Math 447 **Test 2 Solutions** **Fall 2015**

1. Let the random variable Y have distribution function

$$F(y) = \begin{cases} 0, & y \leq 0, \\ \frac{y}{8}, & 0 < y < 2, \\ \frac{y^2}{16}, & 2 \leq y < 4, \\ 1, & y \geq 4. \end{cases}$$

   (a) (5 points) What is the probability density function of Y? (Make sure you account for all the cases in the definition of F!)

   Solution. This is exercise 4.19 (a) from the text.

$$f(y) = F'(y) = \begin{cases} \frac{1}{8} & 0 < y < 2 \\ \frac{y}{8} & 2 < y < 4 \\ 0 & \text{otherwise.} \end{cases}$$

   (b) (5 points) Find the mean of Y.

   Solution. This is exercise 4.25 from the text.

$$E[Y] = \int_{-\infty}^{\infty} y f(y) \mathrm{d}y = \int_0^2 y \cdot \frac{1}{8} \mathrm{d}y + \int_0^2 y \cdot \frac{y}{8} \mathrm{d}y$$

   This integration yields 31/12 after some calculation.

   (c) (5 points) Find the variance of Y.

   Solution. This is exercise 4.25 from the text. Use $V[Y] = E[Y^2] - E[Y]^2$, and first calculate

$$E[Y^2] = \int_{-\infty}^{\infty} y^2 f(y) \mathrm{d}y = \int_0^2 y^2 \cdot \frac{1}{8} \mathrm{d}y + \int_0^2 y^2 \cdot \frac{y}{8} \mathrm{d}y$$

   Then use $E[Y]$, calculated in the previous part, to get $V[Y] = 1.160$.

2. Scores on an examination are assumed to be normally distributed with mean 78 and variance 36.

   (a) (6 points) Suppose that students scoring in the top 10% of this distribution are to receive an A grade. What is the minimum score a student must achieve to earn an A grade?

   Solution. This is exercise 4.74 (b) from the text, which you did for homework. The answer is 85.7.

   (b) (8 points) If it is known that a student's score exceeds 72, what is the probability that his or her score exceeds 84?

   Solution. This is exercise 4.74 (f) from the text, which you did for homework. The answer is 0.1886.

3. Let Y be an exponential random variable with parameter $\beta = 1$.

   (a) (5 points) Find $P(|Y - \mu| \geq 2\sigma)$, where μ and σ are the expectation and standard deviation of Y.

   Solution. This is exercise 4.150 from the text. We know (from the tables, or a direct calculation) that $\mu = 1$ and $\sigma = 1$. We also know $f(y) = e^{-y}$ for $y > 0$ and 0 otherwise. So

$$P(|Y - \mu| \geq 2\sigma) = \int_{-\infty}^{-1} f(y) \mathrm{d}y + \int_3^{\infty} f(y) \mathrm{d}y = 0 + (-e^{-y})|_3^{\infty} = e^{-3}$$

(b) (5 points) What inequality for the probability of the previous part is given by Chebyshev's theorem?

Solution. This is exercise 4.150 from the text. From the statement of Chebyshev's we have

$$P(|Y-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = \frac{1}{4}$$

Note that $e^{-3} < 0.25$.

4. The joint density function of Y_1 and Y_2 is given by

$$f(y_1, y_2) = \begin{cases} 30y_1y_2^2, & y_1 - 1 \leq y_2 \leq 1 - y_1, 0 \leq y_1 \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Solution. This is exercise 5.31 from the text. Most of this (except part (a), the hardest part) was solved in detail in class on November 16th, and you can look at the notes web page for a solution.

(a) (5 points) Find $P(Y_1 > Y_2)$.

Solution. This is 5.13 (c) in the text. The answer is $\frac{21}{32}$. Because the integration is easier this way, we note that $P(Y_1 > Y_2) = 1 - P(Y_2 > Y_1)$, and show that $P(Y_2 > Y_1) = \frac{11}{32}$. Graphing the region shows the relevant integral is

$$\int_0^{0.5} \int_{y_1}^{1-y_1} 30y_1y_2^2 \, dy_2 dy_1$$

The calculuation to get the answer is somewhat tedious, but only requires integrating polynomials.

(b) (5 points) Find the marginal density of Y_1. (Be sure you account correctly for all cases!)

(c) (5 points) Find the the conditional density of Y_2 given $Y_1 = y_1$. (Again, be sure you account correctly for all cases!)

(d) (5 points) Find $P(Y_2 > 0 \mid Y_1 = 0.75)$.

(e) (3 points) Are the random variables Y_1 and Y_2 independent? (Only an answer is necessary.)

Solution. No. See Theorem 5.5.

5. (5 points) Suppose that Y is a random variable that has the beta distribution with parameters $\alpha = \beta = 1$. There is another name for the distribution of Y, i.e. it is one of the distributions we have studied. What is the name of this distribution and what is(are) the relevant parameter(s)?

Solution. This is problem 4.127 from the text. You can plug $\alpha = \beta = 1$ into the density function for the beta distribution and get $f(y) = 1$ in $(0, 1)$ and 0 otherwise. This is the same as the density function for the uniform distribution on $(0, 1)$.

So the answer is "uniform distribution with $\theta_1 = 0$ and $\theta_2 = 1$".

6. Suppose that Y_1 and Y_2 have the joint distribution

$$f(y_1, y_2) = \begin{cases} \frac{1}{8}y_1e^{-(y_1+y_2)/2}, & y_1 > 0 \text{ and } y_2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Remark. This joint density function was repeated in several exercises in the text, including 5.18 and 5.112.

(a) (3 points) Are Y_1 and Y_2 independent? (Only an answer is necessary).

Solution. Yes, by Theorem 5.5.

(b) (4 points) What type of random variable is Y_1? Give the name and relevant parameters.

Solution. The marginal density of Y_1 can be determined by factoring $f(y_1, y_2) = f_1(y_1)f_2(y_2)$, where $f_1(y_1) = \frac{1}{4}y_1e^{-y_1/2}$ (which is $\Gamma(2, 2)$) and $f_2(y_2) = \frac{1}{2}e^{-y_2/2}$ (which is exponential with $\beta = 2$). Then integrating out y_2 yields just the gamma pdf $f_1(y_1)$, because the other pdf integrates to 1. Thus Y_1 is gamma with $\alpha = \beta = 2$.

(c) (4 points) What type of random variable is Y_2? Give the name and relevant parameters.

Solution. As in the previous part, we get that Y_2 is exponential with parameter $\beta = 2$.

(d) (8 points) Let $C = 50 + 2Y_1 + 4Y_2$. Find $V(C)$.

Solution. The constant doesn't matter, and Y_1 and Y_2 are independent. So $V[C] = 2^2V[Y_1]+4^2V[Y_2]$. Since we know the distributions of Y_1 and Y_2, we have $V[C] = 2^2 2 2^2 + 4^2 2^2 = 32 + 64 = 96$.

7. (7 points) Let Y_1 have an exponential distribution with mean λ and the conditional density of Y_2 given $Y_1 = y_1$ be

$$f(y_2 \mid y_1) = \begin{cases} \frac{1}{y_1}, & 0 \le y_2 \le y_1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $E(Y_2)$, the unconditional mean of Y_2.

Solution. This is problem 5.141 in the text. We can use $E[Y_2] = E[E[Y_2 \mid Y_1]]$, which is Theorem 5.14. We compute

$$E[Y_2 \mid Y_1] = \int_{-\infty}^{\infty} y_2 f(y_2 \mid y_1)\mathrm{d}y_2 = \int_0^{y_1} y_2 \frac{1}{y_1}\mathrm{d}y_2 = \frac{1}{y_1}\frac{y_1^2}{2} = \frac{y_1}{2}$$

Then we have $E[E[Y_2 \mid Y_1]] = E[\frac{1}{2}Y_1] = \lambda/2$.

8. (6 points) (Monty Hall, Revisited) On this version of the show, you are, as usual, shown 3 face down cards, one of which is the $Q\heartsuit$ and the other two of which are black spot cards. Your objective is to select the $Q\heartsuit$, which wins \$1,000. As usual, you have to select a face-down card, and then the host turns over one of the other two cards. The twist this time is that the host himself *does not know* where the $Q\heartsuit$ is, and so may reveal the winning card. In this case you would win immediately!

On your first play of the game, you select a face-down card, and the host turns over a black spot card (so no instant win). You are now offered the opportunity to pay \$100 to switch your choice to the remaining face-down card.

Which of the following statements is most accurate, assuming you wish to maximize your expected winnings?

(a) It is to your advantage to switch, and you should pay \$100 to do so.

(b) It is to your advantage to switch, but it is not worth paying \$100 to do so.

(c) It is to your disadvantage to switch; you should not do so even for free.

(d) It makes no difference whether you switch or not.

Solution. (d). There has been some confusion over this, so I will give three versions of the solution.

Version 1: (Enumerate the sample space.) We may assume without loss of generality, that we choose the first card, since we make this choice based on no information. We enumerate the sample space by denoting the spot cards as $x\clubsuit$ and underlining the card revealed by the host. The elements are $Q\heartsuit\underline{x\clubsuit}x\clubsuit$, $Q\heartsuit x\clubsuit\underline{x\clubsuit}$, $x\clubsuit\underline{Q\heartsuit}x\clubsuit$, $x\clubsuit x\clubsuit\underline{Q\heartsuit}$, $x\clubsuit Q\heartsuit\underline{x\clubsuit}$, and $x\clubsuit\underline{x\clubsuit}Q\heartsuit$. Notice that the queen is equally likely ($p = \frac{1}{3}$) to be in any of the three locations, and the host choses a card to reveal at random. So each element of the sample space has probability $\frac{1}{6}$. If we use the switching strategy, we win in the last four cases. If we use the strategy of standing pat, we win in the first four cases. So the probability of winning with either strategy is the same.

Version 2: (Reasoning about sample space without explicit enumeration.) We show that whether we play the game with the switching strategy, or the strategy of sticking with our initial choice, then the probability of winning is 2/3. Note that our probability of winning comes partly from the fact that the host may by chance reveal the prize to us.

If G is the event that our initial selection is the winning $Q\heartsuit$, which happens with probability 1/3, then in case G there is no chance that the winning card will be revealed by the host and switching guarantees a loss.

In case $\bar{G}$ (probability 2/3), there is a 50% chance that the winning card is revealed, and if that doesn't happen, switching guarantees a win. Thus the switching strategy wins in $\bar{G}$, and not otherwise, so has probability of winning 2/3.

The non-switching strategy always wins in case G, and wins 50% of the time in case $\bar{G}$, because the winning card is revealed by the host. So the chance of winning overall is $2/3 = 1/3 + (2/3) \cdot (1/2)$. Thus the answer is (d).

Version 3: (Abstract reasoning.) Another way to look at this is that neither we nor the host have any information about where the prize is, so a strategy that depends on the host's choices shouldn't increase our chances of winning. This is not true for the original game, because in the TV version, the host did know where the prize was, and was revealing some information to us before offering to let us switch.

9. (6 points) A personnel director has two lists of applicants for jobs. List 1 contains the names of five women and two men, whereas list 2 contains the names of two women and six men. A name is randomly selected from list 1 and added to list 2. A name is then randomly selected from the augmented list 2. Given that the name selected is that of a man, what is the probability that a woman's name was originally selected from list 1? Select from the alternatives below the number closest to that probability.

   (a) 0.65

   (b) 0.68

   (c) 0.71

   (d) 0.74

   Solution. (b). This is problem 2.136 from the text. Let A be the event that a woman's name is selected from list 1. Let B be the event that a woman's name is selected from list 2. We have $P(A) = 5/7$, $P(\bar{B} \mid A) = 2/3$, $P(\bar{B} \mid \bar{A}) = 7/9$. Apply Bayes' formula to get $P(A \mid \bar{B}) = 15/22 \approx 0.6818$.

10. (6 points) Suppose that random variables X and Y have correlation $\rho_{XY} = 0.3$. Which of the following numbers is closest to the correlation between $2X - 1$ and $4 - 3Y$?

(a) 0.6

(b) 0.3

(c) 0.15

(d) -0.6

Solution. (d). This is (basically) exercise 5.110 from the text. We can use the bilinearity of covariance to calculate the desired correlation.

$$\begin{aligned}\operatorname{Cov}(2X-1, 4-3Y) &= \operatorname{Cov}(2X, 4) + \operatorname{Cov}(2X, -3Y) + \\ &\operatorname{Cov}(-1, 4) + \operatorname{Cov}(-1, -3Y) = 2 \cdot (-3) \cdot \operatorname{Cov}(X, Y) = -6 \operatorname{Cov}(X, Y)\end{aligned}$$

$$\sigma_{2X-1} = \sqrt{V[2X-1]} = 2\sigma_X \qquad \sigma_{4-3Y} = \sqrt{V[4-3Y]} = \sqrt{(-3)^2 V[Y]} = 3\sigma_Y$$

Thus

$$\rho_{2X-1,4-3Y} = \frac{-6\operatorname{Cov}(X, Y)}{2\sigma_X \cdot 3\sigma_Y} = -\rho_{XY} = -0.3$$