UC Berkeley
Department of Statistics

STAT 210A: Introduction to Mathematical Statistics

Problem Set 8
Fall 2007

Issued: Thursday, October 25 **Due:** Thursday, November 1

Reading: Keener: Chapter 10, 11, §21.6, B & D: §3.5, §4.4

Problem 8.1
Find variance-stabilizing transformations for the following statistics and estimation problems:

(a) The sample mean $\bar{X}_n$ for i.i.d. samples from $\text{Ber}(\theta)$. (Here we want the transformed statistic to have asymptotic variance independent of θ.)

(b) The sample correlation coefficient $r = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2}\sqrt{\sum_{i=1}^n Y_i^2}}$ for i.i.d. samples from the zero-mean bivariate normal with $\sigma_X^2 = \sigma_Y^2 = 1$ and correlation coefficient $\rho \in (-1, 1)$. (Here we want the transformed statistic to have asymptotic variance independent of ρ.)

Problem 8.2
Let $X_1, \ldots, X_n$ be i.i.d. with $\mathbb{E}[X_i] = \theta$, $\text{var}(X_i) = 1$, and $\mathbb{E}[(X_i - \theta)^4] = \mu_4$, and consider the unbiased estimators of θ^2 defined by

$$\delta_1(X) = \frac{1}{n}\sum_{i=1}^n X_i^2 - 1, \qquad \text{and} \qquad \delta_2(X) = (\bar{X}_n)^2 - \frac{1}{n}.$$

(a) Determine the asymptotic relative efficiency (ARE) of δ_1 compared to δ_2.

(b) Show that δ_2 is superior in terms of ARE if the X_i are symmetric about θ.

(c) Find a distribution for which δ_1 is superior in terms of ARE.

Problem 8.3
Recall that a statistical function h is said to be continuous at F in the sup-norm topology if, for any sequence $\{G_n\}$ such that $\|G_n - F\|_\infty \to 0$, there holds $h(G_n) \to h(F)$. Are the following statistical functionals continuous in this sense or not? Prove or disprove.

(a) for fixed $a \in \mathbb{R}$, the evaluation $h(F) = F(a)$.

(b) for fixed CDF F_0 with density f_0, the Cramér-von Mises functional given by $h(F) = \int [F(t) - F_0(t)]^2 f_0(t)dt$.

(c) the mean functional $h(F) = \mathbb{E}_F[X]$.

(d) the quantile functional $h_\alpha(F)$ that returns the level α quantile of F, when the inverse CDF $F^{-1}(t) = \inf\{x \in \mathbb{R} \mid F(x) \geq t\}$ is continuous at α.

Problem 8.4

Let $X_1, \dots, X_n$ be an i.i.d. sample of Uni$[0, 2\theta]$ variables, and let $X_{(1)}, \dots, X_{(n)}$ denote the associated order statistics. Define for any $p \in (0,1)$, the p^{th}-sample quantile $X_{(np)} \equiv X_{(\lceil np \rceil)}$, where for any $x \in \mathbb{R}$, the notation $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

(a) Find the asymptotic distribution of $X_{(\frac{1}{2}n)}$.

(b) Find the asymptotic distribution of $\frac{2}{3}X_{(3n/4)}$.

(c) Find the asymptotic distribution of the midquantile range $\frac{1}{2}[X_{(n/4)} + X_{(3n/4)}]$.

(d) Compare these three statistics as estimators of θ in terms of their asymptotic relative efficiencies.

Note: See §10.4 of Keener for some useful results.

Problem 8.5

Suppose we form an empirical CDF $\widehat{F}_{n-1}$ based on a set of $(n-1)$ samples $x_1, \dots, x_{n-1}$. Now consider the new empirical CDF $\widehat{F}_{n,x}$ formed when the point x is added to $\{x_1, \dots, x_{n-1}\}$ as the n^{th} sample. The *sensitivity function* $S_{h,F}(x) = \left[h(\widehat{F}_{n,x}) - h(\widehat{F}_{n-1})\right]$ measures the effect of adding the single new sample at x.

(a) Compute the sensitivity $S_{h,F}(x)$ when $h(F) = \mathbb{E}_F[X]$ is the mean functional.

(b) Suppose that $n = 2k+1$ is odd, and that the sample median of $\{x_1, \dots, x_{n-1}\}$ is given by $\frac{1}{2}(x_{(k)} + x_{(k+1)})$. Prove that the sensitivity for the median functional is given by

$$S_{h,F}(x) = \begin{cases} x_{(k)} & \text{for } x \leq x_{(k)} \\ x & \text{for } x_{(k)} \leq x \leq x_{(k+1)} \\ x_{(k+1)} & \text{for } x \geq x_{(k+1)} \end{cases}$$

Problem 8.6

Given n i.i.d. samples $X_1, \dots, X_n \sim F$, the *Huber estimate* $\theta(\widehat{F}_n)$ of location is defined implicitly as the solution of the equation $\frac{1}{n}\sum_{i=1}^n \psi_K(X_i - \theta) = 0$, where for each $K \geq 0$, the function ψ_k has the form

$$\psi_K(t) = \begin{cases} t & \text{if } |t| \leq K \\ \text{sign}(t)\, K & \text{if } |t| > K \end{cases}$$

It can be viewed as the empirical version of the statistical function $h(F)$ that outputs the solution θ to the equation $\mathbb{E}_F[\psi_K(X - \theta)] = 0$. Assume that the c.d.f. F has a density f_X that is strictly positive on all $\mathbb{R}$.

(a) Show for $K \to +\infty$, the Huber estimator tends to the sample mean $\bar{X}_n$, and that for $K = 0$, the Huber estimator is the median.

(b) Show that $\theta(F)$ exists and is unique for $K > 0$.

(c) Show that for any $K < +\infty$, then $\lim_{|x| \to +\infty} S_{\theta,F}(x)$ is a finite constant. (The sensitivity S was defined in the previous problem.)