ECON 546: Themes in Econometrics
Supplementary Problems, I

The following questions are designed to give you additional practice with the material that is being covered in class. You should note that many of them are more challenging than the questions that you should expect in the term test or final exam – so don't be put off by this!

A solution sheet is available on request. (This gives you the opportunity to attempt these questions without seeing the answers.)

Question 1

Suppose that the estimator $\hat{\theta}$ is admissible in the class of estimators, *C*, relative to the loss function, *L*. Suppose, further than this estimator's risk function is constant, for all θ. Using a simple diagram, explain why $\hat{\theta}$ is mini-max in *C*, relative to *L*.

Question 2

(a) Let $\{y_1, y_2, \ldots.., y_n\}$ be a random sample from $N[\mu, \sigma^2]$. Prove that $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n} y_i^2$ are jointly sufficient statistics.

(b) Show that $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ and $s^2 = ([1/(n-1)]\sum_{i=1}^{n}(y_i - \bar{y})^2$ are also jointly sufficient statistics.

(c) Use this example to prove that sufficient statistics may be either unbiased or biased estimators of the underlying parameters.

Question 3

Suppose that *Y* is distributed uniformly on the interval [*a*, *b*].

(a) Derive the characteristic function for *Y*, and use it to establish that $E(Y) = (a + b) / 2$, and $\text{var.}(Y) = (b - a)^2 / 12$.

(b) Intuitively, the skewness of this distribution is zero – use the characteristic function to prove that this is the case.

(c) Derive the excess kurtosis for this distribution.

Question 4

Suppose that *Y* follows a (standardized) Cauchy distribution. This distribution arises, for example, if Y is constructed by taking the ratio of two independent standard normal random variables**.** Then, the density of *Y* is:

$$p(y) = 1/[\pi(1+y^2)] \quad ; \qquad -\infty < y < \infty$$

(a) Prove that $\int_a^b y\, p(y)dy = (1/2\pi)[\log_e(1+b^2) - \log_e(1+a^2)]$.

(b) Use this result to show that the first raw moment of $p(y)$ does not exist.

(c) Show that this implies that none of the higher-order moments (raw, or about the mean) exist.

Question 5

Suppose that we want to estimate the variance of the error term in a k-regressor linear regression model that satisfies *all* of the usual assumptions (including Normally distributed errors). Consider the class of estimators:

$$C = \{\hat{\sigma}^2(c) : \hat{\sigma}^2(c) = e'e/(n-k+c)\}$$

where e is the OLS residual vector, n is the sample size, and c is a positive constant.

In keeping with the fact that we have used OLS estimation, we have in mind a loss function that is quadratic. Prove that in order to obtain the estimator in C that minimizes risk, we should set $c = 2$.

[**Hint:** Recall that if $X \sim \chi^2_v$, then $E(X) = v$, and var.$(X) = 2v$.]

Question 6

Sometimes it is important to use a loss function that is bounded (from above). One bounded loss function is the "reflected normal" loss, which is shown in Figure 1 below, and whose mathematical form is:

$$L(\theta, \hat{\theta}) = K\{1 - \exp[-(\hat{\theta} - \theta)^2/(2\gamma^2)]\},$$

where K is the maximum loss and γ is a pre-assigned shape parameter that determines the rate at which the loss approaches its upper bound.

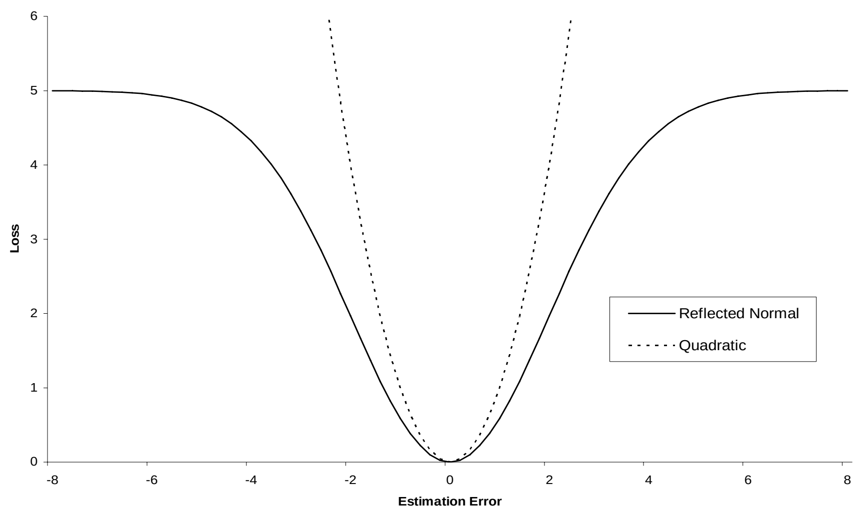
(a) Explain, mathematically, why the two loss functions in Figure 1 are very similar in the region close to the origin on the horizontal axis. Does this similarity depend on K?

(b) Consider the following estimation problem. We have a random sample of n observations from a population that is $N[\mu, \sigma^2]$. Suppose that σ is known and we decide to estimate μ using $\hat{\mu} = \bar{x}$, where $\bar{x}$ is the arithmetic mean of the sample. Show that the risk of this estimator, under "reflected normal" loss, is

$$R(\hat{\mu}) = K\{1 - \gamma/[\frac{\sigma^2}{n} + \gamma^2]\}.$$

(c) Suppose that we also consider the naïve "estimator", $\tilde{\mu} = \mu_0$, where μ_0 is just a pre-assigned constant. Obtain the risk of this estimator under "reflected normal" loss, and draw a diagram that compares the risks of $\hat{\mu}$ and $\tilde{\mu}$.

Figure 1: Reflected Normal & Quadratic Loss Functions
(K = 5; gamma = 1)



Question 7

The so-called "order statistics" are just the n sample values, placed in ascending order. So, we can denote the original (unordered) sample observations as $\{x_1, x_2, x_3, \ldots\ldots., x_n\}$, and the order statistics as $\{y_1, y_2, \ldots., y_n\}$, where $y_1 \le y_2 \le \ldots\ldots \le y_n$. So, $y_1 = \min.\{x_1, x_2, x_3, \ldots\ldots., x_n\}$, and $y_n = \max.\{x_1, x_2, x_3, \ldots\ldots., x_n\}$. Even if the original data are independent, the order statistics are not, and it can be shown that the marginal density for the r-th. order statistic is

$$f_{y_r}(y) = \frac{n!}{(r-1)!(n-r)!}[F(y)]^{r-1}[1-F(y)]^{n-r} f(y)$$

where $F(.)$ and $f(.)$ are the distribution function and density, respectively, for X.

(a) Explain why the order statistics form a set of jointly sufficient statistics for any distribution.

(b) If X is distributed uniformly on [0 , 1], show that $f_{y_1}(y) = n(1-y_1)^{n-1}$.

(c) Determine the expected value of y_1 and explain why this result makes intuitive sense.

(d) What is the bias of y_1 as an estimator of $E(X)$ when $n = 1$? What about as $n \to \infty$?

Question 8

Consider the standard multiple linear regression model, satisfying *all* of the usual assumptions:

$$y = X\beta + \varepsilon \quad ; \qquad \varepsilon \sim N[0, \sigma^2 I_n]$$

The James-Stein (JS) estimator of the coefficient vector is

$$\hat{\beta} = [1 - (ce'e)/(b'X'Xb)]b$$

where 'b' is the OLS estimator, and 'c' is a positive scalar.

(a) Explain why the JS estimator is a 'nonlinear' estimator.

(b) The risk (under quadratic loss) of the JS estimator is approximately

$$Risk(\hat{\beta}) = \sigma^2 tr((X'X)^{-1}) + [nc\sigma^4 / (\beta'X'X\beta)^2][\beta'\beta\{4 + c(n+2)\} - 2(\beta'X'X\beta)tr((X'X)^{-1})]$$

Compare the risks of the JS and OLS estimators when X'X = I_k. In particular, for this situation, show that the OLS estimator is inadmissible by proving that it is risk-dominated by the JS estimator for any choice of 'c' such that

$$0 < c < 2(k-2)/(n+2)$$

What is the minimum number of regressors that must be included in the model for this inadmissibility result to hold?

Question 9

Suppose that we have a sample of *n* independent observations drawn from an underlying population that follows a Logistic distribution, with probability density function given by:

$$f(y) = [\exp(y-a)]/[1+\exp(y-a)]^2 \; ; \qquad -\infty < y < \infty$$

It can be shown that $E(y) = a$, and var.$(y) = (\pi^2 / 3)$.

(a) Show that $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is a mean-square consistent estimator of *a*.

(b) Using the (Lindeberg-Lévy) Central Limit Theorem, write down a careful statement that describes the asymptotic distribution of this estimator.

(c) If $n = 100$, and $\bar{y} = 10.5$, construct an asymptotically valid 95% confidence interval for *a*.