ECON 4130
HG Oct. 05

Lecture Note to Rice Chapter 8

1 Random matrices

Let $Y_{ij},\quad i=1,2,\ldots,m,\quad j=1,2,\ldots,n$ be random variables (r.v.'s). The matrix

$$Y=\begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{m1} & Y_{m2} & \cdots & Y_{mn} \end{pmatrix}$$

is called a random matrix (with a joint *mn*-dimensional distribution, $f(y_{11,}y_{12},\ldots,y_{mn})$).
The expected value of Y is *defined* as

(1) $$\mathrm{E}(Y)\overset{\text{def}}{=}\begin{pmatrix} \mathrm{E}(Y_{11}) & \mathrm{E}(Y_{12}) & \cdots & \mathrm{E}(Y_{1n}) \\ \mathrm{E}(Y_{21}) & \mathrm{E}(Y_{22}) & \cdots & \mathrm{E}(Y_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}(Y_{m1}) & \mathrm{E}(Y_{m2}) & \cdots & \mathrm{E}(Y_{mn}) \end{pmatrix}$$

The expectation satisfies the following rules (which follows directly from the definition (1) combined with the corresponding linear properties for the expectation in the scalar case):

i. $\mathrm{E}(AY+C)=A\cdot\mathrm{E}(Y)+C$
where *A*, *C*, are any matrices of constants with dimensions compatible with *Y* (i.e. $A\sim k\times m$, and $C\sim k\times n$, where *k* is arbitrary).

ii. $\mathrm{E}(AYB+C)=A\cdot\mathrm{E}(Y)\cdot B+C$
where *A*, *B*, *C* are any constant matrices compatible with *Y* in dimension so that the product and sum is well defined..

iii. $\mathrm{E}(Y')=\left[\mathrm{E}(Y)\right]'$ where A' denotes the transposed matrix

If $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ is a n-dimensional random vector, it's expectation, $\mu = \mathrm{E}(Y)$ (sometimes written, μ_Y), is therefore the vector of individual expectations,

$$\mu = \mathrm{E}(Y) = \begin{pmatrix} \mathrm{E}(Y_1) \\ \vdots \\ \mathrm{E}(Y_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

Let $\sigma_{ij} = \mathrm{E}\left[(Y_i - \mu_i)(Y_j - \mu_j)\right] = \sigma_{ji}$ be the covariance between Y_i and Y_j. In particular we have $\sigma_{ii} = \mathrm{E}\left[(Y_i - \mu_i)^2\right] = \mathrm{var}(Y_i)$. The covariance matrix of Y (denoted as $\mathrm{cov}(Y)$) is defined as the matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix} = \begin{pmatrix} \mathrm{var}(Y_1) & \cdots & \mathrm{cov}(Y_1, Y_n) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(Y_n, Y_1) & \cdots & \mathrm{var}(Y_n) \end{pmatrix}$$

which can be expressed as

$$\mathrm{cov}(Y) = \mathrm{E}\left[(Y-\mu)(Y-\mu)'\right] = \mathrm{E}\begin{pmatrix} Y_1 - \mu_1 \\ \vdots \\ Y_n - \mu_n \end{pmatrix}(Y_1 - \mu_1, \ldots, Y_n - \mu_n) =$$

$$= \mathrm{E}\begin{pmatrix} (Y_1 - \mu_1)^2 & \cdots & (Y_1 - \mu_1)(Y_n - \mu_n) \\ \vdots & \ddots & \vdots \\ (Y_n - \mu_n)(Y_1 - \mu_1) & \cdots & (Y_n - \mu_n)^2 \end{pmatrix} \overset{(1)}{=} \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix} = \Sigma$$

Example 1
Suppose that $Y_1, Y_2, \ldots, Y_n$ are *iid* with expectation $\mathrm{E}(Y_i) = \eta$ and $\mathrm{var}(Y_i) = \sigma^2$. Then the vector $Y' = (Y_1, \ldots, Y_n)$ has expectation

$$\mathrm{E}(Y) = \begin{pmatrix} \eta \\ \vdots \\ \eta \end{pmatrix} = \eta \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

and covariance matrix (since $\sigma_{ij} = \text{cov}(Y_i, Y_j) = 0$ for $i \neq j$):

$$\text{cov}(Y) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \sigma^2 I_n$$

where I_n is the *n*-dimensional identity matrix. (*End of example*)

If $Y' = (Y_1, \ldots, Y_n)$ is a random vector, and *A* a $p \times n$ constant matrix, we obtain from **i.-iii.** (and the fact that $(BC)' = C'B'$ for matrices *B* and *C*):

(2) $\quad \text{E}(AY) = A \cdot \text{E}(Y) = A\mu$

and

(3) $\quad \text{cov}(AY) = A \cdot \text{cov}(Y) \cdot A' = A\Sigma A'$ (i.e. a $p \times p$ matrix)

which follows from

$$\text{cov}(AY) = \text{E}\left[(AY - A\mu)(AY - A\mu)'\right] = \text{E}\left[A(Y-\mu)(Y-\mu)'A'\right] = A \cdot \text{E}\left[(Y-\mu)(Y-\mu)'\right]A' = A\Sigma A'$$

In particular, if Z is a linear combination of $Y_1, \ldots, Y_n$, i.e. $Z = a_1 Y_1 + \cdots + a_n Y_n$, then

(4) $\quad \text{var}(Z) = \text{var}(a'Y) = a'\Sigma a$ where $a' = (a_1, \ldots, a_n)$ and $\Sigma = \text{cov}(Y)$.

[**Proof:** Since $Z = (a_1, \ldots, a_n)\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = a'Y$ can be considered a 1×1 matrix, we must have that $Z' = Z$, and, therefore, $\text{cov}(Z) = \text{var}(Z)$ (i.e., $\text{cov}(Z) = \text{E}\left[(Z - \text{E}(Z))(Z - \text{E}(Z))'\right] = \text{E}\left[(Z - \text{E}(Z))^2\right] = \text{var}(Z)$). We then see that (4) is a special case of (3) with $A = a'$]

Example 2 Ordinary least squares (OLS)

Consider the standard multiple regression model with one response, *Y*, and *p* explanatory variables

(5) $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + u_i$ for $i = 1, 2, \ldots, n$

where, for simplicity, all x_{ij} are considered fixed, non random quantities, and the errors, $u_1, u_2, \ldots, u_n$ are assumed to be *iid* and normally distributed with expectation, $\mathrm{E}(u_i) = 0$ and $\mathrm{var}(u_i) = \sigma^2$. We can write (5) in matrix form as follows

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} \\ \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

The three matrices on the right we denote by X, β, and u respectively. The model can now be written as

(6) $Y = X\beta + u$

where *X* is the $n \times p$ (so called) design matrix, β the $(p+1) \times 1$ vector of regression coefficients, and *u* the $n \times 1$ vector of errors. Since

$$\mathrm{E}(u) = \begin{pmatrix} \mathrm{E}(u_1) \\ \mathrm{E}(u_2) \\ \vdots \\ \mathrm{E}(u_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \underline{0}$$

(where $\underline{0}$ denotes a vector of zeroes), we get from **i.** (noting that $X\beta$ is non random)

(7) $\mathrm{E}(Y) = X\beta + \mathrm{E}(u) = X\beta$

The covariance matrix for *Y* becomes, since $Y - X\beta = u$, and using example 1,

(8) $\Sigma_Y = \mathrm{cov}(Y) = \mathrm{E}\left[(Y - X\beta)(Y - X\beta)'\right] = \mathrm{E}\left[\mathrm{uu}'\right] = \mathrm{cov}(u) = \sigma^2 I_n$

The OLS estimator, $\hat{\beta}$, for β is obtained by minimizing the sum of squares

$$Q = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip} \right)^2$$

with respect to β. Differentiating Q with respect to all the β_j's, and setting the derivatives equal to 0, leads to the following system of equations that the $\hat{\beta}_j$'s must satisfy

$$\begin{aligned}
& n\hat{\beta}_0 + \left(\sum\nolimits_i x_{i1}\right)\hat{\beta}_1 + \cdots + \left(\sum\nolimits_i x_{ip}\right)\hat{\beta}_p = \sum\nolimits_i Y_i \\
& \left(\sum\nolimits_i x_{i1}\right)\hat{\beta}_0 + \left(\sum\nolimits_i x_{i1}^2\right)\hat{\beta}_1 + \cdots + \left(\sum\nolimits_i x_{i1}x_{ip}\right)\hat{\beta}_p = \sum\nolimits_i x_{i1}Y_i \\
& \qquad \cdots \\
& \left(\sum\nolimits_i x_{ip}\right)\hat{\beta}_0 + \left(\sum\nolimits_i x_{ip}x_{i1}\right)\hat{\beta}_1 + \cdots + \left(\sum\nolimits_i x_{ip}^2\right)\hat{\beta}_p = \sum\nolimits_i x_{ip}Y_i
\end{aligned}$$

Noting that the coefficients of the left side are exactly the elements in the $(p+1)\times(p+1)$ matrix $X'X$, and that the right side, written as a vector, simply is $X'Y$, we can write the system more compactly as

$$X'X\hat{\beta} = X'Y$$

Assuming that $X'X$ is non singular (which can be shown to be the case if no single *x*-variable can be written exactly as a linear combination of the other *x*-variables, i.e., there is no exact collinearity between the explanatory variables), we obtain the solution (the OLS estimator)

(9) $$\hat{\beta} = (X'X)^{-1}X'Y$$

It is now easy to prove that $\hat{\beta}$ is unbiased since, from **i.** and (7)

(10) $$\mathrm{E}(\hat{\beta}) = \mathrm{E}\left[(X'X)^{-1}X'Y\right] \overset{\text{i.}}{=} (X'X)^{-1}X'\mathrm{E}(Y) \overset{(7)}{=} (X'X)^{-1}X'X\beta = I_p\beta = \beta$$

Writing $C = (X'X)^{-1}X'$, we have $\hat{\beta} = CY$, and obtain the covariance matrix from (3) and (8) [*and also using the rule that the transposed of an inverse square matrix is the inverse of the transposed,* $\left[A^{-1}\right]' = (A')^{-1}$*, which is seen by taking the transposed of the equation,* $A\cdot A^{-1} = I$*. Remember also the* $AI_n = A$ *for any* $p\times n$*-matrix A, and that, if c is a scalar, then c as factor can be taken outside a matrix product,* $A\cdot(cB) = cAB$.].

$$\mathrm{cov}(\hat{\beta}) = \mathrm{cov}(CY) \overset{(3)}{=} C\Sigma_Y C' \overset{(8)}{=} C\left(\sigma^2 I_n\right)C' = \sigma^2 CC' = \sigma^2(X'X)^{-1}XX'(X'X)^{-1}$$

Hence

(11) $\operatorname{cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ (*End of example*.)

2 Multinormal distributions

We say that the vector $X' = (X_1, \ldots, X_n)$ is (multi)normally distributed with expectation $\mu = \mathrm{E}(X)$, and covariance matrix, $\Sigma = \operatorname{cov}(X)$ (written shortly $X \sim N(\mu, \Sigma)$), if the joint pdf is given by

(12) $$f(x_1, \ldots, x_n \mid \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \quad \text{where } x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ and } \det(\Sigma)$$

means the determinant of Σ.

This distribution has a lot of convenient mathematical properties (see e.g. Greene, Econometric Analysis, chapter 3, for a summary), but here we only need the following:

(13) If $X \sim N(\mu, \Sigma)$ and A is a $p \times n$ constant matrix ($p \le n$) and b a $p \times 1$ constant vector, then $Y = AX + b \sim N(\mathrm{E}(Y), \operatorname{cov}(Y)) = N(A\mu + b,\ A\Sigma A')$

[*For proof see e.g. Greene chapter 3.*]

In particular, this shows that all marginal distributions are also normal. For example, the marginal distribution of X_1, X_2 is normal since

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = AX \text{ where } A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \end{pmatrix} \text{ which gives (check!)}$$

(14) $$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(\mathrm{E}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix},\ \operatorname{cov}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}) = N(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}) ,$$

i.e. a bivariate normal distribution

Exercise. Show that the pdf in (14) as defined in (12), reduces to the bivariate normal density as defined in Example F in Rice section 3.3. [**Hint:** Introduce the correlation, ρ, between X_1 and X_2, $\rho = \sigma_{12} \big/ \sqrt{\sigma_{11}\sigma_{22}}$, implying $\sigma_{12}^2 = \sigma_{11}\sigma_{22}\rho$, and the determinant, $\det(\text{cov}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}) = \sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}(1-\rho^2)$ etc.]

Example 3 (Continuation of example 2)

The error vector, *u*, in (6) has expectation $\underline{0}$ and covariance, $\Sigma_{\text{u}} = \text{cov}(u) = \sigma^2 I_n$. We see from (12) that saying that $u \sim N(\underline{0},\ \sigma^2 I_n)$ is the same as saying that $u_1, u_2, \ldots, u_n$ are *iid* and normally distributed with expectation, $\text{E}(u_i) = 0$ and $\text{var}(u_i) = \sigma^2$. In fact, we have the determinant

$$\det(\Sigma_u) = \det(\sigma^2 I_n) = \det\begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^{2n}.$$

and the exponent in (12) reduces to

$$-\frac{1}{2}(u - \text{E}(u))'\Sigma_u^{-1}(u - \text{E}(u)) = -\frac{1}{2}u'\left(\sigma^2 I_n\right)^{-1}u = -\frac{1}{2}u'\left(\frac{1}{\sigma^2}I_n\right)u = -\frac{1}{2\sigma^2}u'u = -\frac{1}{2\sigma^2}\sum_i u_i^2$$

Substituting in (12), shows that the joint distribution (12) reduces to the product of *n* one-dimensional $N(0, \sigma^2)$ -distributions as the *iid* statement would imply.

By (13), (7), and (8) we obtain that *Y* is normally distributed, $Y \sim N(\text{E}(Y), \text{cov}(Y)) = N(X\beta,\ \sigma^2 I_n)$, and, by (13) again, that $\hat{\beta} = (X'X)^{-1}X'Y$ is normally distributed

$$\hat{\beta} \sim N(\text{E}(\hat{\beta}), \text{cov}(\hat{\beta})) = N(\beta,\ \sigma^2(X'X)^{-1})$$

(*End of example*.)

3 On the asymptotic distribution for mle estimators (the multi parameter case)

In this section we will only describe how to determine the asymptotic distribution for the mle estimator in case there are several unknown parameters in the model, without going into details of derivations and proofs. A good summary of the theory can be found in chapter 4 of Greene's book, *Econometric Analysis*. See also Rice at the end of section 8.5.2.

Suppose that $X_1, X_2, \ldots, X_n$ are *iid* with $X_i \sim f(x_i \mid \theta)$ (pdf), where $\theta' = (\theta_1, \theta_2, \ldots, \theta_r)$ is a *r*-dimensional vector of unknown parameters. Then the joint pdf is $\prod_{i=1}^{n} f(x_i \mid \theta)$ and the log likelihood is

$$l(\theta) = \sum_{i=1}^{n} \ln f(x_i \mid \theta)$$

The mle estimator, $\hat{\theta}$, solves r equations

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \ln f(x_i \mid \hat{\theta}) = 0, \qquad j = 1, 2, \ldots, r$$

In order to define the $r \times r$ Fisher information matrix that is needed in the asymptotic distribution of $\hat{\theta}$, we introduce

$$m_{ij}(\theta) = -\mathrm{E}\, \frac{\partial^2 \ln f(X_i \mid \theta)}{\partial \theta_i \partial \theta_j} \qquad i, j = 1, 2, \ldots, r$$

Then the Fisher information matrix for one observation is defined as

$$I(\theta) = \begin{pmatrix} m_{11}(\theta) & \cdots & m_{1r}(\theta) \\ \vdots & \ddots & \vdots \\ m_{r1}(\theta) & \cdots & m_{rr}(\theta) \end{pmatrix}$$

Under regularity conditions similar to the one-parameter case (see Greene for details), we have that the mle satisfies

$$\sqrt{n}(\hat{\theta} - \theta) \underset{n \to \infty}{\overset{D}{\to}} N(\underline{0}, I(\theta)^{-1})$$

The definition of convergence in distribution for random vectors is similar but slightly more technical than the definition for the one-dimensional case, and we skip the details here (see Greene for a precise definition). However, the interpretation of the result is the same as in the one-dimensional case, i.e., that for large *n*,

$$\hat{\theta} \overset{\text{approximately}}{\sim} N\left(\theta,\ \frac{1}{n} I(\theta)^{-1}\right)$$

Hence we can say that $\hat{\theta}$ is asymptotically unbiased with asymptotic covariance matrix, $(1/n)I(\theta)^{-1}$. This matrix is unknown since θ is unknown, but can be consistently estimated by replacing θ by $\hat{\theta}$ (or any other consistent estimator of θ). [That $\hat{\theta}$ is consistent means simply that $\hat{\theta}_j \underset{n\to\infty}{\overset{P}{\to}} \theta_j$ for all $j = 1, 2, \ldots, r$]. A generalization of Slutzki's lemma to the multivariate case (details omitted), now allow us to conclude that, for large *n*

(15) $$\boxed{\hat{\theta} \overset{\text{approximately}}{\sim} N\left(\theta,\ \frac{1}{n} I(\hat{\theta})^{-1}\right)}$$

which is the important result that you should know. Using (13) we also have

(16) $$A\hat{\theta} \overset{\text{approximately}}{\sim} N\left(A\theta,\ \frac{1}{n} A \cdot I(\hat{\theta})^{-1} A'\right)$$

for any constant, $p \times r$ matrix *A*.

From this we get the following: Let $k_{ij}(\theta)$ denote element *i,j* in $I(\theta)^{-1}$. Then the estimated asymptotic variance of $\hat{\theta}_j$ is the *j*-th element on the main diagonal in the estimated covariance matrix, i.e. $k_{jj}(\hat{\theta})/n$.

[Follows from (16). In fact, let $a' = (0, \ldots, 1, \ldots, 0)$ where the 1 is in position *j* and zeroes elsewhere. Then from (16)

$$\hat{\theta}_j = a'\hat{\theta} \overset{\text{approx.}}{\sim} N\left(a'\theta, \frac{1}{n} a' I(\hat{\theta})^{-1} a\right) = N\left(\theta_j, \frac{k_{jj}(\hat{\theta})}{n}\right) \quad]$$

Hence, we obtain an approximate $1-\alpha$ CI for θ_j: $\hat{\theta}_j \pm z_{\alpha/2} \sqrt{k_{jj}(\hat{\theta})} \Big/ \sqrt{n}$
where $z_{\alpha/2}$ is the upper $\alpha/2$-point in $N(0,1)$.

Example 4. Assume we want a CI for the transformed parameter, $\eta = \theta_1 - \theta_2$. This we obtain from (16): Let $b' = (1, -1, 0, \ldots, 0)$. Then, by (16),

$$\hat{\eta} = \hat{\theta}_1 - \hat{\theta}_2 = b'\hat{\theta} \overset{\text{approx.}}{\sim} N\left(b'\theta,\ \frac{b' I(\hat{\theta})^{-1} b}{n} \right) = N\left(\theta_1 - \theta_2,\ \frac{1}{n}(k_{11}(\hat{\theta}) + k_{22}(\hat{\theta}) - 2k_{12}(\hat{\theta})) \right)$$

which leads to the approximate $1-\alpha$ CI for $\theta_1 - \theta_2$:

$$\hat{\theta}_1 - \hat{\theta}_2 \pm z_{\alpha/2} \frac{1}{\sqrt{n}} \sqrt{k_{11}(\hat{\theta}) + k_{22}(\hat{\theta}) - 2k_{12}(\hat{\theta})}$$

[Note that all covariance matrices are symmetric. Hence $k_{12}(\hat{\theta}) = k_{21}(\hat{\theta})$.]

(*End of example*.)

Example 5 (On example C in Rice section 8.5 – precipitation data)

Let X_i be the amount of precipitation for rainstorm no. *i*, $i = 1, 2, \ldots, n$ ($n = 227$ observations).

Model: $X_1, X_2, \ldots, X_n$ are *iid* with $X_i \sim \Gamma(\alpha, \lambda)$. The joint distribution is

$$X_1, X_2, \ldots, X_n \sim \prod_{i=1}^{n} f(x_i \mid \alpha, \lambda) = \frac{\lambda^{n\alpha}}{\Gamma(\alpha)} (x_1 x_2 \cdots x_n)^{\alpha-1} e^{-\lambda \sum x_i}$$

The log likelihood is

(17) $$l(\alpha, \lambda) = n\alpha \ln\alpha + (\alpha - 1)\sum_i \ln x_i - \lambda \sum_i x_i - n \ln\Gamma(\alpha)$$

The first derivatives of *l* are

$$\frac{\partial l}{\partial \alpha} = n \ln\lambda + \sum_i \ln x_i - n \frac{\partial}{\partial \alpha} \ln\Gamma(\alpha)$$

$$\frac{\partial l}{\partial \lambda} = n \frac{\alpha}{\lambda} - \sum_i x_i$$

Setting the derivatives equal to zero and solving with respect to α and λ, gives the mle estimators $\hat{\alpha}$ and $\hat{\lambda}$. [**Note.** There are no explicit formulas for the solution, they must be found by numerical iterations. For example, Excel works well in this case by the Solver module: Choose two cells for the arguments α and λ, with start values e.g. at the moment estimates, and then a third cell for the function (17). Then use Solver to maximize (17). This can also be done in STATA by the ml-command, but slightly more involved.]

Using his program, Rice obtained the mle estimates.

$$\hat{\alpha} = 0,441 \quad \text{and} \quad \hat{\lambda} = 1,96$$

We want approximate 90% CI's for α and λ based on the asymptotic normal distribution of $\hat{\alpha}$ and $\hat{\lambda}$. In order to calculate the asymptotic standard errors we need the so called di- and trigamma functions:

Digamma function: $\psi(\alpha) = \frac{\partial}{\partial \alpha} \ln\Gamma(\alpha)$

Trigamma function: $\psi'(\alpha) = \frac{\partial^2}{\partial \alpha^2} \ln\Gamma(\alpha)$

Both functions can be calculated in STATA (under the names digamma and trigamma).

We need the Fisher information matrix:

$$\ln f(X_i \mid \alpha, \lambda) = \alpha \ln\lambda - \ln\Gamma(\alpha) + (\alpha - 1)\ln X_i - \lambda X_i$$

giving

$$\frac{\partial \ln f}{\partial \alpha} = \ln\lambda - \psi(\alpha) + \ln X_i \quad \text{and} \quad \frac{\partial \ln f}{\partial \lambda} = \frac{\alpha}{\lambda} - X_i$$

Hence

$$\frac{\partial^2 \ln f}{\partial \alpha^2} = -\psi'(\alpha) \quad \text{(trigamma)}$$

$$\frac{\partial^2 \ln f}{\partial \alpha \partial \lambda} = \frac{\partial^2 \ln f}{\partial \lambda \partial \alpha} = \frac{1}{\lambda}$$

$$\frac{\partial^2 \ln f}{\partial \lambda^2} = -\frac{\alpha}{\lambda^2}$$

Hence the Fisher information matrix for one observation

$$I(\alpha,\lambda) = -\mathrm{E}\begin{pmatrix} -\psi'(\alpha) & \frac{1}{\lambda} \\ \frac{1}{\lambda} & -\frac{\alpha}{\lambda^2} \end{pmatrix} = \begin{pmatrix} \psi'(\alpha) & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \frac{\alpha}{\lambda^2} \end{pmatrix}$$

The inverse of a symmetric 2×2 matrix is

$$\begin{pmatrix} a & c \\ c & b \end{pmatrix}^{-1} = \frac{1}{ab-c^2}\begin{pmatrix} b & -c \\ -c & a \end{pmatrix}$$

Hence

$$I(\alpha,\lambda)^{-1} = \frac{1}{\frac{\alpha\psi'(\alpha)}{\lambda^2} - \frac{1}{\lambda^2}}\begin{pmatrix} \frac{\alpha}{\lambda^2} & \frac{1}{\lambda} \\ \frac{1}{\lambda} & \psi'(\alpha) \end{pmatrix} = \frac{1}{\alpha\psi'(\alpha)-1}\begin{pmatrix} \alpha & \lambda \\ \lambda & \lambda^2\psi'(\alpha) \end{pmatrix}$$

We obtain an estimate of this by substituting the mle, $\hat{\alpha} = 0,441$ and $\hat{\lambda} = 1,96$, for α and λ

$$I(\hat{\alpha},\hat{\lambda})^{-1} = \frac{1}{\hat{\alpha}\psi'(\hat{\alpha})-1}\begin{pmatrix} \hat{\alpha} & \hat{\lambda} \\ \hat{\lambda} & \hat{\lambda}^2\psi'(\hat{\alpha}) \end{pmatrix} = \begin{pmatrix} 0,25903 & 1,15123 \\ 1,15123 & 13,82770 \end{pmatrix}$$

Here we found $\psi'(\hat{\alpha}) = 6,128169$ from STATA by the command:

```
di trigamma(0.441)
```

From the theory we have that $\begin{pmatrix} \hat{\alpha} \\ \hat{\lambda} \end{pmatrix} \overset{\text{approx.}}{\sim} N\left(\begin{pmatrix} \alpha \\ \lambda \end{pmatrix},\ C\right)$, where the asymptotic covariance is

$$C = \frac{1}{n} I(\hat{\alpha},\hat{\lambda})^{-1} = \begin{pmatrix} 0,0011411 & 0,0050715 \\ 0,0050715 & 0,0609150 \end{pmatrix}$$

Hence the asymptotic standard errors

$$\mathrm{se}(\hat{\alpha}) = \sqrt{0,0011411} = 0,03378 \text{ and } \mathrm{se}(\hat{\lambda}) = \sqrt{0,060950} = 0,24681$$

According to the theory we then obtain approximate 90% CI for α and λ

$$\hat{\alpha} \pm 1,64 \cdot \text{se}(\hat{\alpha}) = 0,441 \pm (1,64)(0,03378) = [0,386,\ \ 0,496]$$

$$\hat{\lambda} \pm 1,64 \cdot \text{se}(\hat{\lambda}) = 1,96 \pm (1,64)(0,247) = [1,55,\ \ 2,37]$$

Rice (example E, section 8.5.3)) obtains approximate 90% CI's by the parametric bootstrap method:

α : [0,404, 0,523]

λ : [1,46, 2,32]

The difference between the asymptotic intervals and the bootstrap intervals does not appear to be substantial. With as much as 227 observations it is to be expected that the asymptotic theory should work well.