Exam 1 (Take-Home)

Statistics 310

Oct. 12, 1999

Directions

1. This is an open book open – note exam. This refers to your notes only – there is to be no sharing of notes during the exam. You should have your calculator available. There is to be no sharing of calculators during the exam.
2. There are 110 points on the exam. The value of each question is in square brackets after the problem number, and each subsection is similarly labelled. Any points in excess of 100 count as 1/2 point, so the maximum possible score is 105.
3. If you need extra paper to complete a problem, feel free to staple on that which you use.
4. You may take up to 2.5 hours to work this exam. Indicate time started and time finished on the front page of the exam.
5. This exam is due back in class on Thursday, October 14.
6. **Pledge:**

1. [20 points] A box contains 3 coins, one of which is a two-headed coin (the other two are normal). A coin is chosen at random from the box and the chosen coin is tossed three times.

(a) (10pts) What is the probability of obtaining three heads?

Let A be the event that 3 heads are thrown, and B the event that the two-headed coin is drawn. Then, using the Law of Total Probability, we can write

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Here, the chance of drawing the two-headed coin, $P(B) = 1/3$. If the two-headed coin is used, we will always get 3 heads, so $P(A|B) = 1$. If a normal coin is used, the chance of getting 3 heads is $P(A|B^c) = (1/2)^3 = 1/8$. Thus,

$$P(A) = 1 * \frac{1}{3} + \frac{1}{8} * \frac{2}{3} = \frac{5}{12}.$$

(b) (10pts) If a head turns up all three times, what is the probability that we are dealing with the two-headed coin?

Here, we are being asked to find $P(B|A)$, when it is easier to figure things out the other way around. To do this, we will invoke Bayes' theorem, which says that

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}.$$

The denominator is what we found in part (a), so

$$P(B|A) = \frac{1 * (1/3)}{5/12} = \frac{4}{5}.$$

2. [20 points] A man pays \$1 a throw to try to win a Kewpie doll. His probability of winning on each throw is 0.1, and the outcomes of the throws are independent.

(a) (3pts) If we let X denote the number of the toss on which the doll is won, what is the distribution of X? What are its parameter(s)?

We are waiting until the first occurrence of a success (a doll is won) in a series of independent Bernoulli trials (tosses, resulting in won/not won) with the same probability of success on each trial. Thus, X has a Geometric distribution, and the parameter of a geometric is p, the probability of a success on any toss. Here, $p = 0.1$.

(b) (7pts) What is the probability that less than 3 throws are required to win the doll?

In this situation,

$$P(X < 3) = P(X = 1) + P(X = 2).$$

and for a geometric distribution

$$P(X = k) = pq^{k-1}, \quad \text{where } q = 1 - p.$$

Thus,

$$\begin{aligned} P(X < 3) &= p + pq \\ &= 0.1 + 0.1 * 0.9 = 0.19. \end{aligned}$$

(c) (3pts) Now assume that he is on a quest to win dolls for all of his children, of which he has three. If we let Y denote the number of the toss on which he successfully completes his quest, what is the distribution of Y? What are its parameter(s)?

We are now waiting for the kth occurrence of a success, so Y has a negative binomial distribution. The parameters of this distribution are k, the number of successes that we're waiting for, and p, the probability of success on any toss. Here, $k = 3, p = 0.1$.

(d) (7pts) What is the probability that at least 5 throws will be required to complete the quest?

$$\begin{aligned} P(Y \geq 5) &= 1 - P(Y < 5); \\ P(Y < 5) &= P(Y = 3) + P(Y = 4). \end{aligned}$$

Note, we can never have $Y < 3$ as we need at least 3 tosses to get 3 wins. For a negative binomial random variable,

$$P(Y = c) = \binom{c-1}{k-1} p^k q^{c-k}.$$

Thus,

$$\begin{aligned} P(Y < 5) &= \binom{3-1}{3-1}(0.1)^3 + \binom{4-1}{3-1}(0.1)^3(0.9) \\ &= 0.001 + 0.0027 = 0.0037, \end{aligned}$$

and

$$P(Y \geq 5) = 1 - 0.0037 = 0.9963.$$

3. [20 points] The density function of a continuous random variable X is given by

$$f_X(x) \quad = \quad c(x-4)^2, \quad 6 < x < 10.$$

(a) (8pts) Find c.

To do this, we're simply going to integrate the density over its range and set the result equal to 1. Here,

$$\begin{aligned} \int_6^{10} c(x-4)^2 dx &= \frac{c}{3}(x-4)^3\Big|_6^{10} \\ &= \frac{c}{3}(216-8) = \frac{208c}{3} \end{aligned}$$

so $c = 3/208$.

(b) (12pts) Let $Y = 1/X$. Find $f_Y(y)$.

Here, $Y = g(X)$, *where* $g(x) = 1/x$. *To address this, we first find the inverse of this function, so that* $g^{-1}(g(x)) = x$. *Here,* $g^{-1}(x) = 1/x$ *again. Now,*

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y))\left|\frac{\partial}{\partial y} g^{-1}(y)\right| \\ &= \frac{3}{208}(y^{-1}-4)^2 * y^{-2}, \quad \frac{1}{10} < y < \frac{1}{6}. \end{aligned}$$

Note that the Jacobian (change of variables) term has absolute value bars around it! Without these, we will get a "density" function that is negative.

4. [20 points] Let X and Y have the joint pdf

$$f_{XY}(x,y) \quad = \quad x+y, \quad 0<x<1, \quad 0<y<1.$$

(a) (7pts) Find the marginal density of X.

Given the joint density, we find the marginal by integrating over the unwanted variable.

$$\begin{aligned} f_X(x) &= \int f_{XY}(x,y)dy \\ &= \int_0^1 (x+y)dy \\ &= xy + \frac{y^2}{2}\bigg|_0^1 = x + \frac{1}{2}. \end{aligned}$$

(b) (7pts) Find the conditional density of Y given X.

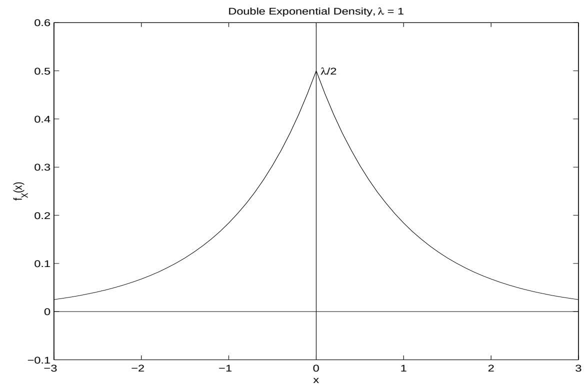
The conditional density is simply the joint density divided by the marginal density, so

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{x+y}{x+(1/2)}, \quad 0<y<1.$$

(c) (6pts) Find the conditional mean of Y given X.

To find the conditional mean of Y given X, we simply take the expectation of Y with respect to the conditional density. Thus,

$$\begin{aligned} E(Y|X) &= \int y f_{Y|X}(y|x)dy \\ &= \int_0^1 \frac{xy+y^2}{x+(1/2)}dy \\ &= \frac{1}{x+(1/2)}\left(x\frac{y^2}{2} + \frac{y^3}{3}\bigg|_0^1 \right) \\ &= \frac{(x/2)+(1/3)}{x+(1/2)} = \frac{3x+2}{6x+3}. \end{aligned}$$



5. [20 points] A random variable X is said to have a double exponential distribution if

$$f_X(x) \;=\; \frac{\lambda}{2}e^{-\lambda|x|}, \quad -\infty < x < \infty, \quad \lambda > 0.$$

(a) (2pts) Sketch the density.

See above. The density is symmetric about 0, and exhibits exponential decay away from the center at the same rate on both sides. The maximal value is attained at the origin, where it is $\lambda/2$*.*

(b) (10pts) Find the MGF of X.

The MGF is the expected value of e^{tX}*, so*

$$\begin{aligned}
E(e^{tX}) &= \int e^{tX} f_X(x)dx \\
&= \int_{-\infty}^{0} e^{tX}\frac{\lambda}{2}e^{\lambda x}dx + \int_{0}^{\infty} e^{tX}\frac{\lambda}{2}e^{-\lambda x}dx \\
&= \int_{-\infty}^{0} \frac{\lambda}{2}e^{(\lambda+t)x}dx + \int_{0}^{\infty} \frac{\lambda}{2}e^{-(\lambda-t)x}dx \\
&= \left(\frac{\lambda}{2}\frac{e^{(\lambda+t)x}}{\lambda+t}\Bigg|_{-\infty}^{0}\right) + \left(-\frac{\lambda}{2}\frac{e^{-(\lambda-t)x}}{\lambda-t}\Bigg|_{0}^{\infty}\right) \\
&= \frac{\lambda}{2}\frac{1}{\lambda+t} + \frac{\lambda}{2}\frac{1}{\lambda-t} \\
&= \frac{\lambda^2 - t\lambda + \lambda^2 + t\lambda}{2(\lambda^2 - t^2)} \\
&= \frac{\lambda^2}{\lambda^2 - t^2}.
\end{aligned}$$

(c) (8pts) Find the mean and variance of X.

$$\begin{aligned}\mu_X &= M_X'(0);\\ M_X'(t) &= \frac{2t\lambda^2}{(\lambda^2 - t^2)^2}\\ \mu_X &= 0.\end{aligned}$$

This should come as no surprise, since the density is symmetric about the origin, suggesting that the mean should be at zero. Now, for the variance, we just need the second moment of X,

$$\begin{aligned}E(X^2) &= M_X''(0);\\ M_X''(t) &= \frac{2\lambda^2}{(\lambda^2 - t^2)^2} + \frac{8t^2\lambda^2}{(\lambda^2 - t^2)^3}\\ \sigma_X^2 &= E(X^2) - E(X)^2 = \frac{2}{\lambda^2}.\end{aligned}$$

6. [10 points] Forty-eight measurements are recorded to several decimal places. Each of these is then rounded to the nearest integer. The sum of the measurements is then approximated by the sum of these integers. We assume that the rounding errors are independent and have Uniform$(-0.5, 0.5)$ distributions. Use the Central Limit Theorem to approximate the probability that the sum of the integers is within 2 units of the true sum.

Ok, let's let X_i denote the rounding error associated with observation i. Then the question is whether the absolute value of the sum (not the average!) of these errors is less than 2; find

$$P\left(\left|\sum_{i=1}^{48} X_i\right| < 2\right) = P\left(-2 < \sum_{i=1}^{48} X_i < 2\right).$$

Now, from the Central Limit Theorem, we are fairly confident that the distribution of the sum of a large number of identical terms is bell-shaped; the sum is normally distributed. To compute probabilities though, we need to find both the mean and variance of the corresponding normal distribution. Here, as the X_i's are independent and identically distributed,

$$\begin{aligned} E(\sum X_i) &= \sum E(X_i) = nE(X) \\ V(\sum X_i) &= \sum V(X_i) = nV(X) \end{aligned}$$

where n is the number of terms in the sum (48 here). Thus, we just need to find the mean and variance of a Uniform$(-0.5, 0.5)$ random variable.

$$\begin{aligned} E(X) &= \int_{-0.5}^{0.5} x\,dx = \left.\frac{x^2}{2}\right|_{-0.5}^{0.5} = 0 \\ E(X^2) &= \int_{-0.5}^{0.5} x^2\,dx = \left.\frac{x^3}{3}\right|_{-0.5}^{0.5} = \frac{1}{12} \end{aligned}$$

and this second moment is the variance as the first moment is zero. The sum of the X_i's is approximately $N(0, 48/12) = N(0, 4)$. Thus,

$$\begin{aligned} P(-2 < \sum_{i=1}^{48} X_i < 2) &= P\left(\frac{-2 - n(0)}{\sqrt{48/12}} < \frac{\sum_{i=1}^{48} X_i - n\mu_X}{\sqrt{n\sigma_X^2}} < \frac{2 - n(0)}{\sqrt{48/12}}\right) \\ &= P(-1 < Z < 1) \\ &= \Phi(1) - \Phi(-1) \\ &= 0.6827. \end{aligned}$$