E1 222 Stochastic Models and Applications
Test I

Time: 90 minutes | Max. Marks:40

Date: 21 Sept 2019

Answer **ALL** questions. All questions carry equal marks
Answers should be written only in the space provided.

1. a. A coin, whose probability of heads is p, is tossed repeatedly till we get a head. Calculate the probability that the number of tosses needed is odd.

Answer: Probability that it takes k tosses is given by $(1-p)^{k-1}p$, $k = 1, 2, \cdots$. We need to sum this over all odd k to get the required probability. We take $k = 2m+1$ and sum over $m = 0, 1, \cdots$, to sum over all odd k. Hence the required probability is

$$\sum_{m=0}^{\infty} (1-p)^{2m} p = \frac{p}{1-(1-p)^2} = \frac{1}{2-p}$$

b. A point S is chosen at random on a line segment AB. Show that the probability that the length of AS divided by the length of SB is smaller than a is $\frac{a}{1+a}$ (where we take $a > 0$). Also show that the probability that the length of the longer segment is greater than three times the length of the shorter segment is 0.5.

Answer: Let $|AS|$ denote length of line segment AS and similarly for others. Then we have

$$\frac{|AS|}{|SB|} < a \;\Rightarrow\; \frac{|AS|}{|AB| - |AS|} < a \;\Rightarrow\; \frac{|AS|}{|AB|} < \frac{a}{1+a}$$

Since the point S is uniformly distributed over line segment AB, the above shows that the probability of $\frac{|AS|}{|SB|} < a$ is $\frac{a}{1+a}$.
Let x denote the length of one segment. Then $L - x$ would be the length of the other segment (assuming L is length of AB). So we need either $x > 3(L-x)$, (which is same as $x > \frac{3}{4}L$) or $(L-x) > 3x$ (which is same as $x < \frac{1}{4}L$). Hence the required probability is $0.25 + 0.25 = 0.5$.

We can also do this more formally using the language of random variables. Let length of AB be L. Let the length of AS be denoted by random variable X. Then S being chosen randomly on line segment AB means X is uniform over $[0,\ L]$. Now for the first part we have

$$P\left[\frac{X}{L-X} < a\right] = P\left[X < \frac{a}{1+a}L\right] = \left(\frac{a}{1+a}L\right)/L = \frac{a}{1+a}$$

For the second part we have

$$\begin{aligned}
P[\max(X, L-X) > 3\min(X, L-X)] &= P\left[X < \frac{L}{2}, \max(X, L-X) > 3\min(X, L-X)\right] \\
&\quad + P\left[X \geq \frac{L}{2}, \max(X, L-X) > 3\min(X, L-X)\right] \\
&= P\left[X < \frac{L}{2}, L-X > 3X\right] + P\left[X \geq \frac{L}{2}, X > 3(L-X)\right] \\
&= P\left[X < \frac{L}{2}, X < \frac{L}{4}\right] + P\left[X \geq \frac{L}{2}, X > \frac{3L}{4}\right] \\
&= 0.25 + 0.25 = 0.5
\end{aligned}$$

(Here we have used the obvious facts that when $X < \frac{L}{2}$, we have $\max(X, L-X) = L - X$ and so on).

2. a. Let X be a continuous random variable with density function

$$f_X(x) = K(1-x)^3,\ 0 \leq x \leq 1$$

Let $Y = X^2$. Find $P[X > 0.5]$, $F_X(x)$, EY and variance of X.

Answer: Since f_X has to be a density, we need $\int_0^1 K(1-x)^3\, dx = 1$ which gives $K = 4$. Since density is zero outside $[0,\ 1]$, we have $F_X(x) = 0$ for $x < 0$ and $F_X(x) = 1$ for $x > 1$. For, $0 \leq x \leq 1$, we have

$$F_x(x) = \int_0^x 4(1-x)^3\, dx = 1 - (1-x)^4$$

Hence, $P[X > 0.5] = 1 - F_X(0.5) = (1-0.5)^4 = 1/16$.
We have $EX = \int_0^1 4x(1-x)^3\, dx$. We can solve this using integration by parts or by writing it as $\int_0^1 4x(1-x^3-3x+3x^2)\, dx$. This gives us $EX = \frac{1}{5}$.

We have $EX^2 = \int_0^1 4x^2(1-x)^3\ dx = \int_0^1 4x^2(1-x^3-3x+3x^2)\ dx$. We get $EX^2 = \frac{1}{15}$.

Hence $EY = EX^2 = \frac{1}{15}$ and $\text{Var}(X) = EX^2 - (EX)^2 = \frac{1}{15} - \frac{1}{25} = \frac{2}{75}$.

b. Consider a random experiment with $\Omega = \{(x,y) \in \Re^2 : -1 \leq x, y \leq +1\}$. Note that each outcome of this experiment is a point in a square of side 2 centered at the origin. This experiment is repeated till the outcome or the point obtained is such that the distance of the point from the origin is less than or equal to one. Let the random variable X denote the number of repetitions needed. What is the probability mass function of X?

Answer: The distance of the point from origin being less than or equal to 1 is same as the point being in or on the circle with center at origin and radius 1. This circle is completely in the unit square. Its area is π. The area of the square is 4. Hence the probability of the point falling in the circle on any repetition of the experiment is $\pi/4$. We would need k repetitions if for the first $k-1$ times the point does not fall in the circle and the k^{th} time it does. Hence we have $f_X(k) = (1-\pi/4)^{k-1}(\pi/4)$, $k = 1, 2, \cdots$.

Comment: Suppose (X, Y) is a two dimensional random vector whose density is constant over the unit disc (that is the region in and on the unit circle) and is zero outside. That is, (X, Y) is uniform over the unit disc. Suppose we want to generate such a two dimensional random vector uniformly distributed over the unit disc. We have access only to a random variable that is uniform over $[0,\ 1]$. Does this problem suggest a method of generating (X, Y) uniform over the unit disc?

3. a. Let X be a continuous random variable that is uniform over $[a,\ b]$. Let $Y = (X-c)^2$ where $\frac{a+b}{2} < c < b$. Find the probability density function of Y.

Answer: The density of X is given as

$$f_x(x) = \frac{1}{(b-a)}, \quad a \leq x \leq b$$

The given condition of c implies that c is closer to b than to a and hence we have $(a-c)^2 > (b-c)^2$. Hence the range of Y is 0 to $(a-c)^2$. For $y \geq 0$, we have

$$\begin{aligned} F_Y(y) &= P[(X-c)^2 \leq y] \\ &= P[-\sqrt{y} \leq (X-c) \leq \sqrt{y}] \\ &= P[c-\sqrt{y} \leq X \leq c+\sqrt{y}] \\ &= F_X(c+\sqrt{y}) - F_X(c-\sqrt{y}) \end{aligned}$$

By differentiating this, we get density of Y as

$$f_Y(y) = \frac{1}{2\sqrt{y}} \left(f_X(c+\sqrt{y}) + f_X(c-\sqrt{y}) \right)$$

Given the condition on c, for $y > (a-c)^2$, $c - \sqrt{y} < a$ and $c + \sqrt{y} > b$. So, $f_X(c+\sqrt{y}) = f_X(c-\sqrt{y}) = 0$ and hence the density of Y is zero in this range, as we already know. Since c is closer to b than a, when $(b-c)^2 \leq y \leq (a-c)^2$, $c+\sqrt{y} > b$ and hence $f_X(c+\sqrt{y}) = 0$. Putting all this together, we can now write density of Y as

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} \left(\frac{1}{b-a} + \frac{1}{b-a} \right) = \frac{1}{\sqrt{y}(b-a)}, \quad 0 < y < (b-c)^2 \\ &= \frac{1}{2\sqrt{y}} \left(0 + \frac{1}{b-a} \right) = \frac{1}{2\sqrt{y}(b-a)}, \quad (b-c)^2 \leq y \leq (a-c)^2 \end{aligned}$$

Comment: Could we have used the formula given in class for finding density of a function of a continuous random variable? The answer is 'No' because the given function is not monotone. (If you take $h(x) = (x-c)^2$ then $h'(x) = 2(x-c)$ and since x takes values that may be greater than or less than c, the derivative can be both positive and negative which means h is not monotone). We know that squaring is not an invertible function because it is not one-to-one. (E.g., both $+1$ and -1 map to $+1$). In class we showed that if $Y = X^2$ (and X is continuous) then $f_Y(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y}))$, $y > 0$. This is essentially same as what we derived above. If we wrongly thought $h(x) = x^2$ is invertible with $h^{-1}(y) = \sqrt{y}$ then we would derive $f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y})$ which is wrong.

b. Let X be a continuous random variable with density: $f_X(x) = 3e^{-3x},\ x > 0$. Let $Y = 1 - e^{-3X}$. Find the density of Y.

Answer: A simple calculation (or recall from memory of the exponential distribution function) gives us $F_X(x) = 1 - e^{-3x},\ x > 0$. Since density of X is zero for negative values, we know $X \geq 0$. Hence $0 \leq e^{-3X} \leq 1$ and hence range of Y is 0 to 1. Take $y \in [0,\ 1]$. Now we have

$$\begin{aligned}
F_Y(y) &= P[1 - e^{-3X} \leq y] \\
&= P[e^{-3X} \geq 1 - y] \\
&= P[-3X \geq \ln(1-y)] \\
&= P[X \leq -\frac{1}{3}\ln(1-y)] \\
&= F_X\left(\frac{-1}{3}\ln(1-y)\right) \\
&= 1 - e^{-3(\frac{-1}{3}\ln(1-y))} = 1 - (1-y) = y
\end{aligned}$$

This shows that Y is uniform over $[0,\ 1]$.

Comment: Suppose F is a continuous strictly monotonically increasing distribution function. In the class we showed the following: if X is uniform over $[0,\ 1]$ and $Y = F^{-1}(X)$ then Y has distribution F. As we saw, this has applications, e.g., for random number generation. This problem shows the following: if X has distribution F and $Y = F(X)$ then Y is uniform over $[0,\ 1]$. This is also useful in applications. One well-known application of this is what is called histogram equalization in image processing.

In this problem we could have used the formula because the function is monotone. Take $h(x) = 1 - e^{-3x}$. This is monotone. (We know e^x is monotone. Even otherwise, we can see $h'(x) = 3e^{-3x}$ which is positive for all x). Simple algebra shows that $h^{-1}(y) = \frac{-1}{3}\ln(1-y)$. Derivative of this is $\frac{1}{3}\frac{1}{1-y}$. Hence we get

$$f_Y(y) = 3e^{-3(\frac{-1}{3}\ln(1-y))}\frac{1}{3}\frac{1}{1-y} = e^{\ln(1-y)}\frac{1}{1-y} = 1$$

Note that our formula does not automatically tell us the range of y for which $f_Y(y)$ is given by the above. From the function h, we can see that the range is [0, 1] and hence that is the range over which density of Y can be nonzero.

4. a. Let X, Y be continuous random variables with joint density

$$f_{XY}(x,y) = 6y, \ \ 0 < y < x < 1$$

Find the marginal densities and $P[X > 0.75 \mid Y = 0.5]$.

Answer: The marginal density for X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dy = \int_0^x 6y\ dy = 3x^2, \ \ 0 < x < 1$$

The marginal density for Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dx = \int_y^1 6y\ dx = 6y(1-y), \ \ 0 < y < 1$$

The conditional density of X given Y is given by

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{6y}{6y(1-y)} = \frac{1}{1-y}, \ \ 0 < y < x < 1$$

(Thus, conditioned on Y, X is uniform from Y to 1).
Now, we can get the conditional probability as

$$P[X > 0.75 \mid Y = 0.5] = \int_{0.75}^{1} f_{X|Y}(x|0.5)\ dx = \int_{0.75}^{1} \frac{1}{1-0.5}\ dx = 0.5$$

b. Suppose two students A and B are solving a problem. The times taken by the two students are independent random variables. The time taken by A is uniformly distributed in [1, 3] while that by B is uniformly distributed in [2, 4]. What is the probability that A takes longer than B to solve the problem.

Answer: For any two independent random variables, X, Y,

$$P[X > Y] = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_X(x) f_Y(y)\ dy\ dx = \int_{-\infty}^{\infty} f_X(x) \int_{-\infty}^{x} f_Y(y)\ dy\ dx$$

We can take X to be time taken by A and Y to be time taken by B. Then, X is uniform over [1, 3]. Hence its density is zero outside [1, 3]. So, now we can write

$$P[X > Y] = \int_1^3 \frac{1}{2} \int_{-\infty}^{x} f_Y(y)\ dy\ dx$$

We are given that Y is uniform over $[2,\ 4]$. Hence, in the above, if $x < 2$ then in the y-integral, throughout the range, $f_Y(y)$ is zero. Hence we get

$$P[X > Y] = \int_2^3 \frac{1}{2} \int_2^x \frac{1}{2}\, dy\, dx = \frac{1}{8}$$

We can also solve this using geometry. Here (X, Y) is uniform over $[1,\ 3] \times [2,\ 4]$. So, actually we can think of this as Ω and take probability of any subset to be area of the subset divided by 4. The subset we want is $\{(x, y) \in \Omega : x > y\}$. It is easily seen to be a triangle with area 0.5.