Final Exam Solution

The exam is for exactly 3 hours.
There are 9 problems.
The maximum score is 100 points.
The exam is open book and notes.

1. *(36 = 6 × 6 points)* For each of the following statements, indicate whether you believe that the statement is true or believe it is false, and give a brief explanation of your reasoning. A correct answer without a valid explanation gets 2 points. A correct answer with a valid explanation gets 6 points.

   (a) If A_1, A_2, and A_3 are events such that

   $$P(A_1 \mid A_2) > P(A_1) \text{ and } P(A_2 \mid A_3) > P(A_2)$$

   then

   $$P(A_1 \mid A_3) > P(A_1) .$$

   Solution :
   The statement is FALSE.
   For example, let the sample space be the unit interval with the probability assigned to an event (a subset of the unit interval) equal to its length, and let $A_1 = [\frac{1}{16}, \frac{1}{4})$, $A_2 = [\frac{1}{8}, \frac{1}{2}]$, and $A_3 = [\frac{1}{4}, \frac{9}{16}]$. Then

   $$P(A_1 \mid A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3} > \frac{3}{16} = P(A_1)$$

   and

   $$P(A_2 \mid A_3) = \frac{P(A_2 \cap A_3)}{P(A_3)} = \frac{\frac{1}{4}}{\frac{5}{16}} = \frac{4}{5} > \frac{3}{8} = P(A_2) .$$

   However $P(A_1 \mid A_3) = 0$ and this is less than $P(A_1) = \frac{3}{16}$.

   (b) Let A_1, A_2, and A_3 be events with $0 < P(A_3) < 1$. Suppose A_1 and A_2 are conditionally independent given A_3 and are also conditionally independent given A_3^c. Then A_1 and A_2 are independent.

Solution :

The statement is FALSE.

For example, suppose the sample space is the unit square, i.e $\{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$, with the probability of an event (a subset of the unit square) equal to its area. Let A_3 be defined by

$$A_3 = \{(x, y) \ : \ 0 \leq x \leq \frac{1}{3}\} .$$

Let

$$A_1 = A_3 \cup \{(x, y) \ : \ \frac{1}{3} \leq x \leq 1, \frac{1}{2} \leq y \leq 1\}$$

and let

$$A_2 = \{(x, y) \ : \ 0 \leq x \leq \frac{2}{3}\} .$$

Then $A_1 \cap A_3 = A_2 \cap A_3 = A_3$. Hence

$$P(A_1 \cap A_2 \mid A_3) = 1 = P(A_1 \mid A_3)P(A_2 \mid A_3) ,$$

so A_1 and A_2 are conditionally independent given A_3.
Also, $P(A_1 \cap A_3^c) = P(A_2 \cap A_3^c) = \frac{1}{3}$, $P(A_3^c) = \frac{2}{3}$, and $P(A_1 \cap A_2 \cap A_3^c) = \frac{1}{6}$. Hence

$$P(A_1 \cap A_2 \mid A_3^c) = \frac{\frac{1}{6}}{\frac{2}{3}} = \frac{1}{4} = P(A_1 \mid A_3^c)P(A_2 \mid A_3^c) ,$$

so A_1 and A_2 are conditionally independent given A_3^c.
However, we have $P(A_1) = P(A_2) = \frac{2}{3}$ and $P(A_1 \cap A_2) = \frac{1}{2}$, so A_1 and A_2 are not independent.

(c) Let X be a nonnegative integer valued random variable and let $G_X(z)$ denote its generating function. If $(G_X(z))^2 = G_X(z^2)$ for all z, then X must be a constant.

Solution :

This statement is TRUE.

Let p_n denote $P(X = n)$, $n = 0, 1, 2, \ldots$. Then $G_X(z) = \sum_{n=0}^{\infty} p_n z^n$. Let k denote the smallest integer such that $p_k > 0$. Then the smallest degree term in $G_X(z)$ is z^k and its coefficient is p_k. Hence the smallest degree term in $(G_X(z))^2$ is z^{2k} and its coefficient is p_k^2. However, we also have that the smallest degree term in $G_X(z^2)$ is z^{2k}, and its coefficient is p_k. The given assumption could hold only if $p_k^2 = p_k$. Since $p_k > 0$, it follows that $p_k = 1$. Hence X equals the constant k.

(d) For any random variable X and any $a > 0$ we have

$$P(\mid X \mid < a) \leq a^2 E[\frac{1}{X^2}] \ .$$

Solution :

This statement is TRUE.

Let Y denote $\frac{1}{X}$. Then Chebyshev's inequality gives

$$P(\mid Y \mid > \frac{1}{a}) \leq a^2 E[Y^2] \ .$$

Since $\{\mid Y \mid > \frac{1}{a}\} = \{\mid X \mid < a\}$ this is the same as

$$P(\mid X \mid < a) \leq a^2 E[\frac{1}{X^2}]$$

which was to be shown.

(e) Let X and Y be random variables, and let $Z = 2Y$. Then $E[X \mid Y] = E[X \mid Z]$.

Solution :

This statement is TRUE.

Since $E[X \mid Y]$ is a function of Y, it is also a function of Z (substitute $\frac{Z}{2}$ for Y). The defining characteristic of $E[X \mid Z]$ is the orthogonality principle, which says that the error term $X - E[X \mid Z]$ should satisfy

$$E[(X - E[X \mid Z])g(Z)] = 0$$

for every function g satisfying $E[g(Z)^2] < \infty$. But this is satisfied by $E[X \mid Y]$, because every function of Z is also a function of Y (substitute $2Y$ for Z).

(f) Let X and Y be random variables, and let $Z = X1(X > 2)$. Then $E[X \mid Y] \geq E[Z \mid Y]$.

Solution :

This statement is TRUE.

Since $X \geq Z$ we have $X - Z \geq 0$. Hence $E[X - Z \mid Y] \geq 0$, which, by linearity of conditional expectation, is the same as what was to be shown.

2. *(8 points)* Alice and Bob play the following game. First, each of them independently picks a random number from $\{1, 2, \ldots, 10\}$. If the number picked by Alice is less than or equal to the number picked by Bob, then Alice loses. Otherwise, Alice picks another number at random from $\{1, 2, \ldots, 10\}$, independently of the previously chosen numbers. If the second number picked by Alice is less than or equal to the number picked by Bob, a draw is declared. If the second number picked by Alice is bigger than the number picked by Bob, then Alice wins.

Find the probability that Alice wins.

Solution :

We may pretend that Alice picks a second time even if there is no need to do so. Let A denote the event that Alice wins.

For $1 \leq k \leq 10$, let B_k denote the event that Bob picks the number k. Let C_k denote the event that Alice picks the number k the first time that she picks. Let D_k denote the event that Alice picks the number k the second time that she picks.

Conditioned on B_k, Alice wins on the event $(\cup_{l=k+1}^{10} C_l) \cap (\cup_{m=k+1}^{10} D_m)$. Thus

$$P(A) = \sum_{k=1}^{10} P(B_k)P(A \mid B_k) = \sum_{k=1}^{10} P(B_k)P((\cup_{l=k+1}^{10} C_l) \cap (\cup_{m=k+1}^{10} D_m) \mid B_k) .$$

The choices of Alice are independent of those of Bob. Hence we may write this as

$$P(A) = \frac{1}{10} \sum_{k=1}^{10} P((\cup_{l=k+1}^{10} C_l) \cap (\cup_{m=k+1}^{10} D_m))$$

where we have used $P(B_k) = \frac{1}{10}$ for each $1 \leq k \leq 10$.

The choices of Alice each time she picks are also independent. Hence we may write this as

$$\begin{aligned} P(A) &= \frac{1}{10} \sum_{k=1}^{10} P(\cup_{l=k+1}^{10} C_l) P(\cup_{m=k+1}^{10} D_m) \\ &= \frac{1}{10} \sum_{k=1}^{10} (1 - \frac{k}{10})^2 \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{1000}\sum_{j=1}^{9} j^2 \\
&= \frac{1}{1000}\frac{9(9+1)(18+1)}{6} \\
&= \frac{57}{200} = 0.285
\end{aligned}
$$

where we have used the formula $\sum_{m=1}^{n} m^2 = \frac{n(n+1)(2n+1)}{6}$.

3. *(6 points)* Let $Y \sim Exp(5)$. Let $Z = Y^{12} + 2Y^6$. Find f_Z.

   Solution : The map $z = g(y) = y^{12} + 2y^6$ has range $[0, \infty)$. Every $z > 0$ has two inverse images

$$(\sqrt{1+z} - 1)^{\frac{1}{6}} \text{ and } -(\sqrt{1+z} - 1)^{\frac{1}{6}} .$$

   Also, we have $g'(y) = 12y^{11} + 12y^5$.

   We have $f_Y(y) = 5e^{-5y}1(y \geq 0)$. Hence we need only worry about the nonnegative inverse image of each $z > 0$. Using the Jacobean rule, we have

$$f_Z(z) = \begin{cases} \frac{1}{12((\sqrt{1+z}-1)^{\frac{11}{6}} + (\sqrt{1+z}-1)^{\frac{5}{6}})} 5e^{-5(\sqrt{1+z}-1)^{\frac{1}{6}}} & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases}$$

4. *(6 points)* The *skewness* of a random variable X is defined to be

$$\frac{E[(X-m)^3]}{\sigma^3}$$

   and its *kurtosis* is defined to be

$$\frac{E[(X-m)^4] - 3\sigma^4}{\sigma^4} .$$

   Here m denotes $E[X]$ and σ^2 denotes $Var(X)$.

   Find the skewness and the kurtosis of a $Unif([a, b])$ random variable.

   Solution : From the definitions, we see that the skewness of any random variable X is the same as that of the random variable $X + c$ for any constant c, and the same is true of the kurtosis. Thus, we may assume without loss of generality that we are dealing with mean zero random variables. Namely, if we let $d = \frac{b-a}{2}$, we may assume that we are considering $X \sim Unif([-d, d])$.

Since x^3 is an odd function and the density of X is an even function, we see immediately that the skewness of X is zero.

For the kurtosis, we observe that it is invariant under scaling, i.e. the kurtosis of cX is the same as that of X, for any constant c. Thus, it suffices to consider $X \sim Unif[-\frac{1}{2}, \frac{1}{2}]$. This has variance $\sigma^2 = \frac{1}{12}$ and

$$E[X^4] = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^4 dx = \frac{1}{80} ,$$

so we see that the kurtosis is

$$\frac{\frac{1}{80}}{\frac{1}{144}} - 3 = -1.2 .$$

5. *(6 points)* Random variables X_1, X_2, X_3, X_4 are i.i.d. exponentially distributed random variables, with each having mean 3.

   Let

$$\begin{aligned} U &= X_1 + 2X_2 + 3X_3 + 4X_4 \\ V &= 4X_1 + 3X_2 + 2X_3 + X_4 \end{aligned}$$

   Find the correlation coefficient of U and V.

   Solution :

   If $X \sim Exp(\lambda)$ then $E[X] = \frac{1}{\lambda}$, so in our problem we have $X_1, \ldots, X_4$ are i.i.d. $\sim Exp(\frac{1}{3})$. Also, if $X \sim Exp(\lambda)$ then $Var(X) = \frac{1}{\lambda^2}$, so in our problem, each X_i has variance 9. This gives

$$Var(U) = Var(V) = (1 + 4 + 9 + 16)Var(X_1) = 270 .$$

   As for the covariance, we have

$$Cov(U, V) = (4 + 6 + 6 + 4)Var(X_1) = 180 .$$

   From this, the correlation coefficient of U and V can be computed as

$$\frac{180}{\sqrt{270}\sqrt{270}} = \frac{2}{3} .$$

6. *(6 points)* Random variables X and Y have joint density

$$f_{XY}(x,y) = \begin{cases} \frac{1}{3\pi} & \text{if } 1 \le (x-1)^2 + (y-3)^2 \text{ and } \frac{(x-1)^2}{9} + \frac{9(y-3)^2}{16} \le 1 \\ 0 & \text{otherwise} \end{cases} .$$

Find $E[X \mid Y]$.

Solution :

The joint density of X and Y is uniform on the region between the ellipse centered at $(1,3)$ with major semi-axis of length 3 along the x-axis and minor semi-axis of length $\frac{4}{3}$ along the y-axis, and the circle of radius 1 centered at $(1,3)$.

If we condition on any value of Y, $\frac{5}{3} < y < \frac{13}{3}$ we see that the conditional density of X is symmetric about $x = 1$. Thus the conditional mean of X conditioned on $Y = y$ would be 1.

It follows that $E[X \mid Y] = 1$.

7. *(6 points)* Either explicitly give an example or describe how to construct an example of random variables X, Y, and Z such that X and Y are jointly Gaussian, Y and Z are jointly Gaussian, and X and Z are jointly Gaussian, but X, Y, and Z are not jointly Gaussian.

Solution :

Let

$$h(x,y,z) = \frac{1}{(2\pi)^{\frac{3}{2}}} e^{-\frac{x^2+y^2+z^2}{2}} .$$

This is the joint density of three independent Gaussian random variables each with mean zero and variance 1. Let

$$a(x,y,z) = \begin{cases} 1 & \text{if } x \in [-\frac{1}{2}, \frac{1}{2}],\ y \in [-\frac{1}{2}, \frac{1}{2}], \text{ and } z \in [-\frac{1}{2}, \frac{1}{2}] \\ 0 & \text{otherwise} \end{cases}$$

Now, define

$$f(x,y,z) = h(x,y,z) + \delta \sum_{s_x} \sum_{s_y} \sum_{s_z} (-1)^{s_x+s_y+s_z} a(x-s_x, y-s_y, z-s_z)$$

where the sum is over the eight possibilities given by $s_x = \pm 1$, $s_y = \pm 1$, and $s_z = \pm 1$, and $\delta > 0$ is sufficiently small such that $f(x,y,z) > 0$ for all (x,y,z). Then f is a valid joint density. If random variables X,

Y, and Z have this joint density then they are not jointly Gaussian. However, by integrating out any one of the components one can see that X and Y are jointly Gaussian, Y and Z are jointly Gaussian, and X and Z are jointly Gaussian.

8. *(10 points)* Random variables X, Y, and Z are jointly Gaussian, with each having mean zero, and with covariance matrix

$$\begin{bmatrix} 2 & 1 & 0 \\ 1 & 10 & 3 \\ 0 & 3 & 5 \end{bmatrix} .$$

Let $U = (X-Y)^5$ and let $V = (Y-3Z)^2$. Find the joint density of U and V.

Solution :

Let W_1 denote $X-Y$ and W_2 denote $Y-3Z$. As linear functions of jointly Gaussian random variables, W_1 and W_2 are jointly Gaussian. Each has mean zero, and their covariance matrix is

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 10 & 3 \\ 0 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ 0 & 37 \end{bmatrix} .$$

Since this is a diagonal matrix, W_1 and W_2 are independent. Since U is a function of W_1 (we have $U = W_1^5$) and V is a function of W_2 (we have $V = W_2^2$) it follows that U and V are independent. Hence

$$f_{UV}(u,v) = f_U(u) f_V(v) .$$

To find $f_U(u)$, note that $U = g(W_1)$, where $g(x) = x^5$. This is a one to one map from $\mathbf{R}$ on to $\mathbf{R}$, with $g^{-1}(u) = u^{\frac{1}{5}}$. Also $g'(x) = 5x^4$. Using the Jacobean rule, we have

$$\begin{aligned} f_U(u) &= \frac{1}{\mid g'(u^{\frac{1}{5}}) \mid} f_{W_1}(u^{\frac{1}{5}}) \\ &= \frac{1}{5u^{\frac{4}{5}}} \frac{1}{\sqrt{20\pi}} e^{-\frac{u^{\frac{2}{5}}}{20}} \end{aligned}$$

where we have used the fact that $W_1 \sim N(0, 10)$.

To find $f_V(v)$ note that $V = h(W_2)$, where $h(x) = x^2$. The range of h is $[0, \infty)$ and every point v in the range other than 0 has two inverse images under h, namely $\pm\sqrt{v}$. Also, $h'(x) = 2x$. Using the Jacobean rule, we have

$$\begin{aligned} f_V(v) &= \frac{1}{\mid h'(\sqrt{v}) \mid} f_{W_2}(\sqrt{v}) + \frac{1}{\mid h'(-\sqrt{v}) \mid} f_{W_2}(-\sqrt{v}) \\ &= \frac{1}{\sqrt{v}} \frac{1}{\sqrt{74\pi}} e^{-\frac{v}{74}} \end{aligned}$$

where we have used the fact that $W_2 \sim N(0, 37)$.

9. *(4 + 4 + 2 + 6 points)*

Six players, numbered $1, 2, \ldots, 6$, play a game involving successive independent rolls of a fair die, as follows. At any given time a subset of the players is considered "active" and the others are considered "dead". The die is rolled. If the die comes up i and if player i is currently active, then player i becomes dead. The status of the other players does not change, i.e. other active players continue to stay active and other dead players continue to stay dead. If the die comes up i and if player i is currently dead, then player i becomes active. The status of the other players does not change.

The initial condition of the game (which players are active at time 0) is assumed to be independent of the rolls of the fair die.

(a) Describe the evolution of the system as a discrete time Markov chain. The state space you choose should be detailed enough that the status of each player at any time can be determined from the state at that time. Use the convention that the status of a player at time n is the status prior to the roll of the die at time n.

(b) Determine the transition probability matrix of the Markov chain.

(c) Is the Markov chain irreducible ?

(d) Determine all possible stationary probability distributions for the Markov chain.

Solution :

(a) We may choose as state space for the Markov chain the set of all subsets of $\{1, 2, \ldots, 6\}$. Thus the state space has cardinality

$2^6 = 64$. State $A \subset \{1, 2, \ldots, 6\}$ is meant to indicate that the players in A are active at time n and the players not in A are dead (by convention, the status of a player at time n is taken to be the one before the roll of the fair die at time n). Additional information about the state of the system prior to time n does not affect the conditional distribution of events relating to the future after time n given the state at time n, because the evolution after time n depends only on the rolls of the die at times $n, n+1, \ldots$, and these are independent of the state at times n and before. Hence if X_n denotes the state at time n then $(X_n, n \geq 0)$ is a Markov chain.

(b) If the current state is $A = \emptyset$ then the next state is $\{i\}$ with probability $\frac{1}{6}$ for each $1 \leq i \leq 6$. If the current state is $A = \{1, 2, \ldots, 6\}$ then the next state is $\{1, 2, \ldots, 6\} - \{i\}$ with probability $\frac{1}{6}$ for each $1 \leq i \leq 6$.

If the current state is A with $1 \leq |A| \leq 5$, then the next state is $A - \{i\}$ for each $i \in A$ with probability $\frac{1}{6}$ and it is $A \cup \{i\}$ for each $i \notin A$ with probability $\frac{1}{6}$.

This completely determines the transition probability matrix of the Markov chain $(X_n, n \geq 0)$.

(c) Every state communicates with every other state, as can be seen, for instance, by observing that every state communicates with the state $\emptyset$. Hence the Markov chain is irreducible.

(d) Since the Markov chain is irreducible, it has a unique stationary distribution. Let $\pi(A)$ note the stationary distribution of state A. By symmetry this can only depend on $|A|$. Thus there are some numbers $\pi_k > 0$, $k = 0, 1, \ldots, 6$ such that

$$\pi(A) = \pi_k \text{ if } |A| = k ,$$

and such that

$$\sum_{k=0}^{6} \binom{6}{k} \pi_k = 1 . \tag{1}$$

Probability balance in stationarity between states of size k and states of size $k+1$, for $0 \leq k \leq 5$, gives

$$\binom{6}{k} \pi_k \frac{6-k}{6} = \binom{6}{k+1} \pi_{k+1} \frac{k+1}{6} .$$

This simplifies to $\pi_k = \pi_{k+1}$. Hence $\pi_0 = \pi_1 = \ldots = \pi_6$. Denoting this common number by π and substituting in equation (1) gives

$$\pi(\sum_{k=0}^{6} \binom{6}{k}) = 1 ,$$

i.e. $\pi = 2^{-6}$.

So we find that all states have equal probability in stationarity and this is 2^{-6}.