UC Berkeley
Department of Statistics

STAT 210A: Introduction to Mathematical Statistics

Problem Set 10
Fall 2007

Issued: Thursday, November 8 **Due:** Thursday, November 15

Reading: Bickel & Doksum: §6.3, 6.4; Keener: Chapter 18

Problem 10.1
Suppose that we use a prior $\lambda = [\lambda_0 \quad (1-\lambda_0)]$ for a binary hypothesis test $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ where $\lambda_0 = \mathbb{P}[\theta = \theta_0]$.

(a) Show that the Bayes risk under $0-1$ loss can be written as

$$r(\lambda, \delta) \quad = \quad \lambda_0 \mathbb{E}_0[\delta(X)] + (1-\lambda_0)\left\{1 - \mathbb{E}_1[\delta(X)]\right\}.$$

(b) Show that any test minimizing the Bayes risk can be expressed as a LRT where the threshold t is a function of λ_0.

(c) Suppose that we are given i.i.d. samples $X_1, \ldots, X_n$ from $\mathbb{P}_\theta$ (where θ is either equal to θ_0 or θ_1). For any fixed $\lambda_0 \in (0,1)$, let $\delta_{n,\lambda}$ denote the Bayes test derived in part (b). Prove that $\lim_{n\to+\infty} r(\lambda, \delta_n) = 0$.

Problem 10.2
For each of the following problems, compute the generalized likelihood ratio test, and compute its asymptotic distribution under the specified hypothesis H_0.

(a) Let $X_1, \ldots, X_n$ be an i.i.d. sample of $N(\mu_x, \sigma_x^2)$ variates, and let $Y_1, \ldots Y_n$ be an i.i.d. sample of $N(\mu_y, \sigma_y^2)$ variates. Consider testing $H_0 : \mu_x = \mu_y$ and $\sigma_x^2 = \sigma_y^2$.

(b) For $i = 1, \ldots, k$, let $X_{i1}, \ldots, X_{in}$ be independent samples from Poisson distributions with means θ_i, respectively. Consider testing $H_0 : \theta_1 = \theta_2 = \ldots \theta_k$.

(c) Let $X_1, \ldots, X_n$ be an i.i.d. sample from the exponential distribution with parameter θ, and $Y_1, \ldots Y_n$ an i.i.d sample from the exponential distribution with parameter μ. Consider testing $H_0 : \mu = 2\theta$.

Problem 10.3
A random sample of n individuals are classified into three groups, with probabilities θ^2, $2\theta(1-\theta)$ and $(1-\theta)^2$ respectively, yielding the trinomial distribution over the group sizes (y_1, y_2, y_3):

$$\mathbb{P}[Y_1 = y_1, Y_2 = y_2, Y_3 = y_3] \quad = \quad \frac{n!}{y_1! y_2! y_3!} \theta^{2y_1} \left[2\theta(1-\theta)\right]^{y_2} (1-\theta)^{2y_3}$$

for $y_i \geq 0$ and $y_1 + y_2 + y_3 = n$. (This is known as the *Hardy-Weinburg* model in genetics.)

(a) Find the MLE of θ, and determine the asymptotic distribution $\sqrt{n}\left(\widehat{\theta}_n - \theta\right)$.

(b) Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. Suppose that for some k, we have $\mathbb{P}_{\theta_0}[2Y_1 + Y_2 \geq k] = \alpha$. Show that the test that rejects H_0 when $2Y_1 + Y_2 \geq k$ is a UMP level α test for H_0 versus H_1.

(c) Suppose that n is large, and $\alpha = 0.05$. Use asymptotic theory to find an approximate value for k in the UMP test from (b).

Problem 10.4
Given samples $X_1^n = \{X_1, \ldots, X_n\}$, suppose that we are testing $H_0 : \theta \in \Omega_0$ versus $H_1 : \theta \in \Omega_1$. A sequence $\{\phi_n\}$ of level α tests is said to be consistent for testing H_0 versus H_1 if $\mathbb{E}_\theta[\phi_n(X_1^n)] \to 1$ for all $\theta \in \Omega_1$.

(a) Suppose that $\Omega_0 = \{\theta_0\}$ and $\Omega_1 = \{\theta_1\}$ and that the samples are i.i.d. Show that a sequence of MP level α tests is consistent.

(b) Suppose that we want to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ using the generalized likelihood ratio $T_n = \sum_{i=1}^n \log[p(X_i; \widehat{\theta}_n)/p(X_i; \theta)]$, where $\widehat{\theta}_n$ is the MLE of θ, and H_0 is rejected for large values of T_n. If T_n/n converges in probability to zero under H_0, show that this sequence of tests is consistent.

Problem 10.5
Suppose that $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ form a collection of independent random variables, with $X_i \sim \text{Exp}(\lambda_i \theta)$ and $Y_i \sim \text{Exp}(\lambda_i)$ for each $i = 1, \ldots n$, where $(\lambda_1, \ldots, \lambda_n) > 0$ and $\theta > 0$ are unknown parameters.

(a) Show that MLE $\widehat{\theta}_n$ based on $(X_1, \ldots X_n)$ and $(Y_1, \ldots, Y_n)$ satisifies the equation

$$\frac{n}{\hat{\theta}_n} - 2\sum_{i=1}^n \frac{R_i}{\hat{\theta}_n R_i + 1} \quad = \quad 0,$$

where $R_i = X_i/Y_i$.

(b) Compute the asymptotic distribution of $\sqrt{n}(\widehat{\theta}_n - \theta)$.

(c) Consider the hypothesis test $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$. Show that the generalized LRT of H_0 versus H_1 is based on thresholding

$$T_n \quad = \quad \sum_{i=1}^n \left\{ \log(\hat{\theta}_n) - 2\log\left(\frac{\hat{\theta}_n R_i + 1}{R_i + 1}\right) \right\}.$$

(d) Assuming that $\theta = 1$, compute the limiting distribution of T_n when $n \to +\infty$. (**Note:** Some care is required since the dimension of the parameter space depends on n.)

Problem 10.6
Consider the location family $p(x; \theta) = g(x - \theta)$ where we assume that g is continuous and $g(t) > 0$ for all $t \in \mathbb{R}$. Show that a necessary and sufficient condition for the family to have a monotone likelihood ratio in x is that the function $-\log g$ is convex.