UNIVERSITY OF LEICESTER

MA7404
All Candidates

Semester 1 Examinations 2020

DO NOT OPEN THE QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY THE CHIEF INVIGILATOR	
School	Mathematics and Actuarial Sciences
Module Code	MA7404
Module Title	Markov processes
Exam Duration	Two hours
CHECK YOU HAVE THE CORRECT QUESTION PAPER	
Number of Pages	9
Number of Questions	4
Instructions to Candidates	Answer all questions
FOR THIS EXAM YOU ARE ALLOWED TO USE THE FOLLOWING:	
Calculators	Approved calculators may be used
Books/Statutes provided by the University	Formulae and tables for actuarial examinations
Are students permitted to bring their own Books/Statutes/Notes?	No
Additional Stationery	No

1. An insurance company is interested in estimating the total size S of claims they will receive next week. They model it as $S = X_1 + X_2 + \dots X_N$, where the number of claims N follows a negative binomial distribution with parameters $k = 3$ and p, and the claim sizes X_i are independent from each other and from N, identically distributed, and follow an exponential distribution with parameter λ. Parameters λ and p are unknown and should be estimated from the past data.

   (i) [5 marks] Formulate the method of moments for estimating unknown parameters of a distribution.

   (ii) [5 marks] The sizes of the last 10 claims were

   $$100, 250, 150, 400, 200, 230, 170, 100, 500 \text{ and } 400.$$

   Use the method of moments to estimate parameter λ.

   (iii) [5 marks] The numbers of claims for the last 10 weeks are $3,5,2,0,8,4,1,0,1$ and 3. Use the method of maximum likelihood to estimate parameter p.

   (iv) [5 marks] Estimate the mean and standard deviation of S.

   (v) [5 marks] What is the most likely number of claims during the next week?

   Total: 25 marks

Answer:

(i) Bookwork (Seen).

The method of moments suggests to select parameters in such a way that first r moments estimated from the data match the first r moments $f_j(a_1, a_2, \dots, a_r)$, $j = 1, 2, \dots, r$ estimated from the formulas for the distribution, that is,

$$m_j = f_j(a_1, a_2, \dots, a_r), \quad j = 1, 2, \dots, r$$

where $m_j = \frac{1}{n}\sum_{i=1}^n x_i^j$, $j = 1, 2, \dots, r$.

(ii) Application (Similar to seen)

If X is the size of a claim, then $E[X] = \int_0^\infty x\lambda e^{-\lambda x}dx = 1/\lambda$. On the other hand, from data $E[X] \approx 250$. Hence, $\lambda \approx 1/250$.

(iii) Application (Similar to seen)

The likelihood function corresponding to a discrete distribution and data $n_1, n_2, \dots, n_m$ is

$$L = \prod_{i=1}^m P(N = n_i)$$

In our case,

$$P(N = n) = \frac{(3+n-1)!}{n!(3-1)!} \cdot p^3(1-p)^n = 0.5(n+2)(n+1)p^3(1-p)^n, \quad n = 0, 1, 2, \dots,$$

so that

$$L = \prod_{i=1}^{10} 0.5(n_i+2)(n_i+1)p^3(1-p)^{n_i} = Cp^{30}(1-p)^{\sum n_i} = Cp^{30}(1-p)^{27},$$

where C is some constant. Then

$$\log L = \log C + 30 \log p + 27 \log(1-p),$$

is maximal when the derivative is 0:

$$0 + \frac{30}{p} - \frac{27}{1-p} = 0,$$

hence $30(1-p) = 27p$, $30 = 57p$, $p = 30/57 = 10/19 \approx 0.526$.

(iv) Application (Similar to seen).

$$E[N] = \frac{k(1-p)}{p} = \frac{3(1-10/19)}{(10/19)} = 27/10 = 2.7,$$

$$Var[N] = \frac{k(1-p)}{p^2} = \frac{3(1-10/19)}{(10/19)^2} = 5.13,$$

Let X be the size of a claim. Then

$$E[X] = \frac{1}{\lambda} = 250$$

$$Var[X] = \frac{1}{\lambda^2} = 62500.$$

Hence,

$$E[S] = E[N] \cdot E[X] = 2.7 \cdot 250 = 675,$$

and

$$\sigma(S) = \sqrt{Var(S)} = \sqrt{E[N]Var(X) + Var[N](E(X))^2} = \sqrt{2.7 \cdot 62500 + 5.13 \cdot (250)^2} \approx 700.$$

(v) Higher skills (Unseen).

Let $p_n = P(N = n)$ be the probability that there will be n claims. Then

$$p_n = \frac{(3+n-1)!}{n!(3-1)!} \cdot p^3 (1-p)^n = \frac{(n+1)(n+2)}{2} \cdot p^3 (1-p)^n, \quad n = 0, 1, 2, \dots$$

Hence, the ratio

$$\frac{p_n}{p_{n+1}} = \frac{(n+1)(n+2)(1-p)^n}{(n+2)(n+3)(1-p)^{n+1}} = \frac{(n+1)}{(n+3)(1-p)}.$$

This ratio is less than 1 if and only if

$$1 - p < \frac{n+1}{n+3} = 1 - \frac{2}{n+3},$$

or equivalently, $p > \frac{2}{n+3}$, or $n+3 > \frac{2}{p}$, or $n > \frac{2}{p} - 3 \approx 0.8$.

Hence, $p_0 < p_1$, but $p_1 > p_2 > p_3 > \dots$.. So, the most likely number of claims is 1.

Syllabus cover

This question covers items 1.1 (loss distributions...) and 1.2 (compound distributions...) of the syllabus.

Feedback

Many students failed to do even bookwork part (i), which required just memorizing and reproducing a small paragraph from lecture notes. Part (ii) was answered well. In part (iii), some students used method of moments while the question clearly requires to use maximal likelihood method. Some other students were not able to do calculations. Part (iv) was answered reasonably well, if we ignore the fact that input from part (iii) was often incorrect. In part (v), surprisingly many students really think that "most likely number" is the same as "expected number"! This is clearly wrong.

2. The densities of non-negative random variables X and Y are $f_X(x) = e^{-x}, x \geq 0$, and $f_Y(y) = \frac{1}{(1+y)^2}, y \geq 0$, respectively, while their copula is $C(u,v) = \frac{uv}{u+v-uv}$.

   (i) [5 marks] Calculate the probability $P[X \leq \ln 2 \text{ and } Y \leq 3]$.

   (ii) [5 marks] Calculate the hazard rates for X and Y.

   (iii) [5 marks] By analysing the hazard rates of X and Y, determine which of these random variables has a heavier tail.

   (iv) [5 marks] Define the coefficients of lower and upper tail dependence for 2 random variables with copula $C(u,v)$.

   (v) [5 marks] Calculate the coefficient of lower tail dependence for X and Y.

   Total: 25 marks

Answer:

(i) Higher skills (Unseen)

By the definition of copula, the joint cdf $F_{XY}(x,y)$ is $F_{XY}(x,y) = C(F_X(x), F_Y(y))$. In particular,

$$P[X \leq \ln 2 \text{ and } Y \leq 3] = F_{XY}(\ln 2, 3) = C(F_X(\ln 2), F_Y(3))$$

The cdf of X is $F_X(x) = \int_0^x f_X(t)dt = \int_0^x e^{-t}dt = 1 - e^{-x}$. In particular, $F_X(\ln 2) = 1 - e^{-(\ln 2)} = 1 - 0.5 = 0.5$.

The cdf of Y is $F_Y(y) = \int_0^y \frac{1}{(1+t)^2}dt = \frac{y}{1+y}$. In particular, $F_Y(3) = \frac{3}{1+3} = 0.75$. Hence,

$$P[X \leq \ln 2 \text{ and } Y \leq 3] = C(0.5, 0.75) = \frac{0.5 \cdot 0.75}{0.5 + 0.75 - 0.5 \cdot 0.75} = \frac{3}{7} \approx 0.4286.$$

(ii) Application (Similar to seen)

The hazard rate of X is

$$h(x) = \frac{f_X(x)}{1 - F_X(x)} = \frac{e^{-x}}{1 - (1 - e^{-x})} = 1.$$

The hazard rate of Y is

$$h(y) = \frac{f_Y(y)}{1 - F_Y(y)} = \frac{1}{(1+y)^2} : \left(1 - \frac{y}{1+y}\right) = \frac{1}{1+y}.$$

(iii) Application (Similar to seen)

For X, the hazard rate $h(x)$ is constant, while for Y, the hazard rate $h(y)$ is a decreasing function. Hence, Y has havier tail than X.

(iv) Bookwork (Seen)

The coefficients of lower and upper tail dependence of 2 random variables are

$$\lambda_L = \lim_{u\to 0+} \frac{C(u,u)}{u}, \quad \text{and} \quad \lambda_U = \lim_{u\to 0+} \frac{\bar{C}(u,u)}{u},$$

where $\bar{C}$ is the survival copula defined by $\bar{C}(1-u,1-v) = 1-u-v+C(u,v)$.

(v) Application (Similar to seen)

$$\lambda_L = \lim_{u\to 0+} \frac{C(u,u)}{u} = \lim_{u\to 0+} \frac{uu}{(u+u-uu)\cdot u} = \lim_{u\to 0+} \frac{1}{2-u} = 0.5.$$

Syllabus cover

This question covers items 1.3 (introduction to copulas) and 1.4 (introduction to extreme value theory) of the syllabus.

Feedback

In part (i), surprisingly many students assumed that X and Y are independent. This is especially surprising because the question is about copulas, and the whole section about copulas was about how random variables DEPEND on each other! They are independent only if copula is $C(u,v) = uv$ (the independence copula), but in this question it is clearly different! For part (ii), surprisingly many students miscalculated cdf F from pdf f. Part (iii) was mostly done well by those who did part (ii). Further, some students failed to do even bookwork part (iv), which required just memorizing and reproducing a couple of formulas from lecture notes. Part (v) was mostly done well by those who did part (iv).

3. A motor insurance company offers discounts of 0%, 20% and 50% of the full premium, determined by the following rules:

   (a) All new policyholders start at the 20% level.

   (b) If no claim is made during the current year the policyholder moves to/stays at the 50% level.

   (c) If one or more claims are made the policyholder moves down one discount level, or remains at the 0% level.

   Since this policy was introduced, there were:

   - $2,000$ cases when a policyholder had no discount. In 500 cases, a claim was made.
   - $5,000$ cases when a policyholder had a 20% discount. In $1,000$ cases, a claim was made.
   - $40,000$ cases when a policyholder had a 50% discount. In $2,000$ cases, a claim was made.

(i) [5 marks] Write down the Chapman-Kolmogorov equations for a Markov chain, explaining all notation used.

(ii) [5 marks] Model the process above as a Markov chain. Determine the state space and transition matrix.

(iii) [5 marks] Estimate the probability that a policyholder initially at the 20% level is at the 20% level after 3 years.

(iv) [5 marks] Calculate the stationary distribution.

(v) [5 marks] Assuming that this policy operates for a long time, estimate the average discount over all policyholders.

Total: 25 marks

Answer:

(i) Bookwork (Seen)

Chapman-Kolmogorov equations are

$$p_{ij}(t_1,t_3) = \sum_{k\in S} p_{ik}(t_1,t_2)\, p_{kj}(t_2,t_3), \quad \text{for} \quad t_1 < t_2 < t_3, \tag{1}$$

where S is the state space and $p_{ij}(s,t)$ is the probability to be in state j by time t given that the process is at state i at time s. In other words, $p_{ij}(s,t) = P[X(t) = j|X(s) = i]$.

(ii) Application (Similar to seen)

The state space is $S = \{0\%, 20\%, 50\%\}$. The transition probability matrix between two states in one year is given by

$$\mathbf{P} = \begin{pmatrix} \frac{500}{2000} & 0 & \frac{2000-500}{2000} \\ \frac{1000}{5000} & 0 & \frac{4000}{5000} \\ 0 & \frac{2000}{40000} & \frac{38000}{40000} \end{pmatrix} = \begin{pmatrix} 0.25 & 0 & 0.75 \\ 0.2 & 0 & 0.8 \\ 0 & 0.05 & 0.95 \end{pmatrix}. \tag{2}$$

(iii) Application (Similar to seen)

To be in the 20%-state after 3 years, a policyholder may either follows the path 20%-50%-50%-20%, the probability of which is $0.8 \cdot 0.95 \cdot 0.05 = 0.038$, or follows the path 20%-0%-50%-20%, the probability of which is $0.2 \cdot 0.75 \cdot 0.05 = 0.0075$. The total probability is 0.0455.

(iv) Application (Similar to seen)

The conditions for a stationary distribution lead to the following expressions

$$\begin{aligned} \pi_0 &= 0.25\pi_0 + 0.2\pi_1, \\ \pi_1 &= 0.05\pi_2, \\ \pi_2 &= 0.75\pi_0 + 0.8\pi_1 + 0.95\pi_2, \\ 1 &= \pi_0 + \pi_1 + \pi_2, \end{aligned}$$

where π_0, π_1, π_2 corresponds to discounts 0%, 20%, and 50%, respectively.

This system is solved with $\pi_0 = \frac{4}{319} \approx 0.0125$, $\pi_1 = \frac{15}{319} \approx 0.047$, $\pi_2 = \frac{300}{319} \approx 0.9404$.

(v) Higher skills (Unseen)

Because the Markov chain is is finite, irreducible (you can reach any state from any other one), and aperiodic (you can return to states 0 and 2 in 1 year, and to state 1 in 2 or 3 years), it converges to its stationary distribution. Because the policy is introduced long ago, we may assume that it is already close to its stationary distribution. Hence, the average discount is

$$\pi_0 \cdot 0\% + \pi_1 \cdot 20\% + \pi_2 \cdot 50\% \approx 0.047 \cdot 20\% + 0.9404 \cdot 50\% \approx 48\%.$$

Syllabus cover

This question covers item 3.2 of the syllabus: Define and apply a Markov chain.

Feedback

Overall, this was the easiest question of the exam. However, surprisingly many students failed to do bookwork part (i), which required just memorizing and reproducing a small paragraph from lecture notes. Many students wrote Kolmogorov forward equations for Markov jump process instead of the requested Chapman-Kolmogorov equations for Markov chain. In part (ii), some students ignored the condition that "If no claim is made during the current year the policyholder moves to/stays at the 50% level" and assumed that they move at most 1 level up. Parts (iii) and (iv) was mostly done well by those who did part (ii) correctly. In part (v), surprisingly many students really think that "most likely number" is the same as "expected number", and claimed that, because 50% is the most likely discount, then it must also be the average one! This is clearly wrong.

4. To model mortality in a group of people aged 60 with no severe health problems known, an insurance company uses a simple two-state model with states A (alive) and D (dead), and transition rate λ from A to D. They would like to estimate parameter λ from past data. Each year from 2010 to 2019, they observed $1,000$ individuals in the given age group and found that the number of deaths was $8,7,5,5,6,4,5,3,4$ and 3 respectively.

   (i) [5 marks] List possible types of stochastic processes with respect to state space and time changes. To which type Markov jump process belongs?

   (ii) [5 marks] Give three different examples of problems which can be solved using machine learning techniques.

   (iii) [5 marks] Use linear regression with sum of squares error minimization to estimate the parameter λ to be used in 2020.

   (iv) [5 marks] Using λ obtained in part (iii), (or, if you cannot solve (iii), use $\lambda = 0.003$), estimate the probability that a person will survive during the whole 2020 year.

   (v) [5 marks] Repeat part (iv) assuming that $\lambda = \lambda(t)$ is not a constant but decreases linearly from $\lambda_0 = 0.004$ at the beginning of 2020 to $\lambda_1 = 0.003$ at the end of the year.

   Total: 25 marks

Answer:

(i) Bookwork (Seen)

Stochastic processes with state space $\mathcal{S}$ and time changes $\mathcal{T}$ can be roughly classified into the following groups:

- Discrete $\mathcal{S}$ and discrete $\mathcal{T}$;
- Continuous $\mathcal{S}$ and discrete $\mathcal{T}$;
- Discrete $\mathcal{S}$ and continuous $\mathcal{T}$;
- Continuous $\mathcal{S}$ and continuous $\mathcal{T}$; or
- Mixed processes.

Markov jump process has Discrete $\mathcal{S}$ and continuous $\mathcal{T}$.

(ii) Bookwork (Seen)

Classification problem. For example, one may want to automatically classify all e-mails into several groups: spam, work e-mails, private e-mails, etc.

Clustering. In some applications of multi-class classification (like spam filtering) it may be clear in advances what classes to consider (spam, non-spam, etc.). However, in some other applications this is unclear. For example, in the image recognition problem, it is difficult to list in advance all possible "objects" the webcam can detect. In this case, we may ask the system to do this automatically. It may represent all observed objects as points in some n-dimensional space, and then automatically detect that these points form m groups, or classes. After this, every new object detected is classified into one of these classes. "Similar" objects go to one class, and "dissimilar" ones - into different classes. This problem is called **clustering**.

Regression. Mathematically, (linear) regression problem is typically reduces to the following task: given m points in the coordinate plane R^{n+1} with coordinates $x_1, x_2, \dots, x_n, y$, approximate the points by linear function $y = w_0 + w_1x_1 + w_2x_2 + \dots + w_nx_n$ as good as possible. Typically, the quality of approximation is measured as the sum of squares of differences between the y coordinates of data points and the function.

(iii) Higher Skills (Unseen)

Let t be the number of years since we have started collect the data in 2010 ($t = 0$ in 2010, $t = 1$ in 2011, and so on), and let $x_0, x_1, \dots, x_9$ be our data. We put points $(0, x_0), (1, x_1), \dots (9, x_9)$ on the coordinate plane (t, x), and approximate them by a line $x = at + b$, to minimize the sum of squares error

$$e(a,b) = \sum_{i=0}^{9} (a \cdot i + b - x_i)^2$$

We have

$$0 = \frac{\partial e(a,b)}{\partial a} = \sum_{i=0}^{9} (2ai^2 + 2ib - 2ix_i) = 2(285a + 45b - 186)$$

and

$$0 = \frac{\partial e(a,b)}{\partial b} = \sum_{i=0}^{9} (2b + 2ai - 2x_i) = 2(10b + 45a - 50)$$

From second equation, $b = 5 - 4.5a$. Substituting in the first one,

$$0 = 285a + 45(5 - 4.5a) - 186 = 285a + 225 - 202.5a - 186 = 82.5a + 39,$$

hence $a = -\frac{78}{165} \approx -0.47$, and $b = 5 - 4.5a \approx 7.13$.

Hence the forecast for 2020 is $x_{10} = 10a + b = 2.4$, and

$$\lambda = \frac{2.4}{1000} = 0.0024.$$

(iv) Application (Similar to seen)

The residual holding time at state A from time t_0 to t is

$$e^{\int_{t_0}^{t_1}(-\lambda)dt} = e^{-\lambda(t_1 - t_0)}.$$

In particular, the chance to survive from time $t_0 = 0$ to time t is

$$e^{-\lambda(t-0)} = e^{-\lambda t}$$

With $t = 1$ (one year), this reduces to $e^{-\lambda}$. With $\lambda = 0.0024$, the answer is ≈ 0.9976. With $\lambda = 0.003$, the answer is ≈ 0.9970.

(v) Application (Similar to seen)

Let t be the time (measured in years) which passed from the beginning of 2020. Hence $t = 0$ at the beginning of the year and $t = 1$ at the end. The have $\lambda(0) = 0.004$, $\lambda(1) = 0.003$. Because $\lambda(t)$ is a linear function, $\lambda(t) = at + b$. Hence, $\lambda(0) = a \cdot 0 + b = 0.004$ and $b = 0.004$. Also, $\lambda(1) = a \cdot 1 + b = 0.003$, hence, $a = 0.003 - b = -0.001$.

The residual holding time at state A from time 0 to 1 is

$$e^{-\int_0^1(-0.001\lambda+0.004)dt} = e^{0.001(1^2-0^2)/2-0.004(1-0)} = e^{-0.0035} \approx 0.9965.$$

Syllabus cover

This question covers item 3.1 of the syllabus: Describe and classify stochastic processes, item 3.3 Define and apply a Markov process, and item 5 Machine Learning.

Feedback

Parts (i) and (ii) were answered reasonably well, part (iii) was difficult computationally, and in parts (iv) and (v) some students forgot the formula with exponent and wrote answer $1 - \lambda$ instead of $e^{-\lambda}$. This answer happened to be very close numerically for this particular λ, but this is still a mistake.