Name: ____

Biostatistics 1st year Comprehensive Examination: Theory

June 5th, 2017: 9am to 5pm

Instructions:

1. There are six questions and 6 pages (not including the cover sheet).
2. Answer each question to the best of your ability. Be as specific as possible and write as clearly as possible.
3. Put your name and problem number on every sheet of paper; **only use one side** of the paper (the exams will be scanned electronically).
4. This is an in-class examination; do not discuss any part of this exam with anyone while you are taking the exam. **NO BOOKS, NO NOTES, NO FRIENDS, NO PETS, NO INTERNET DEVICES, and NO OUTSIDE ASSISTANCE.**
5. You may leave the examination room to use the restroom or to step out into the hallway for a short break. **HOWEVER, YOU MUST LEAVE YOUR CELL PHONE AND ALL EXAM MATERIALS IN THE EXAMINATION ROOM.** If there is an emergency, please discuss this with the exam proctor.
6. Vanderbilt's academic honor code applies; *adhere to the spirit of this code.*

Question	Points	Score	Comments
1	50		
2	50		
3	50		
4	50		
5	50		
6	50		
Total	**300**		

1. Suppose $X_1, \dots, X_n$ are independent exponentially distributed random variables such that

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad \text{for} \quad x > 0, \;\; \theta > 0$$

   a. Define and derive the cumulative distribution function of X_1.

   b. Find the median of the exponential distribution.

   c. Define and derive the moment generating function of X_1. (Hint: $M_{X_1}(t)$ exists for all $t < 1/\theta$).

   d. Find the expectation and variance of X_1.

   e. Derive the distribution of $Y = X_1 + X_2$ and find $E[Y]$ and $Var(Y)$.

   f. Find the covariance between X_1 and Y.

   g. Describe a procedure for randomly generating an exponential X_i using only the uniform distribution.

   h. What is the distribution of $X_{(1)} = \min(X_1, \dots, X_n)$?

   Denote the mean of a sample by $\overline{X} = \sum_{i=1}^{n} X_i / n$.

   i. What is the exact distribution of $\overline{X}$?

   j. Provide an approximate, the large sample distribution for $\log(\overline{X})$. Justify your answer.

2. Let $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where $\sigma^2 > 0$, and let Z_n be a random variable that takes values $\{-an, 0, an\}$ with probability $\{1/n^2, 1 - 2/n^2, 1/n^2\}$ for some constant $a > 0$. Let Y_i and Z_n be independent and define their sum to be:

$$W_n = \overline{Y}_n + Z_n$$

   a. What is the limiting mean of W_n?

   b. Suggest a change to the support of Z_n that would result in W_n having a negative finite bias but no limiting bias.

   c. What is the limiting mean square error of W_n?

   d. Define convergence in probability, convergence almost surely, and convergence in distribution. Draw a diagram that displays at least three relationships among these convergences. That is, show which type of convergence implies another and note any required conditions.

   e. Is W_n is consistent for μ? Justify your answer and show that all necessary conditions are satisfied.

   f. Consider what happens when $a = \sqrt{n}$. Find the MSE Z_n and show that Z_n still converges to 0 in probability.

3. Suppose that $X|p \sim Binomial(n, p)$ and that $p \sim Beta(\alpha, \beta)$ such that

$$f(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \quad \text{for} \quad \alpha, \beta > 0$$

where

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad \text{with } \Gamma(\alpha) = (\alpha-1)!$$

a. Derive $E[p]$ and $Var[p]$.

b. What is marginal mean of X, $E[X]$?

c. What is marginal variance of X, $Var[X]$?

d. Find the marginal distribution of X. When might the marginal distribution of X be used as the statistical model instead of $X|p$?

e. What is the posterior distribution of p?

f. Provide a 95% Bayesian credible interval for p.

g. Provide a competing $1/k$ likelihood support interval for p. Give the definition of the SI and explain your model choice (Note: Simplify when possible; you do not have to find the maximize the likelihood estimate; Just make your notation clear).

4. Suppose we observe pairs $(Y_1, x_1), \dots, (Y_n, x_n)$ where the x_i's are known constants and the Y_i's are independent normal random variables with mean 0 and variance ϕ/x_i^2 for some $\phi > 0$.

   a. Construct the likelihood for ϕ based on observations $y_1, \dots, y_n$.

   b. Find the minimal sufficient statistic for ϕ.

   c. Show that the minimal sufficient statistic is also a complete sufficient statistic.

   d. Find the maximum likelihood estimator of ϕ, denote it by $\hat{\phi}$.

   e. Find the MVUE of ϕ.

   f. Does the variance of the MLE achieve the Cramer-Rao lower bound in finite samples?

   g. Construct an approximate large-sample 95% confidence interval for ϕ. Be sure to explain why and when the confidence interval is valid.

Let $S^2 = \sum_i \left(Y_i - \overline{Y}_n\right)^2 / (n-1)$ be the sample variance.

   h. Does the student's t-interval

$$\overline{Y}_n \pm t_{\alpha/2}^{n-1} \frac{S}{\sqrt{n}}$$

   provide an exact 95% CI for 0? If yes, justify your answer. If not, explain what the problem is and suggest a way to improve the properties of the t-interval.

5. We want to estimate the size of the white tiger population in Mukundpur Zoological Park in India. The plan calls for repeated safaris in the park, recording how many white tigers were observed each time. Suppose we embark on $k > 1$ safaris.

   Consider the model $Y_i \sim Bin(N, \theta)$ for $i = 1, \dots, k$ where $N \geq \max(Y_1, \dots, Y_k)$ is the white tiger population size and θ is the probability of observing a white tiger.

   a. What are the key assumptions of this model for estimating the population size? Briefly explain why these assumptions may or may not be reasonable.

   We execute the plan and observe $y_1, \dots, y_k$ white tigers on k different safaris.

   Suppose it is well known that the probability of observing a white tiger is θ_t.

   b. On the basis of our sample, what is the strength of evidence for a large population size, say $H_l: N = N_l$, versus a small one, say $H_s: N = N_s$?

   c. Find the most powerful α–sized test of $H_l: N = N_l$ versus $H_s: N = N_s$. Justify your answer and explain how to find the critical region for this test. Note: Simplify whenever possible but do not spend time reducing factorials.

   Now suppose the probability of observing a white tiger on safari is unknown.

   d. Find the profile likelihood for N, $L_p(N)$, based on the observed sample.

   e. Suggest a way to measure the strength of evidence for $H_l: N = N_l$ versus $H_s: N = N_s$? Briefly describe your approach and note its strengths and weakness; contrast your approach with your answer from part (b).

   f. Suggest an α–sized hypothesis test of $H_0: N \leq 100$. Provide the test statistic and explain how we could find its distribution under the null hypothesis. Note: Simplify if possible and do not spend time reducing factorials.

6. Dean Balser will select our next department chair by deciding between 3 candidates, persons A, B, and C. He makes the decision by inviting them all to his office to play "odd person wins." The game consists of each candidate flipping a coin. If one coin is different from the other two, that person is the winner (the 'odd' flip). If all three coins are the same, then no one wins and they flip again. Each flip is a round. Let p_A be the probability that person A is the winner on a particular flip, and let p be the probability that there exists an odd person on any particular flip.

   a. Evaluate p_A and p.

   b. Let Y represent the number of flips (i.e., rounds) required until someone wins the game. What is the distribution of Y? (Be sure to define your notation).

   c. What is the mean of Y?

   One day, Dean Balser decides to change the game. Now the candidates play 5 rounds of "odd person out" and the candidate who has won the majority of the 5 rounds will become the chair. If there is a tie after 5 games (i.e., 2 candidates each win 2 games), then all three continue to play until one candidate wins three rounds. Whoever wins three rounds first is named department chair.

   d. What is the probability that candidate A wins 3 of the 5 rounds?

   e. What is the probability that after 5 rounds, there is a tie?

   f. Given that candidate A wins the first round, what is the probability that he/she goes on to become the department chair?

   Now suppose that Dean Balser favors persons B and C. He changes the rules to the following: If candidate B or C is the odd person in the first round, then the winner of that round will be the new chair. However, if candidate A wins the first round, they continue on playing the best out of 5 rounds (as described above) to pick the new chair.

   g. What is the probability that candidate A becomes the new chair?