AMS 310.03 Practice Final Exam — Fall, 2004

Name: **SSN:**

Instructions.
This is a closed book and closed notes exam. You can use two pieces (8×11) of cheat sheet. Please provide detailed solutions. Anyone who cheats on the exam shall receive a score of zero.

1. Heights of American women are normally distributed with a mean of 63.6 in. and a standard deviation of 2.5 in. (based on data from the National Health Survey). In order to fit into the Russian Soyuz spacecraft, an astronaut must have a height between 64.5 in. and 72 in.

   (a) Find the percentage of American women who meet the height requirement.
   Answer: Let X denote the height of a randomly selected woman, we have $X \sim N(\mu = 63.6, \sigma = 2.5)$. Therefore the percent of American women who meet the height requirement is

$$\begin{aligned} P(64.5 < X < 72) &= P\left(\frac{64.5 - 63.6}{2.5} < \frac{X - 63.6}{2.5} < \frac{72 - 63.6}{2.5}\right) \\ &= P(0.36 < Z < 3.36) = 1 - 0.6406 - 0.0004 = 0.359 \end{aligned}$$

   (b) Among 500 randomly selected American women, how many are expected to meet the Soyuz height requirement?
   Answer: Let Y be the number of women among 500 randomly selected who meet the Soyuz height requirement, then

$$Y \sim \mathit{Binomial}(n = 500, p = 0.359)$$

   Therefore, we expect the following number of women to meet the requirement:

$$E(Y) = np = 500 \cdot 0.359 = 179.5$$

2. Let X and Y be continuous random variables with joint pdf

$$f_{X,Y}(x, y) = \frac{1}{8}(6 - x - y), \; 0 < x < 2, \; 2 < y < 4$$

   and 0 elsewhere. Please find (a) $f_X(x)$, (b) $f_{Y|x}(y)$, (c) $P(2 < Y < 3 \mid x = 1)$, (d) $E(Y \mid x = 1)$
   Answer:

   (a)

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_2^4 \frac{1}{8}(6 - x - y)\, dy \\ &= \frac{1}{8}(6 - 2x), \; 0 < x < 2 \end{aligned}$$

   (b)

$$\begin{aligned} f_{Y|x}(y) &= \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{\frac{1}{8}(6 - x - y)}{\frac{1}{8}(6 - 2x)} \\ &= \frac{6 - x - y}{6 - 2x}, \; 0 < x < 2, \; 2 < y < 4 \end{aligned}$$

   (c)

$$P(2 < Y < 3 \mid x = 1) = \int_2^3 f_{Y|1}(y)\, dy = \int_2^3 \frac{6 - 1 - y}{6 - 2 \cdot 1} dy = \frac{5}{8}$$

(d)

$$E(Y \mid x=1) = \int_{-\infty}^{\infty} y \cdot f_{Y|1}(y)\, dy = \int_{2}^{4} y \cdot \frac{6-1-y}{6-2\cdot 1} dy \approx 2.83$$

3. Consider a series of independent trials, each having one of two possible outcomes, success or failure. Let p be the success probability for each trial. Define the random variable X to be the trial at which the first success occurs. Then X follows the Geometric distribution with probability distribution function

$$f(x) = P(X=x) = (1-p)^{x-1} p, \; x = 1, 2, \cdots$$

The mean of X is

$$E(X) = \frac{1}{p}$$

The geometric distribution can be applied to solving the following problem: A grocery store is sponsoring a sales promotion where the cashiers give away one of the letters A, E, L, S, U, and V for each purchase. If a customer collects all six (spelling VALUES), he or she gets \$10 worth of groceries free. What is the expected number of trips to the store a customer needs to make in order to get a complete set? Assume the different letters are given away randomly.
Answer: Let X_i denote the number of purchases necessary to get the ith different letter, $i = 1, 2, \cdots, 6$, and let X denote the number of purchases necessary to qualify for the \$10. Then

$$X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6$$

Clearly, $X_1 = 1$ with probability 1, so $E(X_1) = 1$. Having received the first letter, the chances of getting a different one are $5/6$ for each subsequent trip to the store. Therefore, $X_2 \sim Geometric\,(p = 5/6)$ and $E(X_2) = 6/5$. Similarly, $X_3 \sim Geometric\,(p = 4/6)$ and $E(X_3) = 6/4$; $X_4 \sim Geometric\,(p = 3/6)$ and $E(X_4) = 6/3$; $X_5 \sim Geometric\,(p = 2/6)$ and $E(X_5) = 6/2$; $X_6 \sim Geometric\,(p = 1/6)$ and $E(X_6) = 6/1$. It follows that

$$E(X) = \sum_{i=1}^{6} E(X_i) = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + \frac{6}{1} = 14.7$$

That is, a customer will have to make 14.7 trips to the store, on the average, to collect a complete set of six letters.

4. Suppose that 50 people are to be given a blood test to see who has a certain disease. The obvious laboratory procedure is to examine each person's blood individually, meaning that 50 tests would eventually be run. An alternative strategy is to divide each person's blood samples into two parts – say, A and B. All of the A's would then be mixed together and treated as one sample. If that "pooled" sample proved to be negative for the disease, all 50 individuals must necessarily be free of the infection, and no further testing would need to be done. If the pooled sample gave a positive reading, of course, all 50 B samples would have to be analyzed separately. If the chance of a person being infected is 1 in 1000, for example, how many tests on the average do we have to perform using the pooling strategy?
Answer: Let the random variable X denote the number of tests that will have to be performed if the samples are pooled. Clearly, $X = 1$ if none of the 50 is infected and $X = 51$ if at least one of the 50 is infected. The number of people infected would follow a binomial distribution with parameters $n = 50$ and $p = 1/1000$. Therefore, we have

$$P(X=1) = \binom{50}{0} \left(\frac{1}{1000}\right)^{0} \left(1 - \frac{1}{1000}\right)^{50}$$

$$P(X=51) = 1 - P(X=1)$$

and

$$E(X) = 1 \cdot P(X=1) + 51 \cdot P(X=51) \approx 3.4$$

We conclude that if the chance of a person being infected is 1 in 1000, only 3.4 tests need to be performed on the average using the pooling strategy.

5. An economist wants to estimate the mean income for the first year of work for a college graduate who has the profound wisdom to take a statistics course. How many such incomes must be found if we want to be 95% sure that the sample mean is within \$500 of the true population mean? Assume that a previous study has revealed that for such incomes, $\sigma = \$6250$.
Answer: $E = 500$, $\sigma = 6250$, $Z_{\alpha/2} = Z_{0.025} = 1.96$.

$$n = \frac{\sigma^2 \left(Z_{\alpha/2}\right)^2}{E^2} = \frac{(6250)^2 (1.96)^2}{(500)^2} \approx 600.25$$

Therefore we need to sample at least 601 such incomes.

6. Cars pass by a gas station along a remote freeway at a Poisson rate of 3 per hour.

   (a) Sam is hitch-hiking. When he arrived at the gas station, one car had just passed by. What is the probability that Sam has to wait for more than 30 minutes before any car would come by?
   Answer: Let T denote the inter-arrival time between the last and the next car, then

$$T \sim \textit{Exponential}\,(\lambda = 3)\,.$$

   Therefore the probability that the inter-arrival time will exceed 30 minutes (0.5 hour) is

$$P\,(T > 2) = e^{-\lambda t} = e^{-3 \cdot 0.5} = e^{-1.5} \approx .223.$$

   (b) Suppose one third of the cars passing by are vans. Then vans pass by the gas station following a Poisson process with rate $= \left(3 * \frac{1}{3}\right) = 1$ per hour. What is the chance that you will see two vans passing by within one hour?
   Answer: Let X denote the number of vans passing by within 1 hour, then

$$X \sim \textit{Poisson}\,(\theta = \lambda t = 1 \cdot 1 = 1)$$

   Hence

$$P\,(X = 2) = e^{-\theta}\frac{\theta^2}{2!} = e^{-1}\frac{1}{2} \approx .184$$

   (c) How many cars on the average will pass by the gas station in a 24-hour period?
   Answer: Let X denote the number of cars passing by within a 24-hour period, then

$$X \sim \textit{Poisson}\,(\theta = \lambda t = 3 \cdot 24 = 72)$$

   Hence

$$E\,(X) = \theta = 72$$

   Therefore we expect to see an average of 72 cars passing by per 24-hour period.

7. Urn 1 contains five red chips and four white chips; urn 2 contains four red and five white. Two chips are to be transferred from urn 1 to urn 2. Then a single chip is to be drawn from urn 2. What is the probability that the chip drawn from urn 2 will be white?
Answer: Let B be the event "white chip is drawn from urn 2". Let A_i, $i = 0, 1, 2$, denote the event "i white chips are transferred from urn 1 to urn 2". Then,

$$\begin{aligned}
P\,(B) &= P\,(B \mid A_0)\,P\,(A_0) + P\,(B \mid A_1)\,P\,(A_1) + P\,(B \mid A_2)\,P\,(A_2) \\
&= \left(\frac{5}{11}\right)\frac{\binom{4}{0}\binom{5}{2}}{\binom{9}{2}} + \left(\frac{6}{11}\right)\frac{\binom{4}{1}\binom{5}{1}}{\binom{9}{2}} + \left(\frac{7}{11}\right)\frac{\binom{4}{2}\binom{5}{0}}{\binom{9}{2}} \\
&= \left(\frac{5}{11}\right)\left(\frac{10}{36}\right) + \left(\frac{6}{11}\right)\left(\frac{20}{36}\right) + \left(\frac{7}{11}\right)\left(\frac{6}{36}\right) = \frac{53}{99} \approx 0.535
\end{aligned}$$

8. During a power blackout, 100 persons are arrested on suspicion of looting. Each is given a polygraph test. From past experience it is known that the polygraph is 90% reliable when administered to a guilty suspect and 98% reliable when given to someone who is innocent. Suppose that of the 100 persons taken into custody, only 12 were actually involved in any wrongdoing. What is the probability that a given suspect is innocent given that the polygraph says he is guilty?
Answer: Let B be the event "Polygraph says suspect is guilty" and let A_1 and A_2 be the events "Suspect is guilty" and "Suspect is not guilty", respectively. To say that the polygraph is "90% reliable when administered to a guilty subject" means that $P(B \mid A_1) = 0.90$. Similarly, the 98% reliability for innocent suspects implies that $P(\overline{B} \mid A_2) = 0.98$, or equivalently $P(B \mid A_2) = 0.02$. We also know that $P(A_1) = 12/100$ and $P(A_2) = 88/100$. Using the Bayes' Theorem, we have

$$\begin{aligned} P(A_2 \mid B) &= \frac{P(B \mid A_2) P(A_2)}{P(B \mid A_1) P(A_1) + P(B \mid A_2) P(A_2)} \\ &= \frac{(0.02)(88/100)}{(0.90)(12/100) + (0.02)(88/100)} = 0.14. \end{aligned}$$

9. A fourth-grade teacher thinks that his students are on the whole better spellers than his colleague's students. Six students were selected from each teacher's class and are matched by their IQs. They were then given a standardized spelling test. Use the following results to determine whether the fourth-grade teacher is justified in his claim. Use $\alpha = 0.05$.

IQ-matched Pairs	1	2	3	4	5	6
Foruth-grade class	20	19	16	24	21	20
Colleague's class	15	22	20	16	18	21
Difference (d)	5	-3	-4	8	3	-1

Answer: This is inference on two population means. Paired data. However, by taking the pairwise differences (d), it reduces to an inference on one population mean problem.
$H_0 : \mu_d = 0$; $H_a : \mu_d > 0$. Assume the population of the differences is normal, the test statistic under the null hypothesis is

$$t_0 = \frac{\overline{d} - 0}{s_d/\sqrt{n}} = \frac{1.33}{4.76/\sqrt{6}} \approx 0.68$$

Since $t_0 = 0.68 < t_{5,.05} = 2.015$, we can not reject H_0 at $\alpha = .05$. So there is not sufficient evidence to claim that the fourth grader teacher's students are better spellers. (There are two reasons for me to give this problem. (1), I wish you to know that a paired sample t-test is indeed a one sample t-test on the paired differences. (2) I wish you to know that you need to be able to calculate the sample mean and sample standard deviation for a given sample.)

10. Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5000-mile trip, what is the probability that he or she will be able to complete the trip without having to replace the car battery?
Answer: It follows by the memoryless property of the exponential distribution that the remaining life time X (in thousands of miles) of the battery is exponential with parameter $\lambda = 1/10$. Hence the desired probability is

$$P(X > 5) = e^{-5\lambda} = e^{-0.5} \approx 0.604$$