development questions. The Exam-style questions involve more application and a wider range of ideas and are typically the more challenging questions in the exam.

1 Development questions

Question 3.1

If $\mu_x = 0.01908 + 0.001(x - 70)$ for $x \geq 55$, calculate ${}_5q_{60}$.

[3]

Question 3.2

Let T denote the complete future lifetime of a newborn baby. The hazard rate for mortality at age x is defined to be:

$$\mu_x = \lim_{h \rightarrow 0+} \frac{1}{h} \times P [T \leq x + h \mid T > x]$$

Now let T_x denote the complete future lifetime of a life aged exactly x who is subject to the same mortality rates. Show that, when $t > 0$:

$$\mu_{x+t} = \lim_{h \rightarrow 0+} \frac{1}{h} \times P [T_x \leq t + h \mid T_x > t] \quad [3]$$

Question 3.3

Consider the following expressions:

$$\text{I} \quad \sum_{k=1}^{\omega-x} k p_x$$

$$\text{II} \quad \sum_{k=0}^{\omega-x} k_k p_x$$

$$\text{III} \quad \int_0^{\omega-x} {}_t p_x \mu_{x+t} dt$$

State which of these are correct expressions for calculating the curtate expectation of life for a life aged exactly x . Explain your answers. [3]

Question 3.4

Mortality of a group of lives is assumed to follow Gompertz' law. Calculate μ_x for a 30-year old and a 70-year old, given that μ_x is 0.003 for a 50 year old and 0.01 for a 60-year old. [4]

Question 3.5

For a particular population, $e_{45} = 40.20$ and $e_{46} = 39.27$. Calculate q_{45} . [3]

Question 3.6

Based on the figures in ELT15, calculate the number of birthdays a newborn female baby will expect to have and how many Christmases she will expect to celebrate during her lifetime. [2]

Question 3.7

Express q_{30} , e_{30} and ${}_5 p_{35}$ in terms of probabilities of the random variable K_{30} , which represents the curtate future lifetime of a life aged exactly 30. [4]

Question 3.8

Calculate the complete and curtate expectation of life for an animal subject to a constant force of mortality of 0.05 per annum. [5]

Question 3.9

The “Very-ruthless Management Consultancy Company” pays very high wages but also has a very high failure rate, both from sackings and through people leaving. A life table for a typical new recruit (with durations measured in years) would be:

<i>Duration</i>	<i>No of lives</i>
0	100,000
1	72,000
2	51,000
3	36,000
4	24,000
5	15,000
6	10,000
7	6,000
8	2,500
9	0

75 graduates started working at the company on 1 September this year. Calculate the following:

- (i) The expected number of complete years that a graduate will complete with the company. [2]
 - (ii) A graduate’s expected “lifetime” with the company. [2]
- [Total 4]

Question 3.10

Define Type I and Type II censoring. [2]

Question 3.11

In a medical study of survival rates for patients who have received a new type of treatment, observations have only been recorded weekly. The resulting data show that 5 lives died and 2 lives were censored in week 10. State the convention that is usually applied when calculating the product-limit estimates when such coincident events are present.

[1]

Question 3.12

The integrated hazard for mortality for a group of lives over the period $(0, t)$, where t is measured in weeks, is being modelled by the function:

$$H(t) = 1 - \{1 + \exp[(t - 2)/3]\}^{-1}$$

- (i) Find an expression for $h(t)$, the hazard function at time t . [2]
- (ii) Sketch a graph of $h(t)$. [5]
- (iii) Suggest a context where a hazard function with this shape might be appropriate. [2]
[Total 9]

2 ***Exam-style questions***

Question 3.13

You have been asked to investigate whether the rate of ill-health retirement of the employees of a large company varies with their duration of employment.

The company's records show:

- the date on which an employee was hired
- the calendar year in which they retired, if an employee left employment as a result of ill-health retirement
- the date of retirement, if an employee reached the normal retirement age of 65
- the date of leaving, if an employee left the company for any other reason.

In the context of this investigation consider the following types of censoring and in each case:

- describe the nature of the censoring
 - state whether or not that type of censoring is present in these data
 - if that particular type of censoring is present, explain how it arises.
- (a) Left censoring
(b) Right censoring
(c) Interval censoring
(d) Informative censoring.

[8]

Question 3.14

A life insurance company has carried out a mortality investigation. It followed a sample of independent policyholders aged between 50 and 55 years. Policyholders were followed from their 50th birthday until they died, withdrew from the investigation while still alive, or reached their 55th birthday (whichever of these events occurred first).

- (i) Describe the types of censoring that are present in this investigation. [2]
- (ii) An extract from the data for 12 policyholders is shown in the table below. Use these data to calculate the Nelson-Aalen estimate of the survival function.

Life	Last age at which life was observed (years and months)	Reason for exit
1	50	9
2	51	3
3	51	6
4	51	6
5	51	6
6	52	9
7	53	3
8	54	3
9	54	6
10	55	0
11	55	0
12	55	0

[3]

- (iii) Determine an approximate 95% confidence interval for your estimate of the survival function. [6]

[Total 11]

Question 3.15

You have been asked to advise a sports magazine, as a consultant statistician. You have been asked to investigate the hypothesis that football managers in the Italian “Serie A”, are dismissed more quickly than those in the English premier league.

Each league has twenty teams, each with one manager. During the season the following events happened (at the end of the months indicated) to the twenty managers who started:

Month	Italy	England
1	One dismissed	One died
3		One left of his own accord
5	One left of his own accord	One dismissed
6	Two dismissed	One left of her own accord
8	One died	Two dismissed
11	Two dismissed	One dismissed

Hence there were thirteen of the original twenty managers still employed by the same club at the end of the season, for each of the two leagues.

- (i) Calculate the Kaplan-Meier estimate of the distribution function and its approximate variance for each league separately. [10]
 - (ii) Comment on the hypothesis that Italian managers are dismissed more quickly than those in England. [2]
- [Total 12]

Question 3.16

A clinical trial is being carried out to test the effectiveness of a new drug. Sixty patients were involved in the trial, which followed them for 2 years from the start of their treatment. The following data show the period in complete months from the start of treatment to the end of observation for those patients who died or withdrew from the trial before the end of the 2-year period.

Deaths: 8, 10, 10, 16, 20
 Withdrawals: 2, 6, 9, 16, 18, 22, 22

- (i) Calculate the Kaplan-Meier estimate of the survival function. [4]
 - (ii) Construct an approximate 95% confidence interval for the probability that a patient survives for at least 18 months after the start of the drug treatment. [3]
- [Total 7]

Question 3.17

A medical researcher is investigating the time it takes for their symptoms to disappear after patients start taking a particular drug. The results of a pilot study of 6 men and 8 women were as follows:

Day on which symptoms disappeared

Males: 3, 2, 6, 15, 6, 2
 Females: 5, 2, 2, 3, 7, 2, 6, 6

The researcher is using the following proportional hazards model to describe the hazard rate at time t (measured in days from commencement of treatment) for the i th individual:

$$\lambda_i(t) = \lambda(t) \times k^{x_i}$$

Here $\lambda(t)$ denotes the baseline hazard rate at time t and x_i encodes the sex of the i th individual, with $x_i = 0$ for a male and $x_i = 1$ for a female.

- (i) Without carrying out any calculations, state how the sample mean and standard deviation of the times taken for symptoms to disappear could be used to gauge whether the baseline hazard rate is approximately constant. [2]

- (ii) Show that the partial log-likelihood function can be written in the form:

$$\ell(k) = 8 \log k - 6 \log(0.75 + k) - 2 \log(0.8 + k) - 5 \log(1 + k) + \text{constant} \quad [4]$$

- (iii) An initial estimate of \hat{k} , the maximum likelihood estimate of k , is 1.4. Use one iteration of the Newton-Raphson method to obtain an improved estimate of \hat{k} . [3]

The Newton-Raphson method uses the fact that if x is a good initial approximation to a root of the equation $f(x) = 0$, then $x^ = x - \frac{f(x)}{f'(x)}$ is usually a better approximation.*

- (iv) Assuming that your estimate obtained in (iii) is accurate, evaluate $\ell''(\hat{k})$ and hence carry out an approximate test to determine whether the rates $\lambda_i(t)$ are different for males and females. [4]
- (v) On the assumption that the hazard rates do not differ according to sex, estimate the median time until symptoms disappear using the Nelson-Aalen method. [3]
- [Total 16]

Question 3.18

You want to use Cox regression to estimate the force of mortality for a group of endowment assurance policyholders. You propose using a model that takes account of duration (*i.e.* the time that has elapsed since the policy was issued) and the age and sex of the policyholder. You start by investigating the model:

$$\mu(x, z_1, z_2) = \mu_0(x) e^{\beta_1 Z_1 + \beta_2 Z_2}$$

where x denotes the age of the policyholder,

$$Z_1 = \begin{cases} 0 & \text{if the duration is less than 1 year} \\ 1 & \text{if the duration is at least 1 year} \end{cases}$$

$$Z_2 = \begin{cases} 0 & \text{for males} \\ 1 & \text{for females} \end{cases}$$

You have estimated the values of the parameters β_1 and β_2 , and have obtained the following results:

Covariate	Parameter	Standard error
Duration	0.416	0.067
Sex	-0.030	0.017

- (i) State the class of policyholders to which the baseline hazard refers. [1]
 - (ii) Explain whether the duration covariate is significant in determining mortality. [3]
 - (iii) Compare the force of mortality for a new female policyholder to that of a male policyholder of the same age, who took out a policy 2 years ago. [2]
- [Total 6]

Question 3.19

The Cox proportional hazards model is to be used to model the rate at which students leave a certain profession before qualification. Assuming they stay in the profession, students will qualify three years after joining the profession. In the fitted model, the hazard depends on the time, t , since joining the profession and three covariates. The covariates, their categories and the fitted parameters for each category are shown in the table below:

<i>Covariate</i>	<i>Possibility</i>	<i>Parameter</i>
Size of employer	large	0
	small	0.4
Degree studied	none	0.3
	Science	-0.1
	Arts	0.2
	other	0
Location	London	0
	other UK	-0.3
	overseas	0.4

- (i) Defining clearly all the terms you use, write down an expression for the hazard function in this model. [3]
 - (ii) State the class of students that is most likely to proceed to qualification under this model, and that which is least likely. [2]
 - (iii) A student who has been in the profession for one year moves from a “small” employer to a “large” employer. Express the probability that he will qualify with the “large” employer P_L in terms of the probability that he would have qualified if he had stayed with the “small” employer P_S , all other factors being equal. [2]
- [Total 7]

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Part 3 – Solutions

Solution 3.1

The probability that a 60-year old will survive for 5 years is:

$$\begin{aligned}
 {}_5 p_{60} &= \exp\left(-\int_0^5 \mu_{60+t} dt\right) \\
 &= \exp\left(-\int_0^5 [0.01908 + 0.001(t-10)] dt\right) \\
 &= \exp\left(-\int_0^5 [0.00908 + 0.001t] dt\right) \\
 &= \exp\left(-\left[0.00908t + 0.0005t^2\right]_0^5\right) \\
 &= \exp(-0.00908 \times 5 - 0.0005 \times 25) \\
 &= 0.94374 \quad [2 \frac{1}{2}]
 \end{aligned}$$

So:

$${}_5 q_{60} = 1 - 0.94374 = 0.05626 \quad [\frac{1}{2}]$$

Solution 3.2

Starting with the definition given for μ_x , but with $x+t$ in place of x , we have:

$$\begin{aligned}
 \mu_{x+t} &= \lim_{h \rightarrow 0+} \frac{1}{h} \times P[T \leq x+t+h | T > x+t] \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{P[x+t < T \leq x+t+h]}{P[T > x+t]} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{P[T \leq x+t+h] - P[T < x+t]}{P[T > x+t]} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{P[T \leq x+t+h] - P[T < x] - (P[T < x+t] - P[T < x])}{P[T > x+t]} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{P[x < T \leq x+t+h] - P[x < T < x+t]}{P[T > x+t]} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{\frac{P[x < T \leq x+t+h]}{P[T > x]} - \frac{P[x < T < x+t]}{P[T > x]}}{\frac{P[T > x+t]}{P[T > x]}} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{P[T \leq x+t+h | T > x] - P[T < x+t | T > x]}{P[T > x+t | T > x]} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{P[T_x \leq t+h] - P[T_x < t]}{P[T_x > t]} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times \frac{P[t < T_x \leq t+h]}{P[T_x > t]} \\
 &= \lim_{h \rightarrow 0+} \frac{1}{h} \times P[T_x \leq t+h | T_x > t]
 \end{aligned} \tag{3}$$

We have made extensive use of $P(X < b | X > a) = \frac{P(a < X < b)}{P(X > a)}$ throughout.

Solution 3.3

I is correct. The sum is the total of the probabilities that the life survives to the end of each future year, which gives the curtate future lifetime. [1]

II is not correct. It would be right if $_k p_x$ was replaced by $P(K = k)$ ie $_k p_x q_{x+k}$. [1]

III is not correct. The integral gives the probability of dying. Its value is 1. [1]

Solution 3.4

There are two ways that this may be done.

Gompertz' law is $\mu_x = Bc^x$.

Therefore $0.003 = Bc^{50}$ and $0.01 = Bc^{60}$. [1]

Hence $c^{10} = 3.33333$. So c is 1.128 and B is 7.29×10^{-6} . [2]

This gives Gompertz' law as $\mu_x = 7.29 \times 10^{-6} \times 1.128^x$.

So for a 30-year old, this is 0.00027 and for a 70-year old, this is 0.033. [1]

Alternatively, $\frac{\mu_{60}}{\mu_{50}}$ gives the value of c^{10} .

So:

$$\mu_{70} = \mu_{60} \times c^{10} = \mu_{60} \times \frac{\mu_{60}}{\mu_{50}} = 0.0333$$

Similarly:

$$\mu_{30} = \mu_{50} \times c^{-20} = \mu_{50} \times \left(\frac{\mu_{50}}{\mu_{60}} \right)^2 = 0.00027$$

Solution 3.5

By conditioning on whether a life aged 45 survives the next year, we get:

$$e_{45} = q_{45} \times 0 + p_{45}(1 + e_{46}) = p_{45}(1 + e_{46}) \quad [2]$$

Using the values given for e_{45} and e_{46} :

$$40.20 = p_{45}(1 + 39.27)$$

So:

$$p_{45} = 40.20 / 40.27 = 0.99826 \quad \text{and} \quad q_{45} = 1 - 0.99826 = 0.00174 \quad [1]$$

This type of calculation is quite sensitive to rounding errors.

Solution 3.6

Every time a person lives for a complete year, they register a birthday. So the number of birthdays a person has is equal to the curtate number of years of life that they live, eg if you die aged 79 years and 2 months you will have celebrated 79 birthdays and have a curtate length of life of 79 years.

The expected number of birthdays is the curtate expectation of life:

$$e_0 \approx \dot{e}_0 - \frac{1}{2} = 78.456.$$

So, females will expect to have 78.456 birthdays. [1]

The number of Christmases is approximately $\frac{1}{2}$ more than this, ie $\overset{\circ}{e}_0 = 78.956$ (assuming that births are uniform over the year).

So, females will expect to have 78.956 Christmases. [1]

Here is an alternative way of looking at the difference between these two quantities. Counting from when you're born, you have to wait a full year until your first birthday, whereas typically you only have to wait 6 months till your first Christmas.

Solution 3.7

We can write:

$$q_{30} = P(K_{30} = 0) \quad [1]$$

$$e_{30} = E(K_{30}) = \sum_{k=0}^{\infty} k P(K_{30} = k) \quad [1]$$

and:

$${}_5 p_{35} = P(K_{30} \geq 10 | K_{30} \geq 5) = \frac{P(K_{30} \geq 10)}{P(K_{30} \geq 5)} \quad [2]$$

Solution 3.8

The complete expectation of life is:

$$\mathring{e}_0 = \int_0^{\infty} t p_0 dt = \int_0^{\infty} e^{-0.05t} dt = \frac{1}{0.05} = 20 \quad [2]$$

The curtate expectation of life can be calculated exactly as follows:

$$e_0 = \sum_{k=1}^{\infty} k p_0 = \sum_{k=1}^{\infty} e^{-0.05k} = \frac{e^{-0.05}}{1 - e^{-0.05}} = 19.504 \quad [3]$$

Solution 3.9

- (i) ***Expected number of complete years***

The curtate expectation of life is:

$$\sum_{k=1}^8 k p_0 = \frac{72,000}{100,000} + \frac{51,000}{100,000} + \dots + \frac{2,500}{100,000} = 2.165 \text{ years} \quad [2]$$

- (ii) ***Expected lifetime***

The complete “expectation of life” is equal to the curtate expectation plus $\frac{1}{2}$, ie 2.665 years. [1]

However, this is based on the (quite dubious!) assumption that exits occur evenly over each year. [1]

Solution 3.10***Type I censoring***

If the censoring times are known in advance (a degenerate case of random censoring) then the mechanism is called “Type I censoring”. [1]

Type II censoring

If observation is continued until a predetermined number of deaths has occurred, then “Type II censoring” is said to be present. This can simplify the analysis, because then the number of events of interest is non-random. [1]

Solution 3.11

The usual convention is to assume that the deaths occurred before the censored events.

[1]

Solution 3.12(i) **Hazard function**

We can find an expression for $h(t)$ by differentiating (using the function-of-a-function rule):

$$\begin{aligned} h(t) &= H'(t) = \{1 + \exp[(t-2)/3]\}^{-2} \times \exp[(t-2)/3] \times \frac{1}{3} \\ &= \frac{\exp[(t-2)/3]}{3\{1 + \exp[(t-2)/3]\}^2} \end{aligned} \quad [2]$$

(ii) **Graph of the hazard function**

From the formula we see that:

$$h(0) = \frac{e^{-2/3}}{3(1+e^{-2/3})^2} = 0.075 \quad [\frac{1}{2}]$$

As $t \rightarrow \infty$, the denominator dominates and $h(\infty) = 0$. [\frac{1}{2}]

We can differentiate $h(t)$, using the product rule, to look for maxima or minima:

$$\begin{aligned} h'(t) &= -2\{1 + \exp[(t-2)/3]\}^{-3} \times \{\exp[(t-2)/3] \times \frac{1}{3}\}^2 \\ &\quad + \{1 + \exp[(t-2)/3]\}^{-2} \times \exp[(t-2)/3] \times (\frac{1}{3})^2 \\ &= \{1 + \exp[(t-2)/3]\}^{-3} \exp[(t-2)/3] \times (\frac{1}{3})^2 \\ &\quad \times \{-2 \exp[(t-2)/3] + \{1 + \exp[(t-2)/3]\}\} \\ &= \{1 + \exp[(t-2)/3]\}^{-3} \exp[(t-2)/3] \times (\frac{1}{3})^2 \{1 - \exp[(t-2)/3]\} \end{aligned} \quad [1]$$

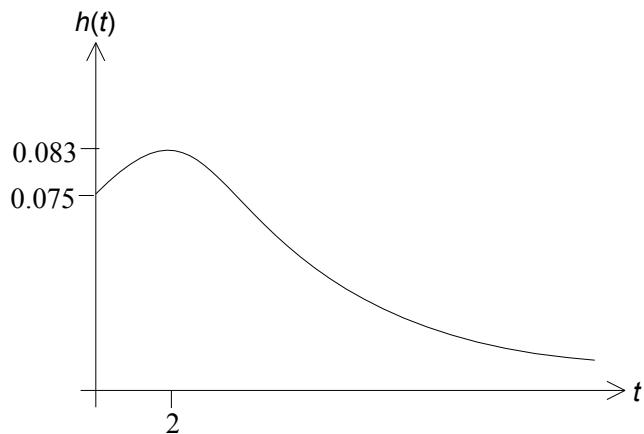
Alternatively, you can differentiate $\log h(t)$.

The derivative will equal zero when the last factor equals zero:

$$1 - \exp[(t-2)/3] = 0 \Rightarrow t = 2 \quad [\frac{1}{2}]$$

The value of the function here is $h(2) = \frac{1}{3(1+1)^2} = 0.083 > h(0)$, so this must be a maximum value. [\frac{1}{2}]

So the graph looks like this:



[2]

This is the log-logistic “humped” hazard function referred to in the notes.

(iii) ***When this might be appropriate***

A hazard function with this shape might be appropriate, for example, for modelling the mortality of patients recovering from an operation where there is a high-risk period (represented by the hump) a couple of weeks after the operation is carried out. [2]

Solution 3.13*Left censoring*

Data in this study would be left censored if the censoring mechanism prevented us from knowing when an employee joined the company. [½]

This is not present because the exact date of joining is given. [½]

Right censoring

Data in this study would be right censored if the censoring mechanism cuts short observations in progress, so that we are not able to discover if and when an employee retired as a result of ill health. [1]

Here there is right censoring of those lives who leave employment before their normal retirement date for reasons other than ill health. [1]

Interval censoring

Data in this study would be interval censored if the observational plan only allows us to say that the duration of employment at the date of ill-health retirement fell within some interval of time (and does not allow us to find the exact duration of employment). [1]

Here we know the calendar year of ill-health retirement and the date of employment, so we will know that the duration of employment falls within a one-year interval. Interval censoring is present. [1]

Informative censoring

Censoring in this study would be informative if the censoring event divided individuals into two groups whose subsequent experience of ill health retirement was thought to be different. [1]

Here the censoring event of leaving the company might be suspected to be informative. Those who leave are more likely to be in better health (less likely to have retired on ill-health grounds had they remained in employment) because they probably left to take another (perhaps better paid and more responsible) job for which they may have been required to pass a medical examination. Similarly, those not resigning their jobs are more likely to retire on ill-health grounds. Informative censoring is present if these groups have different subsequent experience. [2]

Solution 3.14(i) ***Types of censoring present***

Right censoring is present since we don't know the exact future lifetime for the lives that withdrew or left the investigation at age 55. (Right censoring is a special case of interval censoring.) [½]

Random censoring occurs since we don't know the withdrawal times in advance. [½]

Type I censoring occurs since lives that survive to age 55 are certain to be censored at that age. [½]

Non-informative censoring is also present since the withdrawals give us no information about the future mortality of the lives remaining in the investigation. [½]

(ii) ***Nelson-Aalen estimate of the survival function***

The Nelson-Aalen estimate of the survival function is:

$$\hat{S}(t) = e^{-\hat{\Lambda}_t}$$

where:

$$\hat{\Lambda}_t = \sum_{t_j \leq t} \frac{d_j}{n_j} = \begin{cases} 0 & \text{for } 0 \leq t < \frac{9}{12} \\ \frac{1}{12} & \text{for } \frac{9}{12} \leq t < 1\frac{6}{12} \\ \frac{1}{12} + \frac{2}{10} & \text{for } 1\frac{6}{12} \leq t < 4\frac{3}{12} \\ \frac{1}{12} + \frac{2}{10} + \frac{1}{5} & \text{for } 4\frac{3}{12} \leq t < 4\frac{6}{12} \\ \frac{1}{12} + \frac{2}{10} + \frac{1}{5} + \frac{1}{4} & \text{for } 4\frac{6}{12} \leq t \leq 5 \end{cases}$$

$$= \begin{cases} 0 & \text{for } 0 \leq t < \frac{9}{12} \\ 0.08333 & \text{for } \frac{9}{12} \leq t < 1\frac{6}{12} \\ 0.28333 & \text{for } 1\frac{6}{12} \leq t < 4\frac{3}{12} \\ 0.48333 & \text{for } 4\frac{3}{12} \leq t < 4\frac{6}{12} \\ 0.73333 & \text{for } 4\frac{6}{12} \leq t \leq 5 \end{cases}$$

and t is measured in years from age 50. [2]

So:

$$\hat{S}(t) = \begin{cases} 1 & \text{for } 0 \leq t < \frac{9}{12} \\ 0.92004 & \text{for } \frac{9}{12} \leq t < 1\frac{6}{12} \\ 0.75327 & \text{for } 1\frac{6}{12} \leq t < 4\frac{3}{12} \\ 0.61672 & \text{for } 4\frac{3}{12} \leq t < 4\frac{6}{12} \\ 0.48031 & \text{for } 4\frac{6}{12} \leq t \leq 5 \end{cases} \quad [1]$$

(iii) **95% confidence interval**

An approximate 95% confidence interval for the integrated hazard function is:

$$\hat{\Lambda}_t \pm 1.96 \sqrt{\text{var}[\tilde{\Lambda}_t]}$$

$$\text{where } \text{var}[\tilde{\Lambda}_t] \approx \sum_{t_j \leq t} \frac{d_j(n_j - d_j)}{n_j^3}.$$

The variance formula is given on Page 33 of the Tables.

So:

$$\text{var}[\tilde{\Lambda}_t] \approx \begin{cases} 0 & \text{for } 0 \leq t < \frac{9}{12} \\ 0.00637 & \text{for } \frac{9}{12} \leq t < 1\frac{6}{12} \\ 0.02237 & \text{for } 1\frac{6}{12} \leq t < 4\frac{3}{12} \\ 0.05437 & \text{for } 4\frac{3}{12} \leq t < 4\frac{6}{12} \\ 0.10124 & \text{for } 4\frac{6}{12} \leq t \leq 5 \end{cases} \quad [2]$$

An approximate 95% confidence interval for the integrated hazard function is then:

$$\begin{aligned} 0 & \quad \text{for } 0 \leq t < \frac{9}{12} \\ (-0.07305, 0.23971) & \quad \text{for } \frac{9}{12} \leq t < 1\frac{6}{12} \\ (-0.00979, 0.57645) & \quad \text{for } 1\frac{6}{12} \leq t < 4\frac{3}{12} \\ (0.02633, 0.94034) & \quad \text{for } 4\frac{3}{12} \leq t < 4\frac{6}{12} \\ (0.10969, 1.35697) & \quad \text{for } 4\frac{6}{12} \leq t \leq 5 \end{aligned}$$

The integrated hazard must always be a positive number, so we truncate the estimated confidence intervals to reflect this, giving:

$$\begin{aligned}
 0 &\quad \text{for } 0 \leq t < \frac{9}{12} \\
 (0, 0.23971) &\quad \text{for } \frac{9}{12} \leq t < 1\frac{6}{12} \\
 (0, 0.57645) &\quad \text{for } 1\frac{6}{12} \leq t < 4\frac{3}{12} \\
 (0.02633, 0.94034) &\quad \text{for } 4\frac{3}{12} \leq t < 4\frac{6}{12} \\
 (0.10969, 1.35697) &\quad \text{for } 4\frac{6}{12} \leq t \leq 5
 \end{aligned} \tag{3}$$

This truncation will also ensure that the survival probabilities will be between 0 and 1. An approximate 95% confidence interval for the survival function is:

$$\begin{aligned}
 1 &\quad \text{for } 0 \leq t < \frac{9}{12} \\
 (0.78685, 1) &\quad \text{for } \frac{9}{12} \leq t < 1\frac{6}{12} \\
 (0.56189, 1) &\quad \text{for } 1\frac{6}{12} \leq t < 4\frac{3}{12} \\
 (0.39050, 0.97401) &\quad \text{for } 4\frac{3}{12} \leq t < 4\frac{6}{12} \\
 (0.25744, 0.89611) &\quad \text{for } 4\frac{6}{12} \leq t \leq 5
 \end{aligned} \tag{1}$$

Solution 3.15

- (i) **Kaplan-Meier estimates and their variances**

The figures for Italy are as follows:

j	t_j	d_j	n_j	$\hat{\lambda}_j$	$1 - \hat{\lambda}_j$	$\hat{F}(t)$
1	1	1	20	0.050	0.950	0.050
2	6	2	18	0.111	0.889	0.156
3	11	2	15	0.133	0.867	0.268

ie

$$\hat{F}(t) = \begin{cases} 0 & \text{for } 0 \leq t < 1 \\ 0.050 & \text{for } 1 \leq t < 6 \\ 0.156 & \text{for } 6 \leq t < 11 \\ 0.268 & \text{for } 11 \leq t \leq 12 \end{cases} \quad [3]$$

The variance is given by:

j	$\hat{F}(t)$	$\frac{d_j}{n_j(n_j - d_j)}$	$\text{var}[\tilde{F}(t)]$
1	0.050	1/380	0.0024
2	0.156	2/288	0.0068
3	0.268	2/195	0.0106

ie

$$\text{var}[\tilde{F}(t)] = \begin{cases} 0 & \text{for } 0 \leq t < 1 \\ 0.0024 & \text{for } 1 \leq t < 6 \\ 0.0068 & \text{for } 6 \leq t < 11 \\ 0.0106 & \text{for } 11 \leq t \leq 12 \end{cases} \quad [2]$$

The corresponding figures for England are:

j	t_j	d_j	n_j	$\hat{\lambda}_j$	$1 - \hat{\lambda}_j$	$\hat{F}(t)$
1	5	1	18	0.056	0.944	0.056
2	8	2	16	0.125	0.875	0.174
3	11	1	14	0.071	0.929	0.233

ie

$$\hat{F}(t) = \begin{cases} 0 & \text{for } 0 \leq t < 5 \\ 0.056 & \text{for } 5 \leq t < 8 \\ 0.174 & \text{for } 8 \leq t < 11 \\ 0.233 & \text{for } 11 \leq t \leq 12 \end{cases} \quad [3]$$

The variance is given by:

j	$\hat{F}(t)$	$\frac{d_j}{n_j(n_j - d_j)}$	$\text{var}[\tilde{F}(t)]$
1	0.056	1/306	0.0029
2	0.174	2/224	0.0083
3	0.232	2/182	0.0104

ie

$$\text{var}[\tilde{F}(t)] = \begin{cases} 0 & \text{for } 0 \leq t < 5 \\ 0.0029 & \text{for } 5 \leq t < 8 \\ 0.0083 & \text{for } 8 \leq t < 11 \\ 0.0104 & \text{for } 11 \leq t \leq 12 \end{cases} \quad [2]$$

(ii) **Comment**

From the figures given, there is some suggestion that the Italian managers do have a slightly shorter “lifetime”.

However, looking at the variance this is not statistically significant.

[2]

Solution 3.16(i) **Kaplan-Meier estimate of the survival function**

Let t denote time measured in months from the start of treatment. The Kaplan-Meier estimate of the survival function is a step function that starts at 1 and steps down every time a death is observed.

We start with 60 lives, and two of them are censored before the first death, which occurs at time 8. So:

$$\hat{S}(t) = 1 \text{ for } 0 \leq t < 8$$

and letting t_j denote the j th death time, n_j denote the number of lives under observation just before time t_j and d_j denote the number of deaths at time t_j , we have:

j	t_j	n_j	d_j	$\hat{\lambda}_j = \frac{d_j}{n_j}$	$1 - \hat{\lambda}_j$
1	8	58	1	$\frac{1}{58}$	$\frac{57}{58} = 0.98276$
2	10	56	2	$\frac{2}{56}$	$\frac{54}{56} = 0.96429$
3	16	54	1	$\frac{1}{54}$	$\frac{53}{54} = 0.98148$
4	20	51	1	$\frac{1}{51}$	$\frac{50}{51} = 0.98039$

[2]

The Kaplan-Meier estimate of the survival function is then:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \hat{\lambda}_j\right) = \begin{cases} 1 & \text{for } 0 \leq t < 8 \\ 0.98276 & \text{for } 8 \leq t < 10 \\ 0.94766 & \text{for } 10 \leq t < 16 \\ 0.93011 & \text{for } 16 \leq t < 20 \\ 0.91187 & \text{for } 20 \leq t \leq 24 \end{cases}$$

[2]

(ii) **95% confidence interval**

An approximate 95% confidence interval for $S(18)$ is:

$$\hat{S}(18) \pm 1.96 \sqrt{\text{Var}(\tilde{S}(18))} \quad [\frac{1}{2}]$$

From (i):

$$\hat{S}(18) = 0.93011 \quad [\frac{1}{2}]$$

The variance term can be calculated using Greenwood's formula, which is on Page 33 of the *Tables*:

$$\begin{aligned} \text{var}(\tilde{S}(18)) &= (\hat{S}(18))^2 \sum_{t_j \leq 18} \frac{d_j}{n_j(n_j - d_j)} \\ &= 0.93011^2 \left[\frac{1}{58 \times 57} + \frac{2}{56 \times 54} + \frac{1}{54 \times 53} \right] \\ &= 0.001136 \end{aligned} \quad [1]$$

So the required confidence interval is:

$$0.93011 \pm 1.96 \sqrt{0.001136} = (0.8640, 0.9962) \quad [1]$$

Solution 3.17(i) **How to gauge whether baseline hazard is approximately constant**

If the baseline hazard rate is constant ($= \lambda$, say), the times will have an exponential distribution with parameter λ . The mean and standard deviation of this distribution are both equal to $1/\lambda$.

We can gauge whether this assumption is reasonable by comparing the sample mean and standard deviation, which should have similar numerical values. [2]

Note that, if we write the model in the form $\lambda_i(t) = \lambda(t) \times \exp\{(\log k)x_i\}$, we can see that it is just the Cox model in disguise.

(ii) ***Partial log-likelihood***

If we arrange the data values in order of the times, we get the following table:

Day	<i>Male decrements</i>	<i>Female decrements</i>	<i>Males at risk</i>	<i>Females at risk</i>
2	2	3	6	8
3	1	1	4	5
5		1	3	4
6	2	2	3	3
7		1	1	1
15	1			0

[2]

We have a number of ties here (events recorded as occurring at the same time). The usual convention when calculating the partial likelihood is to assume that the number at risk does not change until all the concurrent events have occurred (Breslow's approximation).

The partial likelihood function is then:

$$L(k) = \frac{k^3}{(6+8k)^5} \times \frac{k}{(4+5k)^2} \times \frac{k}{(3+4k)} \times \frac{k^2}{(3+3k)^4} \times \frac{k}{(1+k)} \times \frac{1}{1} \quad [1]$$

If we take out factors from the denominators so that all the k 's have a coefficient of 1, we can write this as:

$$\begin{aligned} L(k) &= \frac{k^3}{(0.75+k)^5} \times \frac{k}{(0.8+k)^2} \times \frac{k}{(0.75+k)} \times \frac{k^2}{(1+k)^4} \times \frac{k}{(1+k)} \times \text{constant} \\ &= \frac{k^8}{(0.75+k)^6 (0.8+k)^2 (1+k)^5} \times \text{constant} \end{aligned}$$

Taking logs to find the partial log-likelihood:

$$\begin{aligned} \ell(k) &= \log L(k) \\ &= 8 \log k - 6 \log(0.75+k) - 2 \log(0.8+k) - 5 \log(1+k) + \text{constant} \end{aligned} \quad [1]$$

(iii) ***Improved estimate***

To find the MLE of k , we need to solve the equation $\ell'(k) = 0$, that is:

$$\ell'(k) = \frac{8}{k} - \frac{6}{0.75+k} - \frac{2}{0.8+k} - \frac{5}{1+k} = 0 \quad [1]$$

If we write $f(x) = \ell'(x)$, we can apply the Newton-Raphson method, starting with $k = 1.4$. To do this, we need the second derivative:

$$\ell''(k) = -\frac{8}{k^2} + \frac{6}{(0.75+k)^2} + \frac{2}{(0.8+k)^2} + \frac{5}{(1+k)^2} \quad [1]$$

So a more accurate approximation to the root of the equation $f(x) = 0$ is:

$$k^* = k - \frac{f(k)}{f'(k)} = 1.4 - \frac{\ell'(1.4)}{\ell''(1.4)} = 1.4 - \left(\frac{-0.0688}{-1.502} \right) = 1.354 \quad [1]$$

(iv) ***Test***

Assuming that $\hat{k} = 1.354$, the second derivative is:

$$\ell''(\hat{k}) = -1.675$$

MLEs are asymptotically unbiased and normally distributed with variance equal to the Cramér-Rao lower bound.

So:

$$\hat{k} \sim N\left(k, \frac{-1}{\ell''(\hat{k})}\right) = N(k, 0.597) \quad [1]$$

The test we require here is:

$$H_0 : k = 1 \text{ versus } H_1 : k \neq 1 \quad [1]$$

The test statistic is:

$$z = \frac{1.354 - 1}{\sqrt{0.597}} = 0.46 \quad [1]$$

This is not significant at the 5% level. So there is no reason to think that the rates differ by sex. [1]

(v) ***Estimate of median time***

The conclusion in (iv) allows us to combine the sexes and assume that a single hazard rate $\lambda^*(t)$ applies to both.

With the Nelson-Aalen method, the estimated distribution function of the times is:

$$\hat{F}(t) = 1 - e^{-\hat{\Lambda}^*(t)} \quad [\frac{1}{2}]$$

where $\hat{\Lambda}^*(t)$ is the estimated integrated hazard, which is calculated as:

$$\hat{\Lambda}^*(t) = \sum_{t_j \leq t} \frac{d_j}{n_j} \quad [\frac{1}{2}]$$

The median time \hat{t}_m will satisfy the equation:

$$F(\hat{t}_m) = 1 - e^{-\hat{\Lambda}^*(\hat{t}_m)} = \frac{1}{2} \Rightarrow \hat{\Lambda}^*(\hat{t}_m) = \log 2 = 0.693$$

So we can estimate the median time by solving the equation:

$$\hat{\Lambda}^*(\hat{t}_m) = \sum_{t_j \leq \hat{t}_m} \frac{d_j}{n_j} = 0.693$$

With $\hat{t}_m = 5$, this works out to:

$$\hat{\Lambda}^*(5) = \frac{5}{14} + \frac{2}{9} + \frac{1}{7} = 0.722 \quad [1]$$

This is close to 0.693. So the median time is approximately 5 days. [1]

Interpolation can be used here to obtain a more accurate answer.

Solution 3.18(i) ***Class of policyholders to which baseline hazard refers***

The baseline hazard refers to male endowment assurance policyholders, who took out their policies less than one year ago. [1]

(ii) ***Is duration significant?***

An approximate 95% confidence interval for the duration parameter is:

$$0.416 \pm (1.96 \times 0.067) = (0.285, 0.547) \quad [2]$$

As this interval does not contain 0, we conclude that the duration covariate is significant in determining mortality. [1]

(iii) ***Comparison of forces of mortality***

According to the model, the force of mortality for a new female policyholder aged x is $\mu_0(x)e^{-0.030}$; the force of mortality for a male policyholder at the same age who took out his policy 2 years ago is $\mu_0(x)e^{0.416}$.

Since:

$$\frac{\mu_0(x)e^{-0.030}}{\mu_0(x)e^{0.416}} = e^{-0.446} = 0.640$$

the model implies that the force of the mortality for the female is 36% less than the force of mortality for the male. [2]

You could also have said that the force of mortality for the male is 56% higher than the force of mortality for the female.

Solution 3.19(i) **Hazard function**

The hazard function is given by:

$$\lambda(t, \mathbf{Z}) = \lambda_0(t) \exp[0.4Z_1 + 0.3Z_2 - 0.1Z_3 + 0.2Z_4 - 0.3Z_5 + 0.4Z_6] \quad [1]$$

where:

$\lambda_0(t)$ = baseline hazard at time t since entry into profession

$\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4, Z_5, Z_6)$

$Z_1 = 1$ if small employer, 0 if not.

$Z_2 = 1$ if no degree, 0 if not

$Z_3 = 1$ if science degree, 0 if not

$Z_4 = 1$ if arts degree, 0 if not

$Z_5 = 1$ if location = UK except London, 0 if not

$Z_6 = 1$ if location = overseas, 0 if not

[2]

(ii) **Most and least likely to qualify**

The students most likely to qualify are those with the lowest hazard function, ie those for which $Z_1 = 0$, $Z_2 = 0$, $Z_3 = 1$, $Z_4 = 0$, $Z_5 = 1$ and $Z_6 = 0$. So the students most likely to qualify are those who work for large employers, have science degrees and work in the UK but outside London.

[1]

The least likely to qualify are those for which $Z_1 = 1$, $Z_2 = 1$, $Z_3 = 0$, $Z_4 = 0$, $Z_5 = 0$ and $Z_6 = 1$, ie those who work for small employers, have no degrees and who work overseas.

[1]

(iii) **Probability of qualifying**

The probability that a student who has been in the profession for one year will qualify is:

$$\exp\left(-\int_1^3 \lambda(t, \mathbf{z}) dt\right)$$

You can think of this as the probability that the student will “survive”, ie avoid leaving the profession, from time 1 to time 3.

If the student works for a large employer, the probability is:

$$\begin{aligned} P_L &= \exp\left(-\int_1^3 \lambda(t)e^{0.3z_2-0.1z_3+0.2z_4-0.3z_5+0.4z_6} dt\right) \\ &= \exp\left[-e^{0.3z_2-0.1z_3+0.2z_4-0.3z_5+0.4z_6} \int_1^3 \lambda(t) dt\right] \end{aligned} \quad [1/2]$$

If the student works for a small employer, the probability is:

$$\begin{aligned} P_S &= \exp\left(-\int_1^3 \lambda(t)e^{0.4+0.3z_2-0.1z_3+0.2z_4-0.3z_5+0.4z_6} dt\right) \\ &= \exp\left[-e^{0.4+0.3z_2-0.1z_3+0.2z_4-0.3z_5+0.4z_6} \int_1^3 \lambda(t) dt\right] \\ &= \exp\left[-e^{0.4} e^{0.3z_2-0.1z_3+0.2z_4-0.3z_5+0.4z_6} \int_1^3 \lambda(t) dt\right] \\ &= \left\{ \exp\left[-e^{0.3z_2-0.1z_3+0.2z_4-0.3z_5+0.4z_6} \int_1^3 \lambda(t) dt\right] \right\}^{\exp(0.4)} \\ &= (P_L)^{\exp(0.4)} \end{aligned} \quad [1]$$

The second last equality follows from the result $e^{AB} = (e^B)^A$.

So:

$$P_L = (P_S)^{\exp(-0.4)} = (P_S)^{0.67032} \quad [1/2]$$

This was Question 5 on the Subject 104, April 2003 exam paper.

Part 4 – Questions

Note that the split between Development questions and Exam-style questions is somewhat subjective. For example, there have been past CT4 exam questions that test knowledge of the Core Reading, and so are similar to what we've included here as Development questions. The Exam-style questions involve more application and a wider range of ideas and are typically the more challenging questions in the exam.

1 **Development questions**

Question 4.1

- (i) State the assumptions underlying the binomial mortality model. [2]
- (ii) A cat has nine lives, so the cat will not die until it has lost all nine of its lives. The probability of a cat losing a life is 20% per week. Assuming that the mortality of each life follows the binomial model, calculate the following:
- (a) The probability that a cat who has currently lost none of its nine lives will die during the next 10 weeks. [3]
- (b) The probability that this cat will die during the fifth week. [2]
- [Total 7]

Question 4.2

On 1 January of a particular year, there were 406 men and 418 women in the age range 25 to 30 living in a small town. If the initial rate of mortality can be assumed to be constant in this age range, 4.2 per 10,000 for men and 3.3 per 10,000 for women, calculate the probability that exactly two of these individuals will die during that year.

[4]

Question 4.3

At a particular hospital, the initial rate of mortality for babies born between 4 and 8 weeks premature is 4% per annum. Using a normal distribution, calculate the approximate probability that, out of 100 such babies, more than 6 will not survive to their first birthday. [4]

Question 4.4

Explain the importance of dividing the data for a mortality investigation into homogeneous classes. [3]

Question 4.5

You have been given the information from the mortality experience of a large life office covering the years 2007 and 2008. Your information includes the number of people aged x nearest birthday at 1 January 2007, 2008 and 2009, and the number of deaths during 2007 and 2008 aged x nearest birthday. You have derived crude central rates of mortality μ_x . State the assumptions that you had to make. [3]

Question 4.6

Explain how the central exposed to risk differs from the initial exposed to risk. [3]

Question 4.7

Explain what is meant by the principle of correspondence. [2]

Question 4.8

Describe briefly some validity checks that can be carried out to ascertain the reliability of the data to be used for a mortality investigation. [3]

Question 4.9

A mortality investigation is being conducted between the dates 1 January 2004 and 31 December 2007. The data from four lives under consideration was as follows:

Life	Date of Birth	Date of Entry	Date of Exit	Mode of Exit
Pele	11.11.73	24.3.02	29.12.07	Death
Johan	1.9.80	30.8.05		Did not leave
Gary	10.2.79	10.10.03	21.6.04	Surrender
Diego	8.2.82	10.8.05		Did not leave

- (i) Assuming that the day of entry, but not the day of exit, counts in the exposed to risk, calculate the number of days of exposure contributed to the central exposed to risk by each life at each age. [12]
 - (ii) State what modifications, if any, you would need to make if you were determining the initial exposed to risk and not the central exposed to risk. [2]
- [Total 14]

Question 4.10

You have been given the following census counts for a population (covering all ages):

$$P_{2004} = \text{Number in population on 1 January 2004} = 20,000$$

$$P_{2005} = \text{Number in population on 1 January 2005} = 40,000$$

$$P_{2006} = \text{Number in population on 1 January 2006} = 30,000$$

Estimate the central exposed to risk (all ages) for this population over each of the following periods, given only the census counts P_{2004} , P_{2005} and P_{2006} . In each case, state any assumptions you have made.

- (i) Period: 1 January 2004 to 31 December 2005 [2]
 - (ii) Period: 1 July 2004 to 30 June 2005 [2]
 - (iii) Period: 1 January 2006 to 31 December 2006 [2]
 - (iv) Period: 1 April 2005 to 31 March 2006 [2]
- [Total 8]

Question 4.11

It has been suggested that the chi square goodness of fit test may fail to detect the following shortcomings of a graduation:

- I A large cumulative deviation over parts of the table.
- II A large cumulative deviation over all of the table.
- III A large number of outliers.

State whether each of the statements above is correct or not.

[3]

Question 4.12

A graphical graduation should be used ...

- A when smooth rates are required, but accuracy is not essential
- B when you don't care about the results
- C where great accuracy and smoothness are not essential
- D where the amount of data is limited.

[2]

Question 4.13

- I When graduating by reference to a standard table you always need to test for smoothness.
- II If the mortality of a whole population was recorded, there would be no need to graduate since random sampling errors would not have occurred.
- III Graduating by reference to a standard table can produce good results even with scanty data.

State whether each of these statements is correct or not.

[3]

Question 4.14

The following numbers represent the individual standardised deviations for five consecutive ages from a graduation.

$$-1.33, +0.22, +0.88, -0.11, +0.62$$

The cumulative deviation in respect of the same five ages is:

- A 0.28
- B 3.16
- C -0.08
- D Can't tell.

[3]

Question 4.15

At a particular age there are 922 deaths, compared to 950 expected. Calculate the approximate individual standardised deviation for that age. [2]

Question 4.16

Name the most commonly used method of graduation when preparing a standard table.

[2]

Question 4.17

- (i) If mortality follows Gompertz' Law such that $\mu_x = 0.00003 \times 1.1^x$, calculate the values of $\mu_{80}, \mu_{81}, \mu_{82}, \mu_{83}$ (to 7 decimal places) and the first, second and third differences derived from these quantities. [3]
- (ii) Recalculate differences using the observed rates $\hat{\mu}_{80} = \mu_{80} - 0.01$, $\hat{\mu}_{81} = \mu_{81}$, $\hat{\mu}_{82} = \mu_{82}$, $\hat{\mu}_{83} = \mu_{83}$. [2]
- (iii) Use the values calculated in (i) and (ii) to explain why the test for smoothness usually requires the *third* differences of the graduated rates to be small, rather than the differences of a lower or a higher order. [4]

[Total 9]

Question 4.18

State the characteristics of a good graduation. [2]

Question 4.19

State six advantages of using the graphical method of graduation. [3]

Question 4.20

State seven disadvantages of using the graphical method of graduation. [3½]

Question 4.21

State four advantages of graduating by reference to a standard table. [2]

Question 4.22

State three disadvantages of graduating by reference to a standard table. [1½]

Question 4.23

State five advantages of graduating by reference to a parametric formula. [2½]

Question 4.24

State two disadvantages of graduating by reference to a parametric formula. [1]

2 Exam-style questions

Question 4.25

Prove each of the following results:

- (i) The probability that exactly x decrements will occur in a population consisting initially of n individuals subject to a single decrement with rate q per annum is:

$$\binom{n}{x} q^x (1-q)^{n-x} \quad [3]$$

- (ii) The maximum likelihood estimator of the parameter q for the binomial model equals the number of decrements divided by the initial population. [4]

- (iii) The maximum likelihood estimator of the parameter μ for the Poisson model is an unbiased estimator of the force of decrement. [3]

[Total 10]

Question 4.26

During a period of length T years, you observe a total of N lives between the ages of x and $x+1$. You do not necessarily observe each life for the entire year of age. The total time spent under observation by the N lives is V . d deaths are observed.

- (i) State the assumptions underlying the Poisson model for d , given that the force of mortality between the ages of x and $x+1$ is a constant, μ . [2]

- (ii) Show that the maximum likelihood estimator of μ is D/V , under the assumptions in (i) above. [2]

- (iii) Show that the maximum likelihood estimator has:

- (a) an expected value of μ

- (b) a variance of μ/V . [2]

[Total 6]

Question 4.27

The table below gives an extract of the data of a mortality investigation, which examines the mortality of lives aged between 70 and 71. For each life, you are given the age at which they were first observed, the age at which they ceased to be observed and their reason for leaving the investigation.

Life	Age at entry	Age at exit	Reason for leaving
1	70 years 0 months	71 years 0 months	censored
2	70 years 0 months	70 years 11 months	died
3	70 years 0 months	70 years 8 months	censored
4	70 years 1 month	70 years 3 months	died
5	70 years 2 months	71 years 0 months	censored
6	70 years 2 months	70 years 10 months	censored
7	70 years 6 months	70 years 10 months	died
8	70 years 6 months	70 years 8 months	censored

- (i) Calculate the Kaplan-Meier estimate of the survival function $S_{T_{70}}(t) = P(T_{70} > t)$ and sketch a graph of this function. [5]
 - (ii) Construct a 90% confidence interval for q_{70} based on your estimate in part (i). [3]
 - (iii) Calculate the total central exposed to risk and the total initial exposed to risk at age 70 last birthday for the group of lives given in the table above. [3]
 - (iv) Calculate the actuarial estimate of q_{70} . [1]
 - (v) Calculate the Poisson estimate of q_{70} and construct a 90% confidence interval for q_{70} based on this estimate. [5]
 - (vi) Comment on the differences between the estimates of q_{70} calculated in (i), (iv) and (v). [4]
- [Total 21]

Question 4.28

In carrying out its mortality investigations, a life insurance company uses census data from its mid-year valuations. For all lines of business in-force policies are classified by age nearest birthday at the valuation date (30 June in each calendar year), where $P_x(T)$ is the number of in-force policies aged x nearest birthday in calendar year T .

For a portfolio of term insurance policies sold in the 1990's the exact date of birth was recorded. The total number of deaths during the calendar years 2007 and 2008, θ_x , have been classified by age x , where:

$$x = \text{age last birthday at the date of death}$$

- (i) Derive an exposed to risk formula for the estimation of the force of mortality at age x . State any assumptions that you make. [4]
 - (ii) State the age to which your estimated rates apply, giving any assumptions that are necessary. [2]
- [Total 6]

Question 4.29

A graduation of a set of mortality rates from age 25 to age 64 has 15 positive individual standardised deviations, which occurred in 8 groups.

Carry out two tests to check the suitability of this graduation. [6]

Question 4.30

Explain why it is necessary to graduate crude rates of mortality for practical use. [4]

Question 4.31

- (i) For a mortality investigation you have been given data relating to the initial exposed to risk and number of deaths for ages 60 to 94. Describe how you would carry out a graphical graduation and show how you would derive approximate 95% confidence intervals for the age specific mortality rates. [8]
- (ii) Describe how you would test your graduated rates for smoothness. [2]
- (iii) In the investigation the following data have been collected.

Age last birthday	Initial exposed to risk	Number of deaths
60-64	20,500	374
65-69	27,800	892
70-74	30,100	1,475
75-79	26,700	2,158
80-84	17,700	2,251
85-89	9,200	1,780
90-94	2,300	630

State, giving reasons, whether the graphical method is appropriate for the graduation of this experience, or whether a different method would be more suitable. [4]

[Total 14]

Question 4.32

- (i) Discuss the advantages and disadvantages of using the graphical method of graduation, explaining under which circumstances it is appropriate to use it. [7]
- (ii) Describe the difficulties that can arise when applying a chi square test to assess a graphical graduation. [3]

[Total 10]

Question 4.33

- (i) In the context of the graduation of mortality data discuss the concepts of “smoothness” and “fidelity to data” and their relationship. [6]
- (ii) Comment on the graduation in the following table, mentioning briefly the limitations of any tests you apply.

Age	Exposed to risk	Actual deaths	Graduated mortality rate \hat{q}_x	Expected deaths
65	60,000	1,370	0.0225	1,350
66	50,000	1,200	0.0250	1,250
67	43,000	1,200	0.0276	1,187
68	37,000	1,090	0.0304	1,125
69	30,000	1,010	0.0334	1,002
70	26,000	950	0.0368	957
71	23,000	980	0.0406	934
72	21,000	950	0.0449	943
73	20,000	1,000	0.0497	994
74	18,000	960	0.0551	992
<hr/>				
	328,000	10,710		10,734

The chi square value of the graduation is 7.29. You may assume that there are 7 degrees of freedom. [20]

[Total 26]

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Part 4 – Solutions

Solution 4.1

(i) ***Assumptions underlying the binomial model***

We observe N identical and independent lives aged x exactly for one year, and record the number d who die. Then d is a sample value of a random variable D . If we suppose that each life dies with probability q_x and survives with probability $1 - q_x$, then D has a binomial distribution with parameters N and q_x . [2]

(ii)(a) ***Probability that cat is dead in ten weeks***

The cat will be dead in 10 weeks if it has lost all 9 lives by then.

The probability that it will NOT lose a particular life during the next 10 weeks is:

$$0.8^{10} = 0.10737 \quad [1]$$

So the probability that it WILL lose a particular life during the next 10 weeks is:

$$1 - 0.10737 = 0.89263 \quad [1]$$

So the probability that it will lose all 9 lives during the 10 weeks is:

$$0.89263^9 = 0.360 \quad [1]$$

(ii)(b) ***Probability that it will die during the fifth week***

We first find the probability that the cat will be dead by the end of the fifth week. Using the same logic as previously, this is:

$$(1 - 0.8^5)^9 = 0.02807 \quad [\frac{1}{2}]$$

Similarly, the probability that the cat is dead by the end of the fourth week is:

$$(1 - 0.8^4)^9 = 0.00872 \quad [\frac{1}{2}]$$

The probability that the cat actually dies during the fifth week is the difference between these probabilities:

$$0.02807 - 0.00872 = 0.019 \quad [1]$$

Solution 4.2

Using M and F to denote male and female deaths, the probability of exactly two deaths is:

$$P(2 \text{ deaths}) = P(MM) + P(MF) + P(FF)$$

Using the probability function for the binomial distribution (and assuming that deaths occur independently):

$$P(MM) = \binom{406}{2} (0.00042)^2 (0.99958)^{404} (0.99967)^{418} = 0.010662 \quad [1]$$

$$\begin{aligned} P(MF) &= \binom{406}{1} (0.00042) (0.99958)^{405} \\ &\times \binom{418}{1} (0.00033)(0.99967)^{417} = 0.017290 \end{aligned} \quad [1]$$

$$P(FF) = \binom{418}{2} (0.00033)^2 (0.99967)^{416} (0.99958)^{406} = 0.006976 \quad [1]$$

So the probability of exactly two deaths during the year is the total of these, *ie* 3.49%.

[1]

Solution 4.3

The mean and variance of the number of deaths (using the binomial model) are:

$$\begin{aligned} E(X) &= 100 \times 0.04 = 4 \\ \text{var}(X) &= 100 \times 0.04 \times 0.96 = 3.84 \end{aligned} \quad [1]$$

Using a normal approximation and applying a continuity correction to take account of the fact that the number of deaths must be a whole number, the probability of more than 6 deaths is approximately:

$$P[N(4, 3.84) \geq 6.5] \quad [1]$$

This is equal to:

$$1 - \Phi(1.276) = 1 - 0.899 = 0.101$$

So the probability of more than 6 deaths is approximately 10.1%. [2]

Solution 4.4

If the groups are not homogeneous, any rates derived will be a weighted average of the underlying rates for the different individuals in the group. The weightings may change with time, which will make it very difficult to establish what patterns are emerging. [2]

If premiums are calculated based on mortality rates derived from heterogeneous groups, then anti-selection may occur, with the more healthy lives choosing to insure themselves with an office where they will not be charged a premium based on others with an inherent higher level of risk. (*We will study the effects of anti-selection in Subject CT5.*) [1]

Solution 4.5

As deaths are recorded as age nearest, everyone will be age x exactly in the middle of the rate interval without any assumptions regarding the distribution of births necessary.

[1]

The exposed to risk is estimated using a census at the start and end of each year. The population is estimated by taking the average of these readings. You must assume that the population varies linearly over each calendar year.

[1]

Note that it is not sufficient to assume that deaths occur uniformly over the calendar year, since there may be other modes of leaving the population, and there will be new entrants too.

[1]

Solution 4.6

The central exposed to risk only takes account of the time up to exit, and thus is independent of the decrement being investigated. The initial exposed to risk counts time until the end of the rate interval if the life leaves by the cause under consideration. The initial exposed to risk can therefore count days outside the period of investigation.

[2]

The central exposed to risk is the population measure for a Poisson model; the initial exposed to risk is the measure for a binomial model.

[1]

Solution 4.7

The principle of correspondence states that if a life would have been included in the deaths figure were it to die on a particular day, then the life should contribute to the exposed to risk for that day.

[1]

This ensures that the deaths and exposed to risk are calculated consistently.

[1]

Solution 4.8

We can:

- check all dates are valid (*eg* 31 September is invalid)
- check all dates are reasonable (*eg* year of birth equal to 1850 is unreasonable)
- check all dates are consistent (*eg* date of issue of policy later than date of birth)
- reconcile the numbers exposed to risk with the numbers from previous investigations and sales information
- compare the results with those of previous studies or other published tables

[3]

Solution 4.9(i) ***Central exposed to risk***

We can consider the central exposed to risk year by year for each person.

Pele

01.01.04 to 10.11.04	315 in E_{30}^c including 29.2.04 as 2004 was a leap year
11.11.04 to 10.11.05	365 in E_{31}^c
11.11.05 to 10.11.06	365 in E_{32}^c
11.11.06 to 10.11.07	365 in E_{33}^c
11.11.07 to 29.12.07	48 in E_{34}^c

Johan

30.08.05 to 31.08.05	2 in E_{24}^c
01.09.05 to 31.08.06	365 in E_{25}^c
01.09.06 to 30.08.07	365 in E_{26}^c
01.09.07 to 31.12.07	122 in E_{27}^c including 31.12.07 as it is not an exit date

Gary

01.01.04 to 09.02.04	40 in E_{24}^c
10.02.04 to 21.06.04	132 in E_{25}^c

Diego

10.08.05 to 07.02.06 182 in E_{23}^c

08.02.06 to 07.02.07 365 in E_{24}^c

08.02.07 to 31.12.07 327 in E_{25}^c including 31.12.07 as it is not an exit date

(ii) ***Modifications required for initial exposed to risk***

The lives receive the same exposure in the initial exposed to risk except for Pele who receives 365 in E_{34} (even though this extends beyond the period of investigation). [2]

(*Gary still receives 132 in E_{25} since he did not die – initial exposed to risk only gets counted to the year-end if the person leaves by the decrement rate being investigated.*)

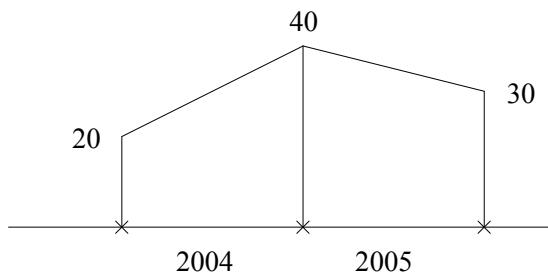
Solution 4.10

For periods that fall between the census dates, the approach to use is to apply the trapezium rule (*Average height \times Width*). For periods that fall outside the census dates, the simplest approach is to assume that the population size has remained constant.

(i) ***Period: 1 January 2004 to 31 December 2005***

Here we would assume that the population has varied linearly over each calendar year, and we would use the approximation:

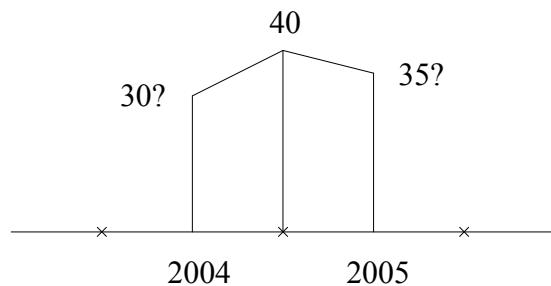
$$\frac{1}{2}(P_{2004} + P_{2005}) + \frac{1}{2}(P_{2005} + P_{2006}) = 30,000 + 35,000 = 65,000 \quad [2]$$



(ii) ***Period: 1 July 2004 to 30 June 2005***

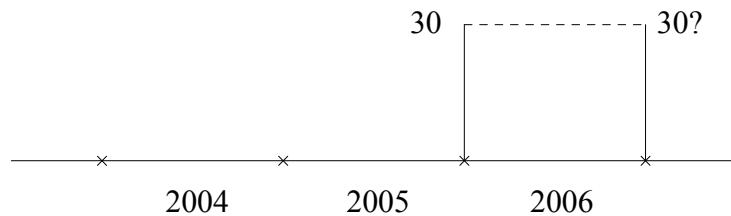
Again we would assume that the population has varied linearly over each calendar year. This means that the estimated population sizes in the middle of 2004 and 2005 would be 30,000 and 35,000. Here, the widths of each section are $\frac{1}{2}$ year, so we would use the approximation:

$$\frac{1}{2}(30,000 + P_{2005}) \times \frac{1}{2} + \frac{1}{2}(P_{2005} + 35,000) \times \frac{1}{2} = 17,500 + 18,750 = 36,250 \quad [2]$$

(iii) ***Period: 1 January 2006 to 31 December 2006***

In the absence of any additional information about the population size after 1 January 2006, we would have to assume that it remained constant. So we would use the approximation:

$$P_{2006} = 30,000 \quad [2]$$

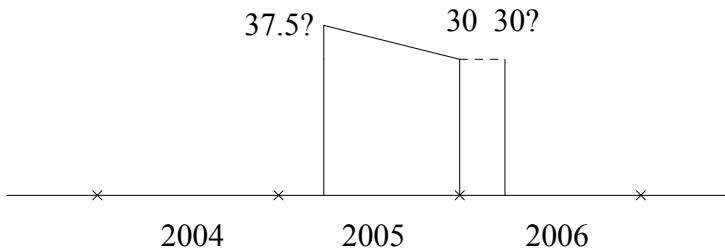


This approach would be unreliable and hence not very satisfactory. For example, the population numbers could continue on a downward “trend” to 20,000, leading to a central exposed to risk of 25,000, which is a significantly different figure.

(iv) ***Period: 1 April 2005 to 31 March 2006***

Here we would assume that the population has varied linearly over the 2005 calendar year and has then remained constant. This means that the estimated population size on 1 April 2005 would be 37,500. So we would use the approximation:

$$\frac{1}{2}(37,500 + P_{2006}) \times \frac{3}{4} + P_{2006} \times \frac{1}{4} = 25,312.5 + 7,500 = 32,812.5 \quad [2]$$



If we had reason to believe that the downward trend seen during 2005 would continue into 2006, we would estimate the population at 31 March 2006 to be 27,500, leading to an estimate of 32,500.

Solution 4.11

The chi square test may miss a small number of outliers, but not a large number.

So I is true, but II and III are false. [3]

Solution 4.12

The answer is C. This is the key point about a graphical graduation. [2]

Solution 4.13

Most standard tables are already smooth. Hence a simple transformation of it will leave smoothness undisturbed, so I is false. [1]

Even if the investigation covers the whole population, the numbers of deaths observed will still be random and we will still need to graduate the crude rates, so II is false. (If we use the whole population the random variations will probably be quite small, but we would need an infinite population to remove them completely!) [1]

III is true. [1]

Solution 4.14

The answer is D.

We can't just calculate the sum of the ISDs given. Although the numerator of the cumulative deviation is based on total actual deaths minus total expected deaths, the denominator isn't additive in this way. [3]

Solution 4.15

The calculation is $\frac{922 - 950}{\sqrt{950}} = -0.908$. [2]

Solution 4.16

By mathematical formula.

The standard tables published in recent years by the CMI have all been graduated by mathematical formula. [2]

Solution 4.17(i) ***Calculations using formula***

The required values are shown in the following table:

x	μ_x	$\Delta\mu_x$	$\Delta^2\mu_x$	$\Delta^3\mu_x$
80	0.0614520	0.0061452	0.0006145	0.0000615
81	0.0675972	0.0067597	0.0006760	
82	0.0743569	0.0074357		
83	0.0817926			

[3]

(ii) ***Calculations using observed rates***

The values based on the observed rates would be:

x	$\hat{\mu}_x$	$\Delta\hat{\mu}_x$	$\Delta^2\hat{\mu}_x$	$\Delta^3\hat{\mu}_x$
80	0.0514520	0.0161452	-0.0093855	0.0100615
81	0.0675972	0.0067597	0.0006760	
82	0.0743569	0.0074357		
83	0.0817926			

[2]

(iii) ***Why use third differences in smoothness test***

The second differences would be exactly zero if the rates followed a linear function. The third differences would be exactly zero if the rates followed a quadratic function. Although these might be satisfactory approximations over some parts of the age range, they would not allow for the rapidly increasing rates observed at the older ages. So we usually need to look beyond the second differences.

[2]

The figures calculated in part (i) show that the third differences are, in fact, relatively small even at older ages (while the second and first differences are approximately 10 and 100 times bigger).

[1]

The figures in part (ii) suggest that when random sampling errors (the -0.01) are present in the data, the differences may actually start to *increase*, rather than *decrease*, as we calculate higher order differences. For this reason, it is usually best not to go beyond the third differences.

[1]

Solution 4.18

It should adhere to the data and be smooth enough for the purposes that it will be used for.

[2]

Solution 4.19***Advantages of the graphical method***

- it can give good results even when data are scanty [½]
- it is easy to make allowance for special features (“intrinsic roughness”) [½]
- it naturally allows weight to be given to those ages where most data is available [½]
- it allows scope for individual judgement (experience) [½]
- it can be done quickly, without the need for a computer [½]
- it involves the actuary in the data, which may give him or her a better understanding of the rates. [½]

Solution 4.20***Disadvantages of the graphical method***

- a relatively high degree of skill is required [½]
- the method only gives results to approximately 3 significant figures [½]
- because of the scales involved, it is often necessary to draw the curve in two parts (although a transformation could be used) [½]
- individual judgement can lead to bias and prejudice [½]
- different results can be obtained from the same data [½]
- it can be difficult to achieve a high degree of smoothness [½]
- it is unclear as to how many degrees of freedom should be used when testing the data using a chi square test for overall adherence to the data. [½]

Solution 4.21***Advantages of the standard table method***

- the method can give good results on very scanty data [½]
- you usually do not have to bother testing for smoothness [½]
- knowledge of other tables is automatically brought into the graduation [½]
- there should be little difficulty with the ends of the table, *ie* amount of extrapolation required is limited. [½]

Solution 4.22***Disadvantages of standard table method***

- reliability of results can be doubtful if there is little data (although this is true for other graduation methods as well) [½]
- it is not always possible to find a suitable standard table (and thus adherence to data would be poor if this were the case) [½]
- any errors in the original table will be repeated. [½]

Solution 4.23***Advantages of parametric formula method***

- the rates will automatically be smooth [½]
- easy to identify mortality trends if the same formula is used [½]
- the goodness of fit is usually satisfactory [½]
- calculations can be computerised [½]
- can give most weight to the ages where most data was available. [½]

Solution 4.24***Disadvantages of parametric formula method***

- it is often difficult to find a single formula that fits over the whole age range [½]
- it can be very time consuming, even with computerisation. [½]

Solution 4.25(i) **Probability**

The probability that a specified x individuals will exit the population (and $n-x$ will not) is:

$$q^x(1-q)^{n-x} \quad [1]$$

There are $\binom{n}{x}$ possible ways of selecting these individuals. [1]

So the probability that exactly x individuals will leave the population is:

$$\binom{n}{x} q^x(1-q)^{n-x} \quad [1]$$

(ii) **Maximum likelihood estimator of q**

The likelihood of recording exactly θ decrements in an initial population of n individuals if the true value of the initial rate of decrement is q is:

$$L(q) = \binom{n}{\theta} q^\theta (1-q)^{n-\theta} \quad [1]$$

This can be maximised by maximising its log:

$$\log L(q) = \log \binom{n}{\theta} + \theta \log q + (n - \theta) \log (1 - q)$$

Differentiating with respect to q :

$$\frac{\partial}{\partial q} \log L(q) = \frac{\theta}{q} - \frac{n - \theta}{1 - q} \quad [1]$$

The MLE is found by equating this to zero:

$$\frac{\theta}{q} = \frac{n - \theta}{1 - q}$$

Rearranging and cancelling:

$$\theta - \theta q = nq - \theta q \Rightarrow \theta = nq$$

So:

$$\hat{q} = \frac{\theta}{n} \quad [1]$$

This is a maximum since $\frac{\partial^2}{\partial q^2} \log L(q) = -\frac{\theta}{q^2} - \frac{n-\theta}{(1-q)^2} < 0$. [1]

(iii) ***Maximum likelihood estimator of μ***

Here we are assuming that $\theta \sim \text{Poisson}(\mu E^c)$.

The expected value of the estimator is:

$$E(\tilde{\mu}) = E(\theta/E^c) = \frac{1}{E^c} E(\theta) = \frac{1}{E^c} \mu E^c = \mu$$

Since this equals the true value, the estimator is unbiased. [3]

Solution 4.26

(i) ***Assumptions underlying the Poisson model***

The lives are assumed to be independent and identical. [½]

The time spent under observation V is known ($= E_x^c$). [½]

The number of deaths is a random variable with a distribution that is $\text{Poisson}(\mu V)$. [1]

(ii) ***Maximum likelihood estimator***

Since D has a Poisson distribution:

$$P(D = d) = \frac{(\mu V)^d e^{-(\mu V)}}{d!}$$

Hence the likelihood of observing d deaths if the true value of the hazard rate is μ is:

$$L(\mu) = \frac{(\mu V)^d e^{-(\mu V)}}{d!} \quad [\frac{1}{2}]$$

We can maximise the log of this function, which is easier:

$$\log L(\mu) = -\mu V + d \log \mu - \text{const} \quad [\frac{1}{2}]$$

Differentiating with respect to μ and setting the derivative to zero:

$$-V + \frac{d}{\mu} = 0 \quad [\frac{1}{2}]$$

The second derivative is $\frac{-d}{\mu^2}$, which is less than zero if $d > 0$. Hence:

$$\hat{\mu} = \frac{d}{V} \quad \text{and} \quad \tilde{\mu} = \frac{D}{V} \quad [\frac{1}{2}]$$

(iii) ***Expected value and variance of maximum likelihood estimator***

Since $D \sim \text{Poisson}(\mu V)$, we have:

$$(a) \quad \tilde{\mu} = \frac{D}{V} \Rightarrow E[\tilde{\mu}] = \frac{1}{V} E[D] = \frac{1}{V} \times \mu V = \mu \quad [1]$$

and:

$$(b) \quad \text{var}[\tilde{\mu}] = \frac{1}{V^2} \text{var}[D] = \frac{\mu V}{V^2} = \frac{\mu}{V} \quad [1]$$

Solution 4.27(i) **Kaplan-Meier estimate of the survival function**

The table below shows the entrants (e), deaths (d) and censorings (c) in this investigation, and the number at risk immediately before the events that happen at time t .

Time, t (months)	0	1	2	3	6	8	10	11	12
Events	e	e	e	d	e	c	c	d	c
	e		e		e	c	d		c
	e								
Number at risk before t	0	3	4	6	5	7	5	3	2

The Kaplan-Meier estimate of the survival is a step function that starts at 1 and steps down every time a death is observed. So, if we measure time in years from the 70th birthday, then:

$$\hat{S}_{T_{70}}(t) = 1 \text{ for } 0 \leq t < \frac{3}{12}$$

Out of the 6 lives at risk at time $\frac{3}{12}$, one is observed to die. There are no more deaths until time $\frac{10}{12}$. So:

$$\hat{S}_{T_{70}}(t) = 1 - \frac{1}{6} = \frac{5}{6} \text{ for } \frac{3}{12} \leq t < \frac{10}{12}$$

Just before time $\frac{10}{12}$, there are 5 lives at risk. We observe one death and one censoring at this time, and we assume that the death occurs first. There are no more deaths until time $\frac{11}{12}$. So:

$$\hat{S}_{T_{70}}(t) = \frac{5}{6} \times \left(1 - \frac{1}{5}\right) = \frac{5}{6} \times \frac{4}{5} = \frac{2}{3} \text{ for } \frac{10}{12} \leq t < \frac{11}{12}$$

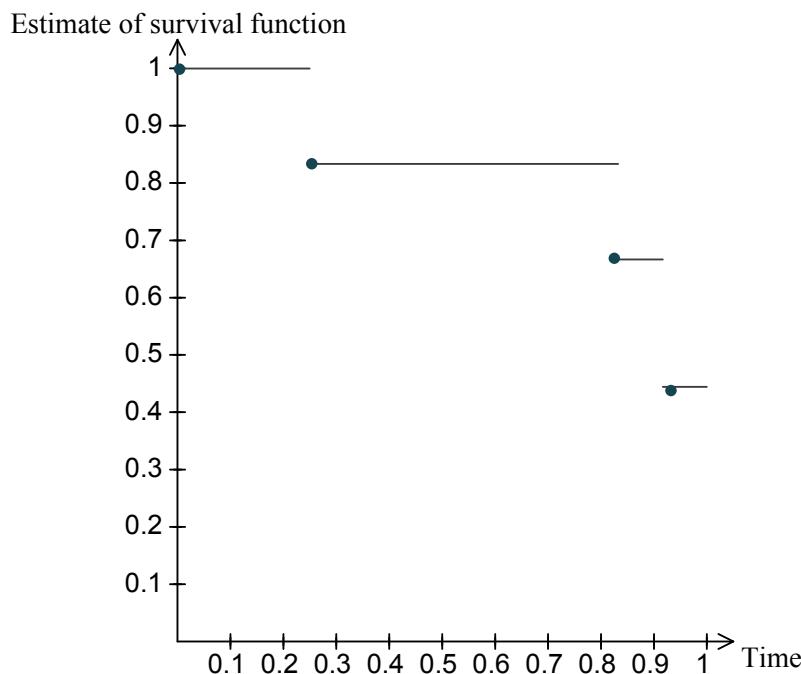
Just before time $\frac{11}{12}$, there are 3 lives at risk. One of these lives dies at time $\frac{11}{12}$. This is the last death to be observed. So:

$$\hat{S}_{T_{70}}(t) = \frac{5}{6} \times \frac{4}{5} \times \left(1 - \frac{1}{3}\right) = \frac{5}{6} \times \frac{4}{5} \times \frac{2}{3} = \frac{4}{9} \text{ for } \frac{11}{12} \leq t \leq 1$$

Summarising this we have:

$$\hat{S}_{T_{70}}(t) = \begin{cases} 1 & \text{for } 0 \leq t < \frac{3}{12} \\ \frac{5}{6} & \text{for } \frac{3}{12} \leq t < \frac{10}{12} \\ \frac{2}{3} & \text{for } \frac{10}{12} \leq t < \frac{11}{12} \\ \frac{4}{9} & \text{for } \frac{11}{12} \leq t \leq 1 \end{cases} \quad [3]$$

The graph of this function is as follows:



[2]

(ii) ***Confidence interval***

We can write:

$$q_{70} = P(T_{70} \leq 1) = F_{T_{70}}(1) = 1 - S_{T_{70}}(1)$$

So a 90% confidence interval for q_{70} is:

$$\hat{F}_{T_{70}}(1) \pm 1.6449 \sqrt{\text{var}(\tilde{F}_{T_{70}}(1))} \quad [1]$$

From part (i):

$$\hat{F}_{T_{70}}(1) = 1 - \hat{S}_{T_{70}}(1) = 1 - \frac{4}{9} = \frac{5}{9}$$

Also, using Greenwood's formula, which is given on Page 33 of the *Tables*:

$$\text{var}(\tilde{F}_{T_{70}}(1)) = (1 - \tilde{F}_{T_{70}}(1))^2 \sum_{t_j \leq 1} \frac{d_j}{n_j(n_j - d_j)}$$

where t_j denotes the j th observed lifetime, n_j denotes the number of lives at risk of dying just before time t_j , and d_j denotes the number of deaths at time t_j . So:

$$\text{var}(\tilde{F}_{T_{70}}(1)) = (1 - \frac{5}{9})^2 \left[\frac{1}{6 \times 5} + \frac{1}{5 \times 4} + \frac{1}{3 \times 2} \right] = 0.049383 \quad [1]$$

and a 90% confidence interval for q_{70} is:

$$\frac{5}{9} \pm 1.6449\sqrt{0.049383} = (0.1900, 0.9211) \quad [1]$$

(iii) ***Central and initial exposed to risk***

The contributions made by each life to the central and initial exposed to risk at age 70 last birthday are as follows:

Life	Contribution to E_{70}^c	Contribution to E_{70}
1	12 months	12 months
2	11 months	12 months
3	8 months	8 months
4	2 months	11 months
5	10 months	10 months
6	8 months	8 months
7	4 months	6 months
8	2 months	2 months
Total	57 months = 4 years, 9 months	69 months = 5 years, 9 months

[3]

Recall that the central exposed to risk is the same as the waiting time. For a censored life, the initial exposed to risk is also just the waiting time for that life. However, for each death, the contribution to the initial exposed to risk at age 70 last birthday is calculated as $71 - (70 + a_i)$, where $70 + a_i$ is the age at which the i th life started to be observed. In calculating the initial exposed to risk, those who die are exposed to the end of the year of death.

(iv) ***Actuarial estimate***

The actuarial estimate of q_{70} is:

$$\hat{q}_{70} = \frac{d_{70}}{E_{70}} = \frac{3}{5\%_{12}} = 0.52174 \quad [1]$$

(v) ***Poisson estimate and confidence interval***

The Poisson model assumes that μ is constant between the ages of 70 and 71. The maximum likelihood estimate (MLE) is given by:

$$\hat{\mu} = \frac{d_{70}}{E_{70}^c} = \frac{3}{4\%_{12}} = 0.63158 \quad [1]$$

The Poisson estimate of q_{70} is then given by:

$$1 - e^{-\hat{\mu}} = 1 - e^{-0.63158} = 0.46825 \quad [1]$$

using the invariance property of MLEs.

A 90% confidence interval for μ based on the Poisson model is:

$$\hat{\mu} \pm 1.6449 \sqrt{\frac{\hat{\mu}}{E_{70}^c}} = 0.63158 \pm 1.6449 \sqrt{\frac{0.63158}{4.75}} = (0.03178, 1.23138) \quad [2]$$

where the standard error of $\hat{\mu}$ is estimated using the Cramér-Rao lower bound (Page 23 of the *Tables*).

So a 90% confidence interval for q_{70} based on this model is:

$$(1 - e^{-0.03178}, 1 - e^{-1.23138}) = (0.03128, 0.70811) \quad [1]$$

(vi) ***Comment***

The Kaplan-Meier estimate of q_{70} is 0.55556.

The actuarial estimate of q_{70} is 0.52174.

The Poisson estimate of q_{70} is 0.46825.

These estimates are different because the underlying models are based on different assumptions. [½]

The Kaplan-Meier estimate is based on a non-parametric approach – that is, it makes no assumption about the distribution of the future lifetime random variable. We use the exact timings of the deaths in the calculation of Kaplan-Meier estimates. [1]

The actuarial estimate is based on method of moments estimation and assumes that the indicator random variables for death are binomially distributed. It does not use the information about the exact timings of the deaths and assumes that the Balducci property holds. The Balducci property implies that the force of mortality is decreasing between integer ages, which is unrealistic for lives aged 70. [1]

The Poisson estimate assumes that the force of mortality is constant between integer ages. This is not very realistic for lives aged 70, but is better than assuming a decreasing force of mortality. In this model, the number of deaths between 70 and 71 is assumed to be a Poisson random variable. This is not exactly true, since a Poisson random variable has no upper limit. However, provided that the probabilities of obtaining large numbers of deaths are small, it should be a reasonable approximation. The Poisson model uses the exact timings of the deaths in the calculation of the estimates, and this is another reason why it should be more reliable than the actuarial estimate. [1½]

Here the number of deaths is large compared to the numbers exposed to risk. So the actuarial estimate will be unsatisfactory. Both the Poisson and the Kaplan-Meier estimates will be satisfactory. [1]

[Maximum 4]

Solution 4.28(i) ***Exposed to risk formula***

Define $P'_x(t)$ to be a census of the number of policies in force at time t after 1 January 2007 who were aged x at time t , where:

$$x = \text{age last birthday at the date of the census.} \quad [\frac{1}{2}]$$

Then the central exposed to risk at age x last birthday during the period of the investigation (the calendar years 2007 and 2008) is:

$$E_x^c = \int_{t=0}^2 P'_x(t) dt \quad [\frac{1}{2}]$$

Assuming that $P'_x(t)$ is linear in t over the intervals $(0,1)$ and $(1,2)$ we can use the mid-ordinate rule to approximate the integral:

$$E_x^c = \int_{t=0}^2 P'_x(t) dt = P'_x\left(\frac{1}{2}\right) + P'_x\left(1\frac{1}{2}\right) \quad [1]$$

The lives in the census $P'_x\left(\frac{1}{2}\right)$ are those aged x last birthday at time $\frac{1}{2}$. The lives in the census $P_x(2007)$ are those age x nearest birthday at time $\frac{1}{2}$. So we can write:

$$P'_x\left(\frac{1}{2}\right) = \frac{1}{2}(P_x(2007) + P_{x+1}(2007))$$

assuming that dates of birth are uniformly distributed over the calendar year. [1]

Similarly:

$$P'_x\left(1\frac{1}{2}\right) = \frac{1}{2}(P_x(2008) + P_{x+1}(2008)) \quad [\frac{1}{2}]$$

Substituting we obtain the exposed to risk formula:

$$E_x^c = \frac{1}{2}(P_x(2007) + P_{x+1}(2007)) + \frac{1}{2}(P_x(2008) + P_{x+1}(2008)) \quad [\frac{1}{2}]$$

(ii) ***Age for estimated rates***

The age at the beginning of the rate year (life year) is x . No assumptions are needed for this result. So:

$$\hat{\mu}_x = \frac{\theta_x}{\frac{1}{2}(P_x(2007) + P_{x+1}(2007)) + \frac{1}{2}(P_x(2008) + P_{x+1}(2008))}$$

estimates μ_{x+f} where $f = +\frac{1}{2}$. [1]

Solution 4.29

The statistical information given allows us to carry out a signs test and a grouping of signs test only.

The null hypothesis is that the observed rates look as if they come from a population in which the graduated rates are the true rates. [½]

Signs test

The observed value of the test statistic, P , is 15. If the null hypothesis is true, then the sampling distribution of the test statistic will be Binomial(40, 0.5). [1]

The Tables do not allow us to carry out an exact test, so we can use the Normal approximation to the binomial sampling distribution. We are using a continuous distribution to approximate a discrete distribution, so we should use a continuity correction.

The observed value of the standard normal variate is:

$$\frac{15\frac{1}{2} - \frac{1}{2} \times 40}{\sqrt{\frac{1}{2} \times \frac{1}{2} \times 40}} = \frac{-4.5}{3.1623} = -1.42 \quad [1]$$

This is a two-tailed test and both small and large values of the test statistic will lead to the rejection of the null hypothesis. At the 5% level of significance the critical values are -1.96 and $+1.96$, so the result is not significant. [1]

Grouping of signs test

The observed value of the test statistic, G , is 8.

[½]

This is a one-tailed test and small values of the test statistic will be significant. The critical values of the test statistics are given on p189 of the *Tables*.

[½]

Here n_1 , the number of positive deviations is 15 and n_2 , the number of negative deviations is 25. The critical value of the test statistics at the 5% level of significance is 6.

[1]

So the observed value of the test statistic does not lie in the critical region and the data support the null hypothesis.

[½]

Solution 4.30

It is necessary to graduate crude rates for practical use:

- to make them fit for the purpose for which they are intended [1]
- to remove random sampling errors, thus better estimating the true underlying mortality rates [1]
- to allow the rate at a particular age to be set with reference to the rates at adjacent ages [1]
- to produce a set of mortality rates that progress smoothly from age to age which allows a practical smooth set of premium rates to be produced. [1]

Solution 4.31(i) ***Graphical graduation and confidence intervals***

First calculate the crude mortality rate at each age from the data and plot these rates on a graph. It may be necessary to use a logarithmic scale if there is a wide range of q_x values.

[1]

Consideration may need to be given to the grouping of data if, due to small exposures to risk at each age, there is a large degree of random error in the crude rates and it is difficult to draw a smooth curve.

[1]

If grouping is necessary, the grouped rate should be determined by dividing the sum of the deaths for the group of ages by the sum of the exposures at these ages. The age for which the grouped rates should be plotted should be the weighted average ages, where the weights are the exposures to risk.

[2]

To derive 95% confidence intervals we use the fact that:

$$\text{var}(\theta_x) = E_x q_x p_x$$

$$\Rightarrow \text{var}(\tilde{q}_x) = q_x p_x / E_x$$

Hence the probability that the interval:

$$\left\{ q_x - 1.96 \sqrt{\frac{q_x p_x}{E_x}}, q_x + 1.96 \sqrt{\frac{q_x p_x}{E_x}} \right\} \quad [2]$$

contains the true underlying value q_x is 0.95.

We use \hat{q}_x (the observed value) as an approximation to q_x , and also use the approximation $p_x \approx 1$.

$$\text{Hence we use the confidence interval } \hat{q}_x \pm 1.96 \frac{\sqrt{\theta_x}}{E_x}. \quad [1]$$

The confidence interval may then be plotted and used as a guide in drawing the curve, which should not fall outside the confidence interval at more than one in twenty ages.

[1]

(ii) ***Testing for smoothness***

To test for smoothness, calculate third differences of the graduated rates. These third differences should be smooth in progression and small in magnitude. [2]

(iii) ***Suitability***

There is a large amount of data and so the crude rates should be fairly reliable. In such cases, it might be more normal to use a standard table or mathematical formula graduation. These have the advantages of automatic smoothness and improved precision. [2]

Also, because there is so much data the need for individual judgement is reduced and this advantage of a graphical graduation is less valuable. [1]

However, it may be difficult to find an appropriate standard table or formula. [1]

Solution 4.32(i) ***Advantages of a graphical graduation***

- can give good results even when data is scanty [½]
- easy to make allowance for special features such as discontinuities [½]
- if confidence intervals can be drawn most weight can be given to the ages where we are more confident of the estimated rates [½]
- allows scope for individual judgement (experience) [½]
- can give an insight into how select rates run into ultimate rates. [½]

Disadvantages of a graphical graduation

- it is usually impossible to obtain sufficient places of decimals [½]
- it can be difficult to achieve a high degree of smoothness [½]
- a relatively high degree of skill is required [½]
- due to the scales involved, it is often necessary to draw the curve in two parts [½]
- and it can be difficult to achieve smoothness at the “joins” [½]
- individual judgement can lead to bias and prejudice [½]
- different results can be obtained from the same data. [½]

Circumstances under which a graphical graduation may be employed

- where a highly accurate answer is not essential, a graphical graduation can be performed very quickly and give a good answer [1]
- if the data are scanty, a graphical graduation may be the only feasible way of graduating [½]
- if the table being prepared is not intended as a standard [½]
- where special features need to be incorporated, eg discontinuities in rates because of pension scheme rules. [½]

[Maximum 7]

(ii) *Difficulties when applying a chi square test to a graphical graduation*

When applying a chi square test to a set of data it is necessary to determine the number of degrees of freedom for the test statistic. The number of degrees of freedom to use when the observed rates have been graduated graphically is not obvious. [1]

The graduating curve has to a certain extent been forced to fit the rough data but it is subjective as to how many degrees of freedom (DF) should be deducted for this. [½]

A reduction of about two or three DF for each section of the curve drawn is generally considered appropriate. [½]

The chi square test is thus approximate and any result should be considered intelligently and not just blindly accepted. [1]

Solution 4.33(i) ***Smoothness and fidelity to data***

Due to natural forces one would expect mortality rates to progress smoothly from age to age, *ie* there should be no sudden change in the curvature of the mortality curve. Death rates are initially produced from raw data by dividing the number of deaths at each age by the corresponding exposed to risk. These death rates will have been obtained from a sample of the population and will contain sampling errors. A graduation attempts to remove random sampling errors by producing a set of rates that progress smoothly from age to age, *ie* in line with what we expect. [2]

There are also practical advantages in having smooth rates in that the resulting actuarial functions will not have any irregularities and will thus be able to produce smooth premium rates. The graduated rates must, however, reasonably represent the underlying experience to have any meaning. Ideally one would like to end up with a smooth curve which fitted the data. [2]

In practice, due to the random fluctuations in the data it is necessary to compromise between smoothness and goodness of fit. Too close a fit will not be smooth and if the graduation is too heavy it will result in genuine features of the experience being lost and unsatisfactory adherence to the observations. (Thus there should be some intrinsic roughness in the male table at about age twenty, for example.) [2]

(ii) ***Chi square test***

The null hypothesis is that the graduation adheres closely to the data. [1]

The value of the chi square statistic is 7.29 on 7 degrees of freedom under the null hypothesis. The mean of the sampling distribution is 7 and so the observed value is not significant and the null hypothesis cannot be rejected. The chi square test does however have limitations, such as the inability to detect some large deviations balanced by lots of smaller ones. Consequently, it is necessary to perform some other statistical tests. [2]

Smoothness

Age	$10,000 q_x$	1st differences	2nd differences	3rd differences
65	225	25	1	1
66	250	26	2	0
67	276	28	2	2
68	304	30	4	0
69	334	34	4	1
70	368	38	5	0
71	406	43	5	1
72	449	48	6	
73	497	54		
74	551			

The graduation is acceptably smooth since the second order differences progress in an orderly fashion and the third order differences are small. [2]

Individual standardised deviations

The standardised deviations are often calculated approximately as:

$$\frac{(\theta_x - E_x q_x)}{\sqrt{E_x q_x}}$$

Age	Actual +	– Expected –	Expected	Standardised Deviation
65	20		1350	0.544
66		50	1250	-1.414
67	13		1187	0.377
68		35	1125	-1.043
69	8		1002	0.253
70		7	957	-0.226
71	46		934	1.505
72	7		943	0.228
73	6		994	0.190
74		32	992	-1.016

[2]

The actual and expected distribution of the standardised deviations is as follows:

Range	Expected	Actual
(-3,-2)	0.2	0
(-2,-1)	1.4	3
(-1,0)	3.4	1
(0,1)	3.4	5
(1,2)	1.4	1
(2,3)	0.2	0

The actual distribution looks a little unusual compared to the expected, hence there is some (weak) evidence of non-adherence to data. [1]

The proportion of absolute standardised deviations between -1 and 1 is expected to be $2/3$. [1]

Hence the graduation passes the test. [1]

This test fails to detect the clumping of deviations of the same sign. [1]

Signs test

If the graduation represents the mortality from which the experience is drawn, the deviations are equally likely to be positive or negative. We therefore expect the number of positive deviations to be a binomial random variable with parameters $n = 10$ and $p = \frac{1}{2}$.

In order of ascending age the signs of the deviations (*ie* A–E) are:

+ - + - + - + + + -

There are 6 positive signs and 4 negative signs which is not significant at any level. [1]

This test fails to detect the clumping of deviations of the same sign, and would not detect anything unusual about the absolute size of the deviations, *ie* would not detect outliers. [1]

Grouping of signs

The number of positive signs $n_1 = 6$, and the number of negative signs $n_2 = 4$.

The number of groups of positive signs is 4, which is greater than the critical value of 1 given in the *Tables*. Hence it is not significant. [2]

This test fails to detect excessive absolute deviations. It can also lead to a different conclusion when groups of negative signs are considered rather than groups of positive signs! [1]

Cumulative deviations

The observed value is -24 with a standard deviation of $103.6 (= 10,734^{1/2})$, which is not significant. (The observed test statistic is -0.23 , which has a standard normal sampling distribution.) [2]

This test will not detect clumping of deviations of the same sign. It will not detect outliers if they lie either side of the curve (hence cancelling each other out). [1]

Conclusion

Overall we do not have sufficient evidence to reject our hypothesis. However, the distribution of the individual standardised deviations looks a little strange. This could be as a result of bad data, eg misstatements of age or misrecording of data. [1]

Part 5 – Revision Questions

This part contains 100 marks of questions testing the material from the whole CT4 course. You may like to try these questions under exam conditions as a mock exam of 3 hours.

Question 5.1

A Markov chain is determined by the transition matrix:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}$$

Find the period of each of the states in this chain.

[2]

Question 5.2

A Markov chain is determined by the transition matrix:

$$P = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Determine which (if any) of the following are stationary distributions for the chain.

I $(1, \frac{3}{4}, -\frac{3}{4})$

II $(0, 0, 1)$

III $(\frac{2}{7}, \frac{3}{14}, \frac{1}{2})$

[3]

Question 5.3

An enterprising scientist is modelling the population of “tribbles”, an alien species renowned for its capacity to reproduce.

The size of the population at time t , X_t is thought to follow a Markov jump process with states $\{1, 2, 3, \dots\}$.

The scientist assumes that the population starts with one individual at time 0, $X_0 = 1$, which causes no problem as the tribble is asexual! The transition rates of the process depend on the current population size. The monthly rates are:

$$\mu_{ij} = \begin{cases} 2^i & \text{if } j = i + 1 \\ -2^i & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

- (i) Write down the generator matrix (transition rate matrix) of the process X_t . [1]
 - (ii) State the distribution of the random variable T_i , which measures the holding time in State i , and its expected value. [1]
 - (iii) Calculate $\sum_{i=1}^{\infty} E[T_i]$. [1]
 - (iv) Comment briefly on your answers. [1]
- [Total 4]

Question 5.4

A mortality table covers the age range 10 to 100, inclusive. The graduated rates have been produced using the parametric formula method – a formula with 3 parameters has been fitted to the crude rates. The deviations of the observed number of deaths from the expected number of deaths at each age have been calculated. You have:

Number of positive deviations: 41

Number of groups of positive deviations: 16

Carry out a grouping of signs test on these graduated rates and state your conclusion. [6]

Question 5.5

- (i) Explain why the criteria of smoothness and adherence to data both need to be met when performing a graduation of mortality data that will be used to set premium rates. [3]
 - (ii) Describe the smoothness test, stating any relevant formulae. [3]
- [Total 6]

Question 5.6

A mortality investigation was held between 1 January 2007 and 1 January 2009. The following information was collected. The figures in the table below are the numbers of lives on each census date with the specified age labels.

Age last birthday	Date		
	1.1.07	1.1.08	1.1.09
48	3,486	3,384	3,420
49	3,450	3,507	3,435
50	3,510	3,595	3,540

During the investigation there were 42 deaths at age 49 nearest birthday. Estimate μ_{49} stating any assumptions that you make. [7]

Question 5.7

You are trying to model the number of people who own mobile phones as a Markov jump process. You assume that once someone has bought a mobile phone they will never give it up, but they can only ever own one mobile phone. You also assume that the number of potential phone owners is infinite and every individual will live forever.

Finally, you assume that only one individual can buy a phone at any instant, and that the average length of time μ_i minutes before the next person buys one depends on the current number of mobile phone owners i , but not the time t .

- (i) Write down the generator matrix (transition rate matrix) for this process. [1]
 - (ii) Write down the integral form of the Kolmogorov forward equations for $P_{ij}(t)$, where $P_{ij}(t)$ denotes the probability that there will be j mobile phone owners by time $s+t$, given that there were i at time s . [3]
 - (iii) Obtain an expression for the probability that exactly one mobile phone will be sold in the next minute if there are currently i people with mobile phones. [2]
 - (iv) After fitting the model to recent data it looks like $\mu_i = (0.9999991705)^i$. Calculate the expected time before the whole population has a mobile phone if 1,000,000 do at present. Comment on your answer. [4]
- [Total 10]

Question 5.8

An insurance company operates a no claims discount system with discount levels of 0%, 30%, 40%, 50% and 60%. The rules are as follows:

- At the end of a claim free year, a policyholder moves up one level (or remains on the maximum discount level).
- At the end of a year in which exactly one claim was made, a policyholder drops back two levels (or moves to zero discount).
- At the end of a year in which more than one claim was made, a policyholder drops back to zero discount.

For a particular driver in any year, the probability of a claim free year is 0.7, the probability of exactly one claim is 0.2, and the probability of more than one claim is 0.1.

- (i) Write down the transition matrix for this time-homogeneous Markov chain. [2]
 - (ii) If the policyholder starts with no discount, calculate the probability that he is at the maximum discount level 6 years later. [2]
 - (iii) If a large number of people having the same claims distribution take out policies at the same time, calculate the proportion you would expect to be in each discount category in the long run. [6]
- [Total 10]

Question 5.9

The following model for the force of mortality for a life insurance company's annuitants has been proposed:

$$\mu(t, i) = (0.015 - 0.0001t) \cdot \exp[\alpha(x_i - 70) + \beta.y_i + \gamma.z_i]$$

where:

$\mu(t, i)$ = force of mortality for the i th life, in calendar year $2000 + t$;

x_i = age of the i th life;

$y_i = 1$ if the i th life is a smoker, or $y_i = 0$ if a non-smoker;

$z_i = 1$ if the i th life is male, or $z_i = 0$ if female; and

α, β, γ are the parameters of the model.

The following data have been observed over the calendar year 2003:

Risk characteristics	Number of annuitants	Number dying
Male non-smoker, average age 65	800	6
Male smoker, average age 60	200	5
Female non-smoker, average age 70	450	2
Female smoker, average age 65	150	1

You can assume that the numbers of annuitants in each class remained constant throughout the investigation period, and that the average age for each class can be treated as representing the value of x_i for each individual in that class.

- (i) Explain why this model is a proportional hazards model. [2]
 - (ii) Explain the importance of subdividing the data by age, sex and smoking status, and explain whether you think each of the parameters α , β and γ would be likely to be positive or negative. [3]
 - (iii) Calculate the force of mortality for female non-smokers with average age 70 in 2007, according to this model. [1]
 - (iv)
 - (a) Obtain an expression for the partial likelihood based on the given data, expressing your answer in terms of α , β and γ only.
 - (b) State how you would estimate the parameters of the model using the partial likelihood. [6]
- [Total 12]

Question 5.10

A pension scheme only allows retirement at exact age 65. An investigation of the mortality of the retired members of the scheme was carried out over the period 1 January 2001 to 31 December 2006.

The following data were obtained:

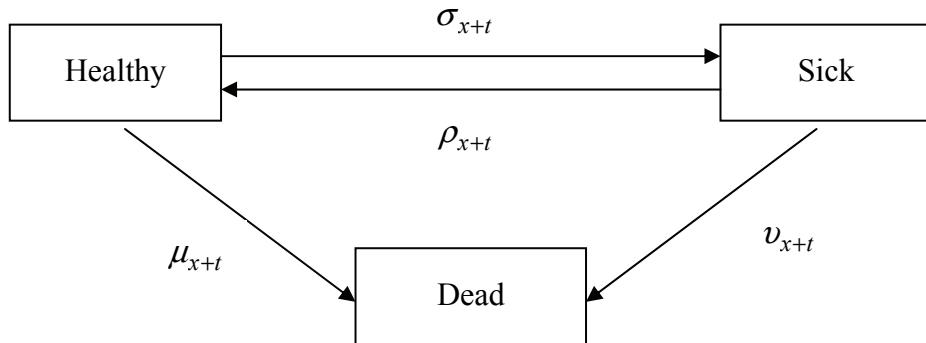
Member	Date of retirement	Date of death (if occurred during the investigation period)
1	1 April 1998	30 April 2005
2	1 August 2000	–
3	1 February 2001	–
4	1 June 2002	31 August 2004
5	1 August 2002	31 December 2006
6	1 March 2004	–
7	1 May 2004	30 November 2006
8	1 January 2005	–

All months should be assumed to be of equal length.

- (i) Explain the form or forms of censoring that are present in these data and in this observational plan. [2]
 - (ii) Calculate the Kaplan-Meier (product-limit) estimate of the survival function $S_{65}(t)$ from these data, stating clearly any additional assumptions that you make. [10]
 - (iii) Estimate the force of transition from alive to dead for a two-state Markov model for age 67 last birthday, using the data given. The exposed to risk should be calculated exactly. State all assumptions required, and state the age to which the estimate would be assumed to apply. [7]
- [Total 19]

Question 5.11

A three-state Markov model is represented by the following transition diagram:



The symbols σ_{x+t} , ρ_{x+t} , μ_{x+t} and ν_{x+t} represent the forces of transition at age $x+t$, where x is an integer and $0 \leq t < 1$. The symbol ${}_t p_x^{\bar{i}}$, $i = H, S, D$, represents the probability that a life, who is in state i at age x , remains in state i until at least age $x+t$.

- (i) Write down the assumptions that are usually made when applying this model. [2]
- (ii) Derive the differential equation:

$$\frac{\partial}{\partial t} {}_t p_x^{\bar{H}} = - {}_t p_x^{\bar{H}} (\sigma_{x+t} + \mu_{x+t})$$

and write down the relevant initial condition. [5]

- (iii) If $\sigma_{x+t} = \sigma$ and $\mu_{x+t} = \mu$ for all $0 \leq t < 1$, show that:

$${}_t p_x^{\bar{H}} = e^{-t(\sigma+\mu)} \quad [3]$$

- (iv) In a mortality and morbidity investigation the following data were recorded for lives between exact ages 50 and 51:

Total time (in years) spent in the healthy state:	750
Total time (in years) spent in the sick state:	32
Number of transitions from the healthy state to the sick state:	15
Number of transitions from the sick state to the healthy state:	35
Number of transitions from the healthy state to the dead state:	3
Number of transitions from the sick state to the dead state:	5

- (a) Write down the likelihood function and derive the maximum likelihood estimate of μ_{50+f} . State the value of f .
- (b) Construct an approximate 90% confidence interval for μ_{50+f} .
- (c) Estimate the value of $p_{50}^{\overline{HH}}$. [8]
- (v) A student has said:

“If you calculate $\hat{\rho}_{50+f}$, you get 1.094. This shows that the model is wrong, since you can’t have a value greater than 1. And anyway, the data must be wrong – how can you have 35 recoveries when only 15 people fell sick?”

Comment on this student’s remarks. [3]
[Total 21]

All study material produced by ActEd is copyright and is sold for the exclusive use of the purchaser. The copyright is owned by Institute and Faculty Education Limited, a subsidiary of the Institute and Faculty of Actuaries.

Unless prior authority is granted by ActEd, you may not hire out, lend, give out, sell, store or transmit electronically or photocopy any part of the study material.

You must take care of your study material to ensure that it is not used or copied by anybody else.

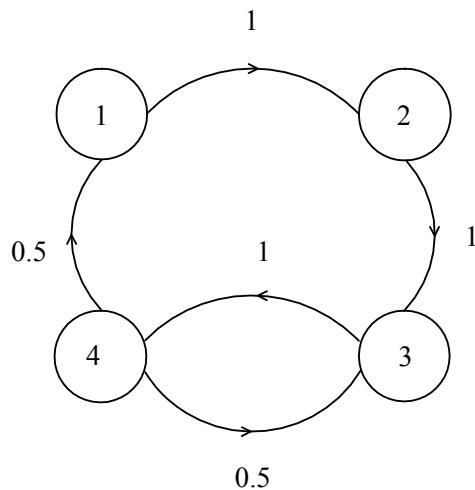
Legal action will be taken if these terms are infringed. In addition, we may seek to take disciplinary action through the profession or through your employer.

These conditions remain in force after you have finished using the course.

Part 5 – Revision Solutions

Solution 5.1

If we draw the transition graph for the Markov chain:



We can see that the chain is irreducible, therefore all the states will have the same period. [1]

We only need to find the period of one of the states. The chain can return to state 1 having started in state 1 after 4, 6, 8, 10, ... moves. The highest common factor of these numbers is 2, so the period of all the states in the chain is 2. [1]

Solution 5.2

Since the process is not irreducible it is possible for more than one solution.

- I This has negative entries so cannot be a distribution. [1]
- II This is a stationary distribution since $(0, 0, 1)P = (0, 0, 1)$. [1]
- III This is also a stationary distribution since $(\frac{2}{7}, \frac{3}{14}, \frac{1}{2})P = (\frac{2}{7}, \frac{3}{14}, \frac{1}{2})$, the entries are non-negative and sum to 1. [1]

Solution 5.3(i) **Generator matrix**

The generator matrix is:

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \dots \\ \left[\begin{matrix} -2 & 2 & 0 & 0 & 0 & 0 & \dots \\ 0 & -4 & 4 & 0 & 0 & 0 & \dots \\ 0 & 0 & -8 & 8 & 0 & 0 & \dots \\ 0 & 0 & 0 & -16 & 16 & 0 & \dots \\ 0 & 0 & 0 & 0 & -32 & 32 & \dots \\ 0 & 0 & 0 & 0 & 0 & -64 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{matrix} \right] \end{array}$$

[1]

(ii) **Expected value of i th holding time**

For any time homogeneous process, the holding time in State i has an exponential distribution with parameter depending on the current state. The value of the parameter is the total force out of State i .

The random variable T_i , measuring the holding time in State i , has an exponential distribution with parameter 2^i .

The expected value of the holding time in State i , $E[T_i]$ is 2^{-i} . [1]

(iii) **Sum of expected values**

We can sum the expected values as an infinite geometric series:

$$\sum_{i=1}^{\infty} E[T_i] = \sum_{i=1}^{\infty} 2^{-i} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1 \quad [1]$$

This result says that starting with one tribble at time $t = 0$, we expect the population to be infinite in 1 month's time! Tribbles breed extremely rapidly, the expected time between births halves with each birth. Such phenomena are called population explosions.

(iv) ***Comment***

Obviously as the population increased the model would cease to be applicable, owing for example, to food shortages. [1]

Solution 5.4

The null hypothesis for this test is:

$$H_0: \text{the graduated rates are the true underlying mortality rates for the population}$$

[1]

If G denotes the number of groups of positive deviations, then approximately we have:

$$G \sim N\left(\frac{n_1(n_2+1)}{n_1+n_2}, \frac{(n_1n_2)^2}{(n_1+n_2)^3}\right)$$

where n_1 and n_2 denote the number of positive and negative deviations, respectively. (This result is given on Page 34 of the *Tables*.)

Using the data given in the question:

$$E(G) = \frac{41 \times 51}{91} = 22.978$$

[1]

$$\text{var}(G) = \frac{(41 \times 50)^2}{91^3} = 5.577$$

[1]

The test statistic (calculated using a continuity correction since we are approximating a discrete random variable with a continuous one) is:

$$Z = \frac{16.5 - 22.978}{\sqrt{5.577}} = -2.74$$

[1]

Since we are testing for grouping of deviations of the same sign, we are only concerned about low values of G . So we compare the test statistic with the lower end of the standard normal distribution. [1]

The lower 5% point of the standard normal distribution is -1.6449 . Since $-2.69 < -1.6449$, we reject the null hypothesis at the 5% significance level and conclude that there is evidence of grouping of signs. [1]

Solution 5.5(i) ***Why both criteria have to be met***

The graduation process aims to produce a set of mortality rates that are the best estimates of the underlying rates. We require these rates to be smooth because:

- Large mortality investigations support the hypothesis that mortality rates vary gradually from age to age. [1]
- We want premiums to increase gradually from age to age. Any jumps or other anomalies would be hard to justify in practice. [1]

At the same time, the graduated rates must be representative of the mortality data from which they were derived. So we also require good adherence to data. [1]

(ii) ***Smoothness test***

The smoothness test examines the third differences in the graduated rates: $\Delta^3 \ddot{q}_x$ or $\Delta^3 \ddot{\mu}_{x+\frac{1}{2}}$. [1]

The third differences measure the change in curvature. [1]

A set of graduated rates is described as smooth if the third differences:

- are small in magnitude compared with the graduated rates themselves [½]
- progress gradually from age to age. [½]

Solution 5.6

The deaths are classified according to age nearest birthday. Lives aged 49 nearest birthday are between the exact ages of $48\frac{1}{2}$ and $49\frac{1}{2}$. The aged in the middle of this rate interval is 49. So:

$$\hat{\mu}_{49} = \frac{d_{49}}{E_{49}^c}$$

estimates μ_{49} . [1]

From the investigation, we have:

$$d_{49} = 42$$

If we define $P_x(t)$ to be the number of lives at time t (measured in years from 1 January 2007) aged x last birthday, then we know the values of $P_x(t)$ for $x = 48, 49, 50$ and $t = 0, 1, 2$.

However, the death data and the census data do not match. So we define a function $P'_x(t)$ that does match with the death data.

Let $P'_x(t)$ denote the number of lives at time t aged x nearest birthday. [½]

Then:

$$E_{49}^c = \int_0^2 P'_{49}(t) dt \quad [½]$$

and assuming that $P'_{49}(t)$ varies linearly between the census dates ... [½]

$$\begin{aligned} E_{49}^c &= \frac{1}{2} [P'_{49}(0) + P'_{49}(1)] + \frac{1}{2} [P'_{49}(1) + P'_{49}(2)] \\ &= \frac{1}{2} P'_{49}(0) + P'_{49}(1) + \frac{1}{2} P'_{49}(2) \end{aligned} \quad [1]$$

Now we need to figure out the numerical values of the terms in the expression above. We have:

$P'_{49}(0)$ = the number of lives at time 0 aged 49 nearest birthday

Lives aged 49 nearest birthday are between the exact ages of 48½ and 49½. So their age last birthday is either 48 or 49. [½]

Assuming that birthdays are uniformly distributed over the calendar year ... [½]

$$P'_{49}(0) = \frac{1}{2} [P_{48}(0) + P_{49}(0)] = \frac{1}{2} [3,486 + 3,450] = 3,468 \quad [½]$$

Similarly:

$$P'_{49}(1) = \frac{1}{2} [P_{48}(1) + P_{49}(1)] = \frac{1}{2} [3,384 + 3,507] = 3,445.5 \quad [½]$$

and:

$$P'_{49}(2) = \frac{1}{2} [P_{48}(2) + P_{49}(2)] = \frac{1}{2} [3,420 + 3,435] = 3,427.5 \quad [½]$$

So:

$$E^c_{49} = \frac{1}{2} \times 3,468 + 3,445.5 + \frac{1}{2} \times 3,427.5 = 6,893.25 \quad [½]$$

and:

$$\hat{\mu}_{49} = \frac{42}{6,893.25} = 0.006093 \quad [½]$$

Solution 5.7(i) **Generator matrix**

Define $\lambda_i = \frac{1}{\mu_i}$.

This process is like a Poisson process but with rates that are state dependent. The transition rates are:

$$\sigma_{i,i+1} = \lambda_i \quad \sigma_{i,i} = -\lambda_i$$

and $\sigma_{i,j} = 0$ otherwise. Note that i and j can take the values $0, 1, 2, \dots$

So the generator matrix is:

$$A = \begin{bmatrix} \dots & i & i+1 & i+2 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & -\lambda_i & \lambda_i & 0 & \dots \\ \dots & 0 & -\lambda_{i+1} & \lambda_{i+1} & \dots \\ \dots & 0 & 0 & -\lambda_{i+2} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \quad [1]$$

(ii) **Integral form of Kolmogorov forward equations**

The only way to go from state i to state i is to remain in state i throughout. Therefore:

$$P_{ii}(t) = e^{-\lambda_i t} \quad [1]$$

(You can quote this result in situations like this.)

When $j < i$, $P_{ij}(t)$ is zero since the process cannot move to a lower state. When $j > i$ we can argue as follows. To end up in state j we must have made our last transition from state $j-1$. The probability of this occurring during an infinitesimal interval at time w is $\sigma_{j-1,j}dw = \lambda_{j-1}dw$. The probability of then remaining in this state until time t is $e^{-\lambda_j(t-w)}$. In addition, the probability of going from state i at time 0 to state $j-1$ at time w is $P_{ij}(w)$. Taking into account the possible times s we have:

$$P_{ij}(t) = \int_0^t P_{ij-1}(w) \lambda_{j-1} e^{-\lambda_j(t-w)} dw \text{ for } j > i \quad [2]$$

Note that we are only told to write down the equation, so the above explanation is not necessary for the marks.

(iii) **Probability**

We can apply (ii):

$$\begin{aligned} P_{i,i+1}(t) &= \int_0^t P_{ii}(s) \lambda_i e^{-\lambda_{i+1}(t-s)} ds \\ &= \int_0^t e^{-\lambda_i s} \lambda_i e^{-\lambda_{i+1}(t-s)} ds \\ &= \lambda_i e^{-\lambda_{i+1} t} \int_0^t e^{-(\lambda_i - \lambda_{i+1})s} ds \\ &= \frac{\lambda_i e^{-\lambda_{i+1} t}}{(\lambda_i - \lambda_{i+1})} \left(1 - e^{-(\lambda_i - \lambda_{i+1})t}\right) \end{aligned}$$

$$\text{Therefore } P_{i,i+1}(1) = \lambda_i \frac{(e^{-\lambda_{i+1}} - e^{-\lambda_i})}{(\lambda_i - \lambda_{i+1})}. \quad [2]$$

(iv) ***Expected time until whole population has a mobile phone***

The expected length of time in state i is $\mu_i = (0.9999991705)^i$. The expected length of time until everyone has a mobile is therefore:

$$\sum_{k=1,000,000}^{\infty} (0.9999991705)^k = \frac{(0.9999991705)^{1,000,000}}{1 - (0.9999991705)} = 1 \text{ year} \quad [2]$$

This model might be OK over short time-scales, but the model may need to be recalibrated or fundamentally changed in the future. It is often unwise to try to project too far into the future. For example, even if future demand was at the level predicted according to the model it is unlikely that supply could meet this demand. The model does also make some simplifying assumptions that cannot be met in practice, eg an infinite population. [2]

Solution 5.8(i) **Transition matrix**

The transition matrix is:

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 & 0 \\ 0.3 & 0 & 0 & 0.7 & 0 \\ 0.1 & 0.2 & 0 & 0 & 0.7 \\ 0.1 & 0 & 0.2 & 0 & 0.7 \end{pmatrix} \quad [2]$$

(ii) **Probability of being on maximum discount in 6 years**

We want:

$$p_{0\%,60\%}^{(6)} = (P^6)_{1,5} = \sum_{k=1}^5 (P^3)_{1,k} (P^3)_{k,5} = \frac{240100}{1000000} = 0.2401. \quad [2]$$

Note that we don't need to work out the full matrix P^6 .

(iii) **Steady-state proportions**

We use the fact that as $n \rightarrow \infty$ we end up in the stationary distribution, which exists and is unique in this case as the process is finite and irreducible. We are looking for a distribution $(\pi_0, \pi_1, \pi_2, \pi_3, \pi_4)$ satisfying:

$$(\pi_0, \pi_1, \pi_2, \pi_3, \pi_4) \begin{pmatrix} 0.3 & 0.7 & 0 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 & 0 \\ 0.3 & 0 & 0 & 0.7 & 0 \\ 0.1 & 0.2 & 0 & 0 & 0.7 \\ 0.1 & 0 & 0.2 & 0 & 0.7 \end{pmatrix} = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4) \quad [1]$$

where we use subscripts 0,1,2,3 and 4 to represent the discount levels.

This is equivalent to the equations:

$$\begin{aligned} 3\pi_0 + 3\pi_1 + 3\pi_2 + \pi_3 + \pi_4 &= 10\pi_0 \\ 7\pi_0 + 2\pi_3 &= 10\pi_1 \\ 7\pi_1 + 2\pi_4 &= 10\pi_2 \\ 7\pi_2 &= 10\pi_3 \\ 7\pi_3 + 7\pi_4 &= 10\pi_4 \end{aligned}$$

We rearrange these and ignore the first since one equation is always redundant:

$$\begin{aligned} 7\pi_0 - 10\pi_1 + 2\pi_3 &= 0 \\ 7\pi_1 - 10\pi_2 + 2\pi_4 &= 0 \\ 7\pi_2 - 10\pi_3 &= 0 \\ 7\pi_3 - 3\pi_4 &= 0 \end{aligned} \quad [1]$$

We will use π_4 as the working variable. The fourth equation gives $\pi_3 = \frac{3}{7}\pi_4$. Substituting into the third equation then gives:

$$\pi_2 = \frac{10}{7}\pi_3 = \frac{10}{7} \times \frac{3}{7}\pi_4 = \frac{30}{49}\pi_4 \quad [1]$$

From the second equation:

$$\pi_1 = \left(\frac{10}{7} \times \frac{30}{49} - \frac{2}{7}\right)\pi_4 = \frac{202}{343}\pi_4 \quad [1]$$

Finally from the first equation:

$$\pi_0 = \left(\frac{10}{7} \times \frac{202}{343} - \frac{2}{7} \times \frac{3}{7}\right)\pi_4 = \frac{1726}{2401}\pi_4 \quad [1]$$

We therefore have the distribution:

$$(1726, 1414, 1470, 1029, 2401) \frac{\pi_4}{2401}$$

By applying the summation condition we get:

$$\frac{1}{8040}(1726, 1414, 1470, 1029, 2401) \quad [1]$$

Solution 5.9(i) ***Proportional hazards***

If we take the ratio $\frac{\mu(t,i)}{\mu(t,j)}$, for example, we obtain:

$$\frac{\exp[\alpha(x_i - 70) + \beta.y_i + \gamma.z_i]}{\exp[\alpha(x_j - 70) + \beta.y_j + \gamma.z_j]} = \text{Constant (independent of } t\text{)}$$

The model therefore predicts that, at any given time t , the force of mortality of a life having any particular set of characteristics (such as x_i , y_i and z_i) is a *constant proportion* of the force of mortality of a life having another particular set of characteristics (such as x_j , y_j and z_j). [2]

(ii) ***Subdividing into homogeneous groups***

The main reason is that the models of mortality that we use assume that all the lives involved experience mortality that is consistent with a particular probability distribution, *i.e.* that the lives are homogeneous with respect to their mortality. [1]

Subdividing the data according to characteristics that are considered to affect mortality aims to achieve homogeneity within the sub-groups obtained. [½]

In this case, older lives usually exhibit higher mortality than younger lives, males usually exhibit higher mortality than females, and smokers usually exhibit higher mortality than non-smokers. Hence we would expect each of the parameters α , β and γ to be positive. [1½]

(iii) ***Calculation of force of mortality***

The value of the exponential equals 1, so the force of mortality during 2007 for this grouping is simply $(0.015 - 0.0001 \times 7) = 0.0143$. [1]

(iv)(a) ***Partial likelihood***

We need:

$$L(\alpha, \beta, \gamma) = \prod_{\text{Deaths}} \frac{\exp[\alpha(x_j - 70) + \beta.y_j + \gamma.z_j]}{\sum \exp[\alpha(x_i - 70) + \beta.y_i + \gamma.z_i]}$$

where the summation in the denominator is the sum over all the lives that could have died at the time of j 's death, and the product is over all the deaths observed.

Given the assumption stated in the question, then the denominator is the same regardless of the time of death, ie:

$$800e^{-5\alpha+\gamma} + 200e^{-10\alpha+\beta+\gamma} + 450 + 150e^{-5\alpha+\beta} = E(\alpha, \beta, \gamma) \text{ (say)} \quad [2]$$

Hence

$$\begin{aligned} L(\alpha, \beta, \gamma) &= \frac{(e^{-5\alpha+\gamma})^6 (e^{-10\alpha+\beta+\gamma})^5 (e^{-5\alpha+\beta})}{(E(\alpha, \beta, \gamma))^{14}} \\ &= \frac{e^{-85\alpha+6\beta+11\gamma}}{(E(\alpha, \beta, \gamma))^{14}} \end{aligned} \quad [3]$$

(iv)(b) ***Estimating the parameters***

The values of α , β and γ would be found that together maximised the value of $L(\alpha, \beta, \gamma)$. This would be done using numerical techniques. [1]

Solution 5.10(i) **Types of censoring**

- Right censoring – cuts short observations that are in progress (here due to the end of the investigation period). [1]
- Because censoring times are known in advance, this is a form of Type I censoring. [1]

Note that while the data are truncated from the left, this is not “left censoring”.

(ii) **Calculating the Kaplan-Meier estimate**

“Duration” in the following table represents the duration since the 65th birthday, in years.

Member	Duration at 1.1.2001	Duration at death	Duration at 31.12.2006
1	2y 9m	7y 1m	–
2	0y 5m	–	6y 5m
3	–	–	5y 11m
4	–	2y 3m	–
5	–	4y 5m	–
6	–	–	2y 10m
7	–	2y 7m	–
8	–	–	2y 0m

Rearranging the data in duration order:

J	t_j	Change	n_j	λ_j	$S(t_j) = \prod_{i=1}^j (1 - \lambda_i)$
0	0		6		
1	0y 5m	+1	7	0	1
2	2y 0m	-1	6	0	1
3	2y 3m	D	5	0.1667	0.8333
4	2y 7m	D	4	0.2	0.6667
5	2y 9m	+1	5	0	0.6667
6	2y 10m	-1	4	0	0.6667
7	4y 5m	D	3	0.25	0.5
8	5y 11m	-1	2	0	0.5
9	6y 5m	-1	1	0	0.5
10	7y 1m	D	0	1.0	0

[3 marks for organising the data, 4 marks for calculations]

In the above:

t_j = duration at time of j th event;

“change” = the change in n_j resulting from the j th event;

n_j = the number alive at duration t_j immediately after the j th event;

λ_j = the number of deaths at duration t_j divided by n_{j-1} ;

$S(t_j)$ = the estimate of the survival function at duration t_j .

Note that the first two lives only came under observation after the study had started. So they need to be included as increments at these times.

The estimate of the survival function is therefore:

t	$S_{65}(t)$
$0 \leq t < 2\frac{3}{12}$	1
$2\frac{3}{12} \leq t < 2\frac{7}{12}$	0.8333
$2\frac{7}{12} \leq t < 4\frac{5}{12}$	0.6667
$4\frac{5}{12} \leq t < 7\frac{1}{12}$	0.5
$t \geq 7\frac{1}{12}$	0

[1 mark for summarising the results correctly (exact ranges are important)]

Assumptions:

- The censoring is non-informative: that is, no information about the probability distribution of the future lifetimes of the individuals can be obtained from the fact that any observation has been censored. [1]
- The lifetimes of each individual are assumed to be independent and identically distributed. [1]

[Total 10 for (ii)]

(iii) ***Estimating the force of mortality***

“67 last birthday” indicates the year of age beginning on the 67th birthday. The following shows the number of lives exposed to risk at each duration (in months) over the year of age [67, 68].

Duration (months)→						
0	6	3m	5	7m	4	9m
					5	10m

The total central exposed to risk, in life years, over the year of age [67, 68] is:

$$\frac{6 \times 3 + 5 \times 4 + 4 \times 2 + 5 \times 1 + 4 \times 2}{12} = \frac{59}{12} = 4.9167 \quad [4]$$

The estimate of the force of mortality is $\tilde{\mu}_{67+f} = \frac{2}{4.9167} = 0.4068$. [1]

$\tilde{\mu}_{67+f}$ would usually be assumed to estimate $\mu_{67+\frac{1}{2}}$. [1]

Assumptions:

- Force of mortality constant over the age range [67, 68]. [½]
 - Lives are independent and homogeneous with regard to their mortality. [½]
 - The probabilities that a life at any given age will be found in either state at any subsequent age depend only on the ages involved and the state currently occupied (the Markov assumption). [½]
 - ${}_h q_{x+t} = h \cdot \mu_{x+t} + o(h)$, ($t \geq 0$); $\lim_{h \rightarrow 0+} \frac{o(h)}{h} = 0$ [½]
- [Maximum 7 for (iii)]

Solution 5.11(i) ***Assumptions***

The assumptions that are usually made are:

- The Markov assumption, *ie* the probability of a life being in any given state at any given time in the future depends only on the current state and the ages involved. [1]
- For small values of h , and any $x, t > 0$:

$${}_h p_{x+t}^{HS} = h\sigma_{x+t} + o(h)$$

$${}_h p_{x+t}^{SH} = h\rho_{x+t} + o(h)$$

$${}_h p_{x+t}^{HD} = h\mu_{x+t} + o(h)$$

$${}_h p_{x+t}^{SD} = hv_{x+t} + o(h)$$

[1]

(ii) ***Differential equation***

Consider a life who is age x at time 0 and consider the time interval $(0, t+h)$. By the Markov property:

$${}_{t+h} \bar{p}_x^{HH} = {}_t \bar{p}_x^{HH} \times {}_h \bar{p}_{x+t}^{HH} \quad [\frac{1}{2}]$$

We can write:

$${}_h \bar{p}_{x+t}^{HH} = 1 - {}_h p_{x+t}^{HS} - {}_h p_{x+t}^{HD} + o(h) \quad [\frac{1}{2}]$$

Note that the $o(h)$ term is needed here to account for the probability of more than one transition (*eg* from healthy to sick and back again) during the interval $(t, t+h)$.

Hence:

$${}_{t+h} \bar{p}_x^{HH} = {}_t \bar{p}_x^{HH} \left[1 - {}_h p_{x+t}^{HS} - {}_h p_{x+t}^{HD} + o(h) \right] \quad [\frac{1}{2}]$$

and, by the second assumption in (i):

$${}_{t+h} \bar{p}_x^{HH} = {}_t \bar{p}_x^{HH} \left[1 - h\sigma_{x+t} - h\mu_{x+t} \right] + o(h) \quad [\frac{1}{2}]$$

This can be rearranged to give:

$$\frac{{}_{t+h} p_x^{\overline{HH}} - {}_t p_x^{\overline{HH}}}{h} = - {}_t p_x^{\overline{HH}} (\sigma_{x+t} + \mu_{x+t}) + \frac{o(h)}{h} \quad [1]$$

Finally, letting $h \rightarrow 0$, we obtain the result:

$$\frac{\partial}{\partial t} {}_t p_x^{\overline{HH}} = - {}_t p_x^{\overline{HH}} (\sigma_{x+t} + \mu_{x+t}) \quad [1]$$

The initial condition is ${}_0 p_x^{\overline{HH}} = 1$. [1]

(iii) ***Solution to differential equation when forces are constant***

If the forces of transition are constant over the year of age, then:

$$\begin{aligned} \frac{\partial}{\partial t} {}_t p_x^{\overline{HH}} &= - {}_t p_x^{\overline{HH}} (\sigma + \mu) \\ \Rightarrow \frac{\partial}{\partial t} \ln({}_t p_x^{\overline{HH}}) &= -(\sigma + \mu) \end{aligned} \quad [1]$$

Integrating both sides:

$$\ln({}_t p_x^{\overline{HH}}) = -(\sigma + \mu)t + C \quad [1]$$

where C denotes a constant of integration. However, since ${}_0 p_x^{\overline{HH}} = 1$, it follows that $C = 0$. So we get:

$${}_t p_x^{\overline{HH}} = e^{-t(\sigma+\mu)} \quad [1]$$

(iv)(a) ***Likelihood function and maximum likelihood estimate***

The likelihood function is:

$$L = e^{-750(\sigma+\mu)} e^{-32(\rho+\nu)} \sigma^{15} \rho^{35} \mu^3 \nu^5 \quad [1]$$

Taking logs:

$$\ln L = -750(\sigma + \mu) - 32(\rho + \nu) + 15 \ln \sigma + 35 \ln \rho + 3 \ln \mu + 5 \ln \nu$$

Differentiating with respect to μ :

$$\frac{\partial}{\partial \mu} \ln L = -750 + \frac{3}{\mu}$$

Setting this equal to 0 and rearranging, we obtain:

$$\mu = \frac{3}{750} = 0.004 \quad [1]$$

To check that this is a maximum, we can differentiate the log-likelihood again and check that it is negative when $\mu = 0.004$:

$$\frac{\partial^2}{\partial \mu^2} \ln L = -\frac{3}{\mu^2} < 0 \text{ for all } \mu$$

So the MLE of μ is $\hat{\mu} = 0.004$. [1]

This force of mortality applies to the age in the middle of the rate interval, ie to age 50½. So $f = \frac{1}{2}$. (No further assumptions are required.) [1]

(iv)(b) ***Confidence interval***

The variance of $\tilde{\mu}$, the maximum likelihood estimator of μ is approximately:

$$\text{var}(\tilde{\mu}) \approx \frac{\hat{\mu}^2}{3} = \frac{0.004^2}{3} \quad [1]$$

Recall that the asymptotic variance of an MLE is equal to the Cramér-Rao lower bound, ie $\left(-\frac{\partial^2}{\partial \mu^2} \ln L \right)^{-1}$.

So an approximate 90% confidence interval for μ is:

$$0.004 \pm 1.645 \sqrt{\frac{0.004^2}{3}} = 0.004 \pm 0.0038 = (0.0002, 0.0078) \quad [1]$$

(iv)(c) ***Estimate of probability***

The required probability is $p_{50}^{\overline{HH}} \approx e^{-(\hat{\sigma}+\hat{\mu})}$. The MLE of σ is:

$$\hat{\sigma} = \frac{15}{750} = 0.02$$

So:

$$p_{50}^{\overline{HH}} \approx e^{-(0.02+0.004)} = 0.97629 \quad [2]$$

(v) ***Comment***

The student has correctly calculated the estimate ($= 35/32$). [½]

ρ is a *rate*, not a probability. So its value *can* exceed 1. [1]

A model is just a proposed mathematical description. It cannot be “wrong” as such, although it may provide a poor description of the situation. [½]

It *is* possible to have more recoveries than deaths because (a) some people may already have been sick at the start of the year and (b) we are looking at a single year of age, which won’t be the same group of people throughout. [1]

Subject CT4: Assignment X1

2013 Examinations

Time allowed: 2½ hours

Instructions to the candidate

1. *Please note that assignment marking is not included in the price of the Course Materials. Please purchase Series Marking or Marking Vouchers before submitting your script.*
2. *We only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2013 exams.*
3. *Attempt all of the questions, leaving space in the margin and beginning your answer to each question on a new page.*
4. *Write in black ink using a medium-sized nib because we will be unable to mark illegible scripts.*
5. ***Leave at least 2cm margin on all borders.***
6. *Attempt the questions as far as possible under exam conditions.*
7. *You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed in the Study Guide for the 2013 exams, on the summary page at the back of this pack and on our website at www.ActEd.co.uk.*

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. It is your responsibility to ensure that scripts reach ActEd in good time. ActEd will not be responsible for scripts lost or damaged in the post or for scripts received after the deadline date. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

At the end of the assignment

If your script is being marked by ActEd, please follow
the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables
and an electronic calculator.

Submission for marking

There are three methods for you to submit your script, namely by *email*, by *fax* or by *post*.

If you are submitting by **email**:

- complete the cover sheet, including the checklist
- scan your script (and Marking Voucher if applicable) to a pdf document, then email it to: **ActEdMarking@bpp.com**.

Please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* “CT4 Assignment X1 No. 12345”, inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF is less than 3 MB in size. **The file size cannot exceed 4 MB.**
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

If you are submitting by **fax**:

- only write on one side of the paper when completing the assignment
- complete the cover sheet, including the checklist
- fax your script (including cover sheet and Marking Voucher if applicable) to **0844 583 4501**.

In addition:

- We recommend that you stay by the fax machine until the fax has been sent so that you can deal with any problems immediately. (If an error occurs, please re-fax the whole script.)
- An email will be sent by the end of the next working day to confirm that we have processed your script. Please do not phone to check progress before then. If the fax was sent without error then it's very unlikely that there will be a problem.

We will **not** accept:

- scripts submitted to other ActEd fax numbers – please use **0844 583 4501**
- scripts that have been split over a number of faxes. (If an error occurs, please re-fax the whole script.)
- more than one script per fax
- jumbled scripts – please fax the pages in the correct order.

If you are submitting by **post**:

- complete the cover sheet, including the checklist.
- we recommend that you photocopy your script before posting, in case your script is lost in the post.
- post your script to: **First Floor, McTimoney House, 1 Kimber Road, Abingdon, Oxfordshire, OX14 1BZ**
- please staple the cover sheet (and Marking Voucher if applicable) to the front of your assignment
- please do not staple more than one assignment together.

Subject CT4: Assignment X1

2013 Examinations

Please complete the following information:

Name:

Address:

ActEd Student Number (see Note below):

--	--	--	--	--

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting this number will help us to process your scripts quickly. If you do not complete this box, your script may be delayed. If you do not know your ActEd Student Number, please email ActEd@bpp.com. **Your ActEd Student Number is not the same as your Faculty/Institute Actuarial Reference Number or ARN.**

Number of following pages: _____

Please put a tick in this box if you have solutions and a cross if you do not:

Please tick here if you are allowed extra time or other special conditions in the Profession's exams:

Time to do assignment (see Note below): _____ hrs _____ mins

Under exam conditions (delete as applicable): yes / nearly / no

Note: If you spend more than 2½ hours on the assignment, you should indicate on the assignment how much you completed within this time so that the marker can provide useful feedback on your chances of success in the exam.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
3	5	6	6	7	7	9	10	12	15	80 = _____ %

Grade: A B C D E

Marker's initials: _____

Please tick the following checklist so that your script can be marked quickly. Have you:

- [] Checked that you are using the latest version of the assignments, eg 2013 for the sessions leading to the 2013 exams?
- [] Written your full name and postal address in the box above?
- [] Completed your ActEd Student Number in the box above?
- [] Recorded your attempt conditions?
- [] Numbered all pages of your script (excluding this cover sheet)?
- [] Written the total number of pages (excluding the cover sheet) in the space above?
- [] Attached your Marking Voucher or ordered Series X Marking?

Please follow the instructions on the previous page when submitting your script for marking.

***This page has been left blank in case you wish to submit your
script by fax.***

Feedback from marker

Notes on marker's section

The marker's main objective is to give you advice on how to improve your answers. The marker will also assess your script quantitatively and qualitatively. The percentage score gives you a quantitative assessment. The grade is a qualitative assessment of how your script might be classified in the exam. The grades are as follows:

A = Clear Pass B = Probable Pass C = Borderline D = Probable Fail E = Clear Fail

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

or when you submit your next script.

***This page has been left blank in case you wish to submit your
script by fax.***

Question X1.1

In the context of a stochastic process $\{X_t : t \in J\}$, explain the meaning of the following conditions:

- (a) strictly stationary
- (b) weakly stationary. [3]

Question X1.2

Teams in an ice hockey league play matches consisting of three 20-minute periods. Each team scores goals independently in accordance with a Poisson process, both with the same rate of λ per hour. The average number of goals scored in a match by both teams combined is 5.

- (i) State the value of λ . [1]
 - (ii) Calculate the probability that, in a given 20-minute period, each team will score exactly one goal. [2]
 - (iii) Calculate the probability that more than 2 goals are scored in a match. [2]
- [Total 5]

Question X1.3

Suppose that $\{Z_n : n = 0, 1, 2, \dots\}$ is a white noise process with a discrete state space.

- (i) Explain what this means. [2]
 - (ii) Explain whether or not $\{Z_n\}$:
 - (a) has independent increments
 - (b) has the Markov property. [4]
- [Total 6]

Question X1.4

- (i) In the context of a stochastic process denoted by $\{X_t : t \in J\}$, define the terms:
- (a) state space
 - (b) time set
 - (c) sample path. [2]
- (ii) Stochastic process models can be placed in one of four categories according to whether the state space is continuous or discrete, and whether the time set is continuous or discrete. For each of the four categories:
- (a) state a stochastic process model of that type
 - (b) give an example of a problem an actuary may wish to study using a model from that category. [4]
- [Total 6]

Question X1.5

- (i) Define the term *stationary distribution* for a Markov chain with transition matrix P . [2]
- (ii) Find all such distributions for the process with $P = \begin{pmatrix} \frac{2}{9} & \frac{1}{3} & \frac{4}{9} \\ \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{9} & \frac{1}{9} & \frac{7}{9} \end{pmatrix}$. [5]
- [Total 7]

Question X1.6

- (i) List the benefits of modelling in actuarial work. [4]
- (ii) Explain the main differences between a deterministic model and a stochastic model. [3]
- [Total 7]

Question X1.7

A time-homogeneous Markov chain has four states numbered 1, 2, 3 and 4. A researcher is considering using one of the following three matrices to model the transition probabilities. In each case, the (i,j) th entry of the matrix represents the probability of a movement from State i to State j at any given time step.

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \quad R = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

The researcher is interested in the long-term behaviour of the resulting processes.

- (i) Explain what is meant by a periodic state and by an irreducible Markov chain. [2]
 - (ii) For each of the three transition matrices P , Q and R , determine the long-term behaviour of the corresponding process, calculating (where appropriate) the equilibrium probabilities corresponding to each state. [7]
- [Total 9]

Question X1.8

A total of N independent lives are observed during a finite period of observation. Between the ages of x and $x+1$, $x+a_i$ is the age at which observation of the i th life starts, and $x+t_i$ is the age at which observation of the i th life ceases. Observation may cease because the life dies, or because the life is censored at age $x+t_i$. For each life, an indicator variable D_i indicates whether life i is observed to die.

$$D_i = \begin{cases} 1 & \text{if life } i \text{ dies at } x + t_i \\ 0 & \text{otherwise} \end{cases} \quad i = 1, 2, \dots, N$$

- (i) If the force of mortality at age $x+t$, $\mu_{x+t} = \mu$ ($0 \leq t < 1$), derive the maximum likelihood estimate of μ . [6]
- (ii) State the asymptotic sampling distribution of the maximum likelihood estimator, $\tilde{\mu}$. [2]
- (iii) Given that 20 deaths are observed and the total waiting time is 780, construct an approximate 95% confidence interval for μ . [2]

[Total 10]

Question X1.9

A company assesses the credit-worthiness of various firms every quarter; the ratings are, in order of decreasing merit, A , B , C and D (default). Historical data support the view that the credit rating of a typical firm evolves as a Markov chain with transition matrix:

$$P = \begin{pmatrix} 1 - \alpha - \alpha^2 & \alpha & \alpha^2 & 0 \\ \alpha & 1 - 2\alpha - \alpha^2 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 - 2\alpha - \alpha^2 & \alpha \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

for some parameter α .

- (i) Draw the transition graph of the chain. [2]
 - (ii) Determine the range of values of α for which the matrix P is a valid transition matrix. [2]
 - (iii) State, with reasons, whether the chain is irreducible and aperiodic. [2]
 - (iv) Derive a stationary probability distribution for the chain and establish whether it is unique. [2]
 - (v) For the value $\alpha = 0.1$, calculate the probability that the company's rating in the third quarter, X_3 , is in the default state D :
 - (a) in the case where the company's rating in the first quarter, X_1 , is equal to A
 - (b) in the case $X_1 = B$
 - (c) in the case $X_1 = C$
 - (d) in the case $X_1 = D$. [4]
- [Total 12]

Question X1.10

A no-claims discount system operated by an insurer selling private medical insurance has four levels of discount:

- Level 1: 0% discount
- Level 2: 10% discount
- Level 3: 20% discount
- Level 4: 25% discount

The insurer operates an accelerated discount scheme with the following rules:

- New policyholders start on Level 1.
- Following a year with one or more claims, move to the next lower level, or remain at Level 1.
- Following a claim-free year:
 - Move up one level, or remain at Level 4, if, in the year before the most recent year, there were one or more claims or no insurance was in force
 - Move up two levels, or move to Level 4 or remain at Level 4 if, in the year before the most recent year, there were no claims.

For any policyholder the probability of a claim-free year is 0.8.

- (i) A stochastic process $X(t)$ is to be used to model the NCD system. $X(t)$ will denote the policyholder's discount Level (1, 2, 3 or 4) in year t . Explain why $X(t)$ is not a Markov chain. [2]
- (ii) Explain how the number of states that $X(t)$ can take can be increased to produce a new process $Y(t)$ that is Markov. [2]
- (iii) Draw and label the transition graph for $Y(t)$. [2]
- (iv) Write down the transition matrix of $Y(t)$. [1]
- (v) Show that the conditions sufficient for $Y(t)$ to have a unique stationary distribution that will be reached are satisfied. [3]
- (vi) Calculate the long-run probability that the policyholder is at discount Level 2. [5]
[Total 15]

Subject CT4: Assignment X2

2013 Examinations

Time allowed: 2½ hours

Instructions to the candidate

1. *Please note that we only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2013 exams.*
2. *Attempt all of the questions, leaving space in the margin and beginning your answer to each question on a new page.*
3. *Write in black ink using a medium-sized nib because we will be unable to mark illegible scripts.*
4. ***Leave at least 2cm margin on all borders.***
5. *Attempt the questions as far as possible under exam conditions.*
6. *You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed in the Study Guide for the 2013 exams, on the summary page at the back of this pack and on our website at www.ActEd.co.uk.*

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. It is your responsibility to ensure that scripts reach ActEd in good time. ActEd will not be responsible for scripts lost or damaged in the post or for scripts received after the deadline date. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

At the end of the assignment

If your script is being marked by ActEd, please follow the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables and an electronic calculator.

Submission for marking

There are three methods for you to submit your script, namely by *email*, by *fax* or by *post*.

If you are submitting by **email**:

- complete the cover sheet, including the checklist
- scan your script (and Marking Voucher if applicable) to a pdf document, then email it to: **ActEdMarking@bpp.com**.

Please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* “CT4 Assignment X2 No. 12345”, inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF is less than 3 MB in size. **The file size cannot exceed 4 MB.**
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

If you are submitting by **fax**:

- only write on one side of the paper when completing the assignment
- complete the cover sheet, including the checklist
- fax your script (including cover sheet and Marking Voucher if applicable) to **0844 583 4501**.

In addition:

- We recommend that you stay by the fax machine until the fax has been sent so that you can deal with any problems immediately. (If an error occurs, please re-fax the whole script.)
- An email will be sent by the end of the next working day to confirm that we have processed your script. Please do not phone to check progress before then. If the fax was sent without error then it's very unlikely that there will be a problem.

We will **not** accept:

- scripts submitted to other ActEd fax numbers – please use **0844 583 4501**
- scripts that have been split over a number of faxes. (If an error occurs, please re-fax the whole script.)
- more than one script per fax
- jumbled scripts – please fax the pages in the correct order.

If you are submitting by **post**:

- complete the cover sheet, including the checklist.
- we recommend that you photocopy your script before posting, in case your script is lost in the post.
- post your script to: **First Floor, McTimoney House, 1 Kimber Road, Abingdon, Oxfordshire, OX14 1BZ**
- please staple the cover sheet (and Marking Voucher if applicable) to the front of your assignment
- please do not staple more than one assignment together.

Subject CT4: Assignment X2

2013 Examinations

Please complete the following information:

Name:

Address:

ActEd Student Number (see Note below):

--	--	--	--	--

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting this number will help us to process your scripts quickly. If you do not complete this box, your script may be delayed. If you do not know your ActEd Student Number, please email ActEd@bpp.com. **Your ActEd Student Number is not the same as your Faculty/Institute Actuarial Reference Number or ARN.**

Number of following pages: _____

Please put a tick in this box if you have solutions and a cross if you do not:

Please tick here if you are allowed extra time or other special conditions in the Profession's exams:

Time to do assignment (see Note below): _____ hrs _____ mins

Under exam conditions (delete as applicable): yes / nearly / no

Note: If you spend more than 2½ hours on the assignment, you should indicate on the assignment how much you completed within this time so that the marker can provide useful feedback on your chances of success in the exam.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
6	7	8	9	9	10	13	18	80 = _____ %

Grade: A B C D E

Marker's initials: _____

Please grade your Assignment X1 marker by ticking the appropriate box.

- [] **Excellent** – the marker's comments were thorough and very helpful
- [] **Good** – the marker's comments were generally helpful
- [] **Acceptable** – please explain below how the marker could have been more helpful
- [] **Poor** – the marker's comments were generally unhelpful; please give details below

Please give any additional comments here (especially if you rate the marker less than good):

Note: Giving feedback on your marker helps us to improve the quality of marking.

Please follow the instructions on the previous page when submitting your script for marking.

***This page has been left blank in case you wish to submit your
script by fax.***

Please tick the following checklist so that your script can be marked quickly. Have you:

- [] Checked that you are using the latest version of the assignments, eg 2013 for the sessions leading to the 2013 exams?
- [] Written your full name and postal address in the appropriate box?
- [] Completed your ActEd Student Number in the appropriate box?
- [] Recorded your attempt conditions?
- [] Numbered all pages of your script (excluding this cover sheet)?
- [] Written the total number of pages (excluding the cover sheet) in the space above?
- [] Attached your Marking Voucher or ordered Series X Marking?
- [] Rated your Assignment X1 marker?

Feedback from marker

Notes on marker's section

The marker's main objective is to give you advice on how to improve your answers. The marker will also assess your script quantitatively and qualitatively. The percentage score gives you a quantitative assessment. The grade is a qualitative assessment of how your script might be classified in the exam. The grades are as follows:

A = Clear Pass B = Probable Pass C = Borderline D = Probable Fail E = Clear Fail

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

or when you submit your next script.

***This page has been left blank in case you wish to submit your
script by fax.***

Question X2.1

The time, in years, until a boiler breaks down is exponentially distributed with parameter λ , where:

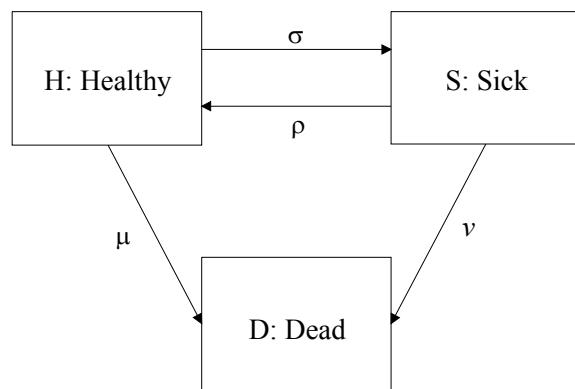
$$\lambda = \begin{cases} \frac{1}{4} & \text{if the boiler has not previously broken down} \\ \frac{1}{2} & \text{if the boiler has broken down once previously} \\ 1 & \text{if the boiler has broken down more than once previously} \end{cases}$$

Once a boiler has broken down 10 times, it is scrapped. If a boiler has broken down fewer than 10 times, it is immediately repaired.

- (i) Calculate the probability that a new boiler will break down more than once in the next 5 years. [5]
- (ii) Calculate the expected lifetime of a new boiler. [1]
[Total 6]

Question X2.2

A life insurance company prices its long-term sickness policies using the following time-homogeneous Markov model:



For a group of policyholders over a 1-year period there are:

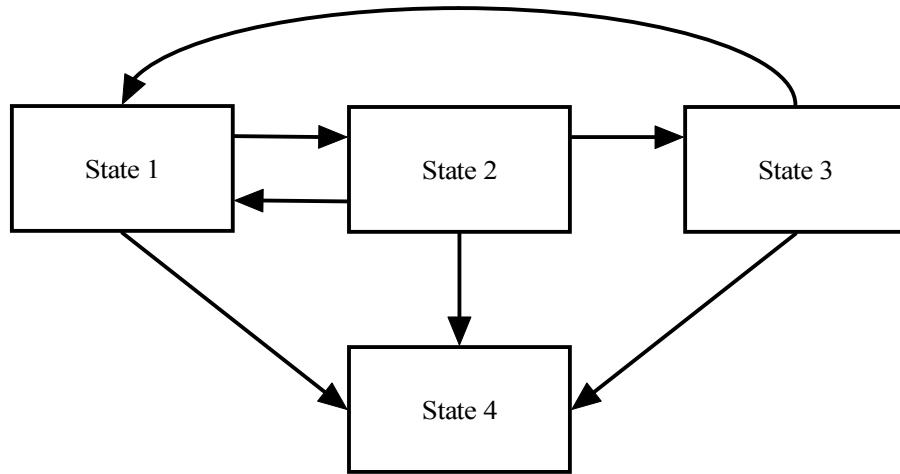
- 34 transitions from State H to State S
- 26 transitions from State S to State H
- 2 deaths from State H
- 7 deaths from State S

The total time spent in State H is 904 years and the total time spent in State S is 112 years.

- (i) Write down the likelihood function for these data. [2]
 - (ii) Show that the maximum likelihood estimate of ρ is 0.23214. (You may assume that this gives a maximum.) [2]
 - (iii) Construct an approximate 95% confidence interval for ρ . [3]
- [Total 7]

Question X2.3

Consider the following multiple-state model in which $S(t)$, the state occupied at time t by a life initially aged x , is assumed to follow a continuous-time Markov process.



Let μ_{x+t}^{ij} denote the force of transition at age $x+t$ ($t \geq 0$) from State i to State j , and let ${}_t p_x^{ij} = P(S(t) = j | S(0) = i)$.

- (i) Derive the forward differential equation:

$$\frac{\partial}{\partial t} {}_t p_x^{21} = {}_t p_x^{22} \mu_{x+t}^{21} + {}_t p_x^{23} \mu_{x+t}^{31} - {}_t p_x^{21} (\mu_{x+t}^{12} + \mu_{x+t}^{14})$$

stating all the assumptions that you make. [5]

- (ii) Write down forward differential equations for ${}_t p_x^{23}$ and ${}_t p_x^{32}$. [3]

[Total 8]

Question X2.4

A 3-state time-homogeneous Markov jump process is determined by the following matrix of transition rates:

$$A = \begin{pmatrix} -3 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

The distribution at time 0 is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Determine the distribution at time 1. [9]

Question X2.5

$\{X_t\}_{t \geq 0}$ is a Poisson process that models the cumulative number of arrivals of insurance claims. The average number of claims per hour is 0.83.

- (i) Write down the transition rates for this process. [1]
 - (ii) State the distribution of each X_t . [1]
 - (iii) Calculate $P[X_3 \leq 2]$ and $P[X_5 - X_2 < 3]$. [3]
 - (iv) Define the first holding time, T_0 , and state its distribution. [1]
 - (v) Show that $P[T_0 > s+t | T_0 > s] = P[T_0 > t]$. [2]
 - (vi) Given that the process is in state 0 after four and a half hours, calculate the probability that the remaining time until the first jump is at least two hours. [1]
- [Total 9]

Question X2.6

A particular machine is in constant use. Regardless of how long it has been since the last repair, it tends to break down once a day and on average it takes the repairman 6 hours to fix. You are modelling the machine's status as a time-homogeneous Markov jump process $\{X(t) : t \geq 0\}$ with two states: "being repaired" denoted by 0, and "working" denoted by 1. Let $P_{i,j}(t)$ denote the probability that the process is in state j at time t given that it was in state i at time 0 and suppose that t is measured in days.

- (i) State the two main assumptions that you make in applying the model and discuss briefly how you could test that each of them hold. [3]
- (ii) Draw the transition graph for the process, showing the numerical values of the transition rates. [2]
- (iii) State Kolmogorov's backward and forward differential equations for the probability $P_{0,0}(t)$. [2]
- (iv) Solve the forward differential equation in (iii) to show that:

$$P_{0,0}(t) = \frac{1}{5} + \frac{4}{5}e^{-5t} \quad [3]$$

[Total 10]

Question X2.7

- (i) Write down the Chapman-Kolmogorov equations for a continuous-time Markov process with discrete state space. [1]
- (ii) Define the transition rates for such a process. [2]
- (iii) Derive from first principles the forward version of the Kolmogorov differential equations:

$$\frac{\partial}{\partial t} P(s,t) = P(s,t) A(t)$$

where A is the matrix of transition rates. [5]

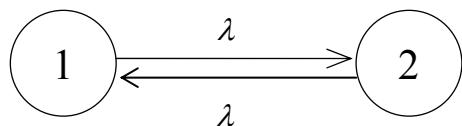
- (iv) An exponential function for any matrix B can be defined by:

$$e^B = \sum_{n=0}^{\infty} \frac{B^n}{n!}$$

where $B^0 = I$, the identity matrix.

Show that $P(s,t) = e^{(t-s)A}$ is a solution of the forward equations for a time-homogeneous process, where A is the matrix of transition rates. [2]

- (v) A time homogeneous process has two states, 1 and 2, with transition rate λ of going in either direction.



Find the transition probability matrix $P(s,t)$. [3]

[Total 13]

Question X2.8

Patients arriving at the Accident and Emergency department of a hospital (state A) wait for an average of one hour before being classified by a junior doctor as requiring in-patient treatment (I), out-patient treatment (O) or further investigation (F). Only one new arrival in ten is classified as an in-patient, five in ten as out-patients.

If needed, further investigation takes an average of 3 hours, after which 50% of cases are discharged (D), 25% are sent to receive out-patient treatment and 25% admitted as in-patients.

Out-patient treatment takes an average of 2 hours to complete, in-patient treatment an average of 60 hours. Both result in discharge.

It is suggested that a time-homogeneous Markov process with states A, F, I, O and D could be used to model the progress of patients through the system, with the ultimate aim of reducing the average time spent in the hospital.

- (i) Write down the matrix of transition rates, $\{\mu_{ij} : i, j = A, F, I, O, D\}$, of such a model. [2]
- (ii) Calculate the proportion of patients who eventually receive in-patient treatment. [1]
- (iii) Derive expressions for the probability that a patient arriving at time $t = 0$ is:
 - (a) yet to be classified by the junior doctor at time t , and
 - (b) undergoing further investigation at time t . [4]

continued ...

- (iv) Let m_i denote the expectation of the time until discharge for a patient currently in state i .

- (a) Explain in words why m_i satisfies the following equation:

$$m_i = \frac{1}{\lambda_i} + \sum_{j \notin \{i, D\}} \frac{\mu_{ij}}{\lambda_i} m_j$$

where $\lambda_i = \sum_{j \neq i} \mu_{ij}$.

- (b) Hence calculate the expectation of the total time until discharge for a newly-arrived patient. [4]

- (v) State the distribution of the time spent in each state visited according to this model. [1]

The average times listed above may be assumed to be the sample mean waiting times derived from tracking a large sample of patients through the system.

- (vi) Describe briefly what additional feature of the data might be used to check that this simple model matches the situation being modelled. [2]

- (vii) The hospital management committee believes that replacing the junior doctor with a more senior doctor will save resources by reducing the proportion of cases sent further investigation. Alternatively, the same resources could go towards reducing out-patient treatment time.

- (a) Outline briefly the calculations that would need to be performed to compare the options.

- (b) Discuss whether the current model is suitable as a basis for making decisions of this nature. [4]

[Total 18]

Subject CT4: Assignment X3

2013 Examinations

Time allowed: 3 hours

Instructions to the candidate

1. *Please note that we only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2013 exams.*
2. *Attempt all of the questions, leaving space in the margin and beginning your answer to each question on a new page.*
3. *Write in black ink using a medium-sized nib because we will be unable to mark illegible scripts.*
4. ***Leave at least 2cm margin on all borders.***
5. *Attempt the questions as far as possible under exam conditions.*
6. *You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed in the Study Guide for the 2013 exams, on the summary page at the back of this pack and on our website at www.ActEd.co.uk.*

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. It is your responsibility to ensure that scripts reach ActEd in good time. ActEd will not be responsible for scripts lost or damaged in the post or for scripts received after the deadline date. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

At the end of the assignment

If your script is being marked by ActEd, please follow the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables and an electronic calculator.

Submission for marking

There are three methods for you to submit your script, namely by *email*, by *fax* or by *post*.

If you are submitting by **email**:

- complete the cover sheet, including the checklist
- scan your script (and Marking Voucher if applicable) to a pdf document, then email it to: **ActEdMarking@bpp.com**.

Please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* “CT4 Assignment X3 No. 12345”, inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF is less than 3 MB in size. **The file size cannot exceed 4 MB.**
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

If you are submitting by **fax**:

- only write on one side of the paper when completing the assignment
- complete the cover sheet, including the checklist
- fax your script (including cover sheet and Marking Voucher if applicable) to **0844 583 4501**.

In addition:

- We recommend that you stay by the fax machine until the fax has been sent so that you can deal with any problems immediately. (If an error occurs, please re-fax the whole script.)
- An email will be sent by the end of the next working day to confirm that we have processed your script. Please do not phone to check progress before then. If the fax was sent without error then it's very unlikely that there will be a problem.

We will **not** accept:

- scripts submitted to other ActEd fax numbers – please use **0844 583 4501**
- scripts that have been split over a number of faxes. (If an error occurs, please re-fax the whole script.)
- more than one script per fax
- jumbled scripts – please fax the pages in the correct order.

If you are submitting by **post**:

- complete the cover sheet, including the checklist.
- we recommend that you photocopy your script before posting, in case your script is lost in the post.
- post your script to: **First Floor, McTimoney House, 1 Kimber Road, Abingdon, Oxfordshire, OX14 1BZ**
- please staple the cover sheet (and Marking Voucher if applicable) to the front of your assignment
- please do not staple more than one assignment together.

Subject CT4: Assignment X3

2013 Examinations

Please complete the following information:

Name:

Address:

ActEd Student Number (see Note below):

--	--	--	--	--

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting this number will help us to process your scripts quickly. If you do not complete this box, your script may be delayed. If you do not know your ActEd Student Number, please email ActEd@bpp.com. **Your ActEd Student Number is not the same as your Faculty/Institute Actuarial Reference Number or ARN.**

Number of following pages: _____

Please put a tick in this box if you have solutions and a cross if you do not:

Please tick here if you are allowed extra time or other special conditions in the Profession's exams:

Time to do assignment (see Note below): _____ hrs _____ mins

Under exam conditions (delete as applicable): yes / nearly / no

Note: If you spend more than 3 hours on the assignment, you should indicate on the assignment how much you completed within this time so that the marker can provide useful feedback on your chances of success in the exam.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total	
3	6	7	12	12	15	18	27	100	= _____ %

Grade: A B C D E

Marker's initials: _____

Please grade your Assignment X2 marker by ticking the appropriate box.

- [] **Excellent** – the marker's comments were thorough and very helpful
- [] **Good** – the marker's comments were generally helpful
- [] **Acceptable** – please explain below how the marker could have been more helpful
- [] **Poor** – the marker's comments were generally unhelpful; please give details below

Please give any additional comments here (especially if you rate the marker less than good):

Note: Giving feedback on your marker helps us to improve the quality of marking.

Please follow the instructions on the previous page when submitting your script for marking.

***This page has been left blank in case you wish to submit your
script by fax.***

Please tick the following checklist so that your script can be marked quickly. Have you:

- [] Checked that you are using the latest version of the assignments, eg 2013 for the sessions leading to the 2013 exams?
- [] Written your full name and postal address in the appropriate box?
- [] Completed your ActEd Student Number in the appropriate box?
- [] Recorded your attempt conditions?
- [] Numbered all pages of your script (excluding this cover sheet)?
- [] Written the total number of pages (excluding the cover sheet) in the space above?
- [] Attached your Marking Voucher or ordered Series X Marking?
- [] Rated your Assignment X2 marker?

Feedback from marker

Notes on marker's section

The marker's main objective is to give you advice on how to improve your answers. The marker will also assess your script quantitatively and qualitatively. The percentage score gives you a quantitative assessment. The grade is a qualitative assessment of how your script might be classified in the exam. The grades are as follows:

A = Clear Pass B = Probable Pass C = Borderline D = Probable Fail E = Clear Fail

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

or when you submit your next script.

***This page has been left blank in case you wish to submit your
script by fax.***

Question X3.1

Consider the following assertions relating to a mortality table. K_x and T_x denote the curtate and complete future lifetimes (respectively) of a particular individual aged exactly x .

- I e_{x+1} can never be greater than e_x
- II $_t p_{x+1}$ can never be greater than $_t p_x$
- III T_x can never be greater than $K_x + 1$

State with reasons whether each of the assertions is true or false. [3]

Question X3.2

A mortality table is defined such that:

$$_t p_x = \left(1 - \frac{t}{110-x}\right)^{\frac{1}{2}} \quad \text{for } x < 110, t < 110-x$$

and:

$$_t p_x = 0 \quad \text{for } t \geq 110-x$$

Calculate:

- (i) the complete expectation of life at exact age 45. [3]
 - (ii) the force of mortality at age 45. [3]
- [Total 6]

Question X3.3

- (i) Explain the differences between random censoring and Type I censoring in the context of an investigation into the mortality of life insurance policyholders. Include in your explanation a statement of the circumstances in which the censoring will be random, and the circumstances in which it will be Type I. [4]
- (ii) Explain what is meant by non-informative censoring in the investigation in (i). Describe a situation in which censoring might be informative in this investigation. [3]

[Total 7]

Question X3.4

As part of a clinical trial, a statistician is studying the survival rates of patients who have undergone a certain surgical procedure. Below is an extract from the statistician's data. Each patient was observed from their date of operation until their date of exit.

Patient number	Date of operation	Date of exit	Reason for exit
1	1 February 2008	1 January 2009	Censored
2	1 April 2008	1 October 2008	Death
3	1 April 2008	1 January 2009	Censored
4	1 July 2007	1 July 2008	Censored
5	1 August 2008	1 January 2009	Censored
6	1 November 2007	1 January 2008	Death
7	1 January 2008	1 January 2009	Censored
8	1 March 2008	1 November 2008	Death
9	1 May 2008	1 November 2008	Death
10	1 June 2008	1 January 2009	Censored

You can assume that the censoring was non-informative with regard to the survival of any individual patient.

- (i) Calculate the Nelson-Aalen estimate of the cumulative hazard function, $\Lambda(t)$, where t is the time in months since the operation. [4]
 - (ii) Hence calculate an estimate of the survival function for patients who have had this operation. [2]
 - (iii) Construct an approximate 95% confidence interval for the probability that a patient survives for at least 10 months after having the operation. [4]
 - (iv) Comment on the statement that at least 90% of patients survive for 10 months or more after having the operation. [2]
- [Total 12]

Question X3.5

An investigation has been carried out into the survival rates of patients who have just undergone a certain medical procedure at one of two major hospitals. The data recorded for each patient were sex, drug treatment received and hospital attended. A Cox proportional hazards model was fitted to the data, and the results are given below.

Covariate	Parameter	Standard error
-----------	-----------	----------------

Sex:

Male	0	
Female	-0.20	0.11

Drug treatment received:

Treatment A	0	
Treatment B	0.12	0.05
Treatment C	-0.05	0.03

Hospital attended:

Hospital A	0	
Hospital B	-0.06	0.04

- (i) Write down a formula for the force of mortality according to this model. You should define all the terms that you use. [3]
- (ii) Explain why this model is:
 - (a) a semi-parametric model
 - (b) a proportional hazards model. [2]
- (iii) In the context of this model state the group of lives:
 - (a) to which the baseline hazard refers
 - (b) with the lowest force of mortality. [2]
- (iv) Explain whether attending Hospital B rather than Hospital A significantly improves the chances of survival. [3]

- (v) Calculate the proportion, according to the fitted model, by which the force of mortality for a male patient on Treatment B who attended Hospital A exceeds that for a female patient on Treatment C who attended Hospital B. [2]
[Total 12]

Question X3.6

- (i) Explain the meaning of the rates of mortality usually denoted q_x and m_x , and the relationship between them. [3]
- (ii) Show that m_x can be expressed as a weighted average of the force of mortality at each age between x and $x+1$, and state what the weights are. [2]
- (iii) In a certain population, $q_x = 0.4$. Calculate the value of m_x assuming:
- that deaths are uniformly distributed between the ages of x and $x+1$
 - a constant force of mortality between the ages of x and $x+1$
 - that the Balducci assumption holds between the ages of x and $x+1$. [7]
- (iv) Comment on your results in part (iii) by considering the force of mortality over the year of age x to $x+1$ in each case. [3]
- [Total 15]

Question X3.7

A study is being conducted, using the Cox regression model, into how certain factors influence a patient's future lifetime after they have had a serious heart attack. Initially, the study has looked at the impact of a patient's smoking habits. The survival times and smoking status for 6 patients are shown in the table below. Patients have been labelled as "censored" if they were still alive at the end of the investigation or if their death was not considered to be attributable to the heart attack.

<i>Patient number</i>	<i>Time to death (weeks)</i>	<i>Smoker (yes/no)</i>	<i>Censored (yes/no)</i>
1	3	Yes	No
2	n/a (still alive)	No	Yes
3	9	No	No
4	10	Yes	Yes
5	8	No	Yes
6	7	No	No

- (i) Explain how you would use the Cox model in this study, assuming that smoking status was the only covariate. Your answer should include a description of the relevant Cox model and you should define any notation that you use. [4]
- (ii) Using the convention that $Z = 1$ for smokers and $Z = 0$ for non-smokers, write down the partial likelihood for these data. Simplify your expression as far as possible. [4]
- (iii) Show that the maximum partial likelihood estimate of the model parameter is 0.21194. [4]
- (iv) You are now given the following additional data:

<i>Patient number</i>	<i>Time to death (weeks)</i>	<i>Smoker (yes/no)</i>	<i>Censored (yes/no)</i>
7	3	Yes	No
8	n/a (still alive)	No	Yes

Write down the partial likelihood using the data from all 8 patients. [3]

- (v) The model is now to be extended so that it also allows for a patient's sex. You are told that lives 1, 4, 5, 6 and 8 are males. Describe how the Cox model could be adapted so that it now allows for gender and state the class of lives to which the baseline hazard refers. [3]

[Total 18]

Question X3.8

A scientist is investigating the average lifetime of a certain type of battery using the Kaplan-Meier method. He has placed ten batteries into similar torches and has observed the number of hours until each battery failed. The experiment was stopped at the end of 16 hours.

- (i) (a) Calculate an estimate of the discrete hazard function for this type of battery, given the data below:

Battery number	Lifetime (hours)
1	14
2	8
3	12
4	Battery did not fail, but bulb failed after 11 hours
5	10
6	Did not fail but spontaneously combusted after 13 hours
7	9
8	7
9	14
10	Did not fail by end of the experiment

- (b) Sketch the estimate of the discrete hazard in (a). [7]

- (ii) Calculate the Kaplan-Meier estimate $\hat{F}(t)$ of the distribution function $F(t)$. [3]

- (iii) Estimate the probability that a new battery dies before the end of eleven hours.

[1]

- (iv) Calculate the variance of $\hat{F}(t)$, the Kaplan-Meier estimator of $F(t)$, at each time t when a failure occurs. [5]

One manufacturer has launched a “long-life battery” which they claim lasts three times as long as the “standard life” battery mentioned above and costs twice as much. As a result the scientist has set up a similar experiment for ten “long-life” batteries.

- (v) Based on the results below, recalculate the statistics and the variance for the long-life battery.

Battery number	Lifetime (hours)
1	14
2	15
3	11
4	Did not fail by end of the experiment
5	Did not fail but spontaneously combusted after 13 hours
6	Did not fail by end of the experiment
7	7
8	Battery did not fail but bulb failed after 12 hours
9	14
10	Did not fail by end of the experiment

[6]

- (vi) Comment on the assertion that the long-life battery lasts three times longer than the standard battery. [3]

- (vii) A second scientist has also tested the difference between the batteries, but he used the Nelson-Aalen estimate, not the Kaplan-Meier estimate. Describe briefly the differences between the two approaches. [2]

[Total 27]

Subject CT4: Assignment X4

2013 Examinations

Time allowed: 3 hours

Instructions to the candidate

1. *Please note that we only accept the current version of assignments for marking, ie you can only submit this assignment in the sessions leading to the 2013 exams.*
2. *Attempt all of the questions, leaving space in the margin and beginning your answer to each question on a new page.*
3. *Write in black ink using a medium-sized nib because we will be unable to mark illegible scripts.*
4. ***Leave at least 2cm margin on all borders.***
5. *Attempt the questions as far as possible under exam conditions.*
6. *You should aim to submit this script for marking by the recommended submission date. The recommended and deadline dates for submission of this assignment are listed in the Study Guide for the 2013 exams, on the summary page at the back of this pack and on our website at www.ActEd.co.uk.*

Scripts received after the deadline date will not be marked, unless you are using a Marking Voucher. It is your responsibility to ensure that scripts reach ActEd in good time. ActEd will not be responsible for scripts lost or damaged in the post or for scripts received after the deadline date. If you are using Marking Vouchers, then please make sure that your script reaches us by the Marking Voucher deadline date to give us enough time to mark and return the script before the exam.

At the end of the assignment

If your script is being marked by ActEd, please follow the instructions on the reverse of this page.

In addition to this paper, you should have available actuarial tables and an electronic calculator.

Submission for marking

There are three methods for you to submit your script, namely by *email*, by *fax* or by *post*.

If you are submitting by **email**:

- complete the cover sheet, including the checklist
- scan your script (and Marking Voucher if applicable) to a pdf document, then email it to: **ActEdMarking@bpp.com**.

Please note the following:

- Please title the email to ensure that the subject and assignment are clear *eg* “CT4 Assignment X4 No. 12345”, inserting your ActEd Student Number for 12345.
- The assignment should be scanned the **right way up** (so that it can be read normally without rotation) and as a single document. We cannot accept individual files for each page.
- Please set the resolution so that the script is legible and the resulting PDF is less than 3 MB in size. **The file size cannot exceed 4 MB.**
- Before emailing to ActEd, please check that your scanned assignment includes all pages and conforms to the above.

If you are submitting by **fax**:

- only write on one side of the paper when completing the assignment
- complete the cover sheet, including the checklist
- fax your script (including cover sheet and Marking Voucher if applicable) to **0844 583 4501**.

In addition:

- We recommend that you stay by the fax machine until the fax has been sent so that you can deal with any problems immediately. (If an error occurs, please re-fax the whole script.)
- An email will be sent by the end of the next working day to confirm that we have processed your script. Please do not phone to check progress before then. If the fax was sent without error then it's very unlikely that there will be a problem.

We will **not** accept:

- scripts submitted to other ActEd fax numbers – please use **0844 583 4501**
- scripts that have been split over a number of faxes. (If an error occurs, please re-fax the whole script.)
- more than one script per fax
- jumbled scripts – please fax the pages in the correct order.

If you are submitting by **post**:

- complete the cover sheet, including the checklist.
- we recommend that you photocopy your script before posting, in case your script is lost in the post.
- post your script to: **First Floor, McTimoney House, 1 Kimber Road, Abingdon, Oxfordshire, OX14 1BZ**
- please staple the cover sheet (and Marking Voucher if applicable) to the front of your assignment
- please do not staple more than one assignment together.

Subject CT4: Assignment X4

2013 Examinations

Please complete the following information:

Name:

Address:

ActEd Student Number (see Note below):

--	--	--	--	--

Note: Your ActEd Student Number is printed on all personal correspondence from ActEd. Quoting this number will help us to process your scripts quickly. If you do not complete this box, your script may be delayed. If you do not know your ActEd Student Number, please email ActEd@bpp.com. **Your ActEd Student Number is not the same as your Faculty/Institute Actuarial Reference Number or ARN.**

Number of following pages: _____

Please put a tick in this box if you have solutions and a cross if you do not:

Please tick here if you are allowed extra time or other special conditions in the Profession's exams:

Time to do assignment (see Note below): _____ hrs _____ mins

Under exam conditions (delete as applicable): yes / nearly / no

Note: If you spend more than 3 hours on the assignment, you should indicate on the assignment how much you completed within this time so that the marker can provide useful feedback on your chances of success in the exam.

Score and grade for this assignment (to be completed by marker):

Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Total
4	5	5	5	6	8	8	8	9	10	16	16	100 = _____ %

Grade: A B C D E

Marker's initials: _____

Please grade your Assignment X3 marker by ticking the appropriate box.

- [] **Excellent** – the marker's comments were thorough and very helpful
- [] **Good** – the marker's comments were generally helpful
- [] **Acceptable** – please explain below how the marker could have been more helpful
- [] **Poor** – the marker's comments were generally unhelpful; please give details below

Please give any additional comments here (especially if you rate the marker less than good):

Note: Giving feedback on your marker helps us to improve the quality of marking.

Please follow the instructions on the previous page when submitting your script for marking.

***This page has been left blank in case you wish to submit your
script by fax.***

Please tick the following checklist so that your script can be marked quickly. Have you:

- [] Checked that you are using the latest version of the assignments, eg 2013 for the sessions leading to the 2013 exams?
- [] Written your full name and postal address in the appropriate box?
- [] Completed your ActEd Student Number in the appropriate box?
- [] Recorded your attempt conditions?
- [] Numbered all pages of your script (excluding this cover sheet)?
- [] Written the total number of pages (excluding the cover sheet) in the space above?
- [] Attached your Marking Voucher or ordered Series X Marking?
- [] Rated your Assignment X3 marker?

Feedback from marker

Notes on marker's section

The marker's main objective is to give you advice on how to improve your answers. The marker will also assess your script quantitatively and qualitatively. The percentage score gives you a quantitative assessment. The grade is a qualitative assessment of how your script might be classified in the exam. The grades are as follows:

A = Clear Pass B = Probable Pass C = Borderline D = Probable Fail E = Clear Fail

Please note that you can provide feedback on the marking of this assignment at:

www.ActEd.co.uk/marking

***This page has been left blank in case you wish to submit your
script by fax.***

Question X4.1

Describe the circumstances under which it would be appropriate to graduate the rates in a mortality investigation using a mathematical function. [4]

Question X4.2

For each of the following observations about the individual standardised deviations in a graduation test based on 100 ages, calculate the probability value for the test, and hence state which result would indicate most strongly that the rates were overgraduated:

- (i) 2 deviations outside the range -3 to $+3$ [1]
 - (ii) 60 positive and 40 negative deviations [1]
 - (iii) a value of the serial correlation coefficient of $r_1 = 0.18$ [1]
 - (iv) 55 absolute deviations exceeding $\frac{2}{3}$. [2]
- [Total 5]

Question X4.3

It has been suggested that a satisfactory graduation could be achieved by drawing a smooth curve that lies in its entirety within the area bounded by the equations:

$$y = q_x \pm \frac{2\sqrt{\theta_x}}{E_x}$$

where θ_x is the number of deaths observed between exact ages x and $x+1$, E_x is the exposed to risk between those ages and q_x is $\frac{\theta_x}{E_x}$.

Explain the reasoning behind this suggestion and state, with reasons, whether or not you agree with it. [5]

Question X4.4

The following data were collected in a mortality investigation covering the period 1 January 2006 to 1 January 2009 in respect of three of the lives concerned.

	<i>Date of birth</i>	<i>Date of joining</i>	<i>Date of exit</i>	<i>Reason for exit</i>
X	1 July 1956	1 January 2000	–	–
Y	1 October 1956	1 January 2007	1 July 2007	Death
Z	1 September 1958	1 October 2008	–	–

Calculate the contribution of the above three lives to the central exposed to risk and the initial exposed to risk at age 50 last birthday. [5]

Question X4.5

In an investigation of mortality during the period 1 January 2008 to 1 July 2009, information is available about the number of lives under observation aged x next birthday on 1 January 2008, 1 January 2009 and 1 July 2009. Information is also available about the number of deaths during the period, classified by age last birthday.

- (i) Derive a formula for the central exposed to risk that corresponds to the death data, stating any assumptions that you make. [5]
- (ii) The force of mortality for deaths with age label x in this investigation estimates μ_{x+f} . Determine the value of f . [1]

[Total 6]

Question X4.6

A graduation of a set of assured lives mortality data has been carried out and you are given the following results:

(1) <i>Age, x</i>	(2) <i>Initial exposed</i>	(3) <i>Actual deaths,</i>	(4) <i>Graduated mortality</i>	(5) <i>Expected deaths,</i>	(6) <i>Standard deviation</i>	(7) <i>Standardised Deviation</i> $\frac{[(3) - (5)]}{(6)}$
	<i>to risk, E_x</i>	θ_x	<i>rate, \hat{q}_x</i>	$E_x \hat{q}_x$	$\sqrt{E_x \hat{q}_x (1 - \hat{q}_x)}$	
50	35,000	161	0.004810	168.3	12.94	-0.5641
51	40,000	205	0.005340	213.6	14.58	-0.5898
52	45,000	260	0.005928	266.8	16.28	-0.4177
53	50,000	344	0.006579	329.0	18.08	0.8296
54	55,000	418	0.007300	401.5	19.96	0.8267
55	60,000	506	0.008097	485.8	21.95	0.9203
56	50,000	463	0.008977	448.8	21.09	0.6733
57	40,000	388	0.009948	397.9	19.85	-0.4987
58	30,000	318	0.011017	330.5	18.08	-0.6914
59	25,000	296	0.012194	304.8	17.35	-0.5072
60	20,000	251	0.013489	269.8	16.31	-1.1527
Total	450,000	3,610		3,616.8		

Carry out a serial correlation test (at lag 1) on these data, and state your conclusions. [8]

Question X4.7

- (i) Explain the terms “undergraduation” and “overgraduation”. [3]
 - (ii) List the possible dangers to a life company of using undergraduated or overgraduated mortality rates. [5]
- [Total 8]

Question X4.8

Describe in detail the exact calculation of the exposed to risk, as it would be used in a mortality investigation. Comment on the advantages and disadvantages of using this exact exposure method, and mention any conventions that you need to adopt. Your answer should include a description of the data required. [8]

Question X4.9

In a mortality investigation, we observe N lives between the ages of x and $x+1$. Let E_x^c denote the total waiting time for these lives, and assume that the force of mortality is constant over the age range x to $x+1$.

- (i) Write down the likelihood function for the force of mortality under the Poisson model. [1]
 - (ii) Explain why the Poisson model is not an exact model for the number of deaths, but why it is often a good approximation. [2]
 - (iii) Derive the maximum likelihood estimate of the force of mortality under the Poisson model. [3]
 - (iv) Over the period of investigation, 52 deaths between the ages of 60 and 61 were observed and the corresponding total waiting time was 8,460 years. Construct an approximate 95% confidence interval for the force of mortality over this age range. [3]
- [Total 9]

Question X4.10

From the following data, calculate estimates of the initial mortality rate (q) and the force of mortality (μ) for those lives aged 63, 64 and 65 last birthday, indicating clearly the ages to which your estimates relate.

Age last birthday	In force on 2 July				Deaths in			
	2003	2004	2005	2006	2003	2004	2005	2006
63	4,192	4,444	4,885	4,889	104	100	117	109
64	3,998	4,200	4,664	4,334	122	114	130	124
65	3,940	4,166	4,321	4,533	118	120	129	140

[10]

Question X4.11

A large life office is investigating the recent mortality experience of its term assurance policyholders. It has been decided to graduate the data by reference to a standard table using the formula:

$$\frac{q_x}{q_x^s} = ax + b$$

where q_x^s is the rate for the standard table.

- (i) Outline the considerations that you would take into account in choosing an appropriate standard table. [5]
 - (ii) Explain how you would check whether the above formula was suitable. [3]
 - (iii) Describe briefly how you would estimate a and b in the formula using:
 - (a) a weighted least squares method
 - (b) a maximum likelihood method. [8]
- [Total 16]

Question X4.12

The following information relates to an investigation of the mortality experience of a large group of lives.

<i>Age</i> <i>x</i>	<i>Exposed</i> <i>to risk</i>	<i>Observed</i> <i>deaths</i>	<i>AM92</i> <i>Ultimate</i>	<i>Expected</i> <i>deaths</i>	<i>Deviation</i> $\theta_x - E_x q_x$
	E_x	θ_x	q_x	$E_x q_x$	
30	7,251	3	0.000590	4.28	-1.28
31	7,344	3	0.000602	4.42	-1.42
32	7,590	2	0.000617	4.68	-2.68
33	7,830	4	0.000636	4.98	-0.98
34	7,998	8	0.000660	5.28	2.72
35	8,280	11	0.000689	5.70	5.30
36	8,370	10	0.000724	6.06	3.94
37	8,805	6	0.000765	6.74	-0.74
38	8,856	3	0.000813	7.20	-4.20
39	8,811	7	0.000870	7.67	-0.67
40	8,730	14	0.000937	8.18	5.82
41	9,024	5	0.001014	9.15	-4.15
42	9,615	2	0.001104	10.61	-8.61
43	9,510	10	0.001208	11.49	-1.49
44	9,483	17	0.001327	12.58	4.42
45	9,855	12	0.001465	14.44	-2.44
Total	137,352	117		123.46	-6.46

It has been suggested that the standard table AM92 Ultimate could be used as a model of the mortality of this population.

- (i) State the null hypothesis and carry out the following tests of this suggestion:
 - (a) chi-squared test
 - (b) signs test
 - (c) grouping of signs test
 - (d) standardised deviations test. [14]

- (ii) Outline the conclusions that can be drawn as a result of your investigations. [2]
[Total 16]

For the session leading to the April 2013 exams – CT Subjects

Marking vouchers

Subjects	Assignments	Mocks
CT1, CT2, CT5, CT8	20 March 2013	26 March 2013
CT3, CT4, CT6, CT7	26 March 2013	3 April 2013

Series X Assignments

Subjects	Assignment	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	X1	14 November 2012	16 January 2013
CT3, CT4, CT6, CT7		21 November 2012	23 January 2013
CT1, CT2, CT5, CT8	X2	28 November 2012	6 February 2013
CT3, CT4, CT6, CT7		5 December 2012	13 February 2013
CT1, CT2, CT5, CT8	X3	23 January 2013	27 February 2013
CT3, CT4, CT6, CT7		30 January 2013	6 March 2013
CT1, CT2, CT5, CT8	X4	13 February 2013	13 March 2013
CT3, CT4, CT6, CT7		20 February 2013	20 March 2013

Mock Exams

Subjects	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	20 March 2013	26 March 2013
CT3, CT4, CT6, CT7	26 March 2013	3 April 2013

We encourage you to work to the recommended submission dates where possible. Please remember that the turnaround of your script is likely to be quicker if you submit it well before the final deadline date.

For the session leading to the September/October 2013 exams – CT Subjects**Marking vouchers**

Subjects	Assignments	Mocks
CT1, CT2, CT5, CT8	28 August 2013	4 September 2013
CT3, CT4, C6, CT7	4 September 2013	11 September 2013

Series X Assignments

Subjects	Assignment	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	X1	12 June 2013	3 July 2013
CT3, CT4, C6, CT7		19 June 2013	10 July 2013
CT1, CT2, CT5, CT8	X2	3 July 2013	24 July 2013
CT3, CT4, C6, CT7		10 July 2013	31 July 2013
CT1, CT2, CT5, CT8	X3	24 July 2013	7 August 2013
CT3, CT4, C6, CT7		31 July 2013	14 August 2013
CT1, CT2, CT5, CT8	X4	7 August 2013	21 August 2013
CT3, CT4, C6, CT7		14 August 2013	28 August 2013

Mock Exams

Subjects	Recommended submission date	Final deadline date
CT1, CT2, CT5, CT8	21 August 2013	4 September 2013
CT3, CT4, C6, CT7	28 August 2013	11 September 2013

We encourage you to work to the recommended submission dates where possible. Please remember that the turnaround of your script is likely to be quicker if you submit it well before the final deadline date.