

1B Markov Chains

Learn the theorems. Do the 2 example sheets. Weber's notes and cross-reference Norris' book.

Definition 1.2 We say that $(X_n)_{n \geq 0}$ is a **Markov chain** with initial distribution λ and transition matrix P if for all $n \geq 0$ and $i_0, \dots, i_{n+1} \in I$,

1. $P(X_0 = i_0) = \lambda_{i_0}$
2. $P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n i_{n+1}}$

Theorem 1.3 $(X_n)_{n \geq 0}$ is *Markov*(λ, P) iff for all $n \geq 0$ and $i_0, \dots, i_n \in I$,

$$P(X_0 = i_0, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} \quad (1.1)$$

Example Sheet 1 Michaelmas 2020 (Bauerschmidt)

Q1 Let $X = (X_n)_{n \geq 0}$ be a Markov chain. Show that, conditioned on $X_m = i$, $Z = (Z_n)_{n \geq 0}$ given by $Z_n = X_{n+m}$ is a Markov chain with starting state i .

For each $j \in I$,

$$P(Z_0 = j) = P(X_m = j) = \delta_{ij} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{else} \end{cases}$$

For each $n \geq 0$ and $i_0, \dots, i_{n+1} \in I$,

$$\begin{aligned} & P(Z_{n+1} = i_{n+1} | Z_0 = i_0, \dots, Z_n = i_n) \\ &= P(X_{m+n+1} = i_{n+1} | X_m = i_0, \dots, X_{m+n} = i_n) \\ &= \frac{P(X_m = i_0, \dots, X_{m+n} = i_n, X_{m+n+1} = i_{n+1})}{P(X_m = i_0, \dots, X_{m+n} = i_n)} \\ &= \frac{\sum_{j_0, \dots, j_{m-1}} P(X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n+1} = i_{n+1})}{\sum_{k_0, \dots, k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\ &= \frac{\sum_{j_0, \dots, j_{m-1}} P(X_{m+n+1} = i_{n+1} | X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n) P(X_0 = j_0, \dots, X_{m-1} = j_{m-1})}{\sum_{k_0, \dots, k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\ &= \frac{\sum_{j_0, \dots, j_{m-1}} P(X_{m+n+1} = i_{n+1} | X_{m+n} = i_n) P(X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)}{\sum_{k_0, \dots, k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\ &= P(X_{m+n+1} = i_{n+1} | X_{m+n} = i_n) \frac{\sum_{j_0, \dots, j_{m-1}} P(X_0 = j_0, \dots, X_{m-1} = j_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)}{\sum_{k_0, \dots, k_{m-1}} P(X_0 = k_0, \dots, X_{m-1} = k_{m-1}, X_m = i_0, \dots, X_{m+n} = i_n)} \\ &= P(X_{m+n+1} = i_{n+1} | X_{m+n} = i_n) = p_{i_n i_{n+1}} = P(Z_{n+1} = i_{n+1} | Z_n = i_n) \end{aligned}$$

Q2 Let $X = (X_n)_{n \geq 0}$ be a sequence of independent random variables. Show that X is a Markov chain. Under what condition is this chain homogeneous?

Since X_0, X_1, \dots, X_n are independent, for all $i_0, \dots, i_{n+1} \in I$,

$$\begin{aligned}
 P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) &= \frac{P(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i_{n+1})}{P(X_0 = i_0, \dots, X_n = i_n)} \\
 &= \frac{P(X_0 = i_0) \dots P(X_n = i_n) P(X_{n+1} = i_{n+1})}{P(X_0 = i_0) \dots P(X_n = i_n)} \\
 &= \frac{P(X_n = i_n) P(X_{n+1} = i_{n+1})}{P(X_n = i_n)} \\
 &= \frac{P(X_n = i_n, X_{n+1} = i_{n+1})}{P(X_n = i_n)} \\
 &= P(X_{n+1} = i_{n+1} | X_n = i_n) = P(X_{n+1} = i_{n+1})
 \end{aligned}$$

so X is a Markov chain. It is homogeneous iff the X_i are identically distributed.

Q3 Let $X = (X_n)_{n \geq 0}$ be a sequence of fair coin tosses (with the two possible outcomes interpreted as 0 and 1) and set $M_n = \max_{k \leq n} X_k$. Show that $(M_n)_{n \geq 0}$ is a Markov chain and find the transition probabilities.

$$P(M_{n+1} = 0 | M_0 = i_0, \dots, M_n = i_n) = \begin{cases} 1/2 & \text{if } i_0 = \dots = i_n = 0 \\ 0 & \text{else} \end{cases}$$

and

$$P(M_{n+1} = 0 | M_n = i_n) = \begin{cases} 1/2 & \text{if } i_n = 0 \\ 0 & \text{else} \end{cases}$$

but $i_0 = \dots = i_n = 0$ iff $i_n = 0$ so the two (conditional) probabilities are the same.

$$P(M_{n+1} = 1 | M_0 = i_0, \dots, M_n = i_n) = \begin{cases} 1/2 & \text{if } i_0 = \dots = i_n = 0 \\ 1 & \text{else} \end{cases}$$

and

$$P(M_{n+1} = 1 | M_n = i_n) = \begin{cases} 1/2 & \text{if } i_n = 0 \\ 1 & \text{else} \end{cases}$$

but $i_0 = \dots = i_n = 0$ iff $i_n = 0$ so the two (conditional) probabilities are the same.

In particular, $p_{01} = p_{00} = 1/2$ and $p_{11} = 1$ and $p_{10} = 0$.

Q4

1B Statistics

Do the 3 example sheets. Weber's notes.

Statistics is a collection of procedures and principles for gaining and processing information in order to make decisions when faced with uncertainty. Two of its principal concerns are **parameter estimation** and **hypothesis testing**.

Data refers to a collection of numbers or other pieces of information to which meaning has been attached. The numbers 1.1, 3.0, 6.5 are not necessarily data. They become so when we are told that they are the muscle weight gains in kg of three athletes who have been trying a new diet.

In statistics, our data are modelled by a vector of random variables

$$X = (X_1, X_2, \dots, X_n)$$

where X_i takes values in \mathbb{Z} or \mathbb{R} .

A **statistic** $T(x)$ is any function of the data. An **estimator** of a parameter θ is a function $T = T(X)$ which we use to estimate θ from an observation of X . T is said to be **unbiased** if

$$\mathbb{E}(T) = \theta$$

The expectation above is taken over X . Once the actual data x is observed, $t = T(x)$ is the **estimate** of θ obtained via the estimator T .

Suppose that the random variable X has probability density function $f(x|\theta)$. Given the observed value x of X , the **likelihood** of θ is defined by

$$lik(\theta) = f(x|\theta)$$

Thus we are considering the density as a function of θ , for a fixed x . In the case of multiple observations, i.e. when $x = (x_1, \dots, x_n)$ is a vector of observed values of X_1, \dots, X_n , we assume unless otherwise stated that X_1, \dots, X_n are IID; in this case $f(x_1, \dots, x_n|\theta)$ is the product of the marginals,

$$lik(\theta) = f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

It makes intuitive sense to estimate θ by whatever value gives the greatest likelihood to the observed data. Thus the **maximum likelihood estimate** $\hat{\theta}(x)$ of θ is defined as the value of θ that maximises the likelihood. Then $\hat{\theta}(X)$ is called the **maximum likelihood estimator (MLE)** of θ .

Of course the maximum likelihood estimator need not exist, but in many examples it does. In practice, we usually find the MLE by maximising $\log f(x|\theta)$, which is known as the **loglikelihood**.

1B Optimisation

Learn the simplex algorithm. Do the 2 example sheets. Weber's notes.

Part 2 Probability & Measure

The relationships between various modes of convergence: almost sure, in probability, in L^p . Strong law of large numbers and central limit theorem.

Part 2 Applied Probability

Review statements and proofs of the basic theorems on queues and the starting point of the story for Poisson processes. Bauerschmidt's notes.

Part 3 Advanced Probability

Would really like to get hold of an interesting almost current concrete research problem to give direction to the theory. Martingale convergence theorems and basic secret examples of stochastic integrals and Feynman-Kac.

Part 3 Stochastic Calculus

Would really like to get hold of an interesting almost current concrete research problem to give direction to the theory.

Number Theory

Greg is taking this in the spring semester though I probably can't fit it into my schedule and the registration is already full last I checked anyways. My main interest in this is to find motivation for commutative algebra from a source other than algebraic geometry, and also to see what the algebraic number theorists have to say about solving diophantine equations, which seems to be a basic and interesting problem, as introduced by Shafarevich.

Not sure what books or notes to start with. Could look into the Cambridge notes and the standard textbook for Greg's course.

Algebraic Geometry

Want to understand (moduli spaces of) J-holomorphic curves in symplectic geometry. This was invented by Gromov and apparently inspired by pre-existing theory of holomorphic curves in algebraic geometry. Sounds to me it is basically about counting how many lines or curves of a particular degree there are in an algebraic variety. Like "27 lines on a cubic hypersurface" kind of thing, which is of course the classical starting point of the mirror symmetry business.

So I should learn about (moduli spaces of) algebraic curves. And probably Riemann surfaces and compare and contrast. "Curves and their Jacobians" seems to be a thing. Riemann surfaces version is in the Rick Miranda book.

Prerequisite for reading about curves and their Jacobians is to review and strengthen general basic algebraic geometry e.g. from Shafarevich, Kempf, Mumford. Would like to also take the opportunity to learn about sheaves, sheaf cohomology, and schemes. Mumford's Red Book is probably the desired "endpoint", though I am not sure how it compares vs Hartshorne. Shafarevich is very nice and concrete but does not talk about sheaves or schemes. Kempf would be a useful stepping stone.

So probably review Shafarevich for concrete elementary examples, read Kempf for the sheaves, read Mumford for the schemes, then read about moduli spaces of curves and Jacobians of curves. I suddenly recall something about Hilbert polynomial or Hilbert scheme which I should also look into. Of course on the Riemann surfaces side there is also the thing about period integrals and Hodge theory that I have wanted to look into for the longest time ever.

Symplectic Geometry

Want to understand framework and applications and computations of Morse theory and Floer homology, in particular J-holomorphic curves for the Floer homology and also more generally in symplectic geometry other than Floer homology. As always, Audin & Damian and the original papers of Floer, but also Abouzaid for the symplectic homology and cohomology and isomorphisms with string or loop homology for the non-compact manifolds.

Other basic concepts: Liouville manifolds, contact manifolds, Legendrian submanifolds, conical ends, Lagrangian torus fibrations over an affine base, almost complex structures which of course reminds one of Kähler manifolds and Calabi-Yau manifolds, the canonical and anticanonical bundles.

3-Manifolds

The basic state of knowledge of 3-manifolds. Required for writing the paper with Nathan when the time comes again. Classification of curves on surfaces and the proof that the algorithm works via flip moves or arcslices.

Differential Geometry

Origins of the subject leading to theories of Riemannian metrics, curvature, and connections. Spivak of course, but also PMH Wilson, which might remind me of some questions regarding Lie groups and hence topological and algebraic groups.

Algebraic Topology

Fundamental class, characteristic classes, computations of homology and cohomology using simplicial and other theories, isomorphisms between these theories with a view towards computations. Spin manifolds.

General brainteasers and problem solving exercises

Literally to prepare for interviews and online assessments. Would like to solve and create my own collection. Start from quant interview guide books and 1A Numbers & Sets but also look into olympiad problems and Putnam and other competitions.

Day 1, 27th December 2024, Friday

Q1

If $p \in Z(I(X))$ then for every polynomial $f \in k[x_1, \dots, x_n]$ such that $f = 0$ on X , have $p \in Z(f)$. So for every Zariski closed subset of \mathbb{A}^n containing X , say defined by (the vanishing of) polynomials f_1, \dots, f_m , have $p \in Z(f_1, \dots, f_m)$, i.e. p is in every Zariski closed subset of \mathbb{A}^n containing X .

Conversely, suppose $p \in \mathbb{A}^n$ is contained in every Zariski closed subset of \mathbb{A}^n containing X . Then for any $f \in I(X)$, since $f = 0$ on X , have that X is contained in the Zariski closed subset $Z(f)$. Hence $p \in Z(f)$. This shows that $f(p) = 0$ for every $f \in I(X)$ so $p \in Z(I(X))$.

Q2

Let $X \subseteq \mathbb{A}^n$ be an irreducible affine variety and $U \subseteq X$ a non-empty open set. Let Y be a closed subset of \mathbb{A}^n containing U , say $Y = Z(f_1, \dots, f_m)$ for some polynomials f_1, \dots, f_m . Suppose there is some $p \in X$ that is not contained in Y , so $f_i(p) \neq 0$ for some $i \in \{1, \dots, m\}$, wlog say $f_1(p) \neq 0$. In particular, $p \notin U$ since $U \subseteq Y$. Write $X_1 = X - U$ and $X_2 = X \cap Z(f_1)$. Clearly these are both closed subsets of X , and since $U \subseteq Z(f_1)$, have that $X = X_1 \cup X_2$. But X_1 is not the whole of X since U is nonempty, and X_2 is not the whole of X since $p \notin X_2$. This gives X as a union of two proper closed subsets, contradicting the irreducibility of X . Thus it is impossible for there to be a closed subset $Y \supseteq U$ but with $p \in X$ and $p \notin Y$. That is, every closed subset of \mathbb{A}^n containing U must contain all of X , i.e. U is dense in X .

Suppose U is not irreducible, so we can write $U = U_1 \cup U_2$ where U_1, U_2 are proper closed subsets of U . Let $X_1, X_2 \subseteq X$ be closed subsets of X s.t. $U_1 = U \cap X_1$ and $U_2 = U \cap X_2$. Clearly $X_1 \neq X$ since $U_1 \neq U$, and similarly $X_2 \neq X$. But $U = U_1 \cup U_2 = (U \cap X_1) \cup (U \cap X_2) = U \cap (X_1 \cup X_2)$ so $X_1 \cup X_2$ is a closed subset of X containing all of U . If $X_1 \cup X_2 = X$, then we would have obtained X as a union of two proper closed subsets, contradicting the irreducibility of X . But if $X_1 \cup X_2$ is not all of X , then writing $X_3 = X - U$ we definitely have that $X = (X_1 \cup X_2) \cup X_3$ as a union of two proper closed subsets, again contradicting the irreducibility of X . Hence it is impossible to write U as a union of proper closed subsets, i.e. U is irreducible.

Suppose X is an irreducible affine variety with at least 2 distinct points p and q . If X is Hausdorff, then there are disjoint open subsets U and V of X containing p and q respectively. Write $Y = X - U$ and $Z = X - V$. These are proper closed subsets of X since $p \notin Y$ and $q \notin Z$. Furthermore, since U and V are disjoint, have $V \subseteq Y$ and $U \subseteq Z$ so $Y \cup Z = X$. This gives X as a union of proper closed subsets, contradicting the irreducibility of X . Hence it is impossible for X to have two distinct points p and q and be Hausdorff.

Q3

Let $X \supseteq X_1 \supseteq X_2 \supseteq X_3 \supseteq \dots$ be a descending chain of closed subsets in an affine variety $X \subseteq \mathbb{A}^n$. Then $I(X_1) \subseteq I(X_2) \subseteq I(X_3) \subseteq \dots$ is an increasing chain of ideals in $k[x_1, \dots, x_n]$, which is a Noetherian ring, so there is some N s.t. $I(X_n) = I(X_N)$ for all $n \geq N$. Since the X_i are closed subsets, $Z(I(X_i)) = X_i$ by Q1, so $X_n = Z(I(X_n)) = Z(I(X_N)) = X_N$ for all $n \geq N$, i.e. the descending chain of closed subsets is eventually constant.

Q4

The coordinate ring of Y is $k[x, y]/(xy - 1)$ but the coordinate ring of \mathbb{A}^1 is $k[z]$. An isomorphism between Y and \mathbb{A}^1 induces an isomorphism between their coordinate rings. In particular, $x, y \in k[x, y]/(xy - 1)$ will be sent under this isomorphism to polynomials $p(z), q(z) \in k[z]$ such that $p(z)q(z) = 1$. But this can only happen if $p(z)$ and $q(z)$ are constants, i.e. if \mathbb{A}^1 is mapped to a single point in Y by the isomorphism, which is impossible (because k is algebraically closed so both Y and \mathbb{A}^1 have infinitely many points).

As argued above, any morphism $\mathbb{A}^1 \rightarrow Y$ induces a pullback of the coordinate functions x and y on Y , sending them to polynomial functions $p(z)$ and $q(z)$ on X , which must satisfy $p(z)q(z) = 1$, which is only possible if x and y are constants. So the only morphism $\mathbb{A}^1 \rightarrow Y$ is the constant map.

Regarding morphisms $Y \rightarrow \mathbb{A}^1$, there are the two obvious projections onto the coordinate axes $(x, y) \mapsto x$ and $(x, y) \mapsto y$. More generally, we can project onto any straight line through the origin, $(x, y) \mapsto ax + by$ for some constants $a, b \in k$, and then compose with any morphism $\mathbb{A}^1 \rightarrow \mathbb{A}^1$, which is simply a polynomial function in one variable, to obtain $(x, y) \mapsto p(ax + by)$ for any $p(z) \in k[z]$.

Day 2, 29th December 2024, Sunday

2.1.c Existence of Pseudo-Gradient Fields

Pseudo-gradient fields exist for all Morse functions on manifolds. This is, for example, a consequence of the existence of Riemannian metrics, and more exactly, of the existence of Riemannian metrics with a prescribed form on a given subset of the manifold. In any case, it is a simple consequence of the existence of partitions of unity.

Proposition 2.1.5 The stable and unstable manifolds of the critical point a are submanifolds of V that are diffeomorphic to open disks. Moreover, we have

$$\dim W^u(a) = \operatorname{codim} W^s(a) = \operatorname{Ind}(a)$$

Proposition 2.1.6 We suppose that the manifold V is compact. Let $\gamma : \mathbb{R} \rightarrow V$ be a trajectory of the pseudo-gradient field X . Then there exist critical points c and d of f such that

$$\lim_{s \rightarrow -\infty} \gamma(s) = c \quad \text{and} \quad \lim_{s \rightarrow +\infty} \gamma(s) = d$$

Theorem 2.2.5 (Smale Theorem) Let V be a manifold with boundary and let f be a Morse function on V with distinct critical values. We fix Morse charts in the neighbourhood of each critical point of f . Let Ω be the union of these charts and let X be a pseudo-gradient field on V that is transversal to the boundary. Then there exists a pseudo-gradient field X' that is close to X (in the \mathcal{C}^1 sense), equals X on Ω and for which we have

$$W_{X'}^s(a) \cap W_{X'}^u(b)$$

for all critical points a, b of f .

Theorem 3.2.2 The space $\overline{\mathcal{L}}(a, b)$ is compact.

Theorem 3.2.7 If $\operatorname{Ind}(a) = \operatorname{Ind}(b) + 2$ then $\overline{\mathcal{L}}(a, b)$ is a compact manifold of dimension 1 with boundary.

Proposition 3.2.8 Let V be a compact manifold, let $f : V \rightarrow \mathbb{R}$ be a Morse function and let X be a pseudo-gradient for f satisfying the Smale property. Let a, c and b be three critical points of indices $k + 1$, k and $k - 1$ respectively. Let $\lambda_1 \in \mathcal{L}(a, c)$ and $\lambda_2 \in \mathcal{L}(c, b)$. There exists a continuous embedding ψ from an interval $[0, \delta)$ onto a neighbourhood of (λ_1, λ_2) in $\overline{\mathcal{L}}(a, b)$ that is differentiable on $(0, \delta)$ and satisfies

$$\psi(0) = (\lambda_1, \lambda_2) \in \overline{\mathcal{L}}(a, b), \quad \psi(s) \in \mathcal{L}(a, b) \quad \text{for} \quad s \neq 0$$

Moreover, if (l_n) is a sequence in $\mathcal{L}(a, b)$ that tends to (λ_1, λ_2) , then l_n is contained in the image of ψ for n sufficiently large.

Lemma Let k be an arbitrary field, $f \in k[x, y]$ an irreducible polynomial, and $g \in k[x, y]$ an arbitrary polynomial. If g is not divisible by f then the system of equations $f(x, y) = g(x, y) = 0$ has only a finite number of solutions.

Day 3, 30th December 2024, Monday

2.3 L^2 bounded martingales

Definition Let

$$M^2 = \{X : \Omega \times [0, \infty) \rightarrow \mathbb{R} : X \text{ is a cadlag martingale with } \sup_{t \geq 0} \mathbb{E}X_t^2 < \infty\} / \sim$$
$$M_c^2 = \{X \in M^2 : X(\omega, \cdot) \text{ is continuous for every } \omega \in \Omega\} / \sim$$

where \sim means that indistinguishable processes are identified. Moreover, set

$$\|X\|_{M^2} = \left(\sup_t \mathbb{E}X_t^2 \right)^{1/2} = (\mathbb{E}X_\infty^2)^{1/2}$$

Proposition M^2 is a Hilbert space and M_c^2 is a closed subspace.

2.4 Quadratic variation

Definition For a sequence of processes (X^n) and a process X , we say $X^n \rightarrow X$ ucp (uniformly on compact sets in probability) if for all $t > 0$ and $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{s \in [0, t]} |X_s^n - X_s| > \epsilon\right) \rightarrow 0$$

as $n \rightarrow \infty$.

Day 4, 31st December 2024, Tuesday

It is 11.30pm on 31st December 2024, Tuesday. The last day of 2024. The new year is approaching and I feel mostly fear for what the future brings. Fear and stress. The past year seems to have gone by quickly and unproductively. I have accomplished little in the past 4 to 5 years but somehow it feels as if I accomplished the least in the past year. Somehow it feels I will accomplish even less in the coming year, perhaps even degenerate and regress.

I have found it hard to focus in the past few days since coming back to Singapore. Many reasons for my depression and unproductivity have occurred to me. The first reason that came to mind was loneliness. Christmas day was one of the most depressing days in recent memory, almost on par with that worst day of the fall 2024 semester. I spent it alone, aimlessly wandering from the Starbucks at Clarke Quay Central to the tables at Guoco tower at Tanjong Pagar, to the Starbucks at Marina Bay Link Mall. Revisiting all the usual old places. These places used to bring joy to me but now they only bring sadness. I saw so many old lonely people eating alone, sitting alone, aimlessly. It feels as though I am destined to become one of them, or perhaps I am already one of them. Lonely and poor with no purpose in life.

But it is also perhaps that I have come back to Singapore too often in the past year. I was only back here in July-August earlier this year, barely 4 months ago. I have accomplished nothing =in the meantime while I was away. It feels as if the city has moved on and left me behind. My peers from Raffles and Cambridge certainly have. I am once again studying at the library. It used to be such a joyful nostalgic thing, learning. But now it seems I am destined, doomed to do this for the rest of my life. Aimlessly, hopelessly studying into oblivion with nothing to show for it. Poor and lonely. Unaccomplished.

Studying math used to bring solace. It was pure, clean, escapism into a beautiful world where I could be myself. But not anymore. I find myself reading the same things again that I was reading 4 months ago, 1 year ago, 2 years ago. It feels like I am going in circles, or perhaps marching on the spot. Going nowhere. And I do have somewhere I need to be, somewhere I want to be. Somewhere else. There here and now do not spark joy. I long for a better life, a better place. But is that even possible at all now. It all feels so hopelessly impossibly out of reach. I am destined to fail. Forever.

Or perhaps it is just the same greed and lack of focus that has always plagued me. I think I have reached my limit. I am being torn apart. My answer to not knowing what I really want to do, all these years, has been to (try to) do everything. It is exhausting. I am exhausted.

It is the new year tomorrow. 1st of January. Just 15 minutes away now. The library won't be open. Sadly? Or perhaps thankfully. I think I will go cycling. I hope the weather will be good. I just want to run away from all of it. But where can I run to? I am stuck. Hopelessly stuck. And alone.

Youth is a wonderful thing. Too bad it is wasted on the young. I wish I could be a secondary school kid at Raffles again. Here I am at 30, turning 31, wishing I was 20 or 15, when one day I shall be 40 wishing I was 30.

$(X_n)_{n \geq 0}$ is *Markov*(λ, P) if and only if for all $n \geq 0$ and $i_0, \dots, i_n \in I$,

$$P(X_0 = i_0, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

Let \mathcal{A} be a π -system. Then any d -system containing \mathcal{A} contains also the σ -algebra generated by \mathcal{A} .

Let X be an integrable random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ -algebra. Then there exists a random variable Y such that

- (a) Y is \mathcal{G} -measurable
- (b) Y is integrable and $\mathbb{E}(X1_A) = \mathbb{E}(Y1_A)$ for all $A \in \mathcal{G}$.

Let X be a continuous local martingale with $X_0 = 0$. If X is also a finite variation process then $X_t = 0 \forall t$ a.s.

1st January 2025, Wednesday

Today was better. Went cycling.

2nd January 2025, Thursday

ALTAM Chapter 2 Survival Models

2.1 Summary

2.2 The future lifetime random variable

2.3 The force of mortality

2.4 Actuarial notation

2.5 Mean and standard deviation of T_x

2.6 Curtate future lifetime

15 regular exercises, 3 excel based exercises

Let (x) denote a life aged x , where $x \geq 0$. The death of (x) can occur at any age greater than x , and we model the future lifetime of (x) by a continuous random variable which we denote by T_x . This means that $x + T_x$ represents the age-at-death random variable for (x) . We write

$$F_x(t) = \mathbb{P}(T_x \leq t)$$

for the lifetime distribution from age x . The survival function

$$S_x(t) = 1 - F_x(t) = \mathbb{P}(T_x > t)$$

is the probability that (x) survives for at least t years.

The fundamental assumption: the events $\{T_x \leq t\}$ and $\{T_0 \leq x + t\}$ are equivalent in the sense that

$$\mathbb{P}(T_x \leq t) = \mathbb{P}(T_0 \leq x + t | T_0 > x)$$

i.e. given that a newborn individual survives to age x , the probability that they die by age $x + t$ is the same as the probability that an individual currently aged x dies within t years. Equivalently,

$$F_x(t) = \mathbb{P}(T_x \leq t) = \mathbb{P}(T_0 \leq x + t | T_0 > x) = \frac{\mathbb{P}(x < T_0 \leq x + t)}{\mathbb{P}(T_0 > x)} = \frac{F_0(x + t) - F_0(x)}{1 - F_0(x)}$$

$$S_x(t) = 1 - F_x(t) = \frac{1 - F_0(x + t)}{1 - F_0(x)} = \frac{S_0(x + t)}{S_0(x)}$$

$$S_0(x+t) = S_0(x)S_x(t)$$

i.e. the probability of a newborn surviving to age $x+t$ is the probability that they survive to x and the probability that a person aged x survives another t years. The events of a newborn surviving two disjoint contiguous time periods are independent. More generally,

$$S_0(x+t+u) = S_0(x+t)S_{x+t}(u) = S_0(x)S_x(t+u) \Rightarrow S_x(t+u) = \frac{S_0(x+t)}{S_0(x)}S_{x+t}(u) = S_x(t)S_{x+t}(u)$$

The events of a life of any age surviving two disjoint contiguous time periods are independent.

We denote the force of mortality at age x by μ_x and define it as

$$\begin{aligned} \mu_x &= \lim_{dx \rightarrow 0^+} \frac{\mathbb{P}(T_x \leq x+dx | T_0 > x)}{dx} \\ &= \lim_{dx \rightarrow 0^+} \frac{\mathbb{P}(T_x \leq dx)}{dx} \\ &= \lim_{dx \rightarrow 0^+} \frac{\mathbb{P}(T_x \leq dx) - \mathbb{P}(T_x \leq 0)}{dx} \\ &= \lim_{dx \rightarrow 0^+} \frac{F_x(dx) - F_x(0)}{dx} \\ &= - \lim_{dx \rightarrow 0^+} \frac{S_x(dx) - S_x(0)}{dx} \\ &= - \lim_{dx \rightarrow 0^+} \frac{S_x(dx) - 1}{dx} \\ &= - \frac{1}{S_0(x)} \lim_{dx \rightarrow 0^+} \frac{S_0(x+dx) - S_0(x)}{dx} = - \frac{1}{S_0(x)} S'_0(x) \end{aligned}$$

or equivalently since $S_0(0) = 1$,

$$S_0(x) = e^{-\int_0^x \mu_y dy}, \quad x \geq 0$$

More generally,

$$S_x(t) = \frac{S_0(x+t)}{S_0(x)} = e^{-\int_x^{x+t} \mu_y dy}$$

In any case, the survival function $S_x(t)$ is kind of exponentially decreasing over time and the force of mortality μ_x contains all of the information about the rate at which people die off at any age so the fraction of people at age x surviving for at least t years is (one over) the proportion beating the (exponential of) the total force of mortality over those years.

Two of the most useful mortality laws are Gompertz's law

$$\mu_x = Bc^x \quad \text{for some constants} \quad B > 0 \quad \text{and} \quad c > 1$$

and Makeham's law, which is a generalisation of Gompertz's law, given by

$$\mu_x = A + Bc^x \quad \text{for some constants} \quad A, B > 0 \quad \text{and} \quad c > 1$$

One of the earliest mortality laws proposed was De Moivre's law

$$\mu_x = \frac{1}{\omega - x}, \quad 0 \leq x < \omega$$

This is unrealistic since it implies that T_x is uniformly distributed on the interval $(0, \omega - x)$, which means that the probability that a life currently aged x dies between ages $x + t$ and $x + t + dt$ is the same for all $t > 0$ as long as $x + t < \omega$.

Another simple mortality law that is very unrealistic for modelling human mortality is the constant force of mortality assumption which states that $\mu_x = \mu$ for all $x \geq 0$. Under this model, T_x has an exponential distribution.

Although the De Moivre and constant force of mortality models are not useful for overall mortality for humans, they may be used in other contexts, such as modelling the failure time of machine components. We also use these models in a limited sense, to model mortality between integer ages.

The actuarial notation for survival and mortality probabilities is

$${}_t p_x = S_x(t), \quad {}_t q_x = F_x(t), \quad {}_u|_t q_x = S_x(u) - S_x(u + t)$$

the last being the probability that a life currently aged x survives for u years but then dies within the t years after that. We drop the subscript t if its value is 1 so p_x is the probability that (x) survives to at least age $x + 1$ and q_x is the probability that (x) dies before age $x + 1$. We call q_x the mortality rate at age x and ${}_u|_t q_x$ the deferred mortality probability.

The complete expectation of life

$$\mathring{e}_x = \mathbb{E}(T_x) = \int_0^\infty t f_x(t) dt = \int_0^\infty t dF_x(t) = - \int_0^\infty t d{}_t p_x = - \left(t {}_t p_x \Big|_0^\infty - \int_0^\infty {}_t p_x dt \right) = \int_0^\infty {}_t p_x dt$$

is the expected future lifetime of (x) .

We are sometimes interested in the future lifetime random variable subject to a cap of n years, which is represented by the random variable $\min(T_x, n)$. The term expectation of life

$$\mathring{e}_{x:n|} = \int_0^n t f_x(t) dt + \int_n^\infty n f_x(t) dt = - \int_0^n t d{}_t p_x + n {}_n p_x = - \left(t {}_t p_x \Big|_0^n - \int_0^n {}_t p_x dt \right) + n {}_n p_x = \int_0^n {}_t p_x dt$$

is the expected value of $\min(T_x, n)$.

The curtate future lifetime random variable

$$K_x = \lfloor T_x \rfloor$$

is the integer part of future lifetime. For $k = 0, 1, 2, \dots$ we note that $K_x = k$ iff (x) dies between the ages of $x + k$ and $x + k + 1$, so

$$\mathbb{P}(K_x = k) = \mathbb{P}(k \leq T_x < k + 1) = {}_k p_x - {}_{k+1} p_x$$

The curtate expectation of life is

$$e_x = \mathbb{E}(K_x) = \sum_{k=0}^{\infty} k \mathbb{P}(K_x = k) = ({}_1 p_x - {}_2 p_x) + 2({}_2 p_x - {}_3 p_x) + 3({}_3 p_x - {}_4 p_x) + \dots = \sum_{k=1}^{\infty} {}_k p_x$$

Similarly we have the random variable $\min(K_x, n)$ with expectation

$$e_{x:n|} = \sum_{k=1}^n {}_k p_x$$

From the trapezium approximation for integrals, we obtain the approximation

$$\dot{e}_x \approx e_x + \frac{1}{2}$$

Exercise 2.1 You are given that

$$p_x = 0.99, \quad p_{x+1} = 0.985, \quad {}_3 p_{x+1} = 0.95, \quad q_{x+3} = 0.02$$

Calculate p_{x+3} , ${}_2 p_x$, ${}_2 p_{x+1}$, ${}_3 p_x$ and ${}_1|_2 q_x$

Solution 2.1

$$x \rightarrow x + 1 \rightarrow x + 2 \quad 1 \rightarrow 0.99 \rightarrow 0.99 \times 0.985$$

$$x + 1 \rightarrow x + 2 \rightarrow x + 3 \rightarrow x + 4 \quad 1 \rightarrow ? \rightarrow ? \rightarrow 0.95$$

$$x + 3 \rightarrow x + 4 \quad 1 \rightarrow 0.98$$

So immediately have $p_{x+3} = 1 - q_{x+3} = 1 - 0.02 = 0.98$

And quite immediately have ${}_2 p_x = p_x p_{x+1} = 0.99 \times 0.985 = 0.97515$

And quite immediately have ${}_2 p_{x+1} = {}_3 p_{x+1} / p_{x+3} = 0.95 / 0.98 = 0.9693877551$

Then we have $p_{x+2} = {}_2p_{x+1}/p_{x+1} = 0.9693877551/0.985 = 0.98415000518$

And hence ${}_3p_x = {}_2p_x p_{x+2} = 0.97515 \times 0.98415000518 = 0.95969387755$

Finally have ${}_1|_2q_x = p_x(1 - {}_2p_{x+1}) = 0.99 \times (1 - 0.9693877551) = 0.03030612245$

Exercise 2.6 The function

$$G(x) = \frac{18000 - 110x - x^2}{18000}$$

has been proposed as the survival function $S_0(x)$ for a mortality model.

What is the implied limiting age ω ? Verify that the function G satisfies the criteria for a survival function. Calculate ${}_{20}p_0$. Determine the survival function for a life aged 20. Calculate ${}_{10}|_{10}q_{20}$. Calculate μ_{50} .

Solution 2.6

$$G(x) = \frac{(200 + x)(90 - x)}{18000}$$

so the limiting age is 90, everyone dies by the time they are 90 years old. G is a survival function since $G(0) = 1$ and G is clearly smooth on $x \in (0, 90)$ and bounded on $x \in [0, 90]$ and hence has finite moments of all orders, and

$$G'(x) = -\frac{110 + 2x}{18000} < 0$$

for $x \in [0, 90]$ so G is decreasing. The probability that a newborn survives to 20 years old is

$${}_{20}p_0 = G(20) = \frac{220 \times 70}{18000} = 0.855555555555$$

The survival function for a life aged 20 is

$${}_tp_{20} = \frac{{}_{20+t}p_0}{{}_{20}p_0} = \frac{G(20+t)}{G(20)} = \frac{(220+t)(70-t)}{220 \times 70} = \frac{15400 - 150t - t^2}{15400}$$

The probability that a life aged 20 survives the next 10 years and then dies within the 10 years after that, is

$${}_{10}|_{10}q_{20} = {}_{10}p_{20}(1 - {}_{10}p_{30}) = \frac{G(30)}{G(20)} \left(1 - \frac{G(40)}{G(30)}\right) = \frac{230 \times 60}{220 \times 70} \left(1 - \frac{240 \times 50}{230 \times 60}\right) = 0.116883117$$

The force of mortality for a life aged 50 is

$$\mu_{50} = -\frac{G'(50)}{G(50)} = \frac{110 + 2x}{(200 + x)(90 - x)} = \frac{210}{250 \times 40} = 0.021$$

3rd January 2025, Friday, 2.40pm at 5th floor of Tampines Library at Tampines Hub

ALTAM Chapter 3 Life tables and selection

3.1 Summary

3.2 Life tables

3.3 Fractional age assumptions

3.4 National life tables

3.5 Survival models for life insurance policyholders

3.6 Life insurance underwriting

3.7 Select and ultimate survival models

3.8 Notation and formulae for select survival models

3.9 Select life tables

3.10 Some comments on heterogeneity in mortality

3.11 Mortality improvement modelling

3.12 Mortality improvement scales

March 2025 FAM syllabus excludes sections 4, 10, 11 and 12.

Given a survival model with survival probabilities ${}_tp_x$, we construct the life table for the model, from some initial age x_0 to a maximum or limiting age ω , using a function $\{l_x\}$, $x_0 \leq x \leq \omega$, where l_{x_0} is an arbitrary positive number called the radix of the table, and

$$l_{x_0+t} = l_{x_0} {}_tp_{x_0}$$

More generally,

$$l_{x+t} = l_{x_0 x+t-x_0} p_{x_0} = l_{x_0 x-x_0} p_{x_0 t} p_x = l_{xt} p_x \quad \Rightarrow \quad {}_tp_x = \frac{l_{x+t}}{l_x}$$

In some cases, a life table tabulated at integer ages also shows values of

$$d_x = l_x - l_{x+1}$$

The uniform distribution of deaths (UDD) assumption is the most common fractional age assumption. It can be formulated in two equivalent ways:

UDD1 For integer x and for $0 \leq s < 1$, assume that

$${}_sq_x = sq_x$$

That is, for any integer x , the mortality probability over $s < 1$ years is s times the one-year mortality probability.

UDD2 For a life (x) where x is an integer, with future lifetime random variable T_x and curtate future lifetime random variable K_x , define a new random variable R_x to represent the fractional part of the future lifetime of (x) lived in the year of death, so that $T_x = K_x + R_x$. We assume

$$R_x \sim U(0, 1) \quad \text{independent of} \quad K_x$$

An immediate consequence is that

$$l_{x+s} = l_x - s d_x, \quad 0 \leq s < 1$$

Thus UDD implies that l_{x+s} is a linearly decreasing function of s between integer ages.

Furthermore,

$$\mu_{x+s} =$$

A second fractional age assumption is that the force of mortality is constant between integer ages. Thus for integer x and $0 \leq s < 1$ we assume that μ_{x+s} does not depend on s , and we denote it μ_x^* . We can obtain the value of μ_x^* from the life table by using the fact that

$$p_x = e^{-\int_0^1 \mu_{x+t} dt} = e^{-\mu_x^*} \Rightarrow \mu_x^* = -\ln p_x$$

Furthermore,

$${}_t p_x = e^{-\int_0^t \mu_{x+s} ds} = e^{-t \mu_x^*} = (p_x)^t$$

More generally for $r, t > 0$ and $r + t < 1$,

$${}_{t+r} p_{x+r} = e^{-\int_0^t \mu_{x+r+s} ds} = e^{-t \mu_x^*} = (p_x)^t$$

Life tables based on the mortality experience of the whole population are regularly produced for countries around the world. Separate life tables are usually produced for males and for females and possibly for some other groups of individuals e.g. by race or socio-economic group.

Suppose we have to choose a survival model appropriate for a man who is currently aged 50 and living in the UK, and who has just purchased a 10-year term insurance policy. We could use a national life table. However, in the UK, as in other countries with well-developed life insurance markets, the mortality experience of people who purchase life insurance policies tends to be different from the population as a whole. The mortality of different types of life insurance policyholders is investigated separately, and life tables appropriate for these groups are published.

The mortality rates in the CMI tables are based on individuals accepted for insurance at normal premium rates, that is, individuals who have passed the required health checks. This means that a man aged 50 who has just purchased a term insurance at the normal premium rate is known to be in good health, assuming the health checks are effective, and so is likely to be much healthier and hence have a lower mortality rate than a man of age 50 picked randomly from the population. When this man reaches age 56, we can no longer be certain he is in good health. All we know is that he was in good health six years ago. Hence his mortality rate at age 56 is higher than that of a man of the same age who has just passed the health checks and been permitted to buy a term insurance policy at normal rates. The effect of passing the health checks at issue eventually wears off, so that at age 62, the force of mortality does not depend on whether the policy was purchased at age 50, 52, 54 or 56.

A feature of the survival models studied in Chapter 2 is that probabilities of future survival depend only on the individual's current age. For example, for a given survival model and a given term t , the probability ${}_t p_x$ that an individual currently aged x will survive to age $x + t$ depends only on the current age x . Such survival models are called aggregate survival models, because lives are all aggregated together.

The difference between an aggregate survival model and the survival model for term insurance policyholders discussed in Section 3.6 is that in the latter case, probabilities of future survival depend not only on current age but also on how long ago the policy was purchased, which was when the policyholder joined the group of insured lives.

The mortality of a group of individuals is described by a “select and ultimate survival model”, usually shortened to “select survival model”, if:

1. Future survival probabilities for an individual in the group depend on the individual's current age and on the age at which the individual joined the group.
2. There is a positive number d , generally an integer, such that if an individual joined the group more than d years ago, future survival probabilities depend only on current age. The initial selection effect is assumed to have worn off after d years.

An individual who enters the group at age x is said to be “select” at age x . The period d after which the age at selection has no effect on future survival probabilities is called the “select period” for the model. The mortality that applies to lives after the select period is complete is called “ultimate mortality”, so that the complete model comprises a select period followed by the ultimate period.

A select survival model represents an extension of the ultimate survival model studied in Chapter 2. This requires the notation

$${}_t p_{[x]+s} = \mathbb{P}(\text{a life } (x + s) \text{ select at aged } x \text{ survives to age } x + s + t)$$

$${}_t q_{[x]+s} = \mathbb{P}(\text{a life } (x + s) \text{ select at aged } x \text{ does not survive to age } x + s + t)$$

$$\mu_{[x]+s} = \lim_{h \rightarrow 0^+} \frac{1 - {}_h p_{[x]+s}}{h} = \text{force of mortality at age } x + s \text{ for a life select at age } x$$

From these definitions we can obtain

$${}_t p_{[x]+s} = e^{-\int_0^t \mu_{[x]+s+u} du}$$

For an ultimate survival model the life table $\{l_x\}$ is useful since it can be used to calculate probabilities of survival. We can construct a select life table in a similar way but we need the table to reflect duration since selection, as well as age, during the select period. Select life tables are usually presented at integer ages only, just as for the aggregated life tables.

14 regular exercises, 0 excel based exercises

Exercise 3.2 You are given the following life table extract:

$$\begin{array}{llll} l_{52} = 89948, & l_{53} = 89089, & l_{54} = 88176, & l_{55} = 87208, \\ l_{56} = 86181, & l_{57} = 85093, & l_{58} = 83940, & l_{59} = 82719 \end{array}$$

Calculate each of the following probabilities assuming (i) uniform distribution of deaths between integer ages, and (ii) a constant force of mortality between integer ages:

$$0.2q_{52.4}, \quad 5.7p_{52.4}, \quad 3.2|_{2.5}q_{52.4}$$

Solution 3.2 Assuming uniform distribution of deaths between integer ages,

$$\begin{aligned} l_{52.4} &= 0.6 \times 89948 + 0.4 \times 89089 = 89604.4, \\ l_{52.6} &= 0.4 \times 89948 + 0.6 \times 89089 = 89432.6, \\ l_{55.6} &= 0.4 \times 87208 + 0.6 \times 86181 = 86591.8, \\ l_{58.1} &= 0.9 \times 83940 + 0.1 \times 82719 = 83817.9 \end{aligned}$$

so

$$\begin{aligned} 0.2q_{52.4} &= 1 - \frac{l_{52.6}}{l_{52.4}} = 0.001917317, \\ 5.7p_{52.4} &= \frac{l_{58.1}}{l_{52.4}} = 0.935421698, \\ 3.2|_{2.5}q_{52.4} &= \frac{l_{55.6}}{l_{52.4}} \left(1 - \frac{l_{58.1}}{l_{55.6}} \right) = 0.030957185 \end{aligned}$$

Assuming a constant force of mortality between integer ages,

$$\begin{aligned} l_{52.4} &= 89948^{0.6}89089^{0.4} = 89603.41054, \\ l_{52.6} &= 89948^{0.4}89089^{0.6} = 89431.61118, \\ l_{55.6} &= 87208^{0.4}86181^{0.6} = 86590.34064, \\ l_{58.1} &= 83940^{0.9}82719^{0.1} = 83817.09332 \end{aligned}$$

so

$$\begin{aligned} 0.2q_{52.4} &= 1 - \frac{l_{52.6}}{l_{52.4}} = 0.001917331, \\ 5.7p_{52.4} &= \frac{l_{58.1}}{l_{52.4}} = 0.935423025, \\ 3.2|_{2.5}q_{52.4} &= \frac{l_{55.6}}{l_{52.4}} \left(1 - \frac{l_{58.1}}{l_{55.6}} \right) = 0.030950243 \end{aligned}$$

Exercise 3.8 A group of 100,000 independent lives each aged 65 purchases one-year term insurance. At the start, 20% of the group are preferred lives with $q_{65} = 0.002$ and 80% of the group are normal lives with $q_{65} = 0.009$.

- (a) Calculate the standard deviation of the number of survivors at the end of the year.
- (b) Calculate the proportion of preferred lives expected in the survivor group.
- (c) Using a normal approximation without continuity correction, calculate the 90th percentile of the number of survivors at the end of the year.

Solution 3.8

$$\mathbb{P}(\text{preferred}) = 0.2, \quad \mathbb{P}(\text{normal}) = 0.8, \quad \mathbb{P}(\text{survive}|\text{preferred}) = 0.998, \quad \mathbb{P}(\text{survive}|\text{normal}) = 0.991$$

$$\mathbb{P}(\text{survive}) = 0.9924, \quad \text{stdev} = \sqrt{100,000 \times 0.9924(1 - 0.9924)} = 27.463$$

Actually maybe we should calculate the standard deviation of each type of life separately and add up within each population as in

$$\sqrt{20,000 \times 0.998 \times 0.002 + 80,000 \times 0.991 \times 0.009} = 27.44886154$$

Expected number of surviving preferred lives = $0.998 \times 20,000 = 19960$

Expected number of surviving normal lives = $0.991 \times 80,000 = 79280$

Proportion of preferred lives expected in survivor group = $19960 / (19960 + 79280) = 0.201128577$

$$\text{mean number of survivors} = 99240, \quad \text{stdev number of survivors} = 27.463$$

Standard normal 90th percentile is 1.28 so our 90th percentile x is given by

$$\frac{x - 99240}{27.463} = 1.28 \quad \Rightarrow \quad x = 99275.15264$$

ALTAM Chapter 4 Insurance benefits

n regular exercises, m excel based exercises

ALTAM Chapter 5 Annuities

n regular exercises, m excel based exercises