Economics 573
Problem Set 5
Fall 2002
Due: 4 October 2002

1. In random sampling from any population with E(X) = μ and Var(X) = σ^2, show (using Chebyshev's inequality) that sample mean converges in probability to μ.

2. In random sampling from any population with E(X) = μ and V(X) = σ^2, show without using Chebyshev's inequality:

   a. The sample mean converges in mean square to μ.

   b. The sample mean converges in probability to the population mean.

3. Prove the following theorem:

   If T_n is a sequence of random variables with lim $E(T_n)$ = c and lim $V(T_n)$ = 0 then T_n converges in mean square to c.

4. Show that $E\left(X_n^2\right) \rightarrow 0 \Rightarrow X_n \xrightarrow{P} 0$

5. Consider a variable T_v which is distributed as a t with v degrees of freedom. The t distribution has mean v and variance 2v. The variable T_v can be written as

$$T_v = \frac{Z}{\sqrt{\chi_v^2/v}}$$

   Using the Chebyshev inequality and Slutsky's theorem show that

$$T_v = \frac{Z}{\sqrt{\chi_v^2/v}} \xrightarrow{d} Z \sim N(0,1)$$

   as v goes to infinity. The Chebyshev inequality is given by

$$P[|X-\mu| \geq \delta\sigma] \leq \frac{1}{\delta^2}$$

$$P[|X-\mu| \geq \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2}$$

$$P[|X-\mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

6. Let $X_1, X_2, \ldots X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Define S_n^2 as

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

Using Chebyshev's inequality, find a sufficient condition that S_n^2 converges in probability to σ^2.

7. Let $X_1, X_2, \ldots X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. It can be shown that $Var(\hat{\sigma}^2)$ can be written as

$$Var(\hat{\sigma}^2) = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}$$

where $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$ and μ_k denotes the kth population moment about the population mean.

a. Show that for a normal population.

$$Var(\hat{\sigma}^2) = \frac{2\mu_2^2 (n-1)}{n^2}$$

The moment generating function for a normal distribution will be useful here.

b. Show that for a normal population

$$Var(\hat{\sigma}^2) = \frac{2\sigma^4 (n-1)}{n^2}$$

c. Show that for a normal population

$$Var(S^2) = \frac{2\sigma^4}{(n-1)}$$

8. Let $x_1, x_2, \ldots x_n$ be the realizations that result from a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Also assume that X_i is distributed normally. Consider the following matrix M, which is $n \times n$

$$M = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{pmatrix} \tag{14}$$

a. Show that M is idempotent..

b. What is the trace of M?

c. Show the general form of the row vector $x'M$.

$$x'M = [x_1,\ x_2,\ x_3,\ \cdots,\ x_n] \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{pmatrix} \tag{15}$$

Hint:

$$\begin{aligned} [x_1,\ x_2,\ x_3,\ \cdots,\ x_n] \begin{pmatrix} 1-\frac{1}{n} \\ -\frac{1}{n} \\ \vdots \\ -\frac{1}{n} \end{pmatrix} &= x_1 - \frac{1}{n}x_1 - \frac{1}{n}x_2 - \frac{1}{n}x_3 + \cdots - \frac{1}{n}x_n \\ &= x_1 - \bar{x} \end{aligned} \tag{16}$$

d. Show the general form of the scalar x′Mx.

e. Show that $\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$

f. Using the information in parts a-e and appropriate theorems show that

$$\frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{\sigma^2} \sim \chi^2(n - 1)$$

9. Let $X_1, X_2, \ldots X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Now define the rth moment of X_i , usually denoted by μ_r', as

$$\begin{aligned} \mu_r' &= E(X_i^r) \\ &= \int_{-\infty}^{\infty} x_i^r f(x_i; \theta_1,\ldots,\theta_K)\, dx_i \end{aligned} \tag{20}$$

The rth central moment of X_i about a is defined as $E[(X_i - a)^r]$. If $a = \mu$, we have the rth central moment of X_i about μ_X, denoted by μ_r, which is

$$\begin{aligned} \mu_r &= E[(X_i - \mu)^r] \\ &= \int_{-\infty}^{\infty} (x_i - \mu)^r f(x_i;\theta_1,\ldots,\theta_K)\, dx_i \end{aligned} \tag{21}$$

The rth sample moment is defined as

$$\hat{\mu}_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r \tag{22}$$

The rth central sample moment is defined as

$$\begin{aligned} C_n^r &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu_1'\right)^r \\ \Rightarrow C_n^1 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu_1'\right) = \bar{X} - \mu \\ \Rightarrow C_n^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu_1'\right)^2 = \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2\right) - \mu^2 \end{aligned}$$

where μ_1' is the first raw moment of the distribution which is $E(X_i) = \mu$. The rth central sample moment about the sample mean is defined as

$$M_n^r = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^r$$

$$\Rightarrow M_n^1 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right) = \bar{X} - \bar{X} = 0$$

$$\Rightarrow M_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2\right) - \bar{X}^2$$

a. Show that

$$E\left(M_n^2\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i^2\right) - Var(\bar{X}) - \left(E\bar{X}\right)^2$$

$$= \mu_2^{'} - \left(\mu_1^{'}\right)^2 - \frac{\mu_2}{n}$$

$$= \frac{n-1}{n}\sigma^2$$

where μ_1' **and** $\mu_2^{'}$ are the first and second population moments, and μ_2 is the second central population moment. Note that this obviously implies

$$E\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = E\left[\left(\sum_{i=1}^{n} X_i^2\right) - n\bar{X}^2\right]$$

$$= (n-1)\,\mu_2$$

b. Now write $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ as follows

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left((X_i - \mu) - (\bar{X} - \mu)\right)^2$$

$$Y_i = X_i - \mu$$

$$\bar{Y} = \bar{X} - \mu$$

Show the following

i) $E(Y_i) = 0$
ii) $Var(Y_i) = \sigma^2$
iii) $E(Y_i^4) = \mu_4$ (fourth central moment of X_i)

c. Now consider computing $Var\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) = Var\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)$. This can be written as follows

$$Var\left[\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right] = E\left[\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)^2\right] - \left[E\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)\right]^2$$

Show that

$$E\left[\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)^2\right] = E\left[\left(\sum_{i=1}^{n} Y_i^2\right)^2\right] - 2nE\left[\bar{Y}^2\sum_{i=1}^{n} Y_i^2\right] + n^2E\left(\bar{Y}^4\right)$$

d. Show that

$$E\left[\left(\sum_{i=1}^{n} Y_i^2\right)^2\right] = E\left[\sum_{i=1}^{n} Y_i^2 \sum_{j=1}^{n} Y_j^2\right] = E\left[\sum_{i=1}^{n} Y_i^4 + \sum_{i \neq j}\sum Y_i^2 Y_j^2\right]$$

$$= n\mu_4 + n(n-1)\mu_2^2$$

$$= n\mu_4 + n(n-1)\sigma^4$$

e. Now show that

$$E\left[\bar{Y}^2\sum_{i=1}^{n} Y_i^2\right] = \frac{1}{n^2}E\left[\sum_{j=1}^{n} Y_j \sum_{k=1}^{n} Y_k \sum_{i=1}^{n} Y_i^2\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n} Y_i^4 + \sum_{i \neq j}\sum Y_i^2 Y_j^2 + \sum_{j \neq k}\sum Y_j Y_k \sum_{\substack{i \neq j \\ i \neq k}} Y_i^2\right]$$

$$= \frac{1}{n^2}\left[n\mu_4 + n(n-1)\mu_2^2\right]$$

$$= \frac{1}{n}\left[\mu_4 + (n-1)\sigma^4\right]$$

f. Now show that

$$E\left[\bar{Y}^4\right] = \frac{1}{n^4}E\left[\sum_{i=1}^{n} Y_i \sum_{j=1}^{n} Y_j \sum_{k=1}^{n} Y_k \sum_{\ell=1}^{n} Y_\ell\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n} Y_i^4 + \sum_{i \neq k}\sum Y_i^2 Y_k^2 + \sum_{i \neq j}\sum Y_i^2 Y_j^2 + \sum_{i \neq j} Y_i^2 Y_j^2 + \cdots\right]$$

where for the first double sum ($i = j \neq k = \ell$), for the second ($i = k \neq j = \ell$), and for the last ($i = \ell \neq j = k$) and ... indicates that all other terms include Y_i.

g. Now show that

$$E\left[\bar{Y}^4\right] = \frac{1}{n^4}\left[n\mu_4 + 3n(n-1)\mu_2^2\right]$$

$$= \frac{1}{n^3}\left[\mu_4 + 3(n-1)\sigma^4\right]$$

h. Now combining the information in all the parts we can show that $Var\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right) = Var\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)$ is as follows

$$Var\left[\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right] = E\left[\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)^2\right] - \left[E\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)\right]^2$$

$$= E\left[\left(\sum_{i=1}^{n} Y_i^2\right)^2\right] - 2nE\left[\bar{Y}^2\sum_{i=1}^{n} Y_i^2\right] + n^2E\left(\bar{Y}^4\right) - \left[E\left(\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right)\right]^2$$

$$= n\mu_4 + n(n-1)\mu_2^2 - 2n\left[\frac{1}{n}\left[\mu_4 + (n-1)\mu_2^2\right]\right] + n^2\left[\frac{1}{n^3}\left[\mu_4 + 3(n-1)\mu_2^2\right]\right] - (n-1)^2\mu_2^2$$

$$= n\mu_4 + n(n-1)\mu_2^2 - 2\left[\mu_4 + (n-1)\mu_2^2\right] + \left[\frac{1}{n}\left[\mu_4 + 3(n-1)\mu_2^2\right]\right] - (n-1)^2\mu_2^2$$

Now show that

$$Var\left[\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right] = \frac{(n-1)^2\mu_4}{n} - \frac{(n-1)(n-3)\mu_2^2}{n}$$

$$= \frac{(n-1)^2\mu_4}{n} - \frac{(n-1)(n-3)\sigma^4}{n}$$

i. Now show that

$$Var\left[S^2\right] = \frac{\mu_4}{n} - \frac{(n-3)\mu_2^2}{n(n-1)}$$

j. Now show that

$$Var\left[\hat{\sigma}^2\right] = \frac{(n-1)^2\mu_4}{n^3} - \frac{(n-1)(n-3)\mu_2^2}{n^3}$$

k. Now show that

$$Var\left[\hat{\sigma}^2\right] = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{(\mu_4 - 3\mu_2^2)}{n^3}$$

10. Let the density function be given by

$$f(x;\ \theta) = \theta x^{\theta - 1} \quad 0<x<1, \quad \theta > 0 \tag{44}$$

a. Find the expected value of x. Also show that

$$\int_0^1 x^{\theta-1} dx = \frac{1}{\theta} \tag{45}$$

b. Show that

$$\int_0^1 x^{\theta-1} \ln x \, dx = \frac{-1}{\theta^2} \tag{46}$$

and then find the E(ln x).

c. By differentiating under the integral sign find

$$\int_0^1 x^{\theta-1} (\ln x)^2 dx = \frac{2}{\theta^3} \tag{47}$$

and then find the Var(ln x).

d. Now consider a random sample $x_1, x_2, \ldots, x_n$. Given the above density function, what is the likelihood function? What is a sufficient statistic for θ?

e. What is the MLE estimator of θ? You can use the log likelihood or remember that

$$\frac{d(p^y)}{dy} = p^y \log p \tag{48}$$

and

$$\frac{d(p^{f(y)})}{dy} = p^{f(y)} \log p \, \frac{df(y)}{dy} \tag{49}$$

f. Now consider the parameter α which is given by $\alpha = 1/\theta$. What is the MLE estimator of α, call it $\hat{\alpha}$. Is it unbiased?

g. What is the variance of $\hat{\alpha}$?

h. What is the Fisher information number of $\hat{\alpha}$?

i. What is the Cramer-Rao lower bound for $\hat{\alpha}$?

11. This problem will analyze an idempotent matrix using MATLAB.

a. Set the length of the vector y equal to n, i.e. n = 4.
b. Create an n vector of ones, call it i.
c. Create an n×n identity matrix.
d. Create the matrix $M = I - 1./n.*i*i'$
e. Check that M is idempotent.
f. Input the vector y

$$y = \begin{pmatrix} -2 \\ 5 \\ 3 \\ -2 \end{pmatrix} \quad \textbf{(55)}$$

g. Find the mean of the elements of the vector y.
h. Find My and comment on what it is.
i. Find y′M′ and comment on what it is.
j. Show that y′M′y = y′M′My
k. Find the eigenvectors and values of the matrix M, call them Q and D respectively.
l. Check to make sure MATLAB computes a Q that is orthogonal.
m. Show that $Q' = Q^{-1}$.
n. Show that Q′MQ = D.
o. Find the trace and rank of M.
p. Define v = Q′y.
q. Show that v′Q′MQv = v′Dv = y′My
r. Which element of v is not included in the summation of squares implied in q?
s. Create the matrix L = ones(3,4)
t. Compute LM.
u. Compute C =LQ.
v. What is the structure of C?
w. Compute Ly
x. Show that Cv = Ly

12. This problem will analyze a projection problem using MATLAB
a. Set the length of the vector y equal to n, i.e. n = 4.
b. Create an n×n identity matrix.
c. Input the vector y and the matrix X

$$y = \begin{pmatrix} -2 \\ 5 \\ 3 \\ -2 \end{pmatrix} \quad X = \begin{pmatrix} 1 & -1 \\ 1 & 4 \\ 1 & 2 \\ 1 & -2 \end{pmatrix} \quad \textbf{(56)}$$

d. Create the projection matrix $P = X(X'X)^{-1}X'$

e. Check that P is idempotent.
f. Check that PX = X.
g. Compute the projection of y on the column space of X, p = Py.
h. Compute the error e1 = y-p.
i. Check that e1 is orthogonal to X.
j. Find the vector c which forms the linear combination of X closest to y, $c=(X'X)^{-1}X'y$
k. Check that Xc = p
l. Compute the least square estimate of c1 in the equation y = Xc1 using c1 = X\y and compare to c.
m. Create the matrix $M = I - X(X'X)^{-1}X'$
n. Check that M is idempotent.
o. Check that PM = 0.
p. Show that $y'M'y = y'M'My$
q. Define e = My.
r. Check that e = e1.
s. Show that $e'e = y'My$
t. Show that MX = 0.

13. Assume that the birth weight in grams of a baby born in the United States is normal with a mean of 3315 and a standard deviation of 525, boys and girls combined. Let X equal the weight of a baby girl who is born at home in Dallas County and assume that the distribution of X is $N(\mu_X, \sigma_X^2)$.

a. Assuming 11 observations of X, give the test statistic and critical region for testing H_0: $\mu_X = 3315$ against the alternative hypothesis H_1: $\mu_X > 3315$ (home-born babies are heavier) if $\alpha = 0.01$.

b. Calculate the value of the test statistic and give your conclusion using the following weights:

3119	2657	3459	3629	3345	3629
3515	3856	3629	3345	3062	

c. What is the p-value of the test?

d. Give the test statistic and critical region for testing H_0: $\sigma_X^2 = 525^2$ against the alternative hypothesis H_1: $\sigma_X^2 < 525^2$ (less variation in the weights of home-born babies) if $\alpha = 0.05$.

e. Calculate the value of your test statistic and state your conclusion.

f. What is the p-value of this test.

14. Let Y equal the weight in grams of a baby boy who is born at home in Dallas County and assume that the distribution of Y is $N(\mu_Y, \sigma_Y^2)$. Using the following weights:

4082	3686	4111	3686	3175	4139
3686	3430	3289	3657	4082	

answer the same questions as in problem 13.

15. Consider generating multivariate normal random variables with the following mean vector and covariance matrix.

$$\mu = \begin{pmatrix} -2 \\ 3 \end{pmatrix} \quad COV = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \tag{61}$$

MATLAB has a function for generating random draws from a normal population, R = normrnd(MU,SIGMA). We can use this to generate multivariate random numbers by factoring the covariance matrix and using a transformation. The following code is helpful where cv is the covariance matrix and n is the number of random numbers we want to create.

```
n=500
mu = [-2;3]
k = length(mu)
cv = [1 0.7;0.7 1]
R = chol(cv)
Z = normrnd(0,1,[n,k]);
X = Z*R + ones(n,1)*mu';
```

Compare the mean and covariance of the sample to that of the posited underlying distribution.

Extra on problem 8.

$$[x_1 - \bar{x}, x_2 - \bar{x}, x_3 - \bar{x}, \cdots, x_n - \bar{x}] \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2 - \bar{x}(x_1 + x_2 + x_3 + \cdots + x_n) \tag{62}$$

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$\begin{aligned} \sum_{i=1}^{n} (x_i - \bar{x})^2 &= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 \\ &= \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \end{aligned} \tag{63}$$

Alternative form of expectation.

$$E\left[\left(M_n^2\right)^2\right] = \frac{1}{n^2}\left[E\left[\left(\sum_{i=1}^{n} X_i^2\right)^2\right] - 2nE\left[\bar{X}^2 \sum_{i=1}^{n} X_i^2\right] - n^2 E\left(\bar{X}^4\right)\right]$$

Derivation of variance

$$Var\left[\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2\right] = n\mu_4 + n(n-1)\mu_2^2 - 2\left[\mu_4 + (n-1)\mu_2^2\right] + \left[\frac{1}{n}\left[\mu_4 + 3(n-1)\mu_2^2\right]\right] - (n-1)^2\mu_2^2$$

$$= \frac{n^2\mu_4 + n^2(n-1)\mu_2^2 - 2n\mu_4 - 2n(n-1)\mu_2^2 + \mu_4 + 3(n-1)\mu_2^2 - n(n-1)^2\mu_2^2}{n}$$

$$= \frac{\mu_4(n^2 - 2n + 1) + \left[(n-1)\mu_2^2\right]\left(n^2 - 2n + 3 - n(n-1)\right)}{n}$$

$$= \frac{\mu_4(n-1)^2 + \left[(n-1)\mu_2^2\right]\left(n^2 - 2n + 3 - n^2 + n\right)}{n}$$

$$= \frac{\mu_4(n-1)^2 - \left[(n-1)\mu_2^2\right](n-3)}{n}$$

$$= \frac{(n-1)^2\mu_4}{n} - \frac{(n-1)(n-3)\mu_2^2}{n}$$

$$= \frac{(n-1)^2\mu_4}{n} - \frac{(n-1)(n-3)\sigma^4}{n}$$

j. Now show that

$$Var\left[\hat{\sigma}^2\right] = \frac{(n-1)^2\mu_4}{n^3} - \frac{(n-1)(n-3)\mu_2^2}{n^3}$$

$$= \frac{n^2\mu_4 - 2n\mu_4 + \mu_4}{n^3} - \frac{n^2\mu_2^2 - 4n\mu_2^2 + 3\mu_2^2}{n^3}$$

$$= \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}$$

k. Now show that

$$Var\left[\hat{\sigma}^2\right] = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{(\mu_4 - 3\mu_2^2)}{n^3}$$