Math 431 An Introduction to Probability

Final Exam — Solutions

1. A continuous random variable X has cdf

$$F(x) = \begin{cases} a & \text{for } x \leq 0, \\ x^2 & \text{for } 0 < x < 1, \\ b & \text{for } x \geq 1. \end{cases}$$

(a) Determine the constants a and b.

(b) Find the pdf of X. Be sure to give a formula for $f_X(x)$ that is valid for *all* x.

(c) Calculate the expected value of X.

(d) Calculate the standard deviation of X.

Answer:

(a) We must have $a = \lim_{x\to-\infty} = 0$ and $b = \lim_{x\to+\infty} = 1$, since F is a cdf.

(b) For all $x \neq 0$ or 1, F is differentiable at x, so

$$f(x) = F'(x) = \begin{cases} 2x & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(One could also use any f that agrees with this definition for all $x \neq 0$ or 1.)

(c) $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^1 x \cdot 2x\, dx = \frac{2}{3}$.

(d) $E(X^2) = \int_0^1 x^2 \cdot 2x\, dx = \frac{1}{2}$, $\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18} \approx 0.0556$, and $\sigma(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{1}{18}} = \frac{1}{3\sqrt{2}} \approx .2357$.

2. Suppose the number of children born in Ruritania each day is a binomial random variable with mean 1000 and variance 100. Assume that the number of children born on any particular day is independent of the numbers of children born on all other days. What is the probability that on at least one day this year, fewer than 975 children will be born in Ruritania?

Answer: We approximate the number of children born each day by a normal random variable. Letting X denote the number of children born on some specified day, and Z denote a standard normal, we have $P(X \geq 975) = P(X \geq 974.5) = P(\frac{X-1000}{10} \geq \frac{974.5-1000}{10} = -2.55) = P(Z \leq +2.55) \approx .9946$. Since each such random variable X (one for each day) is assumed independent of the others, the probability that 975 or more children will be born on every day of this year is $.9946^{365} \approx .1386$, and the probability that, on at least one day this year, fewer than 975 children will be born is close to $1 - .1386 \approx 86\%$.

3. Suppose that the time until the next telemarketer calls my home is distributed as an exponential random variable. If the chance of my getting such a call during the next hour is .5, what is the chance that I'll get such a call during the next two hours?

Answer: First solution: Letting λ denote the rate of this exponential random variable X, we have $.5 = F_X(1) = 1 - e^{-\lambda}$, so $\lambda = \ln 2$ and $F_X(2) = 1 - e^{-2\lambda} = 1 - (e^{-\lambda})^2 = 1 - (.5)^2 = .75$. Second solution: We have $P(X \leq 2) = P(X \leq 1) + P(1 < X \leq 2)$. The first term is .5, and the second can be written as $P(X > 1 \text{ and } X \leq 2) = P(X > 1)P(X \leq 2 \mid X > 1)$. The first of these factors equals $1 - P(X \leq 1) = 1 - .5 = .5$, and the second (by virtue of the memorylessness of the exponential random variable) equals $P(X \leq 1) = .5$. So $P(X \leq 2) = .5 + (.5)(.5) = .75$.

4. Suppose X is uniform on the interval from 1 to 2. Compute the pdf and expected value of the random variable $Y = 1/X$.

Answer: We have

$$f_X(x) = \begin{cases} 1 & \text{if } 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

Putting $g(t) = 1/t$ we have $Y = g(X)$; since g is monotone on the range of X with inverse function $g^{-1}(y) = \frac{1}{y}$, Theorem 7.1 tells us that

$$f_Y(y) = \begin{cases} 1 \cdot |\frac{d}{dy}\frac{1}{y}| = \frac{1}{y^2} & \text{if } \frac{1}{2} < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(Check: $\int_{-\infty}^{\infty} f_Y(y)\,dy = \int_{1/2}^{1} \frac{1}{y^2}\,dy = 1$.) We have $E(Y) = \int_{-\infty}^{\infty} y\,f_Y(y)\,dy = \int_{1/2}^{1} \frac{1}{y}\,dy = \ln 2$. (Check: $E(1/X) = \int_{-\infty}^{\infty} \frac{1}{x} \cdot f_X(x)\,dx = \int_1^2 \frac{1}{x}\,dx = \ln 2$.)

5. I toss 3 fair coins, and then re-toss all the ones that come up tails. Let X denote the number of coins that come up heads on the first toss, and let Y denote the number of re-tossed coins that come up heads on the second toss. (Hence $0 \leq X \leq 3$ and $0 \leq Y \leq 3 - X$.)

(a) Determine the joint pmf of X and Y, and use it to calculate $E(X+Y)$.

(b) Derive a formula for $E(Y|X)$ and use it to compute $E(X+Y)$ as $E(E(X+Y|X))$.

Answer:

(a) $P(X=j, Y=k)$ equals $P(X=j)P(Y=k \mid X=j) = \binom{3}{j}(\frac{1}{2})^j(\frac{1}{2})^{3-j}\binom{3-j}{k}(\frac{1}{2})^k(\frac{1}{2})^{3-j-k} = \binom{3}{j}(\frac{1}{2})^3\binom{3-j}{k}(\frac{1}{2})^{3-j} = \binom{3}{j}\binom{3-j}{k}(\frac{1}{2})^{6-j}$ whenever $0 \le j \le 3$ and $0 \le k \le 3-j$ (and equals zero otherwise), so the joint pmf $f = f_{X,Y}$ has the following values:

$$f(0,0) = \frac{1}{64},\ f(0,1) = \frac{3}{64},\ f(0,2) = \frac{3}{64},\ f(0,3) = \frac{1}{64},\ f(1,0) = \frac{3}{32},$$

$$f(1,1) = \frac{6}{32},\ f(1,2) = \frac{3}{32},\ f(2,0) = \frac{3}{16},\ f(2,1) = \frac{3}{16},\ f(3,0) = \frac{1}{8}.$$

Hence $E(X+Y) = 0\cdot\frac{1}{64} + 1\cdot(\frac{3}{64}+\frac{3}{32}) + 2\cdot(\frac{3}{64}+\frac{6}{32}+\frac{3}{16}) + 3\cdot(\frac{1}{64}+\frac{3}{32}+\frac{3}{16}+\frac{1}{8}) = \frac{9}{4}$. (Alternatively: $X+Y$ is the total number of coins that come up heads on the first toss or, failing that, heads on the re-toss. Each of the three coins has a $\frac{3}{4}$ chance of contributing 1 to this total, so by linearity of expectation, the expected value of the total is $\frac{3}{4}+\frac{3}{4}+\frac{3}{4} = \frac{9}{4}$.

(b) For each fixed x ($0 \le x \le 3$), when we condition on the event $X = x$, Y is just a binomial random variable with $p = \frac{1}{2}$ and $n = 3-x$, and therefore with expected value $pn = \frac{1}{2}(3-x)$. Hence $E(Y|X) = \frac{1}{2}(3-X)$ and $E(X+Y) = E(E(X+Y|X)) = E(E(X|X)+E(Y|X)) = E(X+\frac{1}{2}(3-X)) = E(\frac{1}{2}X+\frac{3}{2}) = \frac{1}{2}E(X)+\frac{3}{2} = \frac{1}{2}\cdot\frac{3}{2}+\frac{3}{2} = \frac{9}{4}$.

6. Let the continuous random variables X, Y have joint distribution

$$f_{X,Y}(x,y) = \begin{cases} 1/x & \text{if } 0 < y < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Compute $E(X)$ and $E(Y)$.

(b) Compute the conditional pdf of Y given $X = x$, for all $0 < x < 1$.

(c) Compute $E(Y|X=x)$ for all $0 < x < 1$, and use this to check your answers to part (a).

(d) Compute $\mathrm{Cov}(X,Y)$.

Answer:

(a) $E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot f_{X,Y}(x,y)\,dy\,dx = \int_0^1\int_0^x x\cdot\frac{1}{x}\,dy\,dx = \int_0^1 x\,dx = \frac{1}{2}$.
$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y \cdot f_{X,Y}(x,y)\,dy\,dx = \int_0^1\int_0^x y\cdot\frac{1}{x}\,dy\,dx = \int_0^1 \frac{1}{2}x\,dx = \frac{1}{4}$.

(b) The marginal pdf for X is $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$, which equals $\int_0^x \frac{1}{x}\,dy = 1$ for $0 < x < 1$ (and equals zero otherwise). That is, X is uniform on the interval from 0 to 1. Hence for each $0 < x < 1$, the conditional pdf for Y given $X = x$ is $f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x)$, which is $\frac{1}{x}$ for $0 < y < x$ and 0 otherwise.

(c) $E(Y|X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\,dy = \int_0^x \frac{y}{x}\,dy = \frac{1}{2}x$. (We can also derive this answer from the fact that the conditional distribution of Y given $X = x$ was shown in (b) to be uniform on the interval $(0, x)$, and from the fact that the expected value of a random variable that is uniform on an interval is just the midpoint of the interval.) To check the formula for $E(Y|X)$, we re-calculate $E(Y) = E(E(Y|X)) = E(\frac{1}{2}X) = \frac{1}{2}E(X)$, which agrees with $E(X) = \frac{1}{2}$, $E(Y) = \frac{1}{4}$.

(d) $E(XY) = \int_0^1 \int_0^x xy \cdot \frac{1}{x}\,dy\,dx = \int_0^1 \frac{1}{2}x^2\,dx = \frac{1}{6}$, so $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{6} - (\frac{1}{2})(\frac{1}{4}) = \frac{1}{24}$.

7. I repeatedly roll a fair die. If it comes up 6, I instantly win (and stop playing); if it comes up k, for any k between 1 and 5, I wait k minutes and then roll again. What is the expected elapsed time from when I start rolling until I win? (Note: If I win on my first roll, the elapsed time is zero.)

Answer: Let T denote the (random) duration of the game, and let X be the result of the first roll. Then $E(T) = E(E(T|W)) = \frac{1}{6}(E(T|W = 1) + E(T|W = 2) + \ldots + E(T|W = 5) + E(T|W = 6)) = \frac{1}{6}((E(T) + 1) + (E(T) + 2) + \ldots + (E(T) + 5) + 0) = \frac{1}{6}(5E(T) + 15)$, so $6E(T) = 5E(T) + 15$ and $E(T) = 15$ (minutes).

8. Suppose that the number of students who enroll in Math 431 each fall is known (or believed) to be a random variable with expected value 90. It does not appear to be normal, so we cannot use the Central Limit Theorem.

(a) If we insist on being 90% certain that there will be no more than 35 students in each section, should UW continue to offer just three sections of Math 431 each fall, or would our level of aversion to the risk of overcrowding dictate that we create a fourth section?

(b) Repeat part (a) under the additional assumption that the variance in the enrollment level is known to be 20 (with *no* other additional assumptions).

Answer:

(a) Since we do not know the variance, the best we can do is use Markov's inequality: $P(X \geq 106) \leq \frac{90}{106} \approx .85$; this is much bigger than .10, so to be on the safe side we should create a fourth section.

(b) Here we know the variance, but since normality is not assumed, we cannot use the Central Limit Theorem; we should use a two-sided or (better still) a one-sided

Chebyshev inequality. The two-sided inequality gives us $P(X \geq 106) \leq P(|X-90| \geq 16) \leq \frac{\sigma^2}{16^2} = \frac{20}{256} \approx .078 < .10$, so we're on the safe side with just three classes. (Or we could use the one-sided Chebyshev inequality: $P(X \geq 106) \leq P(X - 90 \geq 16) \leq \frac{\sigma^2}{\sigma^2+16^2} = \frac{20}{20+256} \approx .072$.)

9.

(a) A coin is tossed 50 times. Use the Central Limit Theorem (applied to a binomial random variable) to estimate the probability that fewer than 20 of those tosses come up heads.

(b) A coin is tossed until it comes up heads for the 20th time. Use the Central Limit Theorem (applied to a negative binomial random variable) to estimate the probability that more than 50 tosses are needed.

(c) Compare your answers from parts (a) and (b). Why are they close but not exactly equal?

Answer:

(a) The number of tosses that come up heads is a binomial random variable, which can be written as a sum of 50 independent indicator random variables. Since 50 is a reasonably large number, it makes sense to use the Central Limit Theorem, and to approximate X (the number of heads in 50 tosses) by a Gaussian with mean $np = 50 \cdot \frac{1}{2} = 25$ and variance $np(1-p) = 50 \cdot \frac{1}{2} \cdot \frac{1}{2} = 12.5$. So $P(X < 20) = P(X \leq 19.5) = P(\frac{X-25}{\sqrt{12.5}} \leq \frac{19.5-25}{\sqrt{12.5}} = -\frac{5.5}{\sqrt{12.5}}) = P(Z \geq \frac{5.5}{\sqrt{12.5}})$, where Z is a standard Gaussian; using $\frac{5.5}{\sqrt{12.5}} \approx 1.56$, we have $P(X < 20) \approx 1 - \Phi(1.56) \approx 6\%$.

(b) The waiting time until the 20th heads-toss is a negative binomial random variable, which can be written as a sum of 20 independent geometric random variables. 20 is a decent-sized number, so, as in part (a), we may apply the Central Limit Theorem and approximate W (the number of tosses required to get heads 20 times) by a Gaussian with mean $\frac{r}{p} = \frac{20}{1/2} = 40$ and variance $\frac{r(1-p)}{p^2} = \frac{20(1/2)}{(1/2)^2} = 40$. So $P(W > 50) = P(W \geq 50.5) = P(\frac{W-40}{\sqrt{40}} \geq \frac{50.5-40}{\sqrt{40}} = \frac{10.5}{\sqrt{40}}) = P(Z \geq \frac{10.5}{\sqrt{40}} \approx 1.66) \approx 1 - \Phi(1.66) \approx 5\%$.

(c) Suppose the coin is tossed until it has been tossed at least 50 times *and* heads has come up at least 20 times. Then the outcomes for which $X < 20$ are precisely those for which $W > 50$, so the two events have equal probability. The reason we did not get the exact same answers in parts (a) and (b) is that the Central Limit Theorem is only an approximation, and when specific numbers are used there is likely to be some error.