Midterm Project

The purpose of the midterm project is to use the theoretical techniques learned in class to derive some large sample and finite sample properties of estimators, and then to use the java applets to check the accuracy of these results and also to compare competitive estimators. This project will be considered as your midterm examination. The project must be handed in by 2:20 p.m., Tuesday, November 13.

A TeXed or typed report should be prepared for each of the two problems. The derivations for the theoretical results can be attached as in a handwritten appendix. A possible outline for each report follows:

a. Introduction
b. Theoretical results
c. Description of the simulation
d. Results of the simulation presented in tables or figures
e. Summary and conclusions. Some questions that you should consider include:
   • How accurate are the results reported from the simulation study.
   • Do the simulation results agree with the theoretical results?
   • Which estimator is best? If none is best, under what conditions are different estimators better?

You should each do your own work for the project. Do not get help from other students or use their worked problems. Be sure to cite any references that you use.

1. Consider the uniform location family where $X_1, \ldots, X_n$ are i.i.d. $f(x - \theta)$, where $-\infty < \theta < \infty$, and $f(x) = \frac{1}{2}I_{[-1,1]}(x)$. Let $Y_p$ denote the sample $p^{th}$ percentile and $X_{(i)}$ denote the $i^{th}$ order statistic.

   (a) Find the minimal sufficient statistic, and show that it is not complete.

   (b) Find the general form of the maximum likelihood estimator, and show that it is not unique. Then show that the sample midrange, $W = (X_{(1)} + X_{(n)})/2$, is an m.l.e.

   (c) We wish to compare the following estimators of $\theta$:

      (i) The sample midrange, $W = (X_{(1)} + X_{(n)})/2$

      (ii) The sample median $Y_{.5}$

      (iii) The sample mean.

      (iv) An L-estimator of the form $V = (Y_{.25} + Y_{.75})/2$. 

Obtain expressions for the mean, variance, and mean squared error for each of these estimators. Then obtain expressions for the relative efficiency of estimators (ii), (iii), and (iv) with respect to estimator (i). What happens as \( n \to \infty \)? Explain why.

(d) Carry out a simulation study to compare the estimators. You can use the applet for “Estimation of the Center of a Symmetric Distribution” to carry this out. You need to design an experiment with various sample sizes. For each sample size, you should generate a large number of samples and compute all the estimates for each sample. Summarize your results by reporting estimated bias, variance, and mean squared error for each estimator for each sample size. Then make some conclusions about your simulation study.

2. Consider the double exponential location family where \( X_1, \ldots, X_n \) are i.i.d. \( f(x - \theta) \), where \(-\infty < \theta < \infty\), and \( f(x) = \frac{1}{2}e^{-|x|} \). We wish to compare the following estimators of \( \theta \):

(i) The sample mean

(ii) The sample median \( Y_{.5} \)

(iii) The 5% trimmed mean

(iv) The 10% trimmed mean

(v) The sample midrange, \( W = (X_{(1)} + X_{(n)})/2 \)

(vi) An L-estimator of the form \( V = (Y_{.25} + Y_{.75})/2 \)

(vii) The MLE of the Cauchy location parameter

(viii) The MLE of the logistic location parameter

(ix) An L-estimator of the form \( V = 0.3Y_{.25} + 0.4Y_{.5} + 0.3Y_{.75} \)

(a) Obtain the the limiting distributions of these estimators and obtain their A.R.E.s.

(b) Carry out a simulation study using the applet to compare the estimators. You need to design an experiment with various sample sizes. For each sample size, you should generate a large number of samples and compute all the estimates for each sample.

(c) Then summarize your results by reporting estimated bias, variance, and mean squared error for each estimator for each sample size.

(d) Make some conclusions about your simulation study.
**Hint:** You may use the following theorem concerning the asymptotic distribution of sample quantiles:

**Theorem:** Suppose that $X_1, \ldots, X_n$ are i.i.d. from a distribution with density $g(x)$ and c.d.f. $G(x)$. Let $Y_{p_1}, \ldots, Y_{p_k}$ be the sample $p_1, \ldots, p_k$ quantiles, respectively, where $p_1 < p_2 < \ldots p_k$. Let $x_{p_j}$ satisfy $G(x_{p_j}) = p_j$. Suppose that $g(x_{p_j}) > 0$ for each $j$. Then

$$
\sqrt{n} \begin{pmatrix}
  Y_{p_1} - x_{p_1} \\
  \vdots \\
  x_{p_i} \\
  \vdots \\
  Y_{p_k} - x_{p_k}
\end{pmatrix} \xrightarrow{d} N_k(0, \Psi)
$$

where $\Psi$ has $(i, j)$ entry

$$
\psi_{i,j} = \frac{p_{\min\{i,j\}} - p_ip_j}{g(x_{p_i})g(x_{p_j})}.
$$