Intro to Econometric Theory
Heinz School, Carnegie Mellon University
90-906, Spring 2003-4

Solution #1

1. Consider a sample of people who were asked to report their sex (Everyone answers, and everyone answers either male or female). From their answers, we construct the following matrix:

$$X = \begin{bmatrix} \mathbf{1} & M & F \end{bmatrix} \tag{1}$$

   In the matrix, each row corresponds to one person's response; $\mathbf{1}$ is a column of 1s, M is a column with a 1 in the ith row if the ith person answered male and a zero otherwise, and F is a column with a 1 in the ith row if person i answered female and a zero otherwise.

   Is $X'X$ invertible. Why or why not?

   No, $X'X$ is not invertible. Notice that $M + F = \mathbf{1}$. This is a linear dependency in the columns of X. Thus, the rank of X must be less than three. Since the rank of $X'X$ (a 3×3 matrix) is the same as the rank of X, its rank must be less than 3. Thus, it is not full rank. Thus it is not invertible.

2. We discussed in class that variance matrixes are always positive semi-definite but not always positive definite. In the case of scalar random variables, a singular variance matrix means a zero variance — that is, that the random variable is not really random at all, but fixed.

   Now, let's think about what a positive semi-definite but not positive definite (hereafter PSDNPD) variance matrix means for the random vector it corresponds to. Let x be a random vector with PSDNPD variance matrix V.

   (a) Must there be a fixed non-zero vector α, such that $\alpha' V\alpha$=0? Why or why not?

   Yes. If there were no such vector, then V would be positive definite, and we are told that it is not.

   (b) Suppose there is such an α, what would the variance of $\alpha' x$ be?

   Zero. $V(\alpha' x) = \alpha' V \alpha = 0$.

   (c) What does this tell us about $\alpha' x$?

   It is fixed, non-random.

(d) Does this mean that $\alpha' x = 0$?

No, it could just as easily be 8 or 0.73 or any fixed number. The variance of 0 is zero, but the variance of 8 or 0.73 are is zero.

(e) So, what does it mean for a random vector to have a PSDNPD variance matrix?

When a random vector has a PSDNDP variance matrix, this means that there are, in some sense, fewer random variables in the vector than appear to be. Consider a scalar random variable z and a random vector:

$$x = \begin{pmatrix} z \\ -z \end{pmatrix}$$

As you can see, although x has two elements it does not *really* contain two random variables, just z.

The variance matrix of x is:

$$V(x) = \begin{bmatrix} \sigma_z^2 & -\sigma_z^2 \\ -\sigma_z^2 & \sigma_z^2 \end{bmatrix}$$

This matrix is singular, and:

$$\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{bmatrix} \sigma_z^2 & -\sigma_z^2 \\ -\sigma_z^2 & \sigma_z^2 \end{bmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0$$

and:

$$V\left((1 \quad 1)x\right) = 0$$

Another way of saying the same thing is that if V is PSDNPD, then you can always write one of the elements of x as a linear function of the other elements plus a fixed number (this follows from $V(\alpha' x) = 0$).

3. We want to estimate the proportion of US households with broadband Internet access. We randomly and independently call up 1,000 households and ask them whether or not they do. We make a $1,000 \times 1$ vector x which has a 1 in the ith row if the ith household answered yes and a 0 if no. Assume that all households in the US have a phone, all answer our call, and all truthfully answer the question.

   Consider the following two estimators for p, the proportion of US households who, in fact, have broadband:

   (a) $\overline{x}$, also known as the sample proportion.

   (b) 0.73 for surveys with sample size $<$ ten million, and $\overline{x}$ for surveys with sample size $>$ ten million.

   Which is unbiased or are both? Which is consistent or are both? What is my point?

Clearly the first is unbiased. $E\{\overline{x}\} = \mu_x = p$. The second is not. Unbiasedness requires that the expected value of the estimator equal the true value of the parameter no matter what the true value of the parameter is. When the true value of the parameter is not 0.73, then, at least sometimes, the second estimator has an expectation different from the true value.

The first is consistent. This is because it is unbiased and its variance goes to zero in n. It variance is $p(1-P)/n$. Thus it converges in mean square to p. Thus it converges in probability to p.

The second is also consistent. Notice that, for n bigger than ten million, the two estimators are the same. Consistency is about *limiting* behavior — behavior for n "near" infinity. Since the two estimators are the same for large n, if one is consistent, then so must the other be.

My point: Consistency is a property which says that an estimator is good in *large* samples and large samples only. A consistent estimator may be very bad, or even very silly, in small samples. Notice that little in my answer would change if the second estimator was -57 for samples less than ten million. Whether a consistent estimator performs well in sample sizes we are likely to encounter in reality is an important question, though not one we are going to treat extensively in the course.

4. Suppose we have an estimator of per-capita income in the US, $\hat{I}$, and an estimator of the share of income saved (NOT the average over people of share of income saved but rather the average over dollars of share of income saved!), $\hat{s}$. The true values of these parameters are μ_I and μ_s. These two estimators are derived somehow from a survey of size n, and we have proved the following result for these estimators:

$$\sqrt{n}\left(\begin{pmatrix} \hat{I} \\ \hat{s} \end{pmatrix} - \begin{pmatrix} \mu_I \\ \mu_s \end{pmatrix}\right) \overset{A}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 400 & 20 \\ 20 & 9 \end{bmatrix}\right) \tag{2}$$

Please construct an approximate (why approximate?) 95% confidence interval for the per-capita savings, $\mu_I \mu_s$.

Approximate because we are using asymptotic (or approximate) distributions, not the true distributions of the estimators to construct the CI — this is true even though the problem I gave did not contain a statement about asymptotic distribution since the theorem we will use has only a conclusion for asymptotic distribution. The CI is right if the sample is very large, and (we hope!) pretty close in sample sizes we are likely to encounter.

Anyway, we have a result from class which says:

If $\sqrt{n}\,(x-\mu_x) \overset{A}{\sim} N(0,\Sigma)$ **then** $\sqrt{n}\,(g(x)-g(\mu_x)) \overset{A}{\sim} N\left(0,\nabla_g'\Sigma\nabla_g\right)$ **In this formula, let's set** $x = \begin{pmatrix} \hat{I} \\ \hat{s} \end{pmatrix}$ **and** $g(x) = x_1 x_2$**. That means that** $\nabla_g = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$ **or** $\nabla_g = \begin{pmatrix} \hat{s} \\ \hat{I} \end{pmatrix}$

Now, applying the theorem, we get:

$$\begin{aligned}
\sqrt{n}\left(\left(\begin{array}{c} \hat{I}\hat{s} \end{array}\right) - \left(\left(\begin{array}{c} \mu_I \mu_s \end{array}\right)\right)\right) &= \sqrt{n}\left((g(x)) - (g(\mu))\right) \\
&\overset{A}{\sim} N\left(0, \begin{pmatrix} \hat{s} & \hat{I} \end{pmatrix} \begin{bmatrix} 400 & 20 \\ 20 & 9 \end{bmatrix} \begin{pmatrix} \hat{s} \\ \hat{I} \end{pmatrix}\right) \\
&= N\left(0, 400\hat{s}^2 + 40\hat{s}\hat{I} + 9\hat{I}^2\right)
\end{aligned}$$

Let's call the rather messy quadratic V**. Had we been given the values of** $\hat{I}$ **and** $\hat{s}$**, we could have calculated** V**. Then, our 95% confidence interval would be:**

$\hat{I}\hat{s} \pm 1.96V$