MS Inclass Qualifying Exam

June 1, 2016

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 **8** 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
June 1, 2016

Exam Number: ________

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

1. Let U be a random variable that is uniformly distributed on the interval $[0, 1]$ or $U \sim unif(0, 1)$, such that $f(u) = 1$ for $0 \leq u \leq 1$.

   (a) Give the cdf for U.

   (b) Derive the expectation of U.

   (c) Derive the variance of U.

   (d) When planning clinical trials we often evaluate the duration of a trial under the assumption that patient enrollment follows a powered uniform distribution. Specifically the enrollment time Y follows a distribution defined by $Y = \lambda U^{\rho}$ where $\lambda > 0$, $\rho > 0$, and $U \sim unif(0, 1)$.

      i. Give the pdf for Y.

      ii. Give the cdf for Y.

      iii. Derive the expectation of Y.

Question-1 Calculations:

2. Sketch the requested power functions. **Make sure to label the graphs clearly with all relevant information.**

   (a) Sketch the **ideal** power function for the hypothesis: $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$.

   (b) Assume that $\alpha = 0.05$, where α is the size of the test of: $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. For a uniformly most powerful unbiased level-α test sketch two power functions on the same graph one assuming a 'small' sample size (dashed line) and the second assuming a 'large' sample size (solid line).

   Note for this problem, you are not expected to derive a test, rather convey the important concepts in your plots.

Graph (a): Ideal Power Function.

Graph (b): Power Functions for 'small' (dashed line) and 'large' (solid line) sample sizes.

Question-2 Calculations:

3. Recall that we write

$$X \sim N\left(\mu, \sigma^2\right) \tag{1}$$

to indicate that X follows a normal distribution with mean μ and variance σ^2. Recall also that the probability density function of X is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \tag{2}$$

with moment generating function

$$\begin{aligned} \mathcal{E}\left(e^{tX}\right) &= M_X(t) && (3)\\ &= \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right] && (4) \end{aligned}$$

Let n be a positive integer. Suppose $\mu_1, \mu_2, \ldots, \mu_n \in \Re$. For $i \in \{1, 2, \ldots, n\}$, suppose that $\sigma_i^2 \in \Re$, and $\sigma_i^2 > 0$. Define a set of mutually stochastically independent random variables $X_1, X_2, \ldots X_n$, where

$$\begin{aligned} X_1 &\sim N\left(\mu_1, \sigma_1^2\right) && (5)\\ X_2 &\sim N\left(\mu_2, \sigma_2^2\right) && (6)\\ &\vdots && (7)\\ X_n &\sim N\left(\mu_n, \sigma_n^2\right). \end{aligned}$$

Finally, define a set of real constants $k_1, k_2, \ldots k_n \in \Re$, and let

$$\begin{aligned} Y &= k_1 X_1 + k_2 X_2 + \cdots + k_n X_n && (8)\\ &= \sum_{i=1}^{n} k_i X_i. \end{aligned}$$

a. Show that the moment generating function of Y is given by

$$M_Y(t) = \exp\left[\left(\sum_{i=1}^{n} k_i \mu_i\right) t + \left(\sum_{i=1}^{n} k_i^2 \sigma_i^2\right) \frac{t^2}{2}\right]. \tag{9}$$

b. Using the moment generating function, or otherwise, show that

$$\mathcal{E}(Y) = \sum_{i=1}^{n} k_i \mu_i. \tag{10}$$

c. Using the moment generating function, or otherwise,show that

$$\mathcal{V}(Y) = \left(\sum_{i=1}^{n} k_i^2 \sigma_i^2\right). \tag{11}$$

d. Show that

$$Y \sim N\left[\left(\sum_{i=1}^{n} k_i \mu_i\right), \left(\sum_{i=1}^{n} k_i^2 \sigma_i^2\right)\right]. \tag{12}$$

e. Define

$$H = X_1 - X_2. \tag{13}$$

Find real constants k_1 and k_2 so that you can write

$$H = \sum_{i=1}^{2} k_i X_i. \tag{14}$$

f. Show that

$$\Pr(H > 0) = \Pr(X_1 > X_2). \tag{15}$$

g. Using the result shown in part d, i.e., that

$$Y \sim N\left[\left(\sum_{i=1}^{n} k_i \mu_i\right), \left(\sum_{i=1}^{n} k_i^2 \sigma_i^2\right)\right], \tag{16}$$

show that

$$\Pr(H > 0) = 1 - \Phi\left[\frac{-(\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)}}\right], \tag{17}$$

where Φ is the cumulative distribution function of the standard normal distribution.

Question-3 Calculations:

Question-3 Calculations:

4. A random sample $X_1, \ldots X_n$ is a random sample drawn from a Normal distribution. Recall a $N(\mu, \sigma^2)$ pdf is $f(x|\,\mu,\,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ for $-\infty < x < \infty$. You may assume that the $E[X] = \mu$ and $Var[X] = \sigma^2$.

   (a) Assuming a $N(\theta, 1)$ distribution

      i. Show that the $N(\theta, 1)$ is an exponential family.

      ii. Find a complete, sufficient statistic.

      iii. Find the Maximum Likelihood Estimate (MLE) of $\hat{\theta}$ of θ.

      iv. Argue that the Best Unbiased Estimator (BUE) of θ is equal to the MLE.

   (b) Now assume that $X_1, \ldots X_n$ is a random sample from a $N(\theta, \theta)$ with $\theta > 0$.

      i. Show that the $N(\theta, \theta)$ is an exponential family.

      ii. Find a complete, sufficient statistic.

      iii. **Assume** the MLE, $\hat{\theta}$, is $(-1 + \sqrt{1 + 4\frac{\sum x_i^2}{n}})/2$. Find the limiting distribution of $\hat{\theta}$.

Question-4 Calculations:

5. Suppose we have observations $X_1, X_2, ...X_n$ all independent and identically distributed random variables with probability mass function:

   $p_N(x) = \frac{2x}{N(N+1)}$ for $x = 1, 2, \ldots N$ and 0 otherwise.

   Assume N is an unknown parameter to be estimated.

   (a) Find the maximum likelihood estimate, $\hat{N}$.

   (b) Does classical likelihood theory apply to this estimate? If so, what is the asymptotic variance of $\hat{N}$? If not, why?

Question-5 Calculations:

6. Suppose that $Y_1, \dots Y_n$ is a random sample from a Weibull distribution with density:

$$f(y|\, S, \lambda) = \frac{S}{\lambda}\left(\frac{y}{\lambda}\right)^{S-1} exp\left[\left(\frac{-y}{\lambda}\right)^{S}\right]$$

with $\lambda > 0$, $S > 0$, and $y \geq 0$.

(a) Is this distribution a member of the exponential family? Show why or why not.

(b) What is the minimal sufficient statistic for the Weibull distribution?

(c) <u>**Bayesian:**</u> Assume that S is known and define $\theta = \lambda^S$, such that

$f(y|\, S, \theta) = \frac{S}{\theta}\, y^{S-1}\, exp\,(\frac{-y^S}{\theta})$ for $\theta > 0,\ S > 0,$ and $y \geq 0$.

i. Show that $\pi(\theta|\, a, b)$ is a conjugate prior, where

$\pi(\theta|\, a, b) = \frac{b^{a-1}\, exp\,(\frac{-b}{\theta})}{\Gamma(a-1)\theta^a}$ for $a > 0$, $b > 0$, $\theta > 0$.

ii. Set up the equations, no need to solve, to find the probability of the null and alternative hypotheses: H_0: $\theta \leq \theta_0$ and H_1: $\theta > \theta_0$.

iii. Set up the equations, no need to solve, to obtain the $(1-\alpha)$ Highest Posterior Density (HPD) credible set for θ. What assumptions must you make?

Question-6 Calculations:

7. Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with mean θ_1 and let $Y_1, Y_2, \ldots, Y_n$ be an independent random sample from an exponential distribution with mean θ_2.

   Note: The exponential pdf is $\frac{1}{\beta} \exp^{-x/\beta}$ for $0 \leq x < \infty$ and $\beta > 0$.

   (a) Derive the likelihood maximum likelihood estimators for θ_1 and θ_2.

   (b) For the test $H_0 : \theta_1/\theta_2 = c_0$ versus $H_1 : \theta_1/\theta_2 \neq c_0$, for some known constant c_0, what is the maximum likelihood estimator of θ_1, say $\hat{\theta}_{10}$ under H_0?

   (c) What is the likelihood ratio test statistic, $\lambda(\boldsymbol{x}, \boldsymbol{y})$, for testing H_0 vs H_1?

   (d) Assuming n is large, describe how you would test H_0 vs H_1. State any additional assumptions.

   (e) Using the results from parts (c) and (d), describe how you would obtain the confidence set for θ_1/θ_2 with approximately 95% coverage probability.

Question-7 Calculations:

8. Let X be an exponential random variable with density $\frac{1}{\theta} e^{-x/\theta}, \;\; 0 \leq x < \infty$, $\theta > 0$.

   (a) What is the distribution of $Y = \frac{X}{\theta}$?

   (b) If $X_1, \ldots X_n$ are iid exponential, find the distribution of $T = \frac{\sum X_i}{\theta}$? You may assume:

      - The moment generating function (MGF) of an *exponential*(β) is $\frac{1}{1-\beta t}$.
      - The MGF of a *gamma*(α, β) is $(\frac{1}{1-\beta t})^{\alpha}$.

   (c) Identify a pivot and derive a 95% confidence interval for θ.

Question-8 Calculations:

Additional Calculations:

Additional Calculations:

Additional Calculations:

Additional Calculations:

Additional Calculations: