MS Inclass Qualifying Exam

May 31, 2017

Exam Number: _____________

NAME: ________________________________

Some Advice: The derivations should not be too long, so if you are proceeding on a path with complicated mathematical computations, regroup and try the problem again. Good Luck!!

Instructions:

1. **Write your name only on this page.**
2. **Write your exam number on every page.**
3. There are **6** problems (all with multiple parts).
4. Show your work so that we can give partial credit where appropriate.
5. Write your answers in the space provided. If you need more space, then use the scratch paper that we provide. We will then insert the extra pages into your exam.
6. You will not need a calculator.
7. The exam is closed book. You may NOT use any notes or other references.
8. Please read and sign the honor code:

   I understand that my participation in this examination and in all academic and professional activities as a CSPH student is bound by the provisions of the CSPH Honor Code. I understand that work on this exam and other assignments are to be done independently unless specific instruction to the contrary is provided.

Signature:

MS Inclass Qualifying Exam
May 31, 2017

Exam Number: ___________

Please write your exam number on all pages (front and back).

1. Question 1 part a. Suppose that two random variables, X and Y, have a bivariate discrete joint probability mass function given by

$$f_{X,Y}(x, y), \tag{1}$$

where $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Also suppose that $E(X) < \infty$ and $E(Y) < \infty$. Show that

$$E(X + Y) = E(X) + E(Y). \tag{2}$$

Question 1 part b. Now suppose that n random variables, say X_1, $X_2, \ldots X_n$ have joint probability distribution $f_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n)$, and that for all $i \in \{1, 2, \ldots, n\}$, we have $E(X_i) < \infty$. We will prove by induction that

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i). \tag{3}$$

I will give you most of the induction steps. You need only provide the proof for the last step.

Step 1. It is a tautology that

$$E(X_1) = E(X_1). \tag{4}$$

Step 2. We showed in part a that

$$E(X + Y) = E(X) + E(Y). \tag{5}$$

Step $(n-1)$. Assume that for $n-1$ random variables, say X_1, $X_2, \ldots X_{n-1}$, which have joint probability distribution $f_{X_1,X_2,\ldots X_{n-1}}(x_1, x_2, \ldots x_{n-1})$, and that for all $i \in \{1, 2, \ldots, n-1\}$, we have $E(X_i) < \infty$. Assume that

$$E\left(\sum_{i=1}^{n-1} X_i\right) = \sum_{i=1}^{n-1} E(X_i). \tag{6}$$

Step n. Consider n random variables, say X_1, $X_2, \ldots X_n$ which have joint probability distribution $f_{X_1,X_2,\ldots X_n}(x_1, x_2, \ldots x_n)$. Suppose that for all $i \in \{1, 2, \ldots, n\}$, we have $E(X_i) < \infty$. Show that

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i). \tag{7}$$

Question 1, part c. (adapted from Ross, A First Course in Probability, p. 249) A group of N men throw their hats in large bin. The hats are mixed up, and each man randomly selects one hat, all at the same time. Each man has an equal chance of getting his own hat. If a man gets his own hat, we deem this to be a match. For $i \in \{1, 2, \ldots, N\}$, let $X_i = 1$ if man i gets his own hat, and 0 otherwise. Let

$$X = \sum_{i=1}^{N} X_i, \tag{8}$$

be the total number of matches, where $0 \leq X \leq N$. Show $E(X) = 1$, using the result proven in part b.

Question 1, part d. By definition, when $E(X^2) < \infty$ and $E(X) < \infty$,

$$V(X) = E\left(X^2\right) - [E(X)]^2. \tag{9}$$

For two random variables, X and Y,

$$Cov(X,Y) = E(XY) - E(X)E(Y). \tag{10}$$

Show

$$V(X+Y) = V(X) + V(Y) + 2Cov(X,Y) \tag{11}$$

Question-1 Calculations:

Question-1 Calculations:

Question-1 Calculations:

2. For the Central Limit Theorem state:

    - The theorem (using equations with text as needed)
    - The assumptions
    - A probability distribution (pdf) where the theorem applies
    - A pdf where the theorem does NOT apply and state the reasons why

Question-2 Calculations:

3. Suppose that $X_1, \dots X_n$ are iid with a $Pareto(\gamma, \beta)$ distribution, where

$$f(x|\gamma,\beta) = \frac{\beta\gamma^\beta}{x^{\beta+1}} \quad , \text{ for } \quad \gamma \le x < \infty, \;\; \gamma > 0, \;\; \beta > 0.$$

You may assume for $\beta > 1$ that $E[X] = \frac{\beta\gamma}{\beta-1}$ and for $\beta > 2$, $Var[X] = \frac{\beta\gamma^2}{(\beta-1)^2\,(\beta-2)}$.

(a) Find a two-dimensional minimal sufficient statistic.

(b) **Assume that γ is fixed and known**.

i. Find a complete, sufficient statistic.

ii. Find the Method of Momments (MOM) estimator for β.

iii. Find the Maximum Likelihood Estimator (MLE) estimator for β.

iv. Find the Uniformly Most Powerful (UMP) test of $H_0 : \beta = 4$ vs. $H_1 : \beta = 2$.
Simplify to determine if you reject when $\prod x_i$ is 'big' or 'small'.
If you were given the distribution of $Y = \prod x_i$, how would you determine the cutoff value.

Question-3 Calculations:

Question-3 Calculations:

Question-3 Calculations:

4. Suppose that $X_1, \ldots X_n$ are iid with a $Pareto(\gamma, \beta)$ distribution, where

$$f(x|\gamma, \beta) = \frac{\beta\gamma^{\beta}}{x^{\beta+1}} \quad \text{, for} \quad \gamma \leq x < \infty, \ \ \gamma > 0, \ \ \beta > 0.$$

You may assume for $\beta > 1$ that $E[X] = \frac{\beta\gamma}{\beta-1}$ and for $\beta > 2$, $Var[X] = \frac{\beta\gamma^2}{(\beta-1)^2(\beta-2)}$.

Assume that β is fixed and known. You may also assume that the CDF is $F(x|\gamma, \beta) = 1 - (\frac{\gamma}{x})^{\beta}$.

(a) Recall that if a pdf has the form: $\frac{1}{\sigma} f(\frac{x}{\sigma})$, then σ is a scale parameter. Show that γ is a scale parameter in the Pareto distribution.

(b) Sketch the likelihood function and show that $\hat{\gamma}$, the mle, is $X_{(1)}$.

(c) Show that the distribution of $X_{(1)}$ is $Pareto(\gamma, n\beta)$. Recall:

$$f_{X(j)} = \frac{n!}{(j-1)!(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j}$$

(d) Is $X_{(1)}$ a consistent estimator of γ?

(e) Find an unbiased estimator of γ that is a function of $X_{(1)}$.

(f) Find the distribution of $Z = \frac{X_{(1)}}{\gamma}$ and argue that it is a pivot. Make sure to identify the range of Z?

(g) Find a $(1 - \alpha)$ confidence interval based on Z. If you are running short on time, you may simply set up the equations.

Question-4 Calculations:

Question-4 Calculations:

Question-4 Calculations:

5. Bayesian: Assume X (one observation), is distributed with density $p(x|\theta) = (2x/\theta^2)$, $0 < x < \theta$ and $\theta > 0$. Let π denote a prior density for θ

   (a) Find the posterior density of θ when $\pi(\theta) = 1,\ 0 \leq \theta \leq 1$.

   (b) Assume that we wish to test $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$. **Set up** (no need to solve) the expressions for the posterior probailities of H_0 and H_1.

Question-5 Calculations:

6. Let the random variable T be the time to an event. Suppose that T follows an exponential distribution with mean $\frac{1}{\lambda}$, and distribution function:

$$f(t) = \lambda e^{-\lambda t}$$

(a) Suppose that a study has n individuals with events at times t_i for $i = 1, ..., n$.

i. Give the likelihood for the data from the sample.

ii. Derive the maximum likelihood estimate for λ.

(b) Suppose that the same study includes m individuals whose event times also follow an exponential distribution with mean $\frac{1}{\lambda}$. Suppose that these individuals have been in the study for times t_j $(j = 1, ..., m)$ without experiencing an event; thus, we only know that their event time will be larger than t_j.

i. Derive the equation for $Pr(T > t)$.

ii. Use this expression to write down the (partial) likelihood for the data from the m individuals who did not experience an event.

(c) What is the (partial) likelihood for the full sample of $n + m$ individuals? Specifically:

i. Use the results from parts (a) and (b) to derive a likelihood for the full sample comprised of the n individuals who have experienced an event and the m individuals who have not experienced an event.

Question-6 Calculations:

Question-6 Calculations:

Question-6 Calculations: