MS Inclass Qualifying Exam

May 30, 2018

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 **7** problems (most 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 30, 2018

Exam Number: ___________

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

1. Exponential families

   (a) Are the following from an exponential family?

      i. Let $X_1, X_2, \ldots X_n$ be iid with $f(x|\theta) = \frac{1}{\sqrt{2\pi}|\theta|} e^{\frac{-(x-\theta)^2}{2\theta^2}}$; $-\infty < x < \infty$; $-\infty < \theta < \infty$.

      ii. Let X be a single observation with density

$$p_\theta(x) = \begin{cases} \frac{1}{2}(1+\theta x), & |x| < 1 \\ 0, & |x| \geq 1 \end{cases}$$

      where $\theta \in [-1, 1]$ is an unknown parameter.

   (b) For any of the above distribution(s) that are exponential family, find a complete, sufficient statistic or argue that one doesn't exist.

   (c) Find a sufficient statistic for any of the above distribution(s) that are not exponential family.

Question-1 Calculations:

2. **Bayesian** Let $X_1, X_2, \ldots X_n$ independent, identically distributed with

$f(x|p) = (1-p)^{x-1}p; \quad x = 1, 2, 3, \ldots \; ; 0 < p < 1$ (Geometric Distribution).

Show that the Beta distribution is the conjugate prior for the Geometric Distribution. Note that the **general form** of the Beta distribution is:

$f(y|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \;\; 0 < y < 1; \quad \alpha > 0, \; \beta > 0.$

Question-2 Calculations:

3. Ladybugs lay eggs in clusters. Assume that the number of eggs in the i^{th} cluster, Y_i, is distributed according to a Poisson probability with mean μ. Also, assume that the probability of finding n clusters in a field of specified area is given by a $NegativeBinomial(r, p)$, such that

$$Pr(N = n) = \binom{r+n-1}{n} p^r q^n \ ,$$

assumptions: $r > 0; 0 \leq p \leq 1; p + q = 1, n = 0, 1, \ldots.$

You may further assume that $E[N] = \frac{r(1-p)}{p}$ and $E[Y_i] = \mu$.

Then the total number of ladybug eggs in the field becomes $T = Y_1 + Y_2 + \ldots + Y_n$, where n is a random variable denoting the number of clusters of eggs in the field.

Find the expected value of $T = \sum_{i=1}^{n} Y_i$, the total number of eggs in the field.

Question-3 Calculations:

4. Suppose we have X_1 and X_2, a random sample of size 2 from a distribution that is normal, with mean μ and variance σ^2. That is, we have $X_1 \sim N(0, \sigma^2)$ and $X_2 \sim N(0, \sigma^2)$, and, because they arise from a random sample, we have that $X_1 \perp X_2$. We read this as X_1 is independent of X_2.

   Recall that if $X_1 \sim f_{X_1}(x_1)$, $X_2 \sim f_{X_2}(x_2)$ and $X_1 \perp X_2$, then

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2).$$

   Using this theorem, we have for $(x_1, x_2) \in \Re^2$

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)\exp\left(-\frac{1}{2}\right)\left(\frac{x_1^2}{\sigma^2} + \frac{x_2^2}{\sigma^2}\right) \qquad (1) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)\exp\left[-\left(\frac{x_1^2 + x_2^2}{2\sigma^2}\right)\right]. \end{aligned}$$

   Define

$$\begin{aligned} g_1(x_1, x_2) &= y_1 \qquad (2) \\ &= x_1 + x_2 \end{aligned}$$

   and

$$\begin{aligned} g_2(x_1, x_2) &= y_2 \qquad (3) \\ &= x_1 - x_2. \end{aligned}$$

   (a) Equations 2 and 3 give two equations in two unknowns. Please solve them, and give expressions for x_1 and x_2.

(b) Show that the Jacobian is equal to 1/2.

(c) Invoke a transformation theorem to show that

$$f_{Y_1,Y_2}\left(y_1, y_2\right) = \left(\frac{1}{4\pi\sigma^2}\right) \exp\left[\left(-\frac{1}{2}\right)\left(\frac{y_1^2 + y_2^2}{2\sigma^2}\right)\right].$$

What domain is this defined on?

(d) Show that y_1 is independent from y_2.

(e) Show $E(y_1) = 0$, and $V(y_1) = 2\sigma^2$.

Question-4 Additional Calculations:

5. Let $X_i = \frac{\theta}{2}t_i^2 + \epsilon_i$, $i = 1, \ldots, n$, where ϵ_i are independent and distributed $N(0, \sigma^2)$ with known variance σ^2. Assume that the MLE for θ is $\widehat{\theta} = \frac{2\sum_{i=1}^{n} t_i^2 X_i}{\sum_{i=1}^{n} t_i^4}$.

   (a) Calculate $E[\hat{\theta}]$ and $Var(\hat{\theta})$ and show that the pivot $V(\theta) = \frac{(\hat{\theta} - E[\hat{\theta}])}{\sqrt{Var(\hat{\theta})}}$ has a known distribution.

Question-5 Continued

(b) Use this pivot to find a $1 - \alpha$ confidence interval for θ.

(c) Invert the confidence interval to define a level α test and report the rejection region.

Question-5 Continued

(d) If n is large, find an approximate confidence interval for $h(\widehat{\theta}) = \sqrt{\widehat{\theta}}$.

6. Suppose that X and Y are jointly distributed from a bivariate normal distribution with correlation ρ, means $E[X] = E[Y] = 0$ and $Var(X) = Var(Y) = \frac{1}{(1-\rho^2)}$.

Note: The general form of a bivariate normal density for $x \in R, y \in R$ is

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp(-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}]),$$

$\sigma_x > 0, \sigma_y > 0, \mu_x \in R, \mu_y \in R$.

(a) Assuming the parameters above, find the joint density for this problem and show that it is an exponential family.

Question-6 Continued

(b) Based on your exponential family, show that $w_1(\theta)$ is a non-decreasing function.

Question-6 Continued

(c) What is the uniformly most powerful test of $H_0 : \rho \leq 0$ versus $H_1 : \rho > 0$ based on (X, Y)? Identify the test statistic and critical (rejection) region.

(d) Describe how you would solve for the critical region to find a level α test. (Solving analytically is tricky, do not attempt).

7. The likelihood function for a single parameter θ, from a sample of n independent observations satisfies the equation

$$\frac{\partial logL}{\partial \theta} = a(\theta)(t - \theta)$$

where t is a statistic, L is the likelihood, and $a(\theta)$ is a function of θ only. Finally, you may assume that the likelihood is based on an exponential family distribution.

   (a) Find the MLE of θ.

Question-7 Continued

(b) Find the MLE of θ^2.

Question-7 Continued

(c) Find a sufficient statistic for θ and show that it is sufficient.

Question-7 Continued

(d) Find an unbiased estimator for θ.

Question-7 Continued

(e) Calculate the expected information.

Question-7 Continued

(f) Does your estimator attain the Cramer-Rao lower bound. Justify your answer.

Question-7 Continued

(g) Suppose the t is a complete statistic. Show whether or not it is the Best Unbiased Estimator (BUE) of θ. Note that the BUE is also known as the UMVUE (Uniform minimum variance unbiased estimator).

Question-7 Calculations:

Additional Calculations:

Additional Calculations:

Additional Calculations:

Additional Calculations:

Additional Calculations: