In-Class Statistics Masters and Ph.D. Qualifying Exam

August 2020

Instructions: *The exam has 7 multi-part problems. All of the problems will be graded. Write your code words on each of your answer sheets. Do not put your name on any of the sheets. Be clear, concise, and complete. All solutions should be rigorously explained.*

Problem 1. Consider a round-robin tournament for 16 sports teams. In a round-robin tournament, n teams are put into $n/2$ pairs. Each pair of teams plays a game in round 1, and the winner moves to the next round. In round 2, the $n/2$ teams are paired, and the winners move on to round 3, which will have $n/4$ teams. This continues until the final round has two teams, and the winner of this final game wins the tournament. In a round-robin tournament with 2^n teams, there are n rounds.

Suppose the round-robin tournament has 3 teams from Albuquerque and 13 teams from elsewhere in New Mexico. For (a)–(b) Assume that the initial pairings of the teams is done at random, so pairings are equally likely. For (a)–(c), assume that each team is equally skilled so that each team has a 50% chance of winning each game, and that all games are independent. Finally, assume that each game results in one team winning, so that there are no ties.

(a)[6pts] What is the probability that one of the teams from Albuquerque wins the tournament?

(b)[6pts] Find the probability that in the first round, two of the teams from Albuquerque play each other.

(c)[6pts] Instead of pairing teams at random, suppose that in round 1, two of the Albuquerque teams are paired, while the third Albuquerque team is paired with another team at random. This guarantees that at least one Albuquerque team wins and at least one Albuquerque team loses in the first round. Also suppose that if there are two Albuquerque teams available in round 2, they are planned to play each other. Thus in round three there are either 0 or 1 Albuquerque teams. Under this setting, find the probability that an Albuquerque team wins the tournament and compare the answer to part (a).

Problem 2. Suppose when I go for an evening walk, I encounter dog-walkers (people walking one or more dogs) according to a Poisson process with rate 1 per kilometer. If I do the full walk, it is two kilometers each way (four kilometers round trip). However, being very afraid of dogs, if I encounter a dog-walker on the first half of the walk, I immediately turn around and come back home. If I encounter a dog-walker on the second half of the walk, I am already on the way home, so it doesn't alter my walk. Recall that for Poisson random variables, $P(X = i) = e^{-\lambda}\lambda^i/i!$, $i = 0, 1, \dots$, and that for a Poisson process, the time t between events is exponential with rate λ, so that for $t > 0$, $P(T < t) = 1 - e^{-\lambda t}$.

(a)[5pts] What is the probability that I do not encounter any dog-walkers on the entire walk?

(b)[5pts] What is the expected length of the walk?

(c)[5pts] Let X be the total length of the walk. Draw a CDF for X and make sure it is clearly labeled.

(d)[5pts] What is the expected number of dog-walkers encountered? (Assume dog-walkers are independently encountered and walk dogs alone.)

Problem 3. Consider two methods of testing for infections in patients. In the first method, each patient is individually tested at a cost c per patient and probability p of infection per patient. Assume all tests are independent. In order to reduce the cost of testing for infections, a second method samples from n patients. The samples are pooled, and the pooled sample is tested once. If the test is negative, no further testing is required. If the test is positive, this means that at least one of the patients is positive, so then each of the n patients has to be separately tested. Again assume that the probability that a patient is positive is p, and that each patient is independent. Assume that both the global test and individual tests have cost c each. Thus, if the pooled test is positive, then the cost for testing is $c + nc$ for the initial test plus the n additional tests.

(a)[6pts] Find the expected costs of testing n patients using both pooled and individual tests.

(b)[6pts] Find a function $f(c, n, p)$ of c, n, and p, such that if $f(c, n, p) > 0$, then the pooled test has lower expected cost than testing each patient independently.

Problem 4. Let $X_1, \ldots, X_n$ be distributed as

$$f(x) = \begin{cases} \theta & 0 \leq x \leq \dfrac{1}{\theta} \\ 0, & \text{elsewhere} \end{cases}$$

(a)[6pts] Find a minimal sufficient statistic for θ.

(b)[6pts] Is the statistic you found in the previous part complete? Show that it is or it is not complete.

(c)[6pts] Find an unbiased estimator of θ utilizing $X_{(n)} = max(X_i, i = 1, \cdots, n)$.

Problem 5. Let $X_1, \ldots, X_n$ be a random sample from a the distribution

$$f(x) = \begin{cases} (2\theta^3)^{-1} x^2 e^{(-x/\theta)} & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

(a)[6pts] Find the CRLB for unbiased estimators of θ.

(b)[6pts] Find a UMVUE of θ.

Problem 6. Let $X_1, \ldots, X_n$ be a random sample from a the distribution given in Problem 5.

(a)[5pts] Find the most powerful test for testing

$$H_0 : \theta = 2 \text{ against } H_\alpha : \theta = 4$$

at level α. Express your decision rule in terms of a "known" distribution such as t, χ^2, F, Normal. Remember to define any notation that you use for the percentiles of the distribution.

(b)[5pts] Find the UMP test for testing:

$$H_0 : \theta \leq 2 \text{ against } H_\alpha : \theta > 2$$

by first finding a MLR statistic.

Problem 7. Let $X_1, \ldots, X_n$ be a random sample from a the distribution given in Problem 5. The following questions pertain to obtaining confidence intervals. Each confidence interval should be in terms of a "known" distribution as mentioned above.

(a)[5pts] Provide an approximate 90% confidence interval for $1/\theta$ (use a CLT type result)

(b)[5pts] Provide an exact 90% confidence interval for $1/\theta$.