Fall 2014 **Statistics 330** **Final Exam**

(120 points)

Please write your first and last name here:

Name ____

Instructions:

- There are **6** questions, which worth in total **120** points.
- Partial credit will be given only if you show your work.
- Reason out your answers. In many cases, a line or two of justification is enough.
- The questions are roughly in the order in which the material is presented in class, so they are not necessarily ordered easiest to hardest.
- If you get stuck on one, it may be a good idea to move on and come back to that question at the end.
- You may use your prepared notes (1 page, both sides) and a calculator only.

1. Assume 80% of all police cars are Ford Crown Victorias and 70% of all Crown Victorias on the road are police cars. (Note that these two are conditional events.) Moreover, 1% of the cars on the road are both Ford Crown Victorias and police cars. Suppose there is a car behind your car. Let A be event that the car is a Crown Victoria and B be the event that it is a police car.

   (a) If you see a Crown Victoria in your rear-view mirror, what is the probability that it is a police car? (5 points)

   Answer: That is, $P(B|A) = 0.7$.

   (b) What is the $P(A)$ and $P(B)$? (8 points)

   Answer: From the question, we know $P(A|B) = 0.8$, $P(B|A) = 0.7$ and $P(A \cap B) = 0.01$.

$$P(A) = P(A \cap B)/P(B|A) = 0.01/0.7 = 0.01429$$
$$P(B) = P(A \cap B)/P(A|B) = 0.01/0.8 = 0.0125$$

   (c) Are event A and B independent? Why or why not (7 points)

   Answer: $P(A)P(B) = 0.0001786 \neq 0.1 = P(A \cap B)$. Thus they are not independent. (alternatives: (1)$P(A|B) \neq P(A)$, (2) $P(B|A) \neq P(B)$)

2. Let X be the proportion of the total time that code spends in a particular function and we assume X has the following probability density function.

$$f(x) = \begin{cases} cx & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value c that makes this a valid probability density function. (4 points)

Answer:

$$1 = \int_0^1 cx dx = c\frac{x^2}{2}|_0^1 = c/2 \implies c = 2$$

(b) What is the probability that less than half time is spent in this function? (4 points)

Answer:

$$\int_0^{0.5} 2x dx = x^2|_0^{0.5} = 0.25$$

(c) What is the expected proportion of time spent in this function? (4 points)

Answer:

$$E[X] = \int_0^1 2x^2 dx = \frac{2x^3}{3}|_0^1 = 2/3$$

(d) What is the variance for the proportion of time spent in this function? (4 points)

Answer:

$$E[X^2] = \int_0^1 2x^3 dx = \frac{x^4}{2}|_0^1 = 1/2$$

and

$$Var[X] = E[X^2] - E[X]^2 = 1/2 - 4/9 - 1/18$$

(e) If a large collection of data, $x_1, \ldots, x_n$ are obtained, what is an approximate distribution for $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$? (4 points)

Answer: The central limit theorem says $\overline{x} \xrightarrow{d} N(2/3, 1/18n)$.

3. There are two servers in a system. The jobs are sent to the system at a rate of 1 per minute. If all two servers are being used, the arrival jobs stay in a queue. The service time of a job follows an Exponential distribution with mean 0.5 minute.

   (a) What is this queuing system? (5 points)

   Answer: It is a M/M/2 queuing system.

   (b) Find the probability that there is zero job in the system. (5 points)

   Answer: Note that $\lambda = 1$, $\mu = 2$, $c = 2$, $a = \lambda/\mu = 0.5$ and $\rho = a/c = 0.25$.

$$p_0 = \left(\sum_{k=0}^{c-1} \frac{a^k}{k!} + \frac{a^c}{c!} \frac{1}{1-\rho} \right)^{-1} = \left(1 + 0.5 + \frac{0.5^2}{2!} \cdot \frac{1}{1-0.25} \right)^{-1} = 3/5 = 0.6.$$

   (c) Find the average length of the queue. (5 points)

   Answer:

$$L_q = p_0 \cdot \frac{a^c}{c!} \cdot \frac{\rho}{(1-\rho)^2} = (0.6) \frac{0.5^2}{2!} \frac{0.25}{(1-0.25)^2} = 1/30 = 0.03333$$

.

   (d) Find the average service time. (5 points)

   Answer:

$$W_s = \frac{1}{\mu} = 0.5$$

4. Let $x_1, x_2, \ldots, x_n$ be a sample from the distribution with following probability mass function:

x	$P(X = x)$
0	θ
1	2θ
2	$1 - 3\theta$

(a) For what values of θ is this a valid probability mass function? (4 points)

Answer: To be a valid probability mass function, we need $0 \leq P(X = x) \leq 1$ for all x. So at the smallest $\theta > 0$ and at the largest $\theta < 1/3$.

(b) Derive the method of moments estimator for the unknown parameter θ. (8 points)

Answer: $E[X] = 0 \times \theta + 1 \times 2\theta + 2 \times (1 - 3\theta) = 2 - 4\theta$. Setting this equal to the first population moment, we have $\overline{x} = E[X] = 2 - 4\theta$ and solving for θ we get $\hat{\theta}_{MoM} = (2 - \overline{x})/4$.

(c) Derive the maximum likelihood estimator (MLE) for the unknown parameter θ. (Hint: The likelihood is $\theta^{n_0}(2\theta)^{n_1}(1 - 3\theta)^{n_2}$ where n_0 is the number of 0s, n_1 is the number of 1s, and n_2 is the number of 2s.) (8 points)

Answer:

$$
\begin{aligned}
L(\theta) &= \theta^{n_0}(2\theta)^{n_1}(1 - 3\theta)^{n_2} \\
\ell(\theta) &= n_0 \log(\theta) + n_1 \log(2\theta) + n_2 \log(1 - 3\theta) \\
\frac{\partial}{\partial \theta}\ell(\theta) &= \frac{n_0}{\theta} + \frac{n_1}{\theta} - 3\frac{n_2}{1 - 3\theta} \overset{set}{=} 0 \implies \\
3\theta n_2 &= n_0(1 - 3\theta) + n_1(1 - 3\theta) \\
&= n_0 + n_1 - 3n_0\theta - 3n_1\theta \implies \\
3\theta(n_0 + n_1 + n_2) &= n_0 + n_1 \implies \\
\hat{\theta}_{MLE} &= \frac{n_0 + n_1}{3(n_0 + n_1 + n_2)} = \frac{n_0 + n_1}{3n}
\end{aligned}
$$

where $n = n_0 + n_1 + n_2$. This should be a bit different.

5. An experiment was conducted to compare two popular pain-killers. The time needed to stop the pain, in minutes, was recorded after each subject took his/her pain-killers:

	Pain-Killer A	Pain-Killer B
sample size	50	50
sample average	15	18
sample variance	36	25

Let μ_A and μ_B be the mean time needed to stop the pain by pain-killer A and B respectively.

(a) Compute a 95% confidence interval for $\mu_A - \mu_B$. (10 points)

Answer: 95% confidence interval for $\mu_A - \mu_B$ is

$$\begin{aligned} &\bar{x} - \bar{y} \pm z_{0.025}\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}} \\ &= (15 - 18) \pm 1.96\sqrt{\frac{36}{50} + \frac{25}{50}} \\ &= -3 \pm 2.16 \end{aligned}$$

That is $(-5.16, -0.84)$.

(b) Perform a hypothesis test on $H_0 : \mu_A - \mu_B = 0$ vs. $H_1 : \mu_A - \mu_B \neq 0$ at significance level 0.1. (10 points)

Answer: The test statistic is:

$$z = \frac{\bar{x} - \bar{y} - 0}{\sqrt{s_x^2/n + s_y^2/m}} = \frac{15 - 18}{\sqrt{36/50 + 25/50}} = -2.72$$

The p-value is $P(|Z| > 2.72) = 2P(Z < -2.72) = 2 \times 0.0033 = 0.0066 < 0.1$. Thus H_0 is rejected at significance level 0.1.

6. A professor is interested in the relationship between how long students work on an exam and the score they get on the exam. For one particular semester, the professor records the time in minutes each student finished the exam and their numeric score (out of 120) on the exam. Here are some summary statistics from the data:

$$\overline{x} = 92 \quad s_x^2 = 350 \quad \overline{y} = 99 \quad s_y^2 = 110 \quad s_{xy} = \frac{1}{55-1}\sum_{i=1}^{55}(x_i - \overline{x})(y_i - \overline{y}) = -164$$

(a) Estimate the slope (β_1) and x the intercept parameter (β_0) of a regression line of exam score (y) on time (x). (6 points)

Answer: $\hat{\beta}_1 = \frac{-164}{350} = -0.469$ and $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 99 + 0.469 \times 92 = 142$.

(b) Compute the sample correlation between x and y. (6 points)

Answer: The sample correlation is $\frac{-164}{\sqrt{350 \times 110}} = -0.836$.

(c) If an individual took 60 minutes on the exam, what is their expected score?(3 points)

Answer: $142 - 0.469 \times 60 = 114$

(d) If an individual took 120 minutes on the exam, what is their expected score?(3 points)

Answer: $142 - 0.469 \times 120 = 86$

(e) Does this suggest that everybody should take less time on their exam? Why or why not? (2 points)

Answer: No. Most likely those who took less time were more prepared.

(This page is intentionally left blank.)