LAST NAME (Please Print): KEY

FIRST NAME (Please Print): ____________________

HONOR PLEDGE (Please Sign): ____________________

Statistics 111

Midterm 2

- This is a closed book exam.
- You may use your calculator and a single page of notes.
- The room is crowded. Please be careful to look only at your own exam. Try to sit one seat apart; the proctors may ask you to randomize your seating a bit.
- Report all numerical answers to at least two correct decimal places or (when appropriate) write them as a fraction.
- All question parts count for 1 point.

The mean and variance of a Poisson r.v. are both λ.

The mean and standard deviation of an Exponential r.v. are both $1/\lambda$.

The mean of a Beta r.v. is $\frac{\alpha}{\alpha+\beta}$ and the variance is $(\alpha\beta)/(\alpha+\beta)^2(\alpha+\beta+1)$.

The mean of a Gamma r.v. is $\frac{\alpha}{\beta}$ and the variance is $\frac{\alpha}{\beta^2}$.

1. You go to a restaurant and get a two-course dinner. The first course lasts for X hours, where X is exponential with parameter 3. The second course takes Y hours, where Y is independent of X and exponential with parameter 2. Let Z be the duration of the meal.

   5/6 or 0.83 What is the expected value of Z?

   $\mathbb{E}[Z] = \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 1/3 + 1/2.$

   0.60 What is the standard deviation of Z?

   $\text{Var}\,[Z] = \text{Var}\,[X + Y] = \text{Var}\,[X] + \text{Var}\,[Y] = (1/3)^2 + (1/2)^2 = 0.361$. Take the square root to get 0.601.

   1/9, or 0.11 What is the covariance between X and Z?

   $\text{Cov}[X, Z] = \mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z] = \mathbb{E}[X(X + Y)] - (1/3)(5/6) = \mathbb{E}[X^2] + \mathbb{E}[XY] - (1/3)(5/6) = \text{Var}\,[X] + (\mathbb{E}[X])^2 + \mathbb{E}[X]\mathbb{E}[Y] - (1/3)(5/6) = (1/3)^2 + (1/3)^2 + (1/3)(1/2) - (1/3)(5/6) = 1/9.$

   0.55 What is the correlation between X and Z?

   $\text{Corr}(X, Z) = \text{Cov}(X, Z)/\sqrt{\text{Var}\,[X]\text{Var}\,[z]} = (1/9)/\sqrt{(1/9)(0.361)} = 0.5547.$

2. Let $X_1, \ldots, X_n$ be a random sample from the distribution with probability mass function

   $$f(x; \theta) = \frac{\theta^x}{x!}\exp(-\theta) \text{ for } x = 0, 1, \ldots$$

   and 0 else, where $\theta \geq 0$.

   Write the likelihood function.

   $f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!}\exp(-\theta)$ for x_i a nonnegative integer and $\theta \geq 0$.

$\bar{x}$ Find the maximum likelihood estimate of θ.

Take the log of the likelihood to get $-n\theta + \ln \sum \frac{1}{x_i!} + (\ln \theta) \sum x_i$. The derivative w.r.t. θ is $-n + \frac{1}{\theta} \sum x_i$. Set to 0 and solve to get $\bar{x}$.

2θ What is the bias of the estimate $\hat{\theta} = 3\bar{X}$?

$\mathbb{E}[3\bar{X}] - \theta = 3\theta - \theta$.

$\frac{9\theta}{n}$ What is the variance of $\hat{\theta} = 3\bar{X}$?

$\text{Var}\,[3\bar{X}] = 9\text{Var}\,[\bar{X}] = 9\frac{\theta}{n}$.

3. 0.60 The number of days Mary takes to read a novel has Gamma distribution with parameters $\alpha = 6$ and $\beta = 2$. What is the approximate probability that she reads more than 10 or more books in January?

   This is a CLT for sums question. What is the chance that the sum of 10 Gamma random variables is less than 31? The sum is approximately normal with mean 10 * (6/2) = 30 and standard deviation $\sqrt{10} * \sqrt{6/(2^2)}$. So the z-transformation is $z = (31 - 30)/\sqrt{10 * (6/4)} = 0.258$. From the table, the chance that the sum is less than 31 is the area under the curve and to the left of 0.258, or 0.6026.

4. 5 Consider the linear congruential generator $X_{n+1} \equiv 11X_n \pmod 7$ with seed (X_0) 6. What is X_2?

   Starting with $X_0 = 6$, 11 * 6 = 66, which is divisble by 7 with a remainder of 3. Now $X_1 = 3$, so 3 * 11 = 33, which is divisible by 7 with a remainder of 5.

5. 0.72 A statistics class has three lab sections, each with 25 students. I give a quiz on confidence intervals to the first section, and 20 students answer correctly. Set a 90% lower confidence bound on the proportion of students in the class who understand confidence intervals.

This is a CI on a probability with an FPCF. The point estimate is $20/25 = 0.8$. The standard error is $\sqrt{0.8 * 0.2/25}$, and this must be multiplied by the FPCF, which is $\sqrt{(75-25)/(75-1)}$. The critical value comes from the z-table: -1.28 has area under the curve and to the right equal to 0.9. Put this together to get 0.7158

6. Mary gets a Poisson number of phone calls in an hour. I think she is quite popular, and my belief about the average number of calls she receives is Gamma(5,2).

   ____________________ I observe that she receives 3, 2, and 1 phone calls over the next three hours. What distribution reflects my new belief about her average? (Include parameter values.)

   This is the Gamma-Poisson conjugate family. The posterior is Gamma$(\alpha + \sum x_i, \beta + n)$ = Gamma(11, 5).

   11/5 or 2.2 What is my best guess for the mean number of phone calls per hour, if my penalty is proportional to the square of my error.

   For squared error loss, use the posterior mean, which is 11/5.

7. You want to place a confidence interval (CI) on the median number of hours slept by Duke students on the night before a statistics exam. You sample 50 random students; each reports the hours they slept and the median of that sample was 7.25 You resample 20 times from that sample (with replacement) and the medians of the 20 resamples are:

   5, 2, 8.5, 7, 4.5, 6, 5.5, 5.5, 7.5, 9, 6, 2.5, 5, 3.5, 8, 3, 4, 6.5, 7, 6.5

   2.25 Set a 95% one-sided lower CI with the percentile bootstrap.

   The lower 95% CI is the 5th percentile, which is any number between the two smallest values, which are 2 and 2.5. We take the average.

Set a 90% CI with the pivot bootstrap: L = 5.75 U = 12.25

First find the 5th and 95th percentiles. This is 2.5, as before, and the average of the two largest values, 8.5 and 9, or 9.75. The $L = 2 * \hat{\theta} - 9.75 = 2 * 7.25 - 9.75 = 5.75$. Similarly $U = 2 * 7.25 - 2.25 = 12.25$.

8. Let $X_1, \ldots X_n$ be a random sample from the exponential distribution with parameter $\lambda = 4$ Write the density function for the sample maximum. (Indicate support.)

$G(y) = \mathbb{P}[\max\{X_1, \ldots, X_n\} \leq y] = \prod \mathbb{P}[X - i \leq y] = [1 - \exp(-4y)]^n$. So take the derivative to get the density: $g(y) = 4n[1 - \exp(-4y)] \exp(-4y)$ for $y \geq 0$.

9. Suppose $f(x, y) = c$ on the region $x^2 + y^2 \leq 1$ with $y > 0$.

$2/\pi$ or 0.64 What is c?

The region is the half circle with radius 1, so the area is $\pi/2$. In order to integrate to 1, $c = 2/\pi$.

What is $f_2(y)$? Indicate support.

Integrate out x where $-\sqrt{1 - y^2} \leq x \leq \sqrt{1 - y^2}$. So

$$f_2(y) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} 2/\pi \, dx = \frac{4}{\pi}\sqrt{1 - y^2} \text{ for } 0 \leq y \leq 1.$$

What is $g_2(y \mid x = 1/2)$? Indicate support.

Since the joint density is constant, then the conditional is uniform on the range between $y = 0$ and $y = \sqrt{1 - x^2} = 0.8660$. So $g_2(y \mid x = 1/2) = 1/0.8860 = 1.15$ on $0 \leq y \leq 0.87$ and is otherwise 0.

0 What is $\mathbb{E}[X]$?

By symmetry, 0.

0 What is the correlation between X and Y?

By symmetry, the covariance is 0, so the correlation is 0.

No Are X and Y independent?

No. Knowing $X = 1/2$ reduced the range of values Y can take.

10. Wally West is pretty fast—I think his probability of winning a race has the Beta(3,1) distribution. He participates in five races and wins each.

Beta(8,1) What is my new belief about p, his probability of winning?

This is the Beta-Binomial conjugate family, so the posterior is Beta$(\alpha + x, \beta + n - x)$ = Beta(8, 1).

8/9 or 0.89 What is my best estimate of p if I pay a penalty proportional to the square of my error?

The posterior mean, which is 8/(8 + 1).

0.92 What is my best estimate of p if I pay a penalty proportional to the absolute value of my error?

This is the posterior median. Call it m. Then $0.5 = \int_0^m 8x^7\,dx = m^8$. The eighth root of 0.5 is 0.917.

5/9 or 0.56 How much weight does the posterior mean place upon the sample mean?

$a * (5/5) + (1 - a)(3/4) = 8/9$. Solving for a gives 5/9.

11. 0.41 The number of alcoholic beverages a Duke freshman consumes in a week has the Poisson distribution with parameter $\lambda = 3.5$. If you sample 20 students, what is the approximate probabilty that $\bar{X}$ is greater than 3.6?

CLT for averages. Since $\bar{X}$ is approximately normal with mean equal to the true mean 3.5 and standard deviation equal the population standard deviation divided by $\sqrt{n}$, or $\sqrt{3.5/20}$. So the z-transformation is $(3.6 - 3.5)/\sqrt{3.5/20}) = 0.239$. From the table, the probability of a value greater than 0.24 is 0.4052.

12. 9.98 Dean Kostyu wants a 95% upper confidence bound on the mean number of ounces of alcohol that Duke freshman consume in a week. A sample of 12 students has average 7.8 and standard deviation 4.2.

This is a CI on an average using the t-table. The point estimate is 7.8, the standard error is $4.2/\sqrt{12}$ and the critical value for a t-distribution with 12 - 1 = 11 degrees of freedom is 1.796. Putting this together gives 9.9775.

13. List all, and only, the true statements. A, B, H, I

    A. Response bias concerns how a survey question is worded.
    B. The continuum of non-response method assumes those who are never reached in a survey are similar to those who are difficult to reach.
    C. Household bias says that large households are over-represented.
    D. Dr. John Snow stopped a smallpox epidemic in London.
    E. William Playfair made a graph showing Napoleon's attack upon and retreat from Moscow.
    F. An age pyramid for Sierra Leone would be tall and narrow.
    G. If X and Y are negatively correlated, then $\text{Var}[X + Y] > \text{Var}[X - Y]$.
    H. The $\text{Var}\left[\sum a_i X_i\right] = \sum a_i^2 \sigma_i^2 + 2 \sum_{i<j} a_i a_j \text{Cov}(X_i, X_j)$.
    I. As confidence increases, so does the width of the confidence interval on μ.
    J. As the sample size increases, so does the width of the confidence interval on μ.