Probability and Mathematical Statistics Comprehensive Exam
August 17, 2015

1. (5pts) Let A, B, and C be events with $P(A) > 0$, $P(B) > 0$, $P(C) > 0$, and $P(B \cap C) > 0$. Show that
$$P(A \cap B|C) = P(A|B \cap C)P(B|C)$$

2. (5pts) Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. For a fixed number x_0 define the function
$$g(x) = \begin{cases} f(x)/(1 - F(x_0)) & x \geq x_0 \\ 0 & x < x_0 \end{cases}$$
Prove that $g(x)$ is a pdf. You can assume that $F(x_0) < 1$.

3. (5pts) In a certain city, 60% of the people possess a gun. Of these 30% favor a new gun control law. If five people are chosen from this city at random, let X be the number of people in the sample who both have a gun and favor the new law. Find the probability distribution of X.

4. (5pts) Two ecologists conducted a study of the demography of grizzly bears (*Ursus arctos*) in the greater Yellowstone area using data from a 20 year period of time (*Pease, C. M., and D. J. Mattson. 1999. Demography of the Yellowstone grizzly bears. Ecology. 80:957-975*). One of the goals of the study was to obtain an estimate of the finite rate of population growth called λ. It is enough to know that $\lambda > 1$ implies an increasing population, $\lambda = 1$ implies a stable population, and $\lambda < 1$ implies a decreasing population. The maximum likelihood estimate was $\widehat{\lambda} = 1.01$ with a standard error of 0.04. The resulting 95% frequentist confidence interval is given as 0.93 to 1.09. The authors wrote:

   "We conclude that within the limits of uncertainty implied by the available data, the size of the population changed little over the time period of interest. In fact, because a Gaussian distribution with mean 1.01 and standard deviation 0.04 has about 40% of its probability mass below 1.00, there is an approximate 40% probability that the number of animals has declined over the 20 years."

   Discuss the validity of this statement to help another ecologist decide whether the conclusions are reasonable and valid from a statistical perspective.

5. You can use well-established results from the theory of mathematical statistics and linear models to answer the following questions.

   Consider the linear model
$$\mathrm{Y} = \mathrm{X}\beta + \epsilon$$
   where $\epsilon \sim (0, \sigma^2\mathrm{I})$. Y is $n \times 1$ random vector, X is a known full column rank $n \times p$ matrix, β is $p \times 1$ unknown vector of parameters, and ϵ is $n \times 1$ random vector. Note that initially we make no specific distributional assumptions about ϵ but do assume that all elements of ϵ have the same distribution and are independent of one another. The ordinary least squares estimator of β is
$$\widehat{\beta} = \left(\mathrm{X}'\mathrm{X}\right)^{-1}\mathrm{X}'\mathrm{Y}$$

(a) (5pts) Is this estimator a UMVUE of β? Justify your answer.

(b) (5pts) Does your answer change if $\epsilon \sim N(0, \sigma^2 \mathrm{I})$? Again justify your answer.

6. (5pts) The following problem was assigned to a class:

   Let X_1 and X_2 be a random sample of size 2 from

$$f(x) = (1-\theta)^x \theta, \quad x = 0, 1, 2, \cdots$$

   and 0 otherwise with $\theta \in (0,1)$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1$. Find the joint distribution of (Y_1, Y_2).

   Attached is an attempt by a student to work this problem. Grade it. (Note: We are not interested in how many points you would assign. We are only interested in whether or not you can determine if the student worked the problem correctly. If they did make mistakes you should identify those and give the correct answer.)

7. Let $X_1, \cdots, X_n$ be a random sample from a continuous distribution with pdf

$$f(x) = \begin{cases} \frac{\theta}{(1+x)^{\theta+1}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

   (a) (5pts) Show that

$$T(\mathbf{X}) = \sum \log(1 + X_i)$$

   Complete?

   is a complete sufficient statistic for θ.

   (b) (10pts) Find the MLE of θ and show that it is a function of the complete sufficient statistic you found in part (a). In the interest of time you can skip the second derivative test confirmation of a max.

   (c) (5pts) Show that $Y = \log(1 + X) \sim Expon(1/\theta)$ [also known as $Expon^*(\theta)$].

   (d) (10pts) Is the MLE a UMVUE of θ? Justify your answer and if it is not find one. (The previous problem may be helpful).

   (e) (5pts) Find the CRLB for all unbiased estimators of θ.

   (f) (6pts) Find the approximate large sample distribution of the MLE and use this to find an approximate $1 - \alpha$ confidence interval for θ.

   (g) (7pts) Give the likelihood ratio test statistic for a test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_1$. Determine the approximate critical value for a size α likelihood ratio test based on a large-sample approximation. (Don't worry about simplifying the LRT test statistic).

8. (7pts) Let X_1 and X_2 be a random sample from a distribution with pdf

$$f(x;\theta) = \exp\left[-(x-\theta)\right] I_{[\theta,\infty)}(x); \quad -\infty < \theta < \infty.$$

   A student argues that $X_1 + X_2$ is sufficient because

$$\begin{aligned} f(x_1;\theta)\, f(x_2;\theta) &= \exp\left[-(x_1-\theta)\right] I_{[\theta,\infty)}(x_1) \exp\left[-(x_2-\theta)\right] I_{[\theta,\infty)}(x_2) \\ &= \left\{\exp\left[-(x_1+x_2-2\theta)\right] I_{[2\theta,\infty)}(x_1+x_2)\right\} \quad (1) \end{aligned}$$

He argues that the Factorization Theorem applies by choosing $h(x_1, x_2) = 1$. Explain why his argument is flawed and give a valid one-dimensional sufficient statistic using the Factorization Theorem.

9. (5pts) Suppose $X \sim \chi^2_m$ and $Y \sim \chi^2_n$ with X and Y independent. Is $Y - X \sim \chi^2$ if $n > m$? Justify your answer and if it is χ^2 give the degrees of freedom.

10. (5pts) Suppose X and Y are continuous random variables with joint pdf

$$f(x, y) = 4(x - xy); 0 < x < 1, 0 < y < 1$$

and 0 elsewhere. Find the correlation of (X, Y).