Math 447 - May 13, 2013 - Final Exam Solutions

Name:______________________________

Read these instructions carefully: The points assigned are *not* meant to be a guide to the difficulty of the problems. If the question is multiple choice, there is a penalty for wrong answers, so that your expected score from guessing at random is zero. No partial credit is possible on multiple-choice and other no-work-required questions.

1. (6 points) We know the following about a colormetric method used to test lake water for nitrates. If water specimens contain nitrates, a solution dropped into the water will cause the specimen to turn red 95% of the time. When used on water specimens without nitrates, the solution causes the water to turn red 10% of the time (because chemicals other than nitrates are sometimes present and they also react to the agent). Past experience in a lab indicates that nitrates are contained in 30% of the water specimens that are sent to the lab for testing. If a water specimen is randomly selected and turns red when tested, which of the following numbers is closest to the probability that it actually contains nitrates?

(a) 0.80
(b) 0.95
(c) 0.30
(d) 0.38

(This is problem 2.152 from the text.)

Solution. (a). Let R denote the event that the specimen turns red and N denote the event that the specimen contains nitrates. Using Bayes' Formula,

$$P(N \mid R) = \frac{P(R \mid N)P(N)}{P(R \mid N)P(N) + P(R \mid \bar{N})P(\bar{N})} = \frac{.95(.3)}{.95(.3) + .1(.7)} \approx 0.803$$

2. (6 points) Many public schools are implementing a "no-pass, no-play" rule for athletes. Under this system, a student who fails a course is disqualified from participating in extracurricular activities during the next grading period. Suppose that the probability is 0.15 that an athlete who has not previously been disqualified will be disqualified next term. For athletes who have been previously disqualified, the probability of disqualification next term is 0.5. If 30% of the athletes have been disqualified in previous terms, which of the following numbers is closest to the probability that a randomly selected athlete will be disqualified during the next grading period?

(a) 0.15
(b) 0.20
(c) 0.25
(d) 0.30

(This is problem 2.174 from the text.)

Solution. (c). Let A be the event that an athlete was disqualified previously and B be the event that an athlete is disqualified next term. We are given $P(B \mid \bar{A}) = 0.15$, $P(B \mid A) = 0.5$, and $P(A) = 0.3$. Using the "Law of Total Probability",

$$P(B) = P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A}) = 0.3(0.5) + 0.7(0.15) = 0.255.$$

3. (12 points) (Monty Hall, Episode 3) On today's episode of the ever-popular game show, you are shown five cards, all face-down, of which one is the $Q\heartsuit$ and four are black spot cards. The objective is to select the $Q\heartsuit$, which wins \$100.

As usual, the host lets you select a face-down card, and then shows you, from one of the other four face-down cards, a black spot card, which he can always do, because he knows where all the cards are. He offers then you the opportunity to switch your choice to one of the three remaining face-down cards.

3.1 (6 points) Suppose you decide to switch, and that you select one of the remaining three face-down cards at random. Let W be the event that your new selection is the (winning) $Q\heartsuit$. Which of the following numbers is closest to $P(W)$?

(a) 0.33
(b) 0.25
(c) 0.23
(d) 0.26

Solution. (d). Let Q denote the event that your initial selection is the $Q\heartsuit$. Then, by the "Law of Total Probability",

$$P(W) = P(W \mid Q)P(Q) + P(W \mid \bar{Q})P(\bar{Q}) = 0 \cdot \frac{1}{5} + \frac{1}{3} \cdot \frac{4}{5} = \frac{4}{15} \approx .0.267$$

3.2 (6 points) Your run of luck continues, and you manage to select the winning $Q\heartsuit$. Your host now offers the opportunity to play again, in a new version of the game. In this version, after making your initial selection, but before deciding whether to switch, you will be shown *two* black spot cards. As in the previous round, the $Q\heartsuit$ wins you \$100. Which of the following numbers is closest to the difference in expected values between this new version of the game and the version played in the previous round (question 3.1)? (You may assume that playing decisions are made in an expectation-maximizing fashion.)

(a) \$10
(b) \$15
(c) \$18
(d) \$20

Solution. (b). In the previous round your expectation was approximately $\$26.67 = 0.2666 \times \$100$, and in this round we can compute $P(W)$ by the same method:

$$P(W) = P(W \mid Q)P(Q) + P(W \mid \bar{Q})P(\bar{Q}) = 0 \cdot \frac{1}{5} + \frac{1}{2} \cdot \frac{4}{5} = \frac{4}{10} = 0.4.$$

Thus the difference in expectation is \$40 − \$26.67 = \$13.33.

5. (6 points) The length of time required for the periodic maintenance of an automobile or another machine usually has a mound-shaped probability distribution. Because some occasional long service times will occur, the distribution tends to be skewed to the right. Suppose that the length of time required to run a 5000-mile check and to service an automobile has mean 1.4 hours and standard deviation .7 hour. Suppose also that the service department plans to service 50 automobiles per 8-hour day and that, in order to do so, it can spend a maximum average service time of only 1.6 hours per automobile. Which of the following numbers is closest to the proportion of all workdays on which the service department will have to work overtime?

(a) 0.045
(b) 0.025
(c) 0.020
(d) 0.005

(This is problem 7.49 from the text.)

Solution. (c) If Y_i is the service time for the i-th car, we may approximate $\bar{Y} = (1/50)(Y_1 + Y_2 + \cdots + Y_{50})$ using the Central Limit Theorem. That is, $\bar{Y}$ may be assumed to be normal with $\mu = 1.4$ and $\sigma = 0.7/\sqrt{50}$. It follows that $P(\bar{Y} > 1.6) \approx P(Z > (1.6 - 1.4)/(0.7/\sqrt{50})) = P(Z > 2.02) \approx 0.0217$.

6. (6 points) Shear strength measurements for spot welds have been found to have standard deviation 10 pounds per square inch (psi). Which of the following answers is closest to the minimum number of test welds that must be sampled if we want the sample mean to be within 1 psi of the true mean with probability 0.99?

(a) 664
(b) 6635
(c) 328
(d) 26

(This is problem 7.51 from the text.)

Solution. (a) If we sample n welds, the sample mean $\bar{Y}$ will be (approximately) normally distributed with mean μ, the population mean, and standard deviation $10/\sqrt{n}$ (by the CLT). We want $P(|\bar{Y} - \mu| < 1) = 0.99$, and we can rewrite this probability as $P(|Z| < 1/(10/\sqrt{n})) = P(-\frac{1}{10/\sqrt{n}} < Z < \frac{1}{10/\sqrt{n}})$. By the symmetry of the normal distribution, we must have $P(Z > \frac{1}{10/\sqrt{n}}) = 0.005$, and looking in the table gives $\frac{1}{10/\sqrt{n}} = 2.576$. Solving, we obtain $n = 663.57$.

7. (6 points) One-hour carbon monoxide concentrations in air samples from a large city average 12 ppm (parts per million) with standard deviation 9 ppm. You are asked to find the probability that the average concentration in 100 randomly selected samples will exceed 14 ppm. Which of the following statements is most accurate?

(a) The distribution of carbon monoxide concentrations in air samples are not likely to be normally distributed, so the central limit theorem cannot be applied to this problem.
(b) The distribution of carbon monoxide concentrations in air samples are likely to be normally distributed, so the central limit theorem can be applied to this problem.

(c) The distribution of carbon monoxide concentrations in air samples are likely to be normally distributed, but the central limit theorem cannot be applied to this problem.

(d) The distribution of carbon monoxide concentrations in air samples are not likely to be normally distributed, but the central limit theorem can be applied to this problem.

(This is based on problem 7.53 from the text.)

Solution. (d). The carbon monoxide concentration cannot be negative, and with the given mean and standard deviation will therefore *not* be anywhere close to normally distributed. The hypotheses of the central limit theorem do not require that the variables involved be normal.

8. (6 points) Suppose that $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ are independent random samples, with the variables X_i normally distributed with mean μ_1 and variance σ_1^2 and the variables Y_i normally distributed with mean μ_2 and variance σ_2^2. Write $\bar{X} = \frac{1}{m}(X_1 + \cdots + X_m)$ and $\bar{Y} = \frac{1}{m}(Y_1 + \cdots + Y_n)$ for the sample means. Which of the following statements is most accurate?

(a) $\bar{X}$ and $\bar{Y}$ are normally distributed, and their difference $\bar{X} - \bar{Y}$ is also normally distributed.

(b) $\bar{X}$ and $\bar{Y}$ are normally distributed, but their difference $\bar{X} - \bar{Y}$ need not be normally distributed.

(c) $\bar{X}$ and $\bar{Y}$ are normally distributed, and if $m = n$, the difference $\bar{X} - \bar{Y}$ will also be normally distributed.

(d) None of $\bar{X}$, $\bar{Y}$, and $\bar{X} - \bar{Y}$ need be normally distributed.

Solution. (a). A linear combination of independent normal random variables is normal. (Theorem 6.3)

9. (6 points) Let $X_1, X_2, \ldots, X_{50}$ be a random sample from a normal population with mean 0 and variance 1. Let $Y = \frac{1}{3}(X_1^2 + X_2^2 + X_3^2)$ and $Z = X_1^2 + X_2^2 + \cdots + X_{50}^2$. Which of the following statements is most accurate?

(a) Y is normally distributed, and Z has the χ^2-distribution.

(b) Y has the chi-square distribution and Z also has the chi-square distribution.

(c) Y has the gamma distribution and the distribution of Z is approximately normal.

(d) None of the above statements is accurate.

Solution. (c). $(X_1^2 + X_2^2 + X_3^2)$ has the χ^2 distribution, by Theorem 7.2. This is a special case of the Gamma distribution, and by the scaling law for the Gamma distribution (which is just a change of variable) $Y \sim \Gamma(\frac{3}{2}, \frac{2}{3})$. Note that this means that Y does *not* have the χ^2 distribution. In addition, Z has the χ^2 distribution with $n = 50$, which is approximately normal. (Use the applets to graph it, if you haven't already.)

10. (6 points) The length of time required to complete a test is found to be normally distributed with mean 70 minutes and standard deviation 12 minutes. When should the test be terminated if we wish to allow sufficient time for 90% of the students to complete the test?

(a) After roughly 85 minutes

(b) After roughly 90 minutes

(c) After roughly 95 minutes

(d) After roughly 98 minutes

(e) Right now!

(This is based on problem 4.161 from the text.)

Solution. (a). Here, $\mu = 70$ and $\sigma = 12$ with the normal distribution. We require $\phi_{0.9}$, the 90th percentile of the distribution of test times. Since for the standard normal distribution, $P(Z < z_0) = 0.9$ for $z_0 = 1.28$, thus $\phi_{0.9} = 70 + 12(1.28) \approx 85$.

11. (24 points) Suppose that X and Y are independent random variables having the Poisson distribution with means λ and ν, respectively. Let $W = X + Y$.

(This is based on problem 6.52 from the text.)

a. (4 points) Find the moment generating function of W.

Solution. Since X and Y are independent, we have $m_W(t) = m_X(t)m_Y(t)$, and m_X and m_Y can be looked up in the tables:

$$m_W(t) = \exp(\lambda(e^t - 1)) \cdot \exp(\nu(e^t - 1)) = \exp((\lambda + \nu)(e^t - 1)).$$

b. (4 points) What is the distribution of W? Give any parameter values that are relevant.

Solution. By uniqueness of moment-generating functions, W has the Poisson distribution with parameter $\lambda + \nu$.

c. (4 points) Apply the definition of the conditional probability function $p_X(x \mid W = w)$ to write this in terms of the probability functions p_X, p_Y and p_W. (To clarify, there should be nothing distribution-specific in your answer here. That's the next part.)

Solution. The definition of conditional probability function is

$$p_X(x \mid W = w) = \frac{p(x, w)}{p_W(w)},$$

where $p(x, w) = P(X = x, W = w)$ is the joint probability function of X and W. But $P(X = x, W = w) = P(X = x, Y = w - x) = p_X(x)p_Y(w - x)$. Thus we have

$$p_X(x \mid W = w) = \frac{p_X(x)p_Y(w - x)}{p_W(w)}.$$

d. (6 points) Use your specific knowledge of the distributions of X, Y, and W to write down and simplify the conditional probability function $p_X(x \mid W = w)$.

Solution. Substituting the Poisson probability functions into the answer of the previous part

$$p_X(x \mid W = w) = \frac{e^{-\lambda}\frac{\lambda^x}{x!}e^{-\nu}\frac{\nu^{w-x}}{(w-x)!}}{e^{-(\lambda+\nu)}\frac{(\lambda+\nu)^w}{w!}},$$

we can simplify to obtain

$$p_X(x \mid W = w) = \binom{w}{x}\left(\frac{\lambda}{\lambda + \nu}\right)^x\left(\frac{\nu}{\lambda + \nu}\right)^{w-x}$$

e. (6 points) The distribution of X, given that $W = w$, is a familiar one. What is it called and what are the values of the associated parameters?

Solution. This is the probability function for a binomal random variable with parameters w and $\frac{\lambda}{\lambda+\nu}$. Thus the distribution of X, given $W = w$, is binomial with these parameters.
11. (14 points) Assume that Y denotes the number of bacteria per cubic centimeter in a particular liquid and that Y has a Poisson distribution with parameter λ. Further assume that λ varies from location to location and has a gamma distribution with parameters α and β, where α is a positive integer. If we randomly select a location, what is the

(This is based on problem 5.138 from the text. See section 5.11 for more explanation of the answers.)
a. (4 points) expected number of bacteria per cubic centimeter?
Solution. $E[Y] = E[E[Y \mid \lambda]] = E[\lambda] = \alpha\beta$.

b. (10 points) standard deviation of the number of bacteria per cubic centimeter?
Solution. $V[Y] = V[E[Y \mid \lambda]] + E[V[Y \mid \lambda]] = \alpha\beta^2 + \alpha\beta$. Thus the standard deviation is $\sqrt{\alpha\beta(1+\beta)}$.

12. (20 points) Suppose that the joint distribution of the random variables X and Y is such that the point (X, Y) is uniformly distributed over the region in the (x, y)-plane given by the inequalities $x \geq 0$, $y \geq 0$, and $x + y \leq 1$. Let $U = X + Y$.

(This is based on problem 6.9 from the text. See section 6.3, particularly the examples therein, for more explanation.)
a. (3 points) Draw the region bounded by these inequalities.
Solution. It's a triangle with vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$. (My formatting language skills aren't up to giving you a picture in pdf form.)
b. (3 points) Dividing into cases if necessary, give a formula for the joint probability density function $f(x, y)$.
Solution.

$$f(x,y) = \begin{cases} 2, & \text{inside the triangle of part (a)} \\ 0, & \text{outside the triangle of part (a)} \end{cases}$$

c. (2 points) Are X and Y independent? (Only an answer is necessary, no reasoning required.)
Solution. No, see Theorem 5.5.
d. (8 points) Dividing into cases if necessary, give a formula for the cumulative distribution function $F_U(u)$.
Solution. $F_U(u) = P(U \leq u)$. Work geometrically and compute the area of the triangle cut off by $x \geq 0$, $y \geq 0$, and $x + y \leq u$, which is $(1/2)u^2$. We obtain

$$F_U(u) = \begin{cases} 0, & u < 0 \\ u^2, & 0 \leq u \leq 1 \\ 1, & u > 1 \end{cases}$$

e. (4 points) Dividing into cases if necessary, give a formula for the probability density function $f_U(u)$.
Solution. Differentiate the previous answer.

$$f_U(u) = \begin{cases} 2u, & 0 \leq u \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

13. (12 points) A large lot of manufactured items contains 10% with exactly one defect, 5% with more than one defect, and the remainder with no defects. Ten items are randomly selected from this lot for sale. If X denotes the number of items with one defect and Y, the number with more than one defect, the repair costs are $C = X + 3Y$.

(This is based on problem 5.126 in the text.)

a. (3 points) Find the probability that $X = Y = 1$. Use your calculator to give a numerical answer, accurate to 3 decimal places.

Solution. This is a multinomial probability, see Section 5.9.

$$p(1,1,8) = \frac{10!}{1!1!8!} \cdot (0.10)^1(0.05)^1(0.85)^8 \approx 0.1226$$

b. (3 points) Find $\mathrm{Cov}(X,Y)$.

Solution. Use the formula for the multinomial covariance: $Cov(X,Y) = -np_X p_Y = -10(0.1)(0.05) = -0.05$.

c. (6 points) Find the variance of C.

Solution. Use the bilinearity of covariance (see Section 5.8).

$$\begin{aligned} V[C] &= \mathrm{Cov}(C,C) = \mathrm{Cov}(X+3Y, X+3Y) = V[X] + 3^2 V[Y] + 2 \cdot 3 \cdot \mathrm{Cov}(X,Y) \\ &= 10(0.1)(0.9) + 9 \cdot 10 \cdot (0.05)(0.95) + 2 \cdot 3 \cdot (-0.05) = 0.9 + 4.275 - 0.3 = 4.875 \end{aligned}$$

14. (12 points) Let X and Y be independent random variables that both have the geometric distribution with parameter p. Find, with proof, the probability that $|X - Y| = 1$.

Solution. Note that $P(|X-Y| = 1) = P(X-Y = 1) + P(X-Y = -1) = 2P(X-Y = 1)$ by symmetry. Now

$$P(X-Y = 1) = P(X = Y+1) = P(Y = 0, X = 1) + P(Y = 1, X = 2) + P(Y = 2, X = 3) + \cdots$$

which, using independence is

$$p \cdot qp + qp \cdot q^2p + q^2p \cdot q^3p + \cdots = p^2q \cdot (1 + q^2 + q^4 + \cdots) = p^2q \cdot \frac{1}{1-q^2},$$

where $q = 1 - p$. So the answer is

$$\frac{2p^2q}{1-q^2}.$$