THE UNIVERSITY OF MELBOURNE

Student Number

Semester 1 Assessment, 2017

School of Mathematics and Statistics

MAST20004 Probability

Writing time: 3 hours

Reading time: 15 minutes

This is NOT an open book exam

This paper consists of 10 pages (including this page)

Authorised Materials

- Mobile phones, smart watches, and internet or communication devices are forbidden.
- Students may bring one double-sided A4 sheet of handwritten notes into the exam room.
- Approved hand-held electronic scientific (but not graphing) calculators may be used.

Instructions to Students

- You must NOT remove this question paper at the conclusion of the exam.
- This paper has 9 questions. Attempt as many questions, or parts of questions, as you can.

  The number of marks allocated to each question is shown in the brackets after the question statement.

  The total number of marks available for this exam is 100.

  There is a table of normal distribution probabilities at the end of this question paper.

  Working and/or reasoning must be given to obtain full credit. Clarity, neatness, and style count.

Instructions to Invigilators

- Students must NOT remove this question paper at the conclusion of the exam.

Blank page (ignored in page numbering)

1. Consider a random experiment with sample space Ω.

   (a) Write down the axioms which must be satisfied by a probability mapping P defined on the events of the experiment.

   (b) Using the axioms, prove that for events A and B where $A \subseteq B$,

$$P(A) \leq P(B).$$

   (c) Using the axioms and part (b), prove that for events A and B,

$$P(A \cup B) \leq P(A) + P(B).$$

[9 marks]

Solution

(a) (3 marks) The axioms are
A1: For events A, $P(A) \geq 0$,
A2: $P(\Omega) = 1$,
A3: For $A_1, A_2, \ldots$ disjoint events,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

(b) (3 marks) If $A \subseteq B$, then $B = A \cup (B \cap A^c)$ and the right hand side is a disjoint union. Therefore by A3 and then A1,

$$P(B) = P(A) + P(B \cap A^c) \geq P(A).$$

(c) (3 marks) $A \cup B = A \cup (B \cap A^c)$, and the right hand side is a disjoint union. Therefore by A3 and then part (b),

$$P(A \cup B) = P(A) + P(B \cap A^c) \leq P(A) + P(B).$$

2. A certain disease affects 3 out of every $1,000$ people in Australia. A test for the disease has a false positive rate of 2%, and a false negative rate of 1% (that is, 2% of those that do not have the disease test positive, and 1% of those that have the disease test negative).

   (a) What proportion of those tested, test positive?

   (b) What is the chance that someone who has had a positive test result will have the disease?

   (c) What is the chance that someone who has had a negative test result will have the disease?

[9 marks]

Solution

Let A be the event that the person tests positive and D the event that they have the disease. We are given $P(D) = 0.003$, $P(A|D) = 0.99$, $P(A|D^c) = 0.02$.

(a) (3 marks) The law of total probability and definition of conditional probability give

$$P(A) = P(A|D)P(D) + P(A|D^c)P(D^c) = 0.99 \cdot 0.003 + 0.02 \cdot 0.997 = 0.02291.$$

(b) (3 marks) The definition of conditional probability implies

$$P(D|A) = \frac{P(A|D)P(D)}{P(A)} = \frac{0.99 \cdot 0.003}{0.02291} = 0.1296.$$

(c) (3 marks) Similar to above,

$$P(D|A^c) = \frac{P(A^c|D)P(D)}{P(A^c)} = \frac{0.01 \cdot 0.003}{1 - 0.02291} = 0.000031.$$

3. Let X and Y have joint probability density function given by

$$f_{(X,Y)}(x,y) = Cxy, \quad 0 < x < y < 1,$$

where C is a constant.

(a) Find the constant C.
(b) Find $f_X(x)$, the marginal density function of X.
(c) Evaluate $P(Y > 3/4|X = 1/2)$.
(d) Evaluate $P(Y > 3/4|X > 1/2)$.
(e) Are X and Y independent? Justify your answer.
(f) Compute $E[(XY)^{-1/2}]$.

[13 marks]

Solution

(a) (2 marks)

$$\int_0^1 \int_0^y Cxydxdy = \frac{C}{2}\int_0^1 y^3 dy = \frac{C}{8}.$$

Therefore $C = 8$.

(b) (2 marks) For $0 < x < 1$,

$$f_X(x) = \int_x^1 8xydy = 4x(1 - x^2).$$

(c) (3 marks) First find $f_{Y|X}(y|x)$, the conditional density function of Y given X. For $0 < x < y < 1$,

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{(X,Y)}(x,y)}{f_X(x)} \\ &= \frac{8xy}{4x(1-x^2)} \\ &= \frac{2y}{1-x^2}. \end{aligned}$$

Now,

$$P(Y > 3/4|X = 1/2) = \frac{8}{3}\int_{3/4}^1 ydy = \frac{7}{12} = 0.5833.$$

(d) (3 marks)

$$\begin{aligned}
P\left(Y>3/4|X>1/2\right) &= \frac{P\left(Y>3/4, X>1/2\right)}{P(X>1/2)} \\
&= \frac{\displaystyle\int_{3/4}^{1}\int_{1/2}^{y} 8xy\,dx\,dy}{\displaystyle\int_{1/2}^{1} 4x(1-x^2)\,dx} \\
&= \frac{\displaystyle\int_{3/4}^{1}(4y^3-y)dy}{(9/16)} \\
&= \frac{(119/256)}{(9/16)} \\
&= \frac{119}{144} = 0.8264.
\end{aligned}$$

(e) (1 mark) They are not independent since the support is not rectangular. (Alternatively note that parts (c) and (d) have different answers; or that $f_{Y|X}$ clearly depends on x.)

(f) (2 marks) We compute

$$E[(XY)^{-1/2}] = \int_0^1\int_0^y 8\sqrt{xy}dxdy = 16/9 = 1.7778.$$

4. For parameter $a>1$, let X have density function $f_X(x) = ax^{-(a+1)}$ on the interval $[1,\infty)$. Let $Y=\sqrt{X}$ and $Z=\log(Y)$.

(a) Find $F_Y(y)$, the distribution function of Y.

(b) Find $f_Y(y)$, the density function of Y, state the values of y for which it is defined, and show that it is a density function.

(c) Compute the expected value of Y, or show that it does not exist.

(d) Compute the variance of Y, or show that it does not exist.

(e) Approximate $E[Z]$ and $Var(Z)$ using suitable Taylor series expansions.

(f) Compute the exact value of $E[Z]$.

[15 marks]

Solution

Note first that $F_X(x) = P(X\le x) = 1-x^{-a}$, for $x\ge 1$, and zero otherwise.

(a) (2 marks) $F_Y(y)=0$ for $y\in(-\infty,1]$. For $y\in(1,\infty)$,

$$\begin{aligned}
F_Y(y) &= P(Y\le y) \\
&= P(\sqrt{X}\le y) \\
&= P(X\le y^2) \\
&= 1-P(X>y^2) \\
&= 1-y^{-2a}.
\end{aligned}$$

(b) (3 marks) Differentiating the expression for $F_Y(y)$ gives, for $y \in (1, \infty)$,

$$f_Y(y) \ = 2ay^{-2a-1}.$$

Clearly, for $y \in (1, \infty)$, $f_Y(y) \geq 0$. Also,

$$\int_1^\infty 2ay^{-2a-1}dy = -y^{-2a}\big|_{y=1}^\infty = 1$$

Thus f_Y is a density function.

(c) (2 marks)

$$E(Y) = \int_1^\infty (2a)y^{-2a}dy = \frac{2a}{2a-1};$$

the integral is finite since $a > 1$.

(d) (2 marks)

$$E(Y^2) = \int_1^\infty (2a)y^{-2a+1}dy = \frac{2a}{2a-2};$$

the integral is finite since $a > 1$. Therefore,

$$Var(Y) = \frac{2a}{2a-2} - \left(\frac{2a}{2a-1}\right)^2 = \frac{2a}{(2a-1)^2(2a-2)}.$$

(e) (4 marks) Use the approximation for functions ψ:

$$E[\psi(Y)] \approx \psi(E[Y]) + Var(Y)\psi''(E[Y])/2$$
$$Var(\psi(Y)) \approx \psi'(E[Y])^2 Var(Y).$$

Here $\psi(y) = \log(y)$ and so $\psi'(y) = y^{-1}$ and $\psi''(y) = -y^{-2}$. Thus

$$E[Z] \approx \log\left(\frac{2a}{2a-1}\right) - \frac{2a}{2(2a-1)^2(2a-2)}\left(\frac{2a-1}{2a}\right)^2 = \log\left(\frac{2a}{2a-1}\right) - \frac{1}{4a(2a-2)},$$

and

$$Var(Z) \approx \left(\frac{2a-1}{2a}\right)^2 \frac{2a}{(2a-1)^2(2a-2)} = \frac{1}{2a(2a-2)}.$$

(f) (2 marks) Note first that for $z > 0$,

$$P(Z > z) = P(Y > e^z) = e^{-2az},$$

so using the formula $E[Z] = \int_0^\infty P(Z > z)dz$ easily shows $E[Z] = 1/(2a)$. (Also can be done directly from $E[Z] = \int \log(y) f_Y(y)dy$.)

5. Let U, V, and W be independent random variables, all distributed uniformly on the interval $(0, 1)$.

   (a) Find the probability density functions for

   (i) $X = \max\{U, V\}$;

   (ii) $Y = \min\{U, V\}$;

   (iii) $Z = \max\{U, V, W\}$.

   (b) Find the joint probability density function of (X, Y).

   (c) Compute the moment generating function of $X + Y$.

[10 marks]

Solution

(a) (i) (1 mark) For $0 < x < 1$, $P(X \leq x) = P(U \leq x, V \leq x) = x^2$. Thus

$$f_X(x) = 2x, \quad 0 < x < 1.$$

(ii) (2 marks) For $0 < y < 1$, $P(X > y) = P(U > y, V > y) = (1 - y)^2$. Thus $F_Y(y) = 1 - (1 - y)^2$ and

$$f_Y(y) = 2(1 - y), \quad 0 < y < 1.$$

(iii) (1 mark) For $0 < z < 1$, $P(Z \leq z) = P(U \leq z, U \leq z, V \leq z) = z^3$. Thus

$$f_Z(z) = 3z^2, \quad 0 < z < 1.$$

(b) (3 marks) For $0 \leq y < x \leq 1$,

$$P(X \leq x, Y > y) = P(y < U \leq x, y < U \leq x) = (x - y)^2.$$

Thus the density is

$$f_{(X,Y)}(x, y) = 2, \quad 0 \leq y < x \leq 1.$$

An alternative derivation is to note the intuitively clear fact that $Y|X = x$ is uniform on $(0, x)$, so that $f_{Y|X}(y|x) = 1/x$ on $0 < y < x$.

(c) (3 marks) From the previous item, for all $t \neq 0$,

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = \int_0^1 \int_0^x 2e^{t(x+y)} dy dx = \frac{2}{t} \int_0^1 e^{tx}(e^{tx} - 1) dx = \frac{(e^t - 1)^2}{t^2}.$$

An alternative derivation could be had by noting $X + Y$ has the same distribution as $U + V$.

6. The price of a stock at the beginning of the trading day is 50 dollars. The price of the same stock one hour into the trading day and four hours into the trading day is $(S_1, S_4) = (50e^X, 50e^Y)$, where (X, Y) is a bivariate normal random variable with mean parameters $(\mu_X, \mu_Y) = (1, 4)$, variance parameters $(\sigma_X^2, \sigma_Y^2) = (2, 8)$, and $Cov(X, Y) = 2$.

   (a) What is the probability the stock has a higher price four hours into the trading day than at the beginning?

   (b) Given the price of the stock four hours into the trading day is $50e^8$ dollars, what is the probability that the price one hour into the trading day was greater than $50e^3$ dollars?

   (c) What is the mean of S_1?

(d) What is the mean of S_4?

(e) What is the covariance of S_1 and S_4?

[15 marks]

Solution Let Z denote a standard normal variable.

(a) (3 marks)

$$P(S_4 > 50) = P(Y > 0) = P\left(\frac{Y-4}{2\sqrt{2}} > -\sqrt{2}\right) = P(Z > -\sqrt{2}) = 0.9207.$$

(b) (3 marks) Noting that $X|Y = 8$ is normal with mean 2 and variance $3/2$,

$$\begin{aligned} P(S_1 > 50e^3 | S_4 = 50e^8) &= P(X > 3 | Y = 8) \\ &= P\left(\frac{X-2}{\sqrt{3/2}} > \frac{1}{\sqrt{3/2}} \Big| Y = 8\right) \\ &= P(Z > 0.82) = 0.2061. \end{aligned}$$

(c) (2 marks) Using the formula $E[e^{tZ}] = e^{t^2/2}$,

$$E[S_1] = 50E[e^X] = 50eE[e^{\sqrt{2}Z}] = 50e^2.$$

(d) (2 marks) Using the formula $E[e^{tZ}] = e^{t^2/2}$,

$$E[S_4] = 50E[e^Y] = 50e^4E[e^{2\sqrt{2}Z}] = 50e^8.$$

(e) (5 marks) The usual decomposition of the bivariate normal implies

$$X = \mu_X + \frac{\sigma_X \rho}{\sigma_Y}(Y - \mu_Y) + \sigma_X\sqrt{1-\rho^2}Z,$$

where ρ is the correlation of X and Y and Z is standard normal, independent of Y. Specialising to our parameters, we have $\rho = 1/2$ and

$$X = \frac{Y}{4} + \sqrt{\frac{3}{2}}Z,$$

and so by independence,

$$E[S_1S_4] = 50^2E[e^{X+Y}] = 50^2E[e^{(5/4)Y+\sqrt{3/2}Z_1}] = 50^2e^5E[e^{(5/\sqrt{2})Z}]E[e^{\sqrt{3/2}Z}] = 50^2e^{12}.$$

Therefore the covariance of S_1 and S_4 is

$$Cov(S_1, S_4) = E[S_1S_4] - E[S_1]E[S_4] = 50^2e^{12} - 50^2e^{10}.$$

7. Consider a roulette wheel with 37 slots, each having a number $0 - 36$. There are 18 red slots, 18 black slots, and 1 green slot. Each roulette game consists of choosing a slot uniformly at random. For a given game, we consider two kinds of bets.

- If you make a bet of $\$w$ on a specific number and the ball lands on that number, then you keep your bet and win an additional $\$35w$. Otherwise you lose your bet (so your winnings are either $-\$w$ or $\$35w$).
- If you make a bet of $\$w$ on "black" and the ball lands on a black slot, then you keep your bet and win an additional $\$w$. Otherwise you lose your bet (so your winnings are either $-\$w$ or $\$w$).

Let W_1 be the winnings from betting \$5 on a single number in 25 consecutive games, and let W_2 be the winnings from betting \$125 on "black" in a single game.

(a) Find $E[W_1]$ and $Var(W_1)$.

(b) Find $E[W_2]$ and $Var(W_2)$.

(c) Compute $P(W_1 > 0)$ using an appropriate approximation.

(d) Compute $P(W_2 > 0)$.

(e) If you had to choose between betting \$5 on a single number in 25 consecutive games and betting \$125 on "black" in a single game, using the information from the previous parts of the question, which would you choose and why?

[11 marks]

Solution

(a) (3 marks) Write $W_1 = \sum_{i=1}^{25} X_i$ where the X_i are the winnings on the ith game. By the description, X_i equals 175 with probability 1/37 and -5 otherwise. Thus

$$E[X_i] = \frac{175 - 180}{37} = -\frac{5}{37} = -0.135135,$$

and

$$Var(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{175^2 + 5^2 \cdot 36}{37} - \frac{25}{37^2} = 852.01.$$

Therefore,

$$E[W_1] = 25E[X_1] = -\frac{125}{37} = -3.378,$$

and by the independence of the X_i,

$$Var(W_1) = 25Var(X_1) = 21300.219.$$

(b) (2 marks) By the description, W_2 equals 125 with probability 18/37, and -125 with probability 19/37. Thus

$$E[W_2] = -\frac{125}{37} = -3.378,$$

and

$$Var(W_2) = (125)^2 - (E[W_2])^2 = 15613.5866.$$

(c) (3 marks) The event $\{W_1 > 0\}$ corresponds to winning at least one of the 25 games. If X is the number of times a game is won, then X is binomial with parameters 25 and 1/37, which is approximately Poisson with mean $\lambda = 25/37$. Therefore

$$P(W_1 > 0) = P(X > 0) = 1 - e^{-25/37} = 0.491187.$$

(d) (1 mark) $P(W_2 > 0) = 18/37 = 0.486486$.

(e) (2 marks) The average winning of the two bets is the same, but betting on a single number 25 times gives a slightly higher chance of walking away with more money than you came with, so that is better.

8. (a) Let U be uniformly distributed on the interval $(0, 1)$ and let X have density function $3(1-x)^2$ on $0 < x < 1$. Find a function ψ so that $\psi(U)$ has the same distribution as X.

   (b) Let X be as in part (a) and let $U_1, U_2, \ldots, U_{1000}$ be independent and each uniformly distributed on the interval $(0, 1)$. Write down a function of $U_1, \ldots, U_{1000}$ which would make a good estimate of $E[e^{X^2}]$.

   (c) Let Y be a Bernoulli random variable with parameter p with $0 < p < 1$. If V is uniformly distributed on the interval $(0, 1)$, find a function ϕ such that $\phi(V)$ has the same distribution as Y.

   (d) Let X be as in part (a) and let the distribution of W given $X = x$ be Bernoulli with parameter x. Let U, V be independent and uniformly distributed on the interval $(0, 1)$. Find a function χ such that $\chi(U, V)$ has the same distribution as W.

[8 marks]

Solution

(a) (2 marks) For $0 < x < 1$, $F_X(x) = (1-x)^3$. We know $F_X^{-1}(U)$ has the same distribution as X, and so

$$\psi(U) = 1 - U^{1/3}$$

has the same distribution as X.

(b) (2 marks) If $X_1, \ldots, X_{1000}$ are independent and distributed as X, then

$$\frac{1}{1000}\sum_{i=1}^{1000} e^{X_i^2} \approx E[e^{X^2}].$$

Therefore from the previous part,

$$E[e^{X^2}] \approx \frac{1}{1000}\sum_{i=1}^{1000} e^{\psi(U_i)^2}.$$

(c) (2 marks) Since $P(V \leq p) = p$, we set, for $0 < v < 1$,

$$\phi(v) = \begin{cases} 1 & v \leq p \\ 0 & v > p. \end{cases}$$

(d) (2 marks) Combining the parts above, we have for $0 < u, v < 1$

$$\chi(u, v) = \begin{cases} 1 & v \leq \psi(u) \\ 0 & v > \psi(u). \end{cases}$$

9. Consider the branching process $\{X_n, n = 0, 1, 2, \ldots\}$ where X_n is the population size of the nth generation. Assume $P(X_0 = 1) = 1$ and that the probability generating function of the offspring distribution is $A(z) = C(1 + 2z + 3z^2)$ for some constant C.

   (a) What is the probability mass function of the offspring distribution?

   (b) Find the expected value of the offspring distribution.

   (c) Find a simple expression for $E[X_n]$.

   (d) If $q_n = P(X_n = 0)$ for $n = 0, 1, \ldots$, write down an equation relating q_n and q_{n+1}.

   (e) Find the extinction probability $q = \lim_{n\to\infty} q_n$.

   (f) Find a simple expression for $P(X_2 > 0)$.

[10 marks]

Solution

(a) (2 marks) Since A is a probability generating function, $1 = A(1) = 6C$. Thus $C = 1/6$. Denote by N a variable having the offspring distribution. By the definition of the probability generating function, the coefficient of z^n is $P(N = n)$.Thus $P(N = 0) = 1/6$, $P(N = 1) = 1/3$, and $P(N = 2) = 1/2$.

(b) (1 mark) $E[N] = 1/3 + 1 = 4/3$.

(c) (1 marks) Since the expectation of N is 4/3 and the sequence $(E[X_n])_{n\geq 0}$ satisfies $E[X_n] = E[N]^n$, we have $E[X_n] = (4/3)^n$.

(d) (1 mark) If A_n denotes the generating function of X_n, then $A_n(0) = q_n$. Moreover, lecture facts imply $A_{n+1} = A(A_n)$ and so

$$q_{n+1} = A(q_n).$$

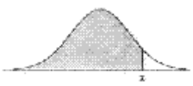
(e) (2 marks) We know from lectures that $\lim_{n\to\infty} q_n = q$ is the minimum non-negative solution of $A(z) = z$. The equation $z = (1/6) + z/3 + z^2/2$, has solutions 1/3 and 1, and so $q = 1/3$.

(f) (3 marks) We know $P(X_2 = 0) = A_2(0) = A(A(0)))$. Note first

$$A_2(z) = \frac{1}{6} + \frac{1}{3}\left(\frac{1}{6} + \frac{1}{3}z + \frac{1}{2}z^2\right) + \frac{1}{2}\left(\frac{1}{6} + \frac{1}{3}z + \frac{1}{2}z^2\right)^2,$$

so that $A_2(0) = 17/72$ and $P(X_2 > 0) = 1 - P(X_2 = 0) = 55/72 = 0.7639$.

Tables of the Normal Distribution



Probability Content from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817