UNIVERSITY OF MASSACHUSETTS
Department of Mathematics and Statistics
Basic Exam - Statistics
August, 2021

Show all work in your solution to each problem. 60 points are needed to pass at the Masters Level and 75 to pass at the Ph.D. level.

1. Suppose that $X_1, \dots, X_n$ are IID observations and that the density function of the observations is given by

$$f_\theta(x) = \frac{1}{2\sqrt{\theta}} e^{-|x|/\sqrt{\theta}}$$

for all $x \in \mathbb{R}$ and some unknown parameter $\theta > 0$.

(a) (6 points) Find the maximum likelihood estimator (MLE) of θ.

(b) (6 points) Use the fact that $E[|X|] = \sqrt{\theta}$ and $\mathrm{Var}(|X|) = \theta < \infty$ (which you do not have to prove) to determine whether the MLE is consistent.

(c) (6 points) What is the MLE of $\sqrt{\theta}$?

(d) (6 points) Use the fact that $E[|X|] = \sqrt{\theta}$ and $\mathrm{Var}(|X|) = \theta$ (which you do not have to prove) to find the bias of the MLE of θ. If the MLE is biased, adjust it to make it unbiased.

(e) (6 points) Find a minimal sufficient statistic $T(X_1, \dots, X_n)$ for θ, and justify that it is indeed a minimal sufficient statistic.

2. Suppose that $X_1, \dots, X_n$ are IID random variables with CDF

$$P(X \leq x; \lambda, k) = F(x; \theta, k) = 1 - e^{-x^k/\theta}$$

for $x > 0$, and 0 otherwise, where $\theta \in (0, \infty)$ is the unknown parameter of interest and $k \in (0, \infty)$ is assumed to be known.

(a) (6 points) Show that $\sum_{i=1}^n 2X_i^k/\theta$ is a pivotal quantity, and derive its sampling distribution. (Hint: what is the distribution of $2X_i^k/\theta$?)

(b) (6 points) Derive a pivotal two-sided $100\times(1-\alpha)$-level confidence interval for θ.

Suppose that we assign a prior distribution $\theta \sim \mathrm{InverseGamma}(\alpha_0, \beta_0)$ for θ, where the density of the inverse gamma distribution is given by

$$\pi(\theta; \alpha_0, \beta_0) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-\alpha_0 - 1} e^{-\beta_0/\theta}.$$

The mean of the $\mathrm{InverseGamma}(\alpha_0, \beta_0)$ distribution is $\beta_0/(\alpha_0 - 1)$.

(c) (6 points) Derive the posterior distribution of θ.

(d) (6 points) Derive the Bayes estimator minimizing squared-error risk.

(e) (6 points) Describe a $100\times(1-\alpha)$-level credible interval for θ.

3. An urn contains five balls: r red and $5-r$ white. The null hypothesis states that all balls are of the same color (i.e. $H_0 : r = 0$ *or* $r = 5$). Suppose that we take a sample of size 2 and reject H_0 if the balls are of different colors.

   (a) (6 points) Find the power of this test for $r = 0, 1, \ldots, 5$ if the sample is drawn without replacement.

   (b) (6 points) Find the power of this test for $r = 0, 1, \ldots, 5$ if the sample is drawn with replacement.

   (c) (6 points) Find the the probably of type-I error for each of these cases.

4. A sample of 200 trees in a forest have been inspected for the presence of an infestation (infestation means some bugs living in the tree). 37 were found to be infested.

   (a) (6 points) Assuming a binomial model, give a consistent and unbiased point estimate $\hat{p}$ and the form of a 90% confidence interval (based on a normal approximation) for the probability p of a tree being infested. (express this using the numbers given and quantiles of known distributions, but you need not compute the values).

   If a tree is infested, one might expect some of the neighboring trees to be infested. If the trees were selected from close to each other, this would induce dependence between the observed trees. Suppose the marginal distribution of each tree is still Bernoulli, but the joint distribution is not independent.

   (b) (5 points) Express $\hat{p}$ as the sum of $N = 200$ dependent random variables. Suppose the trees were selected in clusters of size 4. Assume the correlation between each pair of trees in the same cluster is ρ, $0 < \rho < 1$. Assume clusters are independent. Derive an expression for the variance of $\hat{p}$ and show that $\hat{p}$ is still unbiased and consistent.

   (c) (5 points) Is the 90% interval from part (a) still valid? Why or why not? If not, is it conservative (too wide) or anti-conservative (too narrow)?

5. (6 points) The sample space of a test statistic X has five values, $\{a, b, c, d, e\}$. We are testing that the distribution of X is f_o against the alternative f_1, where f_0 and f_1 are given by the table:

X	a	b	c	d	e
f_0	0.2	0.2	0	0.1	0.5
f_1	0.2	0.4	0.3	0	0.1

   Find the most powerful test with $\alpha = 0.2$ against alternative f_1.