Study Guide for Final
BIOSTAT 602

Test Date

- Thursday, April 27, 1:30 PM–3:30 PM

Venue

- 3695 Medical Sciences II (NLH)

Syllabus

- Chapters 7 (7.2.2, 7.2.3, 7.3.1), 8 (8.1, 8.2.1, 8.3.1, 8.3.2), 9 (9.1, 9.2.1, 9.2.2, 9.2.3, 9.3.1), 10 (10.1.1, 10.1.2, 10.1.3, 10.3.1, 10.3.2) from Casella and Berger

Focus of Topics

- Maximum Likelihood Estimation, Bayesian Estimation, Loss Function
- Calculation of Risk, Mean Squared Error
- Likelihood Ratio Tests, UMP tests, Neyman Pearson Lemma, Karlin-Rubin Theorem
- Large-sample tests; Score and Wald Tests
- Large-sample inference; Consistency, efficiency, large-sample results for MLEs
- Interval Estimation; Inverting a test statistic, Pivotal quantities

Some Pointers

The test will involve some small applications or theoretical derivations similar to those you did in the homework or those you saw in class. The questions will be directly from material in the above-mentioned sections. However, the connection to he material before Midterm will be implicit and cannot be ignored. While you will not be asked to formally derive any theoretical result, you will be expected to identify the appropriate results that are applicable in working out specific examples or verify/disprove some conceptual statements. In this regard homework as well as practice problems, illustrations shown in class, the quiz questions from Friday discussions are all good examples for studying.

Studying important shortcuts will be extremely useful. For example, establishing that a pdf/pmf belongs to exponential family significantly simplifies the problem for searching optimal estimators. Similarly, verifying that a pdf belongs to a scale or a location family facilitates the search for a pivotal.

This is a closed book, closed note examination apart from the test aid you shall be given during the examination. You will also not be allowed to use any laptop, or other devices such as iphone, ipad etc.

Below is a selection of some problems that is designed to give you some practice. The problems are only supposed to provide you a review of the topics.

Practice Problems

1. Let X be a random variable whose pmf under $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ is given by

x	1	2	3	4	5	6	7
$f(x\|\theta_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$f(x\|\theta_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

   (a) Find the UMP test for H_0 versus H_1 with size $\alpha = 0.05$.

   (b) Compute the probability of Type II Error for this test.

2. Let $X_1, \ldots, X_n$ be *i.i.d.* random variables from Uniform$(0, \theta)$ distribution with pdf

$$f_X(x|\theta) = \frac{1}{\theta} I(0 \leq x \leq \theta), \qquad \theta > 0$$

   (a) Show that the $T = X_{(n)} = \max(X_1, \cdots, X_n)$ has Monotonic Likelihood Ratio (MLR) property.

   (b) Show that the hypothesis testing procedure specified by rejection region

$$R = \{\mathbf{X} : X_{(n)} = \max(X_1, \cdots, X_n) > k\theta_0\}$$

   is the UMP level α test for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ for its size. Represent k in terms of α in a closed form.

   (c) Find the $100(1-\alpha)\%$ confidence interval of θ obtained from the hypothesis testing procedure above.

{ **Hint:** The mean, and variance of Uniform$(0, \theta)$ distribution is

$$\mathrm{E}X = \frac{\theta}{2}, \qquad \mathrm{Var}X = \frac{\theta^2}{12}$$

You may use the fact that $\frac{X_{(n)}}{\theta} \sim$ Beta$(n, 1)$ distribution. The pdf, mean and variance of Beta(α, β) is

$$f_X(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \le x \le 1,\ \alpha > 0,\ \beta > 0$$

$$\mathrm{E}X = \frac{\alpha}{\alpha+\beta}, \qquad \mathrm{Var}X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\}$$

3. Let $X_1, \ldots, X_n$ be *i.i.d.* random variables from Logistic$(\theta, 1)$ distribution with pdf

$$f_X(x|\theta) = \frac{e^{-(x-\theta)}}{1+e^{-(x-\theta)}}, \qquad -\infty < x < \infty, -\infty < \theta < \infty$$

   (a) Show that

$$V_n = \frac{1}{1+\exp(\overline{X})}$$

   are consistent estimator for $\Pr(X \le 0) = \frac{1}{1+e^{\theta}}$.

   (b) We know that

$$W_n = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le 0)$$

   follows an asymptotic distribution

$$W_n \sim \mathcal{AN}\left(\frac{1}{1+e^{\theta}}, \frac{e^{\theta}}{n(1+e^{\theta})^2}\right)$$

   Compute ARE(V_n, W_n) and determine whether V_n is asymptotically more efficient than W_n. Justify your answer.

   (c) When $n = 1$, construct a UMP level α test for testing $H_0 : \theta = 0$ against $H_1 : \theta = 1$.

{ **Hint:** The cdf, mean and variance of Logistic$(\theta, 1)$ distribution are

$$\Pr(X \le x) = \int_{-\infty}^{x} f_X(X|\theta)dX = \frac{1}{1+e^{-(x-\theta)}}$$

$$\mathrm{E}(X) = \theta, \qquad \mathrm{Var}(X) = \frac{\pi^2}{3}\}$$

4. Let $X_1, \ldots, X_n$ be *i.i.d.* random variables from the following pdf

$$f_X(x|\theta) = \frac{1}{x \log \theta} \qquad 1 \leq x \leq \theta, \ \theta > 1$$

   (a) Construct a size α likelihood ratio test (LRT) for testing $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$ for arbitrary $\theta_0 > 1$.

   (b) Calculate the confidence coefficient of an interval estimator $[X_{(n)}, X_{(n)}^2]$ for θ.

   (c) Consider a prior distribution of θ from the following pdf

$$f_Y(y|\alpha, \beta) = \frac{\beta\alpha^\beta}{y(\log y)^{\beta+1}}, \quad y > \alpha, \ \alpha > 0, \ \beta > 0$$

   Calculate the posterior distribution of θ and justify whether the family of $f_Y(\theta|\alpha, \beta)$ is conjugate for $f_X(x|\theta)$ or not.

{ **Hint:** The cdf, mean, and variance of $f_X(x|\theta)$ are

$$\Pr(X \leq x) = \frac{\log x}{\log \theta}$$

$$\mathrm{E}(X) = \frac{\theta - 1}{\log \theta}, \qquad \mathrm{Var}(X) = \frac{\frac{1}{2}\theta^2 \log \theta - \log \theta - (\theta - 1)^2}{[\log \theta]^2}\}$$

5. Let $X_1, \cdots, X_n$ be iid samples from the following geometric distribution

$$f(x|\theta) \ = \ \theta(1-\theta)^x \qquad x \in \{0, 1, 2, \cdots\}, \ 0 < \theta \leq 1$$

whose expectation and variance is known to be $\mathrm{E}(X) = (1-\theta)/\theta$ and $\mathrm{Var}(X) = (1-\theta)/\theta^2$.

   (a) Let $\pi(\theta) \sim Beta(\alpha, \beta)$ be the prior distribution, where α and β are known constants. Find the posterior distribution. Explain whether $\pi(\theta)$ is a conjugate family for $f(x|\theta)$ or not.

   (b) Compute the Bayes' rule estimator of θ for loss function $L = (\theta - \hat{\theta})^2$

   (c) Is the Bayes estimator in part (b) a consistent estimator? Justify your answer.

{ **Hint:** You may use the fact that the pdf of Beta distribution is defined as

$$f_{Beta}(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} \qquad 0<x<1,\ \alpha>0,\ \beta>0$$

with mean $\alpha/(\alpha+\beta)$ and variance $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. $\Gamma(n)=(n-1)!$.}