Appendix A

Matrix Algebra

1. For the matrices $\mathbf{A} = \begin{bmatrix} 1 & 3 & 3 \\ 2 & 4 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 4 \\ 1 & 5 \\ 6 & 2 \end{bmatrix}$ compute **AB**, **A′B′**, and **BA**.

$$\mathbf{AB} = \begin{bmatrix} 23 & 25 \\ 14 & 30 \end{bmatrix},\ \mathbf{BA} = \begin{bmatrix} 10 & 22 & 10 \\ 11 & 23 & 8 \\ 10 & 26 & 20 \end{bmatrix},\ \mathbf{A}'\mathbf{B}' = (\mathbf{BA})' = \begin{bmatrix} 10 & 11 & 10 \\ 22 & 23 & 26 \\ 10 & 8 & 20 \end{bmatrix}.$$

2. Prove that $tr(\mathbf{AB}) = tr(\mathbf{BA})$ where **A** and **B** are any two matrices that are conformable for both multiplications. They need not be square.

The ith diagonal element of **AB** is $\sum_j a_{ij}b_{ji}$. Summing over i produces $tr(\mathbf{AB}) = \sum_i \sum_j a_{ij}b_{ji}$. The jth diagonal element of **BA** is $\sum_j b_{ji}a_{ij}$. Summing over i produces $tr(\mathbf{BA}) = \sum_i \sum_j b_{ji}a_{ij}$.

3. Prove that $tr(\mathbf{A}'\mathbf{A}) = \sum_i \sum_j a_{ij}^2$.

The jth diagonal element of **A′A** is the inner product of the jth column of **A**, or $\sum_i a_{ij}^2$. Summing over j produces $tr(\mathbf{A}'\mathbf{A}) = \sum_j \sum_i a_{ij}^2 = \sum_i \sum_j a_{ij}^2$.

4. Expand the matrix product $\mathbf{X} = \{[\mathbf{AB} + (\mathbf{CD})'][(\mathbf{EF})^{-1} + \mathbf{GH}]\}'$. Assume that all matrices are square and **E** and **F** are nonsingular.

In parts, $(\mathbf{CD})' = \mathbf{D}'\mathbf{C}'$ and $(\mathbf{EF})^{-1} = \mathbf{F}^{-1}\mathbf{E}^{-1}$. Then, the product is

$$\{[\mathbf{AB} + (\mathbf{CD})'][(\mathbf{EF})^{-1} + \mathbf{GH}]\}' = (\mathbf{ABF}^{-1}\mathbf{E}^{-1} + \mathbf{ABGH} + \mathbf{D}'\mathbf{C}'\mathbf{F}^{-1}\mathbf{E}^{-1} + \mathbf{D}'\mathbf{C}'\mathbf{GH})'$$
$$= (\mathbf{E}^{-1})'(\mathbf{F}^{-1})'\mathbf{B}'\mathbf{A}' + \mathbf{H}'\mathbf{G}'\mathbf{B}'\mathbf{A}' + (\mathbf{E}^{-1})'(\mathbf{F}^{-1})'\mathbf{CD} + \mathbf{H}'\mathbf{G}'\mathbf{CD}.$$ □

5. Prove for that for $K\times 1$ column vectors, $\mathbf{x}_i$ $i = 1,...,n$, and some nonzero vector, **a**,

$$\sum_{i=1}^n (\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})' = \mathbf{X}'\mathbf{M}^0\mathbf{X} + n(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})'.$$

Write $\mathbf{x}_i - \mathbf{a}$ as $[(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]$. Then, the sum is

$$\sum_{i=1}^n [(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]\,[(\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{a})]' =$$
$$\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' + \sum_{i=1}^n (\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})'$$
$$+ \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}} - \mathbf{a})' + \sum_{i=1}^n (\bar{\mathbf{x}} - \mathbf{a})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

Since $(\bar{\mathbf{x}} - \mathbf{a})$ is a vector of constants, it may be moved out of the summations. Thus, the fourth term is $(\bar{\mathbf{x}} - \mathbf{a})\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' = \mathbf{0}$. The third term is likewise. The first term is $\mathbf{X}'\mathbf{M}^0\mathbf{X}$ by the definition while the second is $n(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})'$. □

6. Let **A** be any square matrix whose columns are $[\mathbf{a}_1,\mathbf{a}_2,...,\mathbf{a}_M]$ and let **B** be any rearrangement of the columns of the $M\times M$ identity matrix. What operation is performed by the multiplication **AB**? What about **BA**?

B is called a permutation matrix. Each column of **B**, say, $\mathbf{b}_i$, is a column of an identity matrix. The jth column of the matrix product **AB** is $\mathbf{A}\,\mathbf{b}_i$ which is the jth column of **A**. Therefore, post multiplication of **A** by **B** simply rearranges (permutes) the columns of **A** (hence the name). Each row of the product **BA** is one of the rows of **A**, so the product **BA** is a rearrangement of the rows of **A**. Of course, **A** need not be square for us to permute its rows or columns. If not, the applicable permutation matrix will be of different orders for the rows and columns.

7. Consider the 3×3 case of the matrix **B** in Exercise 6. For example, $\mathbf{B} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ Compute $\mathbf{B}^2$ and $\mathbf{B}^3$. Repeat for a 4×4 matrix. Can you generalize your finding?

$$\mathbf{B}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{B}^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since each power of **B** is a rearrangement of **I**, some power of **B** will equal **I**. If n is this power, we also find, therefore, that $\mathbf{B}^{n-1} = \mathbf{B}^{-1}$. This will hold generally.

8. Calculate $|\mathbf{A}|$, $tr(\mathbf{A})$ and $\mathbf{A}^{-1}$ for $\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 3 & 2 & 5 \\ 5 & 2 & 8 \end{bmatrix}$.

$|\mathbf{A}|$ = 1(2)(8)+4(5)(5)+3(2)(7)-5(2)(7)-1(5)(2)-3(4)(8) = -18,
$tr(\mathbf{A})$ = 1 + 2 + 8 = 11

$$\mathbf{A}^{-1} = \frac{-1}{18}\begin{bmatrix} \det\begin{pmatrix} 2 & 5 \\ 2 & 8 \end{pmatrix} & -\det\begin{pmatrix} 4 & 7 \\ 2 & 8 \end{pmatrix} & \det\begin{pmatrix} 4 & 7 \\ 2 & 5 \end{pmatrix} \\ -\det\begin{pmatrix} 3 & 5 \\ 5 & 8 \end{pmatrix} & \det\begin{pmatrix} 1 & 7 \\ 5 & 8 \end{pmatrix} & -\det\begin{pmatrix} 1 & 7 \\ 3 & 5 \end{pmatrix} \\ \det\begin{pmatrix} 3 & 2 \\ 5 & 2 \end{pmatrix} & -\det\begin{pmatrix} 1 & 4 \\ 5 & 2 \end{pmatrix} & \det\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix} \end{bmatrix} = \begin{bmatrix} -6/18 & 18/18 & -6/18 \\ -1/18 & 27/18 & -16/18 \\ 4/18 & -18/18 & 10/18 \end{bmatrix}. \quad \square$$

9. Obtain the Cholesky decomposition of the matrix $\mathbf{A} = \begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}$.

Recall that the Cholesky decomposition of a matrix, **A**, is the matrix product **LU** = **A** where **L** is a lower triangular matrix and **U** = **L′**. Write the decomposition as $\begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}. = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \end{bmatrix}. \begin{bmatrix} \lambda_{11} & \lambda_{21} \\ 0 & \lambda_{22} \end{bmatrix}$. By direct multiplication, 25 = λ_{11}^2 so $\lambda_{11} = 5$. Then, $\lambda_{11}\lambda_{21} = 7$, so $\lambda_{21} = 7/5 = 1.4$. Finally, $\lambda_{21}^2 + \lambda_{22}^2 = 13$, so $\lambda_{22} = 3.322$.

10. A symmetric positive definite matrix, **A**, can also be written as **A** = **UL**, where **U** is an upper triangular matrix and **L** = **U′**. This is not the Cholesky decomposition, however. Obtain this decomposition of the matrix in Exercise 9.

Using the same logic as in the previous problem, $\begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}. = \begin{bmatrix} \mu_{11} & \mu_{12} \\ 0 & \mu_{22} \end{bmatrix}. \begin{bmatrix} \mu_{11} & 0 \\ \mu_{12} & \mu_{22} \end{bmatrix}$. Working from the bottom up, $\mu_{22} = \sqrt{13} = 3.606$. Then, $7 = \mu_{12}\mu_{22}$ so $\mu_{12} = 7/\sqrt{13} = 1.941$. Finally, $25 = \mu_{11}^2 + \mu_{12}^2$ so $\mu_{11}^2 = 25 - 49/13 = 21.23$, or $\mu_{11} = 4.61$.

11. What operation is performed by postmultiplying a matrix by a diagonal matrix? What about premultiplication?

The columns are multiplied by the corresponding diagonal element. Premultiplication multiplies the rows by the corresponding diagonal element.

12. Are the following quadratic forms positive for all values of **x**?

(a) $y = x_1^2 - 28x_1x_2 + (11x_2^2)$,

(b) $y = 5x_1^2 + x_2^2 + 7x_3^2 + 4x_1x_2 + 6x_1x_3 + 8x_2x_3$?

The first may be written $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -14 \\ -14 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. The determinant of the matrix is 121 - 196 = -75, so it is not positive definite. Thus, the first quadratic form need not be positive. The second uses the matrix $\begin{bmatrix} 5 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 7 \end{bmatrix}$. There are several ways to check the definiteness of a matrix. One way is to check the signs of the principal minors, which must be positive. The first two are 5 and 5(1)-2(2)=1, but the third, the determinant, is -34. Therefore, the matrix is not positive definite. Its three characteristic roots are 11.1, 2.9, and -1. It follows, therefore, that there are values of x_1, x_2, and x_3 for which the quadratic form is negative.

13. Prove that $tr(\mathbf{A}\otimes\mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$.

The jth diagonal block of the product is $a_{jj}\mathbf{B}$. Its ith diagonal element is $a_{jj}b_{ii}$. If we sum in the jth block, we obtain $\sum_i a_{jj}b_{ii} = a_{jj}\sum_i b_{ii}$. Summing down the diagonal blocks gives the trace, $\sum_j a_{jj}\sum_i b_{ii}$ = $tr(\mathbf{A})tr(\mathbf{B})$.

14. A matrix, **A**, is *nilpotent* if $\lim_{k\to\infty} \mathbf{A}^k = \mathbf{0}$. Prove that a necessary and sufficient condition for a symmetric matrix to be nilpotent is that all of its characteristic roots be less than one in absolute value.

Use the spectral decomposition to write **A** as $\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$ where Λ is the diagonal matrix of characteristic roots. Then, the Kth power of **A** is $\mathbf{C}\boldsymbol{\Lambda}^K\mathbf{C}'$. Sufficiency is obvious. Also, since if some λ is greater than one, Λ^K must explode, the condition is necessary as well.

15. Compute the characteristic roots of $\mathbf{A} = \begin{bmatrix} 2 & 4 & 3 \\ 4 & 8 & 6 \\ 3 & 6 & 5 \end{bmatrix}$.

The roots are determined by $|\mathbf{A} - \lambda\mathbf{I}| = 0$. For the matrix above, this is

$$|\mathbf{A} - \lambda\mathbf{I}| = (2-\lambda)(8-\lambda)(5-\lambda) + 72 + 72 - 9(8-\lambda) - 36(2-\lambda) - 16(5-\lambda)$$
$$= -\lambda^3 + 15\lambda^2 - 5\lambda = -\lambda(\lambda^2 - 15\lambda + 5) = 0.$$

One solution is obviously zero. (This might have been apparent. The second column of the matrix is twice the first, so it has rank no more than two, and therefore no more than two nonzero roots.) The other two roots are $\left(15 \pm \sqrt{205}\right)/2$ = .341 and 4.659.

16. Suppose $\mathbf{A} = \mathbf{A}(z)$ where z is a scalar. What is $\partial\mathbf{x}'\mathbf{A}\mathbf{x}/\partial z$? Now, suppose each element of **x** is also a function of z. Once again, what is $\partial\mathbf{x}'\mathbf{A}\mathbf{x}/\partial z$?

The quadratic form is $\sum_i\sum_j x_ix_ja_{ij}$, so

$\partial\mathbf{x}'\mathbf{A}(z)\mathbf{x}/\partial z = \sum_i\sum_j x_ix_j(\partial a_{ij}/\partial z) = \mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x}$ where $\partial\mathbf{A}(z)/\partial z$ is a matrix of partial derivatives.

Now, if each element of **x** is also a function of z, then,

$$\partial\mathbf{x}'\mathbf{A}\mathbf{x}/\partial z = \sum_i\sum_j x_ix_j(\partial a_{ij}/\partial z) + \sum_i\sum_j (\partial x_i/\partial z)x_ja_{ij} + \sum_i\sum_j x_i(\partial x_j/\partial z)a_{ij}$$
$$= \mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x} + (\partial\mathbf{x}(z)/\partial z)'\mathbf{A}(z)\mathbf{x}(z) + \mathbf{x}(z)'\mathbf{A}(z)(\partial\mathbf{x}(z)/\partial z)$$

If **A** is symmetric, this simplifies a bit to $\mathbf{x}'(\partial\mathbf{A}(z)/\partial z)\mathbf{x} + 2(\partial\mathbf{x}(z)/\partial z)'\mathbf{A}(z)\mathbf{x}(z)$.

17. Show that the solutions to the determinantal equations $|\mathbf{B} - \lambda\mathbf{A}| = 0$ and $|\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}| = 0$ are the same. How do the solutions to this equation relate to those of the equation $|\mathbf{B}^{-1}\mathbf{A} - \mu\mathbf{I}| = 0$?

Since **A** is assumed to be nonsingular, we may write

$$\mathbf{B} - \lambda\mathbf{A} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}). \text{ Then, } |\mathbf{B} - \lambda\mathbf{A}| = |\mathbf{A}|\times|\mathbf{A}^{-1}\mathbf{B} - \lambda\mathbf{I}|.$$

The determinant of **A** is nonzero if **A** is nonsingular, so the solutions to the two determinantal equations must be the same. $\mathbf{B}^{-1}\mathbf{A}$ is the inverse of $\mathbf{A}^{-1}\mathbf{B}$, so its characteristic roots must be the reciprocals of those of $\mathbf{A}^{-1}\mathbf{B}$. There might seem to be a problem here since these two matrices need not be symmetric, so the roots could be complex. But, for the application noted, both **A** and **B** are symmetric and positive definite. As such, it can be shown tat the solution is the same as that of a third determinantal equation involving a symmetric matrix.

18. Using the matrix **A** in Exercise 9, find the vector **x** that minimizes $y = \mathbf{x}'\mathbf{A}\mathbf{x} + 2x_1 + 3x_2 - 10$. What is the value of y at the minimum? Now, minimize y subject to the constraint $x_1 + x_2 = 1$. Compare the two solutions.

The solution which minimizes $y = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + d$ will satisfy $\partial y\partial\mathbf{x} = 2\mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{0}$. For this problem, $\mathbf{A} = \begin{bmatrix} 25 & 7 \\ 7 & 13 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and $\mathbf{A}^{-1} = \begin{bmatrix} 13/276 & -7/276 \\ -7/276 & 25/276 \end{bmatrix}$, so the solution is x_1 =-5/552 = -.0090597 and x_2 = -61/552 = -.110507.

The constrained maximization problem may be set up as a Lagrangean,
$L^* = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{b}'\mathbf{x} + d + \lambda(\mathbf{c}'\mathbf{x} - 1)$ where $\mathbf{c} = [1,1]'$. The necessary conditions for the solution are

$$\partial L^*/\partial\mathbf{x} = 2\mathbf{A}\mathbf{x} + \mathbf{b} + \lambda\mathbf{c} = \mathbf{0}$$

$$\partial L^*/\partial\lambda = \mathbf{c}'\mathbf{x} - 1 = 0,$$

or,

$$\begin{bmatrix} 2\mathbf{A} & \mathbf{c} \\ \mathbf{c}' & 0 \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} -\mathbf{b} \\ 1 \end{bmatrix}.$$

Inserting **A, b,** and **c** produces the solution $\begin{bmatrix} 50 & 14 & 1 \\ 14 & 26 & 1 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \\ 1 \end{bmatrix}$. The solution to the three equations is obtained by premultiplying the vector on the right by the inverse of the matrix on the left. The solutions are 0.27083, 0.72917, and, -25.75. The function value at the constrained solution is 4.240, which is larger than the unconstrained value of -10.00787.

19. What is the Jacobian for the following transformations?

$$y_1 = x_1/x_2,$$

$$\ln y_2 = \ln x_1 - \ln x_2 + \ln x_3,$$

and

$$y_3 = x_1x_2x_3.$$

Let capital letters denote logarithms. Then, the three transformations can be written as

$$Y_1 = X_1 - X_2$$

$$Y_2 = X_1 - X_2 + X_3$$

$$Y_3 = X_1 + X_2 + X_3.$$

This linear transformation is $\mathbf{Y} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\mathbf{X} = \mathbf{J}\mathbf{X}$. The inverse transformation is

$\mathbf{X} = \begin{bmatrix} 1 & -1/2 & 1/2 \\ 0 & -1/2 & 1/2 \\ 1 & 1 & 0 \end{bmatrix}\mathbf{Y} = \mathbf{J}^{-1}\mathbf{Y}$. In terms of the original variables, then, $x_1 = y_1(y_2/y_3)^{1/2}$, $x_2 = (y_3/y_2)^{1/2}$, and
$x_3 = y_1y_2$. The matrix of partial derivatives can be obtained directly, but an algebraic shortcut will prove useful for obtaining the Jacobian. Note first that $\partial x_i/\partial y_j = (x_i/y_j)(\partial\log x_i/\partial\log y_j)$. Therefore, the elements of the partial derivatives of the inverse transformations are obtained by multiplying the ith row by x_i, where we will substitute the expression for x_i in terms of the ys, then multiplying the jth column by $(1/y_j)$. Thus, the result of Exercise 11 will be useful here. The matrix of partial derivatives will be

$$\begin{bmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 & \partial x_1/\partial y_3 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 & \partial x_2/\partial y_3 \\ \partial x_3/\partial y_1 & \partial x_3/\partial y_2 & \partial x_3/\partial y_3 \end{bmatrix} = \begin{bmatrix} x_1 & 0 & 0 \\ 0 & x_2 & 0 \\ 0 & 0 & x_3 \end{bmatrix} \begin{bmatrix} 1 & -1/2 & 1/2 \\ 0 & -1/2 & 1/2 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/y_1 & 0 & 0 \\ 0 & 1/y_2 & 0 \\ 0 & 0 & 1/y_3 \end{bmatrix}.$$

The determinant of the product matrix is the product of the three determinants. The determinant of the center matrix is -1/2. The determinants of the diagonal matrices are the products of the diagonal elements. Therefore, the Jacobian is $J = \text{abs}(|\partial\mathbf{x}/\partial\mathbf{y}'|) = ½(x_1x_2x_3)/(y_1y_2y_3) = 2(y_1/y_2)$ (after making the substitutions for x_i).

20. Prove that exchanging two columns of a square matrix reverses the sign of its determinant. (**Hint:** use a permutation matrix. See Exercise 6.)

Exchanging the first two columns of a matrix is equivalent to postmultiplying it by a permutation matrix $\mathbf{B} = [\mathbf{e}_2,\mathbf{e}_1,\mathbf{e}_3,\mathbf{e}_4,...]$ where $\mathbf{e}_i$ is the ith column of an identity matrix. Thus, the determinant of the matrix is $|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}|$. The question turns on the determinant of **B**. Assume that **A** and **B** have n columns. To obtain the determinant of **B**, merely expand it along the first row. The only nonzero term in the determinant is $(-1)|\mathbf{I}_{n-1}| = -1$, where $\mathbf{I}_{n-1}$ is the $(n-1)\times(n-1)$ identity matrix. This completes the proof.

21. Suppose $\mathbf{x}=\mathbf{x}(z)$ where z is a scalar. What is $\partial[(\mathbf{x}'\mathbf{A}\mathbf{x})/(\mathbf{x}'\mathbf{B}\mathbf{x})]/z$?

The required derivatives are given in Exercise 16. Let $\mathbf{g} = \partial\mathbf{x}/\partial z$ and let the numerator and denominator be a and b, respectively. Then,

$$\partial(a/b)/\partial z = [b(\partial a/\partial z) - a(\partial b/\partial z)]/b^2$$
$$= [\mathbf{x}'\mathbf{B}\mathbf{x}(2\mathbf{x}'\mathbf{A}\mathbf{g}) - \mathbf{x}'\mathbf{A}\mathbf{x}(2\mathbf{x}'\mathbf{B}\mathbf{g})] / (\mathbf{x}'\mathbf{B}\mathbf{x})^2 = 2[\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{B}\mathbf{x}][\mathbf{x}'\mathbf{A}\mathbf{g}/\mathbf{x}'\mathbf{A}\mathbf{x} - \mathbf{x}'\mathbf{B}\mathbf{g}/\mathbf{x}'\mathbf{B}\mathbf{x}].$$

22. Suppose **y** is an $n\times 1$ vector and **X** is an $n\times K$ matrix. The projection of **y** into the column space of **X** is defined in the text after equation (2-55), $\hat{\mathbf{y}} = \mathbf{Xb}$. Now, consider the projection of $\mathbf{y}^* = c\mathbf{y}$ into the column space of $\mathbf{X}^* = \mathbf{XP}$ where c is a scalar and **P** is a nonsingular $K\times K$ matrix. Find the projection of $\mathbf{y}^*$ into the column space of $\mathbf{X}^*$. Prove that the cosine of the angle between $\mathbf{y}^*$ and its projection into the column space of $\mathbf{X}^*$ is the same as that between **y** and its projection into the column space of **X**. How do you interpret this result?

The projection of $\mathbf{y}^*$ into the column space of $\mathbf{X}^*$ is $\mathbf{X}^*\mathbf{b}^*$ where $\mathbf{b}^*$ is the solution to the set of equations $\mathbf{X}^{*\prime}\mathbf{y}^* = \mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{b}^*$ or $\mathbf{P}'\mathbf{X}'(c\mathbf{y}) = \mathbf{P}'\mathbf{X}'\mathbf{XPb}^*$. Since **P** is nonsingular, $\mathbf{P}'$ has an inverse. Premultiplying the equation by $(\mathbf{P}')^{-1}$, we have $c\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}(\mathbf{Pb}^*)$ or $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}[(1/c)\mathbf{Pb}^*]$. Therefore, in terms of the original **y** and **X**, we see that $\mathbf{b} = (1/c)\mathbf{Pb}^*$ which implies $\mathbf{b}^* = c\mathbf{P}^{-1}\mathbf{b}$. The projection is $\mathbf{X}^*\mathbf{b}^* = (\mathbf{XP})(c\mathbf{P}^{-1}\mathbf{b}) = c\mathbf{Xb}$. We conclude, therefore, that the projection of $\mathbf{y}^*$ into the column space of $\mathbf{X}^*$ is a multiple c of the projection of **y** into the space of **X**. This makes some sense, since, if **P** is a nonsingular matrix, the column space of $\mathbf{X}^*$ is exactly the same as the same as that of **X**. The cosine of the angle between $\mathbf{y}^*$ and its projection is that between $c\mathbf{y}$ and $c\mathbf{Xb}$. Of course, this is the same as that between **y** and **Xb** since the length of the two vectors is unrelated to the cosine of the angle between them. Thus,
$\cos\theta = (c\mathbf{y})'(c\mathbf{X}\text{b}))/(||c\mathbf{y}||\times||c\mathbf{Xb}||) = (\mathbf{y}'\mathbf{Xb}))/(||\mathbf{y}||\times||\mathbf{Xb}||)$.

23. For the matrix $\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & -2 & 3 & -5 \end{bmatrix}$, compute $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{M} = (\mathbf{I} - \mathbf{P})$. Verify that $\mathbf{MP} = \mathbf{0}$.

Let $\mathbf{Q} = \begin{bmatrix} 1 & 3 \\ 2 & 8 \end{bmatrix}$ (**Hint:** Show that **M** and **P** are idempotent.)

(a) Compute the **P** and **M** based on **XQ** instead of **X**.
(b) What are the characteristic roots of **M** and **P**?

First, $\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 0 \\ 0 & 54 \end{bmatrix}$, $(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/54 \end{bmatrix}$,

$$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{bmatrix} 1 & 4 \\ 1 & -2 \\ 1 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 1/4 & 0 \\ 0 & 1/54 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & -2 & 3 & -5 \end{bmatrix} = \frac{1}{108}\begin{bmatrix} 59 & 11 & 51 & -13 \\ 11 & 35 & 15 & 47 \\ 51 & 15 & 45 & -3 \\ -13 & 47 & -3 & 77 \end{bmatrix} = \mathbf{P}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{P} = \frac{1}{108}\begin{bmatrix} 49 & -11 & -51 & 13 \\ -11 & 73 & -15 & -47 \\ -51 & -15 & 63 & 3 \\ 13 & -47 & 3 & 31 \end{bmatrix}$$

(a) There is no need to recompute the matrices **M** and **P** for **XQ**, they are the same. Proof: The counterpart to **P** is $(\mathbf{XQ})[(\mathbf{XQ})'(\mathbf{XQ})]^{-1}(\mathbf{XQ})' = \mathbf{XQ}[\mathbf{Q}'\mathbf{X}'\mathbf{XQ}]^{-1}\mathbf{Q}'\mathbf{X}' =$
$\mathbf{XQQ}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{Q}')^{-1}\mathbf{Q}'\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. The **M** matrix would be the same as well. This is an application of the result found in the previous exercise. The **P** matrix is the projection matrix, and, as we found, the projection into the space of **X** is the same as the projection into the space of **XQ**.

(b) Since **M** and **P** are idempotent, their characteristic roots must all be either 0 or 1. The trace of the matrix equals the sum of the roots, which tells how many are 1 and 0. For the matrices above, the traces of both **M** and **P** are 2, so each has 2 unit roots and 2 zero roots.

24. Suppose that **A** is an $n\times n$ matrix of the form $\mathbf{A} = (1\text{-}\rho\mathbf{I}) + \rho\mathbf{ii}'$, where **i** is a column of 1s and $0 < \rho < 1$. Write out the format of **A** explicitly for $n = 4$. Find all of the characteristic roots and vectors of **A**. (**Hint:** There are only two distinct characteristic roots, which occur with multiplicity 1 and n-1. Every **c** of a certain type is a characteristic vector of **A**.) For an application which uses a matrix of this type, see Section 14.5 on the random effects model.

For $n = 4$, $\mathbf{A} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$. There are several ways to analyze this matrix. Here is a simple shortcut. The characteristic roots and vectors satisfy $[(1\text{-}\rho)\mathbf{I} + \rho\mathbf{ii}']\mathbf{c} = \lambda\mathbf{c}$. Multiply this out to obtain $(1\text{-}\rho)\mathbf{c} + \rho\mathbf{ii}'\mathbf{c} = \lambda\mathbf{c}$ or $\rho\mathbf{ii}'\mathbf{c} = [\lambda\text{- }(1\text{-}\rho)]\mathbf{c}$. Let $\mu = \lambda - (1\text{-}\rho)$, so $\rho\mathbf{ii}'\mathbf{c} = \mu\mathbf{c}$. We need only find the characteristic roots of $\rho\mathbf{ii}'$,μ. The characteristic roots of the original matrix are just $\lambda = \mu + (1\text{-}\rho)$. Now, $\rho\mathbf{ii}'$ is a matrix with rank one, since every column is identical. Therefore, n-1 of the μs are zero. Thus, the original matrix has n-1 roots equal to $0 + (1\text{-}\rho) = (1 \text{-}\rho)$. We can find the remaining root by noting that the sum of the roots of $\rho\mathbf{ii}'$ equals the trace of $\rho\mathbf{ii}'$. Since $\rho\mathbf{ii}'$ has only one nonzero root, that root is the trace, which is $n\rho$. Thus, the remaining root of the original matrix is $(1 - \rho + n\rho)$. The characteristic vectors satisfy the equation $\rho\mathbf{ii}'\mathbf{c} = \mu\mathbf{c}$. For the nonzero root, we have $\rho\mathbf{ii}'\mathbf{c} = n\rho\mathbf{c}$. Divide by $n\rho$ to obtain $\mathbf{i}(1/n)\mathbf{i}'\mathbf{c} = \mathbf{c}$. This equation states that for each element in the vector, $c_i = (1/n)\sum_i c_i$. This implies that every element in the characteristic vector corresponding to the root $(1\text{-}\rho+n\rho)$ is the same, or **c** is a multiple of a column of ones. In particular, so that it will have unit length, the vector is $(1/\sqrt{n})\,\mathbf{i}$. For the remaining zero roots, the characteristic vectors must satisfy $\rho\mathbf{i}(\mathbf{i}'\mathbf{c}) = 0\mathbf{c} = \mathbf{0}$. If the characteristic vector is not to be a column of zeroes, the only way to make this an equality is to require $\mathbf{i}'\mathbf{c}$ to be zero. Therefore, for the remaining n-1 characteristic vectors, we may use *any* set of orthogonal vectors whose elements sum to zero and whose inner products are one. There are an infinite number of such vectors. For example, let **D** be any arbitrary set of n-1 vectors containing n elements. Transform all columns of **D** into deviations from their own column means. Thus, we let $\mathbf{F} = \mathbf{M}^0\mathbf{D}$ where $\mathbf{M}^0$ is defined in Section 2.3.6. Now, let $\mathbf{C} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-2}$. **C** is a linear combination of the columns of **F**, so its columns sum to zero. By multiplying it out and using the results of Section 2.7.10, you will find that $\mathbf{C}'\mathbf{C} = \mathbf{I}$, so the columns are orthogonal and have unit length.

25. Find the inverse of the matrix in Exercise 24. [Hint: Use (A-66).]

Using the hint, the inverse is

$$[(1-\rho)\mathbf{I}]^{-1} \ - \ \frac{[(1-\rho)\mathbf{I}]^{-1}[\rho\mathbf{ii}'][(1-\rho)\mathbf{I}]^{-1}}{1+\left(\sqrt{\rho}\mathbf{i}\right)'[(1-\rho)\mathbf{I}]^{-1}\left(\sqrt{\rho}\mathbf{i}\right)} \ = \ \frac{1}{1-\rho}\{\mathbf{I} \ - \ [\rho/(1-\rho+n\rho)]\mathbf{ii}'\}$$

26. Prove that every matrix in the sequence of matrices $\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{d}_i\mathbf{d}_i'$, where $\mathbf{H}_0 = \mathbf{I}$, is positive definite. For an extension, prove that every matrix in the sequence of matrices in (E-22) is positive definite if $\mathbf{H}_0 = \mathbf{I}$.

By repeated substitution, we find $\mathbf{H}_{i+1} = \mathbf{I} + \sum_{j=1}^{i}\mathbf{d}_j\mathbf{d}_j'$. A quadratic form in $\mathbf{H}_{i+1}$ is, therefore

$$\mathbf{x}'\mathbf{H}_{i+1}\mathbf{x} \ = \ \mathbf{x}'\mathbf{x} \ + \ \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)(\mathbf{d}_j'\mathbf{x}) \ = \mathbf{x}'\mathbf{x} \ + \ \sum_{j=1}^{i}(\mathbf{x}'\mathbf{d}_j)^2$$

This is obviously positive for all $\mathbf{x}$. A simple way to establish this for the matrix in (E-22) is to note that in spite of its complexity, it is of the form $\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{d}_i\mathbf{d}_i' + \mathbf{f}_i\mathbf{f}_i'$. If this starts with a positive definite matrix, such as $\mathbf{I}$, then the identical argument establishes its positive definiteness.

27. What is the inverse matrix of $\mathbf{P} = \begin{bmatrix}\cos(x) & \sin(x)\\ -\sin(x) & \cos(x)\end{bmatrix}$? What are the characteristic roots of $\mathbf{P}$?

The determinant of $\mathbf{P}$ is $\cos^2(x) + \sin^2(x) = 1$, so the inverse just reverses the signs of the two off diagonal elements. The two roots are the solutions to $|\mathbf{P}-\lambda\mathbf{I}| = 0$, which is $\cos^2(x) + \sin^2(x) - 2\lambda\cos(x) + \lambda^2 = 0$. This simplifies because $\cos^2(x) + \sin^2(x) = 1$. Using the quadratic formula, then, $\lambda = \cos(x) \pm (\cos^2(x) - 1)^{1/2}$. But, $\cos^2(x) - 1 = -\sin^2(x)$. Therefore, the imaginary solutions to the resulting quadratic are $\lambda_1,\lambda_2 = \cos(x) \pm i\sin(x)$.

28. Derive the off diagonal block of $\mathbf{A}^{-1}$ in Section B.6.4.

For the simple 2×2 case, $\mathbf{F}_2$ is derived explicitly in the text, as $\mathbf{F}_2 = (\mathbf{x}'\mathbf{M}^0\mathbf{x})^{-1} = 1/\sum_i(x_i-\bar{x})^2$. Using (2-74), the off diagonal element is just $\mathbf{F}_2(\sum_i x_i)/n = \bar{x}/\sum_i(x_i-\bar{x})^2$. To extend this to a matrix containing a constant and K-1 variables, use the result at the end of the section. The off diagonal vector in $\mathbf{A}^{-1}$ when there is a constant and K-1 other variables is $-\mathbf{F}_2\mathbf{A}_{21}(\mathbf{A}_{11})^{-1} = [\mathbf{X}'\mathbf{M}^0\mathbf{X}]^{-1}\bar{\mathbf{x}}$. In all cases, $\mathbf{A}_{11}$ is just n, so $(\mathbf{A}_{11})^{-1}$ is $1/n$.

29. (This requires a computer.) For the $\mathbf{X}'\mathbf{X}$ matrix at the end of Section 2.4.1,

(a) Compute the characteristic roots of $\mathbf{X}'\mathbf{X}$.

(b) Compute the condition number of $\mathbf{X}'\mathbf{X}$. (Do not forget to scale the columns of the matrix so that the diagonal elements are 1.)

$$\text{The matrix is} \quad \begin{bmatrix} 15.000 & 120.00 & 19.310 & 111.79 & 99.770\\ 120.00 & 1240.0 & 164.30 & 1035.9 & 875.60\\ 19.310 & 164.30 & 25.218 & 148.98 & 131.22\\ 111.79 & 1035.9 & 148.98 & 943.86 & 799.02\\ 99.770 & 875.60 & 131.22 & 799.02 & 716.67 \end{bmatrix}$$

Its characteristic roots are 2486, 72.96, 19.55, 2.027, and .007354. To compute the condition number, we first extract $\mathbf{D} = \text{diag}(15,1240,25.218,943.86,716.67)$. To scale the matrix, we compute $\mathbf{V} = \mathbf{D}^{-2}\mathbf{X}'\mathbf{X}\mathbf{D}^{-2}$.

$$\text{The resulting matrix is} \begin{bmatrix} 1 & .8798823 & .992845 & .939515 & .962265\\ .879883 & 1 & .929119 & .957532 & .928828\\ .992845 & .929119 & 1 & .965648 & .976079\\ .939515 & .957532 & .965648 & 1 & .971503\\ .962265 & .928828 & .976079 & .971503 & 1 \end{bmatrix}.$$

The characteristic roots of this matrix are 4.801, .1389, .03716, .02183, and .0003527. The square root of the largest divided by the smallest is 116.675. These data are highly collinear by this measure.

Appendix B

Probability and Distribution Theory

1. How many different 5 card poker hands can be dealt from a deck of 52 cards?

There are $\binom{52}{5}$ = (52×51×51...×1)/[(5×4×3×2×1)(47×46×...×1)] = 2,598,960 possible hands. □

2. Compute the probability of being dealt 4 of a kind in a poker hand.

There are 48(13) possible hands containing 4 of a kind and any of the remaining 48 cards. Thus, given the answer to the previous problem, the probability of being dealt one of these hands is 48(13)/2598960 =.00024, or less than one chance in 4000.

3. Suppose a lottery ticket costs $1 per play. The game is played by drawing 6 numbers without replacement from the numbers 1 to 48. If you guess all six numbers, you win the prize. Now, suppose that N = the number of tickets sold and P = the size of the prize. N and P are related by

$$N = 5 + 1.2P$$
$$P = 1 + .4N$$

N and P are in millions. What is the expected value of a ticket in this game? (Don't forget that you might have to share the prize with other winners.)

The size of the prize and number of tickets sold are jointly determined. The solutions to the two equations are N = 11.92 million tickets and P = $5.77 million. The number of possible combinations of 48 numbers without replacement is $\binom{48}{6}$ = (48×47×46...×1)/[(6×5×4×3×2×1)(42×41×...×1)] = 12,271,512 so the probability of making the right choice is 1/12271512 = .000000081. The expected number of winners is the expected value of a binomial random variable with N trials and this success probability, which is N times the probability, or 11.92/12.27 = .97, or roughly 1. Thus, one would not expect to have to share the prize. Now, the expected value of a ticket is Prob[win](5.77 million - 1) + Prob[lose](-1) . -53 cents.

4. If x has a normal distribution with mean 1 and standard deviation 3, what are

(a) Prob[$|x| > 2$].
(b) Prob[$x > -1 \mid x < 1.5$].

Using the normal table,

(a) Prob[$|x| > 2$] = 1 - Prob[$|x| \le 2$]
= 1 - Prob[$-2 \le x \le 2$]
= 1 - Prob[$(-2-1)/3 \le z \le (2-1)/3$]
= 1 - [F(1/3) - F(-1)] = 1 - .6306 + .1587 = .5281

(b) Prob[$x > -1 \mid x < 1.5$] = Prob[$-1 < x < 1.5$] / Prob[$x < 1.5$]
Prob[$-1 < x < 1.5$] = Prob[$(-1-1)/3 < z < (1.5-1)/3)$]
= Prob[$z < 1/6$] - Prob[$z < -2/3$]
= .5662 - .2525 = .3137.

The conditional probability is .3137/.5662 = .5540.

5. Approximately what is the probability that a random variable with chi-squared distribution with 264 degrees of freedom is less than 297?

We use the approximation in (3-37), $z = [2(297)]^2 - [2(264) - 1]^2 = 1.4155$, so the probability is approximately .9215. To six digits, the approximation is .921539 while the correct value is .921559.

6. **Chebychev Inequality** For the following two probability distributions, find the lower limit of the probability of the indicated event using the Chebychev inequality and the exact probability using the appropriate table:

(a) $x \sim$ Normal$[0,3^2]$, and $-4 < x < 4$.

(b) $x \sim$ chi-squared, 8 degrees of freedom, $0 < x < 16$.

The inequality given in (3-18) states that Prob$[|x - \mu| \leq k\sigma] \geq 1 - 1/k^2$. Note that the result is not informative if k is less than or equal to 1.

(a) The range is 4/3 standard deviations, so the lower limit is $1 - (3/4)^2$ or 7/16 = .4375. From the standard normal table, the actual probability is 1 - 2Prob[$z < -4/3$] = .8175.

(b) The mean of the distribution is 8 and the standard deviation is 4. The range is, therefore, $\mu \pm 2\sigma$. The lower limit according to the inequality is $1 - (1/2)^2 = .75$. The actual probability is the cumulative chi-squared(8) at 16, which is a bit larger than .95. (The actual value is .9576.)

7. Given the following joint probability distribution,

		X		
		0	1	2
Y	0	.05	.1	.03
	1	.21	.11	.19
	2	.08	.15	.08

(a) Compute the following probabilities: Prob[$Y < 2$], Prob[$Y < 2, X > 0$], Prob[$Y = 1, X \geq 1$].

(b) Find the marginal distributions of X and Y.

(c) Calculate $E[X]$, $E[Y]$, Var[X], Var[Y], Cov[X,Y], and $E[X^2Y^3]$.

(d) Calculate Cov[Y,X^2].

(e) What are the conditional distributions of Y given $X = 2$ and of X given $Y > 0$?

(f) Find $E[Y|X]$ and Var[$Y|X$]. Obtain the two parts of the variance decomposition
Var[Y] = E_x[Var[$Y|X$]] + Var$_x$[$E[Y|X]$].

We first obtain the marginal probabilities. For the joint distribution, these will be

X: P(0) = .34, P(1) = .36, P(2) = .30

Y: P(0) = .18, P(1) = .51, P(2) = .31

Then,

(a) Prob[$Y < 2$] = .18 + .51 = .69.
Prob[$Y < 2, X > 0$] = .1 + .03 + .11 + .19 = .43.
Prob[Y = 1, X $ 1] = .11 + .19 = .30.

(b) They are shown above.

(c) $E[X]$ = 0(.34) + 1(.36) + 2(.30) = .96
$E[Y]$ = 0(.18) + 1(.51) + 2(.31) = 1.13
$E[X^2]$ = $0^2(.34) + 1^2(.36) + 2^2(.30)$ = 1.56
$E[Y^2]$ = $0^2(.18) + 1^2(.51) + 2^2(.31)$ = 1.75
Var[X] = $1.56 - .96^2$ = .6384
Var[Y] = $1.75 - 1.13^2$ = .4731
$E[XY]$ = 1(1)(.11)+1(2)(.15)+2(1)(.19)+2(2)(.08) = 1.11
Cov[X,Y] = 1.11 - .96(1.13) = .0252
$E[X^2Y^3]$ = .11 + 8(.15) + 4(.19) + 32(.08) = 4.63.

(d) E[YX^2] = 1(12).11+1(22).19+2(12).15+2(22).08 = 1.81
Cov[Y,X^2] = 1.81 - 1.13(1.56) = .0472.

(e) Prob[Y = 0 * X = 2] = .03/.3 = .1
Prob[Y = 1 * X = 2] = .19/.3 = .633
Prob[Y = 1 * X = 2] = .08/.3 = .267
Prob[X = 0 * Y > 0] = (.21 + .08)/(.51 + .31) = .3537
Prob[X = 1 * Y > 0] = (.11 + .15)/(.51 + .31) = .3171
Prob[X = 2 * Y > 0] = (.19 + .08)/(.51 + .31) = .3292.

(f) $E[Y * X$=0] = 0(.05/.34)+1(.21/.34)+2(.08/.34) = 1.088
$E[Y^2 * X$=0] = $1^2(.21/.34)+2^2(.08/.34)$ = 1.559
Var[$Y * X$=0] = $1.559 - 1.088^2$ = .3751
$E[Y*X$=1] = 0(.1/.36)+1(.11/.36)+2(.15/.36) = 1.139
$E[Y^2*X$=1] = $1^2(.11/.36)+2^2(.15/.36)$ = 1.972
Var[$Y*X$=1] = $1.972 - 1.139^2$ = .6749
$E[Y*X$=2] = 0(.03/.30)+1(.19/.30)+2(.08/.30) = 1.167

$E[Y^2*X=2]$ = $1^2(.19/.30)+2^2(.08/.30) = 1.700$
Var$[Y*X=2]$ = $1.700 - 1.167^2 = .6749 = .3381$
$E[\text{Var}[Y*X]]$ = $.34(.3751)+.36(.6749)+.30(.3381) = .4719$
Var$[E[Y*X]]$ = $.34(1.088^2)+.36(1.139^2)+.30(1.167^2) - 1.13^2 = 1.2781 - 1.2769 = .0012$
$E[\text{Var}[Y*X]] + \text{Var}[E[Y*X]] = .4719 + .0012 = .4731 = \text{Var}[Y]$. ~

8. **Minimum mean squared error predictor**. For the joint distribution in Exercise 7, compute $E[y - E[y|x]]^2$. Now, find the a and b which minimize the function $E[y - a - bx]^2$. Given the solutions, verify that $E[y - E[y|x]]^2 \leq E[y - a - bx]^2$. The result is fundamental in least squares theory. Verify that the a and b which you found satisfy (3-68) and (3-69).

$$
\begin{aligned}
& \qquad\qquad\qquad\quad (x=0) \qquad\qquad (x=1) \qquad\qquad (x=2)\\
E[y - E[y|x]]^2 = \;& (y=0) \quad .05(0 - 1.088)^2 + .10(0 - 1.139)^2 + .03(0 - 1.167)^2 \\
& (y=1) \quad + .21(1 - 1.088)^2 + .11(1 - 1.139)^2 + .19(1 - 1.167)^2 \\
& (y=2) \quad + .08(2 - 1.088)^2 + .15(2 - 1.139)^2 + .08(2 - 1.167)^2 \\
= \;& .4719 = E[\text{Var}[y|x]].
\end{aligned}
$$

The necessary conditions for minimizing the function with respect to a and b are

$$\partial E[y - a - bx]^2/\partial a = 2E\{[y - a - bx](-1)\} = 0$$
$$\partial E[y - a - bx]^2/\partial b = 2E\{[y - a - bx](-x)\} = 0.$$

First dividing by -2, then taking expectations produces

$$E[y] - a - bE[x] = 0$$
$$E[xy] - aE[x] - bE[x^2] = 0.$$

Solve the first for $a = E[y] - bE[x]$ and substitute this in the second to obtain

$$E[xy] - E[x](E[y] - bE[x]) - bE[x^2] = 0$$

or $\quad (E[xy] - E[x]E[y]) = b(E[x^2] - (E[x])^2)$

or $\quad b = \text{Cov}[x,y] / \text{Var}[x] = -.0708 / .4731 = -.150$

and $\quad a = E[y] - bE[x] = 1.13 - (-.1497)(.96) = 1.274.$

The linear function compared to the conditional mean produces

	x=0	x=1	x=2
$E[y\|x]$	1.088	1.139	1.167
$a + bx$	1.274	1.124	.974

Now, repeating the calculation above using $a + bx$ instead of E$[y|x]$ produces

$$
\begin{aligned}
& \qquad\qquad\qquad\quad (x=0) \qquad\qquad (x=1) \qquad\qquad (x=2)\\
E[y - a - bx]^2 = \;& (y=0) \quad .05(0 - 1.274)^2 + .10(0 - 1.124)^2 + .03(0 - .974)^2 \\
& (y=1) \quad + .21(1 - 1.274)^2 + .11(1 - 1.124)^2 + .19(1 - .974)^2 \\
& (y=2) \quad + .08(2 - 1.274)^2 + .15(2 - 1.124)^2 + .08(2 - .974)^2 \\
= \;& .4950 > .4719.
\end{aligned}
$$

9. Suppose x has an exponential distribution, $f(x) = \theta e^{-\theta x}$, $x \geq 0$. Find the mean, variance, skewness, and kurtosis of x. The Gamma integral will be useful for finding the raw moments.)

In order to find the central moments, we will use the raw moments, $E[x^r] = \int_0^\infty \theta x^r e^{-\theta x} dx$. These can be obtained by using the gamma integral. Making the appropriate substitutions, we have

$$E[x^r] = [\theta\Gamma(r+1)]/\theta^{r+1} = r!/\theta^r.$$

The first four moments are: $E[x] = 1/\theta$, $E[x^2] = 2/\theta^2$, $E[x^3] = 6/\theta^3$, and $E[x^4] = 24/\theta^4$. The mean is, thus, $1/\theta$ and the variance is $2/\theta^2 - (1/\theta)^2 = 1/\theta^2$. For the skewness and kurtosis coefficients, we have

$$E[x - 1/\theta]^3 = E[x^3] - 3E[x^2]/\theta + 3E[x]/\theta^2 - 1/\theta^3 = 2/\theta^3.$$

The normalized skewness coefficient is 2. The kurtosis coefficient is

$$E[x - 1/\theta]^4 = E[x^4] - 4E[x^3]/\theta + 6E[x^2]/\theta^2 - 4E[x]/\theta^3 + 1/\theta^4 = 9/\theta^4.$$

The degree of excess is 6.

10. For the random variable in Exercise 9, what is the probability distribution of the random variable $y = e^{-x}$? What is $E[y]$? Prove that the distribution of this y is a special case of the beta distribution in (3-40).

If $y = e^{-x}$, then $x = -\ln y$, so the Jacobian is $|dx/dy| = 1/y$. The distribution of y is, therefore,

$$f(y) = \theta e^{-\theta(-\ln y)}(1/y) = (\theta y^\theta)/y = \theta y^{\theta-1} \text{ for } 0 < y < 1.$$

This is in the form of (3-40) with y instead of x, $c = 1$, $\beta = 1$, and $\alpha = \theta$.

11. If the probability density of y is $\alpha y^2(1-y)^3$ for y between 0 and 1, what is α? What is the probability that y is between .25 and .75?

This is a beta distribution of the form in (3-40) with $\alpha = 3$ and $\beta = 4$. Therefore, the constant is $\Gamma(3+4)/(\Gamma(3)\Gamma(4)) = 60$. The probability is

$$\int_{.25}^{.75} 60y^2(1-y)^3 dy = 60\int_{.25}^{.75} (y^2 - 3y^3 + 3y^4 - y^5)dy = 60(y^3/3 - 3y^4/4 + 3y^5/5 - y^6/6)\Big|_{.25}^{.75} = .79296.$$

12. Suppose x has the following discrete probability distribution:

X	1	2	3	4
Prob[$X = x$]	.1	.2	.4	.3.

Find the exact mean and variance of X. Now, suppose $Y = 1/X$. Find the exact mean and variance of Y. Find the mean and variance of the linear and quadratic approximations to $Y = f(X)$. Are the mean and variance of the quadratic approximation closer to the true mean than those of the linear approximation?

We will require a number of moments of x, which we derive first:

$$E[x] = .1(1) + .2(2) + .4(3) + .3(4) = 2.9 = \mu$$
$$E[x^2] = .1(1) + .2(4) + .4(9) + .3(16) = 9.3$$
$$\text{Var}[x] = 9.3 - 2.9^2 = .89 = \sigma^2.$$

For later use, we also obtain

$$E[x - \mu]^3 = .1(1 - 2.9)^3 + ... = -.432$$
$$E[x - \mu]^4 = .1(1 - 2.9)^4 + ... = 1.8737.$$

The approximation is $y = 1/x$. The exact mean and variance are

$$E[y] = .1(1) + .2(1/2) + .4(1/3) + .3(1/4) = .40833$$
$$\text{Var}[y] = .1(12) + .2(1/4) + .4(1/9) + .3(1/16) - .40833^2 = .04645.$$

The linear Taylor series approximation around μ is $y \approx 1/\mu + (-1/\mu^2)(x - \mu)$. The mean of the linear approximation is $1/\mu = .3448$ while its variance is $(1/\mu^4)\text{Var}[x-\mu] = \sigma^2/\mu^4 = .01258$. The quadratic approximation is

$$y \approx 1/\mu + (-1/\mu^2)(x - \mu) + (1/2)(2/\mu^3)(x - \mu)^2$$
$$= 1/\mu - (1/\mu^2)(x - \mu) + (1/\mu^3)(x - \mu)^2.$$

The mean of this approximation is $E[y] \approx 1/\mu + \sigma^2/\mu^3 = .3813$ while the variance is approximated by the variance of the right hand side,

$$(1/\mu^4)\text{Var}[x - \mu] + (1/\mu^6)\text{Var}[x - \mu]^2 - (2/\mu^5)\text{Cov}[(x-\mu),(x-\mu)^2]$$
$$= (1/\mu^4)\sigma^2 + (1/\mu^6)(E[x - \mu]^4 - \sigma^4] - (2/\mu^5)E[x - \mu]^3$$
$$= .01498.$$

Neither approximation provides a close estimate of the variance. Note that in both cases, it would be possible simply to evaluate the approximations at the four values of x and compute the means and variances directly. The virtue of the approach above is that it can be applied when there are many values of x, and is necessary when the distribution of x is continuous.

13. **Interpolation in the chi-squared table.** In order to find a percentage point in the chi-squared table which is between two values, we interpolate linearly between the *reciprocals* of the degrees of freedom. The chi-squared distribution is defined for noninteger values of the degrees of freedom parameter [see (3-39)], but your table does not contain critical values for noninteger values. Using linear interpolation, find the 99% critical value for a chi-squared variable with degrees of freedom parameter 11.3.

The 99% critical values for 11 and 12 degrees of freedom are 24.725 and 26.217. To interpolate linearly between these values for the value corresponding to 11.3 degrees of freedom, we use

$$c = 26.217 + \frac{(111.3 - 1/12)}{(1/11 - 1/12)}(24.725 - 26.217) = 25.2009. \ \square$$

14. Suppose x has a standard normal distribution. What is the pdf of the following random variable?

$y = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, 0 < y < \frac{1}{\sqrt{2\pi}}$. [**Hints:** You know the distribution of $z = x^2$ from (C-30). The density of this z is given in (C-39). Solve the problem in terms of $y = g(z)$.]

We know that $z = x^2$ is distributed as chi-squared with 1 degree of freedom. We seek the density of $y = ke^{-z/2}$ where $k = (2\pi)^{-2}$. The inverse transformation is $z = 2\ln k - 2\ln y$, so the Jacobian is $|-2/y| = 2/y$. The density of z is that of Gamma with parameters 1/2 and 1/2. [See (C-39) and the succeeding discussion.] Thus,

$$f(z) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} e^{-z/2} z^{-1/2}, z > 0.$$

Note, $\Gamma(1/2) = \sqrt{\pi}$. Making the substitution for z and multiplying by the Jacobian produces

$$f(y) = \frac{(1/2)^{1/2}}{\Gamma(1/2)} \frac{2}{y} e^{(-1/2)(2\ln k - 2\ln y)} (2\ln k - 2\ln y)^{-1/2}$$

The exponential term reduces to y/k. The scale factor is equal to $2k/y$. Therefore, the density is simply
$f(y) = 2(2\ln k - 2\ln y)^{-1/2} = \sqrt{2}\ (\ln k - \ln y)^{-1/2} = \{2/[\ln(1/(y(2\pi)^{1/2}))]\}$, $0 < y < (2\pi)^{-1/2}$.

15. **The fundamental probability transformation**. Suppose that the continuous random variable x has cumulative distribution $F(x)$. What is the probability distribution of the random variable $y = F(x)$? (**Observation:** This result forms the basis of the simulation of draws from many continuous distributions.)

The inverse transformation is $x(y) = F^{-1}(y)$, so the Jacobian is $dx/dy = F^{-1\prime}(y) = 1/f(x(y))$ where $f(.)$ is the density of x. The density of y is $f(y) = f[F^{-1}(y)] \times 1/f(x(y)) = 1$, $0 \leq y \leq 1$. Thus, y has a continuous uniform distribution. Note, then, for purposes of obtaining a random sample from the distribution, we can sample $y_1,...,y_n$ from the distribution of y, the continuous uniform, then obtain $x_1 = x_1(y_1)$, ... $x_n = x_n(y_n)$.

16. **Random number generators**. Suppose x is distributed uniformly between 0 and 1, so $f(x) = 1$, $0 \leq x \leq 1$. Let θ be some positive constant. What is the pdf of $y = -(1/\theta)\ln x$. (**Hint**: See Section 3.5.) Does this suggest a means of simulating draws from this distribution if one has a random number generator which will produce draws from the uniform distribution? To continue, suggest a means of simulating draws from a logistic distribution, $f(x) = e^{-x}/(1+e^{-x})^2$.

The inverse transformation is $x = e^{-\theta y}$ so the Jacobian is $dx/dy = \theta e^{-\theta y}$. Since $f(x) = 1$, this Jacobian is also the density of y. One can simulate draws y from any exponential distribution with parameter θ by drawing observations x from the uniform distribution and computing $y = -(1/\theta)\ln x$. Likewise, for the logistic distribution, the CDF is $F(x) = 1/(1 + e^{-x})$. Thus, draws y from the uniform distribution may be taken as draws on $F(x)$. Then, we may obtain x as $x = \ln[F(x)/(1 - F(x)] = \ln[y/(1 - y)]$.

17. Suppose that x_1 and x_2 are distributed as independent standard normal. What is the joint distribution of $y_1 = 2 + 3x_1 + 2x_2$ and $y_2 = 4 + 5x_1$? Suppose you were able to obtain two samples of observations from independent standard normal distributions. How would you obtain a sample from the bivariate normal distribution with means 1 and 2 variances 4 and 9 and covariance 3?

We may write the pair of transformations as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{b} + \mathbf{Ax}.$$

The problem also states that $\mathbf{x} \sim N[\mathbf{0},\mathbf{I}]$. From (C-103), therefore, we have $\mathbf{y} \sim N[\mathbf{b} + \mathbf{A0}, \mathbf{AIAN}]$ where

$$E[\mathbf{y}] = \mathbf{b} + \mathbf{A0} = \mathbf{b} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \text{Var}[\mathbf{y}] = \mathbf{AA}' = \begin{bmatrix} 13 & 15 \\ 15 & 25 \end{bmatrix}.$$

For the second part of the problem, using our result above, we would require the $\mathbf{A}$ and $\mathbf{b}$ such that

$\mathbf{b} + \mathbf{A0} = (1,2)'$ and $\mathbf{AA}' = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$. The vector is obviously $\mathbf{b} = (1,2)'$. In order to find the elements of $\mathbf{A}$, there are a few ways to proceed. The Cholesky factorization used in Exercise 9 is probably the simplest. Let $y_1 = 1 + 2x_1$. Thus, y_1 has mean 1 and variance 4 as required. Now, let $y_2 = 2 + w_1x_1 + w_2x_2$. The covariance between y_1 and y_2 is $2w_1$, since x_1 and x_2 are uncorrelated. Thus, $2w_1 = 3$, or $w_1 = 1.5$. Now, $\text{Var}[y_2] = w_1^2 + w_2^2 = 9$, so $w_2^2 = 9 - 1.5^2 = 6.75$. The transformation matrix is, therefore, $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1.5 & 2.598 \end{bmatrix}$. This is the Cholesky factorization of the desired $\mathbf{AA}'$ above. It is worth noting, this provides a simple method of finding the requisite $\mathbf{A}$ matrix for any number of variables. Finally, an alternative method would be to use the characteristic roots and vectors of **AA′**. The inverse square root defined in Section B.7.12 would also provide a method of transforming **x** to obtain the desired covariance matrix.

18. The density of the standard normal distribution, denoted $\phi(x)$, is given in (C-28). The function based on the ith derivative of the density given by $H_i = [(-1)^i d^i\phi(x)/dx^i]/\phi(x)$, $i = 0,1,2,...$ is called a *Hermite polynomial*. By definition, $H_0 = 1$.

(a) Find the next three Hermite polynomials.

(b) A useful device in this context is the differential equation

$$d^r\phi(x)/dx^r + xd^{r-1}\phi(x)/dx^{r-1} + (r-1)d^{r-2}\phi(x)/dx^{r-2} = 0.$$

Use this result and the results of part a. to find H_4 and H_5.

The crucial result to be used in the derivations is $d\phi(x)/dx = -x\phi(x)$. Therefore,

$$d^2\phi(x)/dx^2 = (x^2 - 1)\phi(x)$$

and $$d^3\phi(x)/dx^3 = (3x - x^3)\phi(x).$$

The polynomials are $H_1 = x$, $H_2 = x^2 - 1$, and $H_3 = x^3 - 3x$.

For part (b), we solve for $d^r\phi(x)/dx^r = -xd^{r-1}\phi(x)/dx^{r-1} - (r-1)d^{r-2}\phi(x)/dx^{r-2}$

Therefore, $d^4\phi(x)/dx^4 = -x(3x - x^3)\phi(x) - 3(x^2 - 1)\phi(x) = (x^4 - 6x^2 + 3)\phi(x)$

and $d^5\phi(x)/dx^5 = (-x^5 + 10x^3 - 15x)\phi(x)$.

Thus, $H_4 = x^4 - 6x^2 + 3$ and $H_5 = x^5 - 10x^3 + 15x$. □

19. Continuation: *orthogonal polynomials*: The Hermite polynomials are orthogonal if x has a standard normal distribution. That is, $E[H_iH_j] = 0$ if $i \neq j$. Prove this for the H_1, H_2, and H_3 which you obtained above.

$$E[H_1(x)H_2(x)] = E[x(x^2 - 1)] = E[x^3 - x] = 0$$

since the normal distribution is symmetric. Then,

$$E[H_1(x)H_3(x)] = E[x(x^3 - 3x)] = E[x^4 - 3x^2] = 0.$$

The fourth moment of the standard normal distribution is 3 times the variance. Finally,

$$E[H_2(x)H_3(x)] = E[(x^2 - 1)(x^3 - 3x)] = E[x^5 - 4x^3 + 3x] = 0$$

because all odd order moments of the normal distribution are zero. (The general result for extending the preceding is that in a product of Hermite polynomials, if the sum of the subscripts is odd, the product will be a sum of odd powers of x, and if even, a sum of even powers. This provides a method of determining the higher moments of the normal distribution if they are needed. (For example, $E[H_1H_3] = 0$ implies that $E[x^4] = 3E[x^2]$.)

20. If x and y have means μ_x and μ_y and variances σ_x^2 and σ_y^2 and covariance σ_{xy}, what is the approximation of the covariance matrix of the two random variables $f_1 = x/y$ and $f_2 = xy$?

The elements of **JΣJ**N are $(1,1) = \frac{\sigma_x^2}{\mu_y^2} + \frac{\sigma_y^2\mu_2^x}{\mu_y^4} - \frac{2\sigma_{xy}\mu_x}{\mu_y^3}$

$$(1,2) = \sigma_x^2 - \sigma_y^2\,\mu_x^2 / \mu_y^4$$

$$(2,2) = \sigma_x^2\,\mu_y^4 + \sigma_y^2\,\mu_x^2 + 2\sigma_{xy}\mu_x\mu_y.$$

21. ***Factorial Moments.*** For finding the moments of a distribution such as the Poisson, a useful device is the factorial moment. (The Poisson distribution is given in Example 3.1.) The density is

$$f(x) = e^{-\lambda}\lambda^x / x!,\ x = 0,1,2,...$$

To find the mean, we can use

$$E[x] = \sum_{x=0}^{\infty} xf(x) = \sum_{x=0}^{\infty} xe^{-\lambda}\lambda^x / x!$$

$$= \sum_{x=1}^{\infty} e^{-\lambda}\lambda^{x-1} / (x-1)!$$

$$= \lambda\sum_{y=0}^{\infty} e^{-\lambda}\lambda^y / y!$$

$$= \lambda,$$

since the probabilities sum to 1. To find the variance, we will extend this method by finding $E[x(x-1)]$, and likewise for other moments. Use this method to find the variance and third central moment of the Poisson distribution. (Note that this device is used to transform the factorial in the denominator in the probability.)

Using the same technique,

$$E[x(x-1)] = \sum_{x=0}^{\infty} x(x-1)f(x) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\lambda^{x}/x!$$
$$= \sum_{x=2}^{\infty} e^{-\lambda}\lambda^{x-2}/(x-2)!$$
$$= \lambda^2 \sum_{y=0}^{\infty} e^{-\lambda}\lambda^{y}/y!$$
$$= \lambda^2$$
$$= E[x^2] - E[x]$$

So, $E[x^2] = \lambda^2 + \lambda$.

Since $E[x] = \lambda$, it follows that $\text{Var}[x] = (\lambda^2 + \lambda) - \lambda^2 = \lambda$. Following the same pattern, the preceding produces

$$E[x(x-1)(x-2)] = E[x^3] - 3E[x^2] + 2E[x].$$
$$= \lambda^3.$$

Therefore, $E[x^3] = \lambda^3 + 3(\lambda + \lambda^2) - 2\lambda$
$$= \lambda^3 + 3\lambda^2 + \lambda.$$

Then, $E[x - E[x]]^3 = E[x^3] - 3\lambda E[x^2] + 3\lambda^2 E[x] - \lambda^3$
$$= \lambda. \quad \square$$

22. If x has a normal distribution with mean μ and standard deviation σ, what is the probability distribution of $y = e^x$?

If $y = e^x$, then $x = \ln y$ and the Jacobian is $dx/dy = 1/y$. Making the substitution,

$$f(y) = \frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{1}{2}\left[(\ln y - \mu)/\sigma\right]^2}$$

This is the density of the lognormal distribution.

23. If y has a lognormal distribution, what is the probability distribution of y^2?

Let $z = y^2$. Then, $y = \sqrt{z}$ and $dy/dz = 1/(2\sqrt{z})$. Inserting these in the density above, we find

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\sqrt{z}} \frac{1}{2\sqrt{z}} e^{-\frac{1}{2}\left[\left(\frac{1}{2}\ln z - \mu\right)/\sigma\right]^2}, z > 0$$
$$= \frac{1}{(2\sigma)z\sqrt{2\pi}} e^{-\frac{1}{2}\left[(\ln z - 2\mu)/(2\sigma)\right]^2}, z > 0.$$

Thus, z has a lognormal distribution with parameters 2μ and 2σ. The general result is that if y has a lognormal distribution with parameters μ and σ, y^r has a lognormal distribution with parameters $r\mu$ and $r\sigma$.

24. Suppose y, x_1, and x_2 have a joint normal distribution with parameters $\boldsymbol{\mu}' = [1, 2, 4]$

and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 5 & 2 \\ 1 & 2 & 6 \end{bmatrix}$

(a) Compute the intercept and slope in the function $E[y|x_1]$, $\text{Var}[y|x_1]$, and the coefficient of determination in this regression. (**Hint**: See Section 3.10.1.)

(b) Compute the intercept and slopes in the conditional mean function, $E[y|x_1,x_2]$. What is $E[y|x_1=2.5,x_2=3.3]$? What is $\text{Var}[y|x_1=2.5,x_2=3.3]$?

First, for normally distributed variables, we have from (3-102),

$$E[y|\mathbf{x}] = \mu_y + \text{Cov}[y,\mathbf{x}]\{\text{Var}[\mathbf{x}]\}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$$

and $\text{Var}[y|\mathbf{x}] = \text{Var}[y] - \text{Cov}[y,\mathbf{x}]\{\text{Var}[\mathbf{x}]\}^{-1}\text{Cov}[\mathbf{x},y]$

and $COD = \text{Var}[E[y|\mathbf{x}]] / \text{Var}[y]$
$$= \text{Cov}[y,\mathbf{x}]\{\text{Var}[\mathbf{x}]\}^{-1}\text{Cov}[\mathbf{x},y] / \text{Var}[y].$$

We may just insert the figures above to obtain the results.

$$E[y|x_1] = 1 + (3/5)(x_1 - 2) = -.2 + .6x_1,$$
$$\text{Var}[y|x_1] = 2 - 3(1/5)3 = 1/5 = .2$$

$COD \quad = .6^2(5)/2 = .9$

$E[y*x_1,x_2] \quad = 1 + \begin{bmatrix}3 & 1\end{bmatrix}\begin{bmatrix}5 & 2\\ 2 & 6\end{bmatrix}^{-1}\begin{bmatrix}3\\ 1\end{bmatrix}$

$= -.4615 + .6154x_1 - .03846x_2,$

$Var[y*x_1,x_2] \quad = 2 - (.6154,-.03846)(3,1)N = .1923.$

$E[y*x_1=2.5,x_2=3.3] = 1.3017.$

The conditional variance is not a function of x_1 or x_2.

25. What is the density of $y = 1/x$ if x has a chi-squared distribution?

The density of a chi-squared variable is a gamma variable with parameters 1/2 and $n/2$ where n is the degrees of freedom of the chi-squared variable. Thus,

$$f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2}x} x^{\frac{n}{2}-1}, x > 0.$$

If $y = 1/x$ then $x = 1/y$ and $|dx/dy| = 1/y^2$. Therefore, after multiplying by the Jacobian,

$$f(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2y}} \left(\frac{1}{y}\right)^{\frac{n}{2}+1}, y > 0. \quad \square$$

26. What is the density and what are the mean and variance of $y = 1/x$ if x has the gamma distribution described in Section C.4.5.

The density of x is $f(x) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda x} x^{P-1}, x > 0$. If $y = 1/x$, then $x = 1/y$, and the Jacobian is $|dx/dy| = 1/y^2$. Using the change of variable formula, as usual, the density of y is $f(y) = \frac{\lambda^P}{\Gamma(P)} \frac{1}{y^2} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1}, y > 0$. The mean is $E(y) = \int_0^\infty y \frac{\lambda^P}{\Gamma(P)} \frac{1}{y^2} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1} dy$. This is a gamma integral (see Section 5.2.4b). Combine terms to obtain $E(y) = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P} dy$. Now, in order to use the results for the gamma integral, we will have to make a change of variable. Let $z = 1/y$, so $|dy/dz| = 1/z^2$. Making the change of variable, we find $E(y) = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^P \left(\frac{1}{z^2}\right) dz = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-2} dz$. Now, we can use the gamma integral directly, to find $E(y) = \frac{\lambda^P}{\Gamma(P)} \times \frac{\Gamma(P-1)}{\lambda^{P-1}} = \frac{\lambda}{P-1}$. Note that for this to exist, P must be greater than one. We can use the same approach to find the variance. We start by finding $E[y^2]$. First, $E(y^2) = \int_0^\infty y^2 \frac{\lambda^P}{\Gamma(P)} \frac{1}{y^2} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1} dy = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda/y} \left(\frac{1}{y}\right)^{P-1} dy$. Once again, this is a gamma integral, which we can evaluate by first making the change of variable to $z = 1/y$. The integral is $E(y^2) = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-1} \left(\frac{1}{z^2}\right) dz = \int_0^\infty \frac{\lambda^P}{\Gamma(P)} e^{-\lambda z} z^{P-3} dz$. This is $\frac{\lambda^P}{\Gamma(P)} \times \frac{\Gamma(P-2)}{\lambda^{P-2}} = \frac{\lambda^2}{(P-1)(P-2)}$.

Now, $Var[y] = E[y^2] - E^2[y] = \frac{\lambda^3}{(P-1)^2(P-2)}, P > 2.$

27. Suppose x_1 and x_2 have the bivariate normal distribution described in Section 3.8. Consider an extension of Example 3.4, where the bivariate normal distribution is obtained by transforming two independent standard normal variables. Obtain the distribution of $z = \exp(y_1)\exp(y_2)$ where y_1 and y_2 have a bivariate normal distribution and are correlated. Solve this problem in two ways. First, use the

transformation approach described in Section C.6.4. Second, note that $z = \exp(y_1+y_2) = \exp(w)$, so you can first find the distribution of w, then use the results of Section 3.5 (and, in fact, Section 3.4.4 as well).

The (extremely) hard way to proceed is to define the joint transformations $z_1 = \exp(y_1)\exp(y_2)$ and $z_2 = \exp(y_2)$. The Jacobian is $1/(z_1z_2)$. The joint distribution is the Jacobian times the bivariate normal distribution, evaluated at $y_1 = \log z_1 - \log z_2$ and $y_2 = \log z_2$, from which it is now necessary to integrate out z_2. Obviously, this is going to be tedious, but the hint gives a much simpler way to proceed. The variable $w = y_1+y_2$ has a normal distribution with mean $\mu = \mu_1+\mu_2$ and variance $\sigma^2 = (\sigma_1^2 + \sigma_2^2 + 2\sigma_{12})$. We already have a simple result for $\exp(w)$ in Exercise 22; this has a lognormal distribution.

28. **Probability Generating Function**. For a discrete random variable, x, the function

$$E[t^x] = \sum_{x=0}^{\infty} t^x \text{Prob}[X = x]$$

is called the **probability generating function** because in the function, the coefficient on t^i is Prob[$X=i$]. Suppose that x is the number of the repetitions of an experiment with probability π of success upon which the first success occurs. The density of x is the *geometric distribution*,

$$\text{Prob}[X=x] = (1 - \pi)^{x-1}\pi.$$

What is the probability generating function?

$$\begin{aligned} E[t^x] &= \sum_{x=0}^{\infty} t^x (1-\pi)^{x-1}\pi \\ &= \frac{\pi}{(1-\pi)}\sum_{x=0}^{\infty}[t(1-\pi)]^x \\ &= \frac{\pi}{(1-\pi)}\frac{1}{1-t(1-\pi)}. \quad \square \end{aligned}$$

29. **Moment Generating Function**. For the random variable X, with probability density function $f(x)$, if the function $M(t) = E[e^{tx}]$ exists, it is the moment generating function. Assuming the function exists, it can be shown that $d^rM(t)/dt^r|t=0 = E[x^r]$. Find the moment generating functions for

(a) The Exponential distribution of Exercise 9.
(b) The Poisson distribution of Exercise 21.

For the continuous variable in (a), For $f(x) = \theta\exp(-\theta x)$, $M(t) = \int_0^{\infty} e^{tx}\theta e^{-\theta x}dx = \int_0^{\infty}\theta e^{-(\theta - t)x}dx$.

This is θ times a Gamma integral (see Section 5.4.2b) with p=1, c=1, and $a = (\theta-t)$. Therefore, $M(t) = \theta/(\theta - t)$.

For the Poisson distribution,

$$\begin{aligned} M(t) &= \sum_{x=0}^{\infty} e^{tx}e^{-\lambda}\lambda^x / x! = \sum_{x=0}^{\infty} e^{-\lambda}(\lambda e^t)^x / x! \\ &= \sum_{x=0}^{\infty} e^{-\lambda}e^{\lambda e^t}e^{-\lambda e^t}(\lambda e^t)^x / x! \\ &= e^{-\lambda+\lambda e^t}\sum_{x=0}^{\infty} e^{-\lambda e^t}(\lambda e^t)^x / x! \end{aligned}$$

The sum is the sum of probabilities for a Poisson distribution with parameter λe^t, which equals 1, so the term before the summation sign is the moment generating function, $M(t) = \exp[\lambda(e^t - 1)]$. $\square$

28. **Moment generating function for a sum of variables.** When it exists, the moment generating function has a one to one correspondence with the distribution. Thus, for example, if we begin with some random variable and find that a transformation of it has a particular MGF, we may infer that the function of the random variable has the distribution associated with that MGF. A useful application is the following:
If x and y are independent, the MGF of $x + y$ is $M_x(t)M_y(t)$.

(a) Use this result to prove that the sum of Poisson random variables has a Poisson distribution.
(b) Use the result to prove that the sum of chi-squared variables has a chi-squared distribution. [Note, you must first find the MGF for a chi-squared variate. The density is given in (3-39).]
(c) The MGF for the standard normal distribution is $M_z = \exp(-t^2/2)$. Find the MGF for the N[μ,σ^2] distribution, then find the distribution of a sum of normally distributed variables.

(a) From the previous problem, $M_x(t) = \exp[\lambda(e^t - 1)]$. Suppose y is distributed as Poisson with parameter μ. Then, $M_y(t)=\exp[\mu(e^t-1)]$. The product of these two moment generating functions is $M_x(t)M_y(t)= \exp[\lambda(e^t - 1)]\exp[\mu(e^t - 1)] = \exp[(\lambda+\mu)(e^t - 1)]$, which is the moment generating function of the Poisson distribution with parameter $\lambda+\mu$. Therefore, on the basis of the theorem given in the problem, it follows that $x+y$ has a Poisson distribution with parameter $\lambda+\mu$.

(b) The density of the Chi-squared distribution with n degrees of freedom is [from (C-39)]

$$f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-\frac{1}{2}x} x^{\frac{n}{2}-1}, x > 0.$$

Let the constant term be k for the present. The moment generating function is

$$M(t) = k\int_0^\infty e^{tx}e^{-x/2}x^{(n/2)-1}dx$$

$$= k\int_0^\infty e^{-x(1/2-t)}x^{(n/2)-1}dx\ .$$

This is a gamma integral which reduces to $M(t) = k(1/2 - t)^{-n/2}\Gamma(n/2)$. Now, reinserting the constant k and simplifying produces the moment generating function $M(t) = (1 - 2t)^{-n/2}$. Suppose that x_i is distributed as chi-squared with n_i degrees of freedom. The moment generating function of $\Sigma_i x_i$ is

$$\Pi_i M_i(t) = (1-2t)^{-\sum_i n_i/2}$$

which is the MGF of a chi-squared variable with $n = \Sigma_i n_i$ degrees of freedom.

(c) We let $y = \sigma z + \mu$. Then, $M_y(t) = E[\exp(ty)] = E\left[e^{t(\sigma z+\mu)}\right] = e^{t\mu}E\left[e^{\sigma tz}\right] = e^{t\mu}E\left[e^{(\sigma t)z}\right]$

$$= e^{\mu t}e^{-(\sigma t)^2/2} = \exp\left[\mu t - (\sigma^2 t^2)/2\right]$$

Using the same approach as in part b., it follows that the moment generating function for a sum of random variables with means μ_i and standard deviations σ_i is

$$M_{\sum_i x_i} = \exp\left[\sum_i \mu_i - \frac{1}{2}\left(\sum_i \sigma_i^2\right)t^2\right].\ \square$$

Appendix C

Estimation and Inference

1. The following sample is drawn from a normal distribution with mean μ and standard deviation σ:

x = 1.3, 2.1, .4, 1.3, .5, .2, 1.8, 2.5, 1.9, 3.2.

Compute the mean, median, variance, and standard deviation of the sample.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 1.52,$$

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} = .9418,$$

s = .97

median = 1.55, midway between 1.3 and 1.8.

2. Using the data in the previous exercise, test the following hypotheses:

(a) $\mu > 2$.
(b) $\mu < .7$.
(c) $\sigma^2 = .5$.
(d) Using a likelihood ratio test, test the following hypothesis $\mu = 1.8$, $\sigma^2 = .8$.

(a) We would reject the hypothesis if 1.52 is too small relative to the hypothesized value of 2. Since the data are sampled from a normal distribution, we may use a t test to test the hypothesis. The t ratio is

$$t[9] = (1.52 - 2) / [.97/\sqrt{10}] = -1.472.$$

The 95% critical value from the t distribution for a one tailed test is -1.833. Therefore, we would not reject the hypothesis at a significance level of 95%.

(b) We would reject the hypothesis if 1.52 is excessively large relative to the hypothesized mean of .7. The t ratio is $t[9] = (1.52 - .7) / [.97/\sqrt{10}] = 2.673$. Using the same critical value as in the previous problem, we would reject this hypothesis.

(c) The statistic $(n\text{-}1)s^2/\sigma^2$ is distributed as χ^2 with 9 degrees of freedom. This is 9(.94)/.5 = 16.920. The 95% critical values from the chi-squared table for a two tailed test are 2.70 and 19.02. Thus we would not reject the hypothesis.

(d) The log-likelihood for a sample from a normal distribution is

$$\ln L = -(n/2)\ln(2\pi) - (n/2)\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

The sample values are $\hat{\mu} = \bar{x} = 1.52$, $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n} = .8476$.

The maximized log-likelihood for the sample is -13.363. A useful shortcut for computing the log-likelihood at the hypothesized values is $\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n\left(\bar{x} - \mu\right)^2$. For the hypothesized value of $\mu = 1.8$, this is $\sum_{i=1}^{n}(x_i - 1.8)^2 = 9.26$. The log-likelihood is -5(ln(2π) - 5(ln.8) - (1/1.6)9.26 = -13.861. The likelihood ratio statistic is $-2(\ln L_r - \ln L_u) = .996$. The critical value for a chi-squared with 2 degrees of freedom is 5.99, so we would not reject the hypothesis.

3. Suppose that the following sample is drawn from a normal distribution with mean μ and standard deviation σ: y = 3.1, -.1, .3, 1.4, 2.9, .3, 2.2, 1.5, 4.2, .4. Test the hypothesis that the mean of the distribution which produced these data is the same as that which produced the data in Exercise 1. Test the hypothesis assuming that the variances are the same. Test the hypothesis that the variances are the same using an F test and using a likelihood ratio test. (Do not assume that the means are the same.)

If the variances are the same,

$$\bar{x}_1 \sim N[\mu_1, \sigma_1^2 / n_1] \text{ and } \bar{x}_2 \sim N[\mu_2, \sigma_2^2 / n_2],$$
$$\bar{x}_1 - \bar{x}_2 \sim N[\mu_1 - \mu_2, \sigma^2\{(1/n_1) + (1/n_2)\}],$$
$$(n_1\text{-}1)s_1^2/\sigma^2 \sim \chi^2[n_1\text{-}1] \text{ and } (n_2\text{-}1)s_2^2/\sigma^2 \sim \chi^2[n_2\text{-}1]$$
$$(n_1\text{-}1)s_1^2/\sigma^2 + (n_2\text{-}1)s_2^2/\sigma^2 \sim \chi^2[n_1 + n_2 - 2]$$

Thus, the statistic
$$t = \frac{\left\{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right\} / \sqrt{\sigma^2\left[(1/n_1) + (1/n_2)\right]}}{\sqrt{\left\{(n_1 - 1)s_1^2 / \sigma^2 + (n_2 - 1)s_2^2 / \sigma^2\right\} / (n_1 + n_2 - 2)}}$$
is the ratio of a standard normal variable to the square root of a chi-squared variable divided by its degrees of freedom which is distributed as t with $n_1 + n_2 - 2$ degrees of freedom. Under the hypothesis that the means are equal, the statistic is
$$t = \frac{(\bar{x}_1 - \bar{x}_2) / \sqrt{(1/n_1) + (1/n_2)}}{\sqrt{\left\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\right\} / (n_1 + n_2 - 2)}}$$
The sample statistics are
$$n_1 = 10,\ \bar{x}_1 = 1.52,\ s_1^2 = .9418$$
$$n_2 = 10,\ \bar{x}_2 = 1.62,\ s_2^2 = 2.0907$$
so $t[18] = .1816$. This is quite small, so we would not reject the hypothesis of equal means.

For random sampling from two normal distributions, under the hypothesis of equal variances, the statistic $F[n_1\text{-}1,n_2\text{-}1] = \dfrac{\left[(n_1 - 1)s_1^2 / \sigma^2\right] / (n_1 - 1)}{\left[(n_2 - 1)s_2^2 / \sigma^2\right] / (n_2 - 1)}$ is the ratio of two independent chi-squared variables, each divided by its degrees of freedom. This has the F distribution with n_1-1 and n_2-1 degrees of freedom. If $n_1 = n_2$, the statistic reduces to $F[n_1\text{-}1,n_2\text{-}1] = s_1^2 / s_2^2$. For our purposes, it is more convenient to put the larger variance in the denominator. Thus, for our sample data, $F[9,9] = 2.0907 / .9418 = 2.2199$. The 95% critical value from the F table is 3.18. Thus, we would not reject the hypothesis of equal variances.

The likelihood ratio test is based on the test statistic $\lambda = -2(\ln L_r - \ln L_u)$. The log-likelihood for the joint sample of 20 observations is the sum of the two separate log-likelihoods if the samples are assumed to be independent. A useful shortcut for computing the log-likelihood arises when the maximum likelihood estimates are inserted: At the maximum likelihood estimates, $\ln L = (-n/2)[1 + \ln(2\pi) + \ln \hat{\sigma}^2]$. So, the log-likelihood for the sample is $\ln L_2 = (-5/2)[1 + \ln(2\pi) + \ln((9/10)2.0907)] = -17.35007$. (Remember, we don't make the degrees of freedom correction for the variance estimator.) The log-likelihood function for the sample of 20 observations is just the sum of the two log-likelihoods if the samples are completely independent. The unrestricted log-likelihood function is, thus, -13.363+(-17.35001) = -30.713077. To compute the restricted log-likelihood function, we need the pooled estimator which does not assume that the means are identical. This would be
$$\hat{\sigma}^2 = [(n_1\text{-}1)\, s_1^2 + (n_2\text{-}1)\, s_2^2]/[n_1 + n_2]$$
$$= [9(.9418) + 9(2.0907)]/20 = 1.36463.$$
So, the restricted log-likelihood is $\ln L_r = (-20/2)[1 + \ln(2\pi) + \ln(1.36463)] = -31.4876$. Minus twice the difference is $\lambda = -2[-31.4876 - (-30.713077)] = 1.541$. This is distributed as chi-squared with one degree of freedom. The critical value is 3.84, so we would not reject the hypothesis.

4. A common method of simulating random draws from the standard normal distribution is to compute the sum of 12 draws from the uniform [0,1] distribution and subtract 6. Can you justify this procedure?

The uniform distribution has mean 2 and variance 1/12. Therefore, the statistic $12(\bar{x} - 1/2) = \sum_{i=1}^{12} x_i - 6$ is equivalent to $z = \sqrt{n}\,(\bar{x} - \mu) / \sigma$. As $n \to \infty$, this converges to a standard normal variable. Experience suggests that a sample of 12 is large enough to approximate this result. However, more recently developed random number generators usually use different procedures based on the truncation error which occurs in representing real numbers in a digital computer.

5. Using the data in Exercise 1, form confidence intervals for the mean and standard deviation.

Since the underlying distribution is normal, we may use the t distribution. Using (4-57), we obtain a 95% confidence interval for the mean of $1.52 - 2.262[.97/\sqrt{10}] \leq \mu \leq 1.52 + 2.262[.97/\sqrt{10}]$ or $.826 \leq \mu \leq 2.214$. Using the procedure in Example 4.30, we obtain a 95% confidence for σ^2 of $9(.941)/19.02 \leq \sigma^2 \leq 9(.941)/2.70$ or $.445 \leq \sigma^2 \leq 3.137$. Taking square roots gives the confidence interval for σ, $.667 \leq \sigma \leq 1.771$.

6. Based on a sample of 65 observations from a normal distribution, you obtain a *median* of 34 and a standard deviation of 13.3. Form a confidence interval for the mean. (**Hint**: Use the asymptotic distribution. See Example 4.15.) Compare your confidence interval to the one you would have obtained had the estimate of 34 been the sample mean instead of the sample median.

The asymptotic variance of the median is $\pi\sigma^2/(2n)$. Using the asymptotic normal distribution instead of the t distribution, the confidence interval is $34 - 1.96(13.3^2\pi/130)^2 \leq \mu \leq 34 + 1.96(13.3^2\pi/130)^2$ or $29.95 \leq \mu \leq 38.052$. Had the estimator been the mean instead of the median, the appropriate asymptotic variance would be σ^2/n, instead, which we would estimate with $13.3^2/65 = 2.72$ compared to 4.274 for the median. The confidence interval would have been (30.77,37.24), which is somewhat narrower.

7. The random variable x has a continuous distribution $f(x)$ and cumulative distribution function $F(x)$. What is the probability distribution of the sample maximum? (**Hint:** In a random sample of n observations, x_1, x_2, ..., x_n, if z is the maximum, then every observation in the sample is less than or equal to z. Use the cdf.)

If z is the maximum, then every sample observation is less than or equal to z. The probability of this is $\text{Prob}[x_1 \# z, x_2 \# z, ..., x_n \# z] = F(z)F(z)...F(z) = [F(z)]^n$. The density is the derivative, $n[F(z)]^{n-1}f(z)$.

8. Assume the distribution of x is $f(x) = 1/\theta$, $0 \leq x \leq \theta$. In random sampling from this distribution, prove that the sample maximum is a consistent estimator of θ. Note: you can prove that the maximum is the maximum likelihood estimator of θ. But, the usual properties do not apply here. Why not? (**Hint:** Attempt to verify that the expected first derivative of the log-likelihood with respect to θ is zero.)

Using the result of the previous problem, the density of the maximum is

$$n[z/\theta]^{n-1}(1/\theta),\ 0 < z < \theta.$$

Therefore, the expected value is $E[z] = \int_0^\theta z^n dz = [\theta^{n+1}/(n+1)][n/\theta^n] = n\theta/(n+1)$. The variance is found likewise, $E[z^2] = \int_0^\theta z^2 n(z/n)^{n-1}(1/\theta)dz = n\theta^2/(n+2)$ so $\text{Var}[z] = E[z^2] - (E[z])^2 = n\theta^2/[(n+1)^2(n+2)]$. Using mean squared convergence we see that $\lim_{n\to\infty} E[z] = \theta$ and $\lim_{n\to\infty} \text{Var}[z] = 0$, so that plim $z = \theta$. □

9. In random sampling from the exponential distribution, $f(x) = \frac{1}{\theta}e^{\frac{-x}{\theta}}$, $x > 0$, $\theta > 0$, find the maximum likelihood estimator of θ and obtain the asymptotic distribution of this estimator.

The log-likelihood is $\ln L = -n\ln\theta - (1/\theta)\sum_{i=1}^n x_i$. The maximum likelihood estimator is obtained as the solution to $\partial\ln L/\partial\theta = -n/\theta + (1/\theta^2)\sum_{i=1}^n x_i = 0$, or $\hat{\theta}_{ML} = (1/n)\sum_{i=1}^n x_i = \bar{x}$. The asymptotic variance of the MLE is $\{-E[\partial^2\ln L/\partial\theta^2]\}^{-1} = \{-E[n/\theta^2 - (2/\theta^3)\sum_{i=1}^n x_i]\}^{-1}$. To find the expected value of this random variable, we need $E[x_i] = \theta$. Therefore, the asymptotic variance is θ^2/n. The asymptotic distribution is normal with mean θ and this variance.

10. Suppose in a sample of 500 observations from a normal distribution with mean μ and standard deviation σ, you are told that 35% of the observations are less than 2.1 and 55% of the observations are less than 3.6. Estimate μ and σ.

If 35% of the observations are less than 2.1, we would infer that

$$\Phi[(2.1 - \mu)/\sigma] = .35, \text{ or } (2.1 - \mu)/\sigma = -.385 \Rightarrow 2.1 - \mu = -.385\sigma.$$

Likewise,

$$\Phi[(3.6 - \mu)/\sigma] = .55, \text{ or } (3.6 - \mu)/\sigma = .126 \Rightarrow 3.6 - \mu = .126\sigma.$$

The joint solution is $\hat{\mu}$ = 3.2301 and $\hat{\sigma}$ = 2.9354. It might not seem obvious, but we can also derive asymptotic standard errors for these estimates by constructing them as method of moments estimators. Observe, first, that the two estimates are based on moment estimators of the probabilities. Let x_i denote one of the 500 observations drawn from the normal distribution. Then, the two proportions are obtained as follows: Let $z_i(2.1) = \mathbf{1}[x_i < 2.1]$ and $z_i(3.6) = \mathbf{1}[x_i < 3.6]$ be indicator functions. Then, the proportion of 35% has been obtained as $\bar{z}(2.1)$ and .55 is $\bar{z}(3.6)$. So, the two proportions are simply the means of functions of the sample observations. Each z_i is a draw from a Bernoulli distribution with success probability $\pi(2.1) = \Phi((2.1-\mu)/\sigma)$ for $z_i(2.1)$ and $\pi(3.6) = \Phi((3.6-\mu)/\sigma)$ for $z_i(3.6)$. Therefore, $E[\bar{z}(2.1)] = \pi(2.1)$, and $E[\bar{z}(3.6)] = \pi(3.6)$. The variances in each case are $\text{Var}[\bar{z}(.)] = 1/n[\pi(.)(1-\pi(.))]$. The covariance of the two sample means is a bit trickier, but we can deduce it from the results of random sampling. $\text{Cov}[\bar{z}(2.1), \bar{z}(3.6)]]$
$= 1/n\ \text{Cov}[z_i(2.1),z_i(3.6)]$, and, since in random sampling sample moments will converge to their population counterparts, $\text{Cov}[z_i(2.1),z_i(3.6)] = \text{plim}\ [\{(1/n)\sum_{i=1}^{n} z_i(2.1)z_i(3.6)\} - \pi(2.1)\pi(3.6)]$. But, $z_i(2.1)z_i(3.6)$ must equal $[z_i(2.1)]^2$ which, in turn, equals $z_i(2.1)$. It follows, then, that
$\text{Cov}[z_i(2.1),z_i(3.6)] = \pi(2.1)[1 - \pi(3.6)]$. Therefore, the asymptotic covariance matrix for the two sample proportions is $Asy.Var[p(2.1), p(3.6)] = \mathbf{\Sigma} = \frac{1}{n}\begin{bmatrix} \pi(2.1)(1-\pi(2.1)) & \pi(2.1)(1-\pi(3.6)) \\ \pi(2.1)(1-\pi(3.6)) & \pi(3.6)(1-\pi(3.6)) \end{bmatrix}$. If we insert our sample estimates, we obtain $Est.Asy.Var[p(2.1), p(3.6)] = \mathbf{S} = \begin{bmatrix} 0.000455 & 0.000315 \\ 0.000315 & 0.000495 \end{bmatrix}$. Now, ultimately, our estimates of μ and σ are found as functions of $p(2.1)$ and $p(3.6)$, using the method of moments. The moment equations are

$$m_{2.1} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(2.1)\right] - \Phi\left[\frac{2.1-\mu}{\sigma}\right] = 0,$$

$$m_{3.6} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(3.6)\right] - \Phi\left[\frac{3.6-\mu}{\sigma}\right] = 0.$$

Now, let $\mathbf{\Gamma} = \begin{bmatrix} \partial m_{2.1}/\partial\mu & \partial m_{2.1}/\partial\sigma \\ \partial m_{3.6}/\partial\mu & \partial m_{3.61}/\partial\sigma \end{bmatrix}$ and let $\mathbf{G}$ be the sample estimate of Γ. Then, the estimator of the asymptotic covariance matrix of $(\hat{\mu}, \hat{\sigma})$ is $[\mathbf{GS}^{-1}\mathbf{G}']^{-1}$. The remaining detail is the derivatives, which are just $\partial m_{2.1}/\partial\mu = (1/\sigma)\phi((2.1-\mu)/\sigma)$ and $\partial m_{2.1}/\partial\sigma = (2.1-\mu)/\sigma[\text{M}m_{2.1}/\text{M}\sigma]$ and likewise for $m_{3.6}$. Inserting our sample estimates produces $\mathbf{G} = \begin{bmatrix} 0.37046 & -0.14259 \\ 0.39579 & 0.04987 \end{bmatrix}$. Finally, multiplying the matrices and computing the necessary inverses produces $[\mathbf{GS}^{-1}\mathbf{G}']^{-1} = \begin{bmatrix} 0.10178 & -0.12492 \\ -0.12492 & 0.16973 \end{bmatrix}$. The asymptotic distribution would be normal, as usual. Based on these results, a 95% confidence interval for μ would be $3.2301 \pm 1.96(.10178)^2$ = 2.6048 to 3.8554.

11. For random sampling from a normal distribution with nonzero mean μ and standard deviation σ, find the asymptotic joint distribution of the maximum likelihood estimators of σ/μ and μ^2/σ^2.

The maximum likelihood estimators, $\hat{\mu} = (1/n)\sum_{i=1}^{n} x_i$ and $\hat{\sigma}^2 = (1/n)\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ were given in (4-49). By the invariance principle, we know that the maximum likelihood estimators of μ/σ and μ^2/σ^2 are $\hat{\mu}/\hat{\sigma}$ and $\hat{\mu}/\hat{\sigma}^2$ and the maximum likelihood estimate of σ is $\sqrt{\hat{\sigma}}$. To obtain the asymptotic joint distribution of the two functions of $\hat{\mu}$ and $\hat{\sigma}$, we first require the asymptotic joint distribution of $\hat{\mu}$ and $\hat{\sigma}^2$. This is normal with mean vector (μ,σ^2) and covariance matrix equal to the inverse of the information matrix. This is the inverse of

$$-E\begin{bmatrix} \partial^2 \log L / \partial\mu^2 & \partial^2 \log L / \partial\mu\partial\sigma^2 \\ \partial^2 \log L / \partial\sigma^2\partial\mu & \partial^2 \log L / \partial(\sigma^2)^2 \end{bmatrix} = \begin{bmatrix} -n/\sigma^2 & -(1/\sigma^3)\sum_{i=1}^{n}(x_i-\mu) \\ -(1/\sigma^3)\sum_{i=1}^{n}(x_i-\mu) & n/(2\sigma^4)-(1/\sigma^6)\sum_{i=1}^{n}\left(x_i-\mu\right)^2 \end{bmatrix}$$

The off diagonal term has expected value 0. Each term in the sum in the lower right has expected value σ^2, so, after collecting terms, taking the negative, and inverting, we obtain the asymptotic covariance matrix,

$\mathbf{V} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}$. To obtain the asymptotic joint distribution of the two nonlinear functions, we use the multivariate version of Theorem 4.4. Thus, we require $\mathbf{H} = \mathbf{JVJ'}$ where

$\mathbf{J} = \begin{bmatrix} \partial(\mu/\sigma)/\partial\mu & \partial(\mu/\sigma)/\partial\sigma^2 \\ \partial(\mu^2/\sigma^2)/\partial\mu & \partial(\mu^2/\sigma^2)/\partial\sigma^2 \end{bmatrix} = \begin{bmatrix} 1/\sigma & -\mu/(2\sigma^3) \\ 2\mu/\sigma^2 & -\mu/\sigma^4 \end{bmatrix}$. The product is

$\mathbf{H} = \dfrac{1}{n}\begin{bmatrix} 1+\mu^2/(2\sigma^2) & 2\mu/\sigma+(\mu/\sigma)^3 \\ 2\mu/\sigma+(\mu/\sigma)^3 & 4\mu^2/\sigma^2+2\mu^4/\sigma^4 \end{bmatrix}$.

12. The random variable x has the following distribution: $f(x) = e^{-\lambda}\lambda^x / x!$, $x = 0,1,2,...$ The following random sample is drawn: 1,1,4,2,0,0,3,2,3,5,1,2,1,0,0. Carry out a Wald test of the hypothesis that $\lambda = 2$.

For random sampling from the Poisson distribution, the maximum likelihood estimator of λ is $\bar{x}$ = 25/15. (See Example 4.18.) The second derivative of the log-likelihood is $-\sum_{i=1}^{n} x_i / \lambda^2$, so the the asymptotic variance is λ/n. The Wald statistic would be

$$W = \frac{\left(\bar{x}-2\right)^2}{\hat{\lambda}/n} = [(25/15 - 2)^2]/[(25/15)/15] = 1.0.$$

The 95% critical value from the chi-squared distribution with one degree of freedom is 3.84, so the hypothesis would not be rejected. Alternatively, one might estimate the variance of with s^2/n = 2.38/15 = 0.159. Then, the Wald statistic would be $(1.6 - 2)^2/.159 = 1.01$. The conclusion is the same. ~

13. Based on random sampling of 16 observations from the exponential distribution of Exercise 9, we wish to test the hypothesis that $\theta = 1$. We will reject the hypothesis if $\bar{x}$ is greater than 1.2 or less than .8. We are interested in the power of this test.

(a) Using the asymptotic distribution of $\bar{x}$ graph the asymptotic approximation to the true power function.

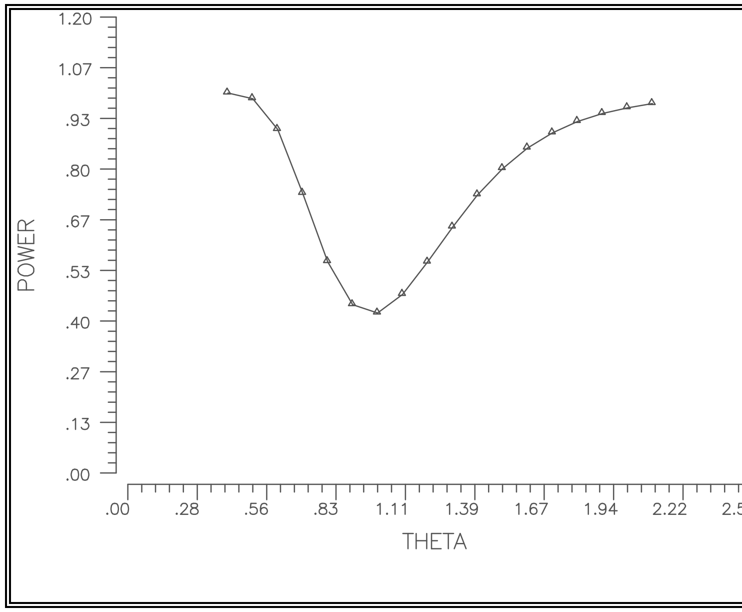
(b) Using the result discussed in Example 4.17, describe how to obtain the true power function for this test.

The asymptotic distribution of $\bar{x}$ is normal with mean θ and variance θ^2/n. Therefore, the power function based on the asymptotic distribution is the probability that a normally distributed variable with mean equal to θ and variance equal to θ^2/n will be greater than 1.2 or less than .8. That is,

$$\text{Power} = \Phi[(.8 - \theta)/(\theta/4)] + 1 - \Phi[(1.2 - \theta)/(\theta/4)].$$

Some values of this power function and a sketch are given below:

θ	Approx. Power	True Power
.4	1.000	1.000
.5	.992	.985
.6	.908	.904
.7	.718	.736
.8	.522	.556
.9	.420	.443
1.0	.423	.421
1.1	.496	.470
1.2	.591	.555
1.3	.685	.647
1.4	.759	.732
1.5	.819	.801
1.6	.864	.855
1.7	.897	.895
1.8	.922	.925
1.9	.940	.946
2.0	.954	.961
2.1	.963	.972



Note that the power function does not have the symmetric shape of Figure 4.7 because both the variance and the mean are changing as θ changes. Moreover, the power is not the lowest at the value of $\theta = 1$, but at about $\theta = .9$. That means (assuming that the normal distribution is appropriate) that the test is slightly biased. The size of the test is its power at the hypothesized value, or .423, and there are points at which the power is less than the size.

According to the example cited, the true distribution of $\bar{x}$ is that of $\theta/(2n)$ times a chi-squared variable with $2n$ degrees of freedom. Therefore, we could find the true power by finding the probability that a chi-squared variable with $2n$ degrees of freedom is less than $.8(2n/\theta)$ or greater than $1.2(2n/\theta)$. Thus,

$$\text{True power} = F(25.6/\theta) + 1 - F(38.4/\theta)$$

where $F(.)$ is the CDF of the chi-squared distribution with 32 degrees of freedom. Values for the correct power function are shown above. Given that the sample is only 16 observations, the closeness of the asymptotic approximation is quite impressive.

14. For the normal distribution, $\mu_{2k} = \sigma^{2k}(2k)!/(k!2^k)$ and $\mu_{2k+1} = 0$, $k = 0,1,...$ Use this result to show that in Example 4.27, $\theta_1 = 0$ and $\theta_2 = 3$, and $\mathbf{JVJ'} = \begin{bmatrix} 6 & 0 \\ 0 & 24 \end{bmatrix}$.

For θ_1 and θ_2, just plug in the result above using $k = 2$, 3, and 4. The example involves 3 moments, m_2, m_3, and m_4. The asymptotic covariance matrix for these three moments can be based on the formulas given in Example 4.26. In particular, we note, first, that for the normal distribution, Asy.Cov$[m_2,m_3]$ and Asy.Cov$[m_3,m_4]$ will be zero since they involve only odd moments, which are all zero. The necessary even moments are $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$. $\mu_6 = 15\sigma^6$, $\mu_8 = 105\sigma^8$. The three variances will be

$$n[\text{Asy.Var}(m_2)] = \mu_4 - \mu_2^2 = 3\sigma^4 - (\sigma^2)^2 = 2\sigma^4$$
$$n[\text{Asy.Var}(m_3)] = \mu_6 - \mu_3^2 - 6\mu_4\mu_2 + 9\mu_2^3 = 6\sigma^6$$
$$n[\text{Asy.Var}(m_4)] = \mu_8 - \mu_4^2 - 8\mu_5\mu_3 + 16\mu_2\mu_3^2 = 96\sigma^8$$

and $$n[\text{Asy.Cov}(m_2,m_4)] = \mu_6 - \mu_2\mu_4 - 4\mu_3^2 = 12\sigma^6.$$

The elements of $\mathbf{J}$ are given in Example 4.27. For the normal distribution, this matrix would be $\mathbf{J} = \begin{bmatrix} 0 & 1/\sigma^3 & 0 \\ -6/\sigma^2 & 0 & 1/\sigma^4 \end{bmatrix}$. Multiplying out **JVJ/N** produces the result given above. □

15. Testing for normality. One method that has been suggested for testing whether the distribution underlying a sample is normal is to refer the statistic $L = n\{\text{skewness}^2/6 + (\text{kurtosis-3})^2/24\}$ to the chi-squared distribution with 2 degrees of freedom. Using the data in Exercise 1, carry out the test.

The skewness coefficient is .14192 and the kurtosis is 1.8447. (These are the third and fourth moments divided by the third and fourth power of the sample standard deviation.) Inserting these in the expression above produces $L = 10\{.14192^2/6 + (1.8447 - 3)^2/24\} = .59$. The critical value from the chi-squared distribution with 2 degrees of freedom (95%) is 5.99. Thus, the hypothesis of normality cannot be rejected.

16. Suppose the joint distribution of the two random variables x and y is

$$f(x,y) = \theta e^{-(\beta+\theta)y}(\beta y)^x / x! \quad \beta,\theta \; 0,\; y \$ 0,\; x = 0,1,2,...$$

(a) Find the maximum likelihood estimators of β and θ and their asymptotic joint distribution.
(b) Find the maximum likelihood estimator of θ/(β+θ) and its asymptotic distribution.
(c) Prove that $f(x)$ is of the form $f(x) = \gamma(1-\gamma)^x$, $x = 0,1,2,...$
Then, find the maximum likelihood estimator of γ and its asymptotic distribution.
(d) Prove that f(y*x) is of the form $\lambda e^{-\lambda y}(\lambda y)^x/x!$ Prove that $f(y|x)$ integrates to 1. Find the maximum likelihood estimator of λ and its asymptotic distribution. (**Hint**: In the conditional distribution, just carry the xs along as constants.)
(e) Prove that $f(y) = \theta e^{-\theta y}$ then find the maximum likelihood estimator of θ and its asymptotic variance.
(f) Prove that $f(x|y) = e^{-\beta y}(\beta y)^x/x!$. Based on this distribution, what is the maximum likelihood estimator of β?

The log-likelihood is $\ln L = n\ln\theta - (\beta+\theta)\sum_{i=1}^n y_i + \ln\beta\sum_{i=1}^n x_i + \sum_{i=1}^n x_i \log y_i - \sum_{i=1}^n \log(x_i!)$

The first and second derivatives are

$$\partial\ln L/\partial\theta = n/\theta - \sum_{i=1}^n y_i$$
$$\partial\ln L/\partial\beta = -\sum_{i=1}^n y_i + \sum_{i=1}^n x_i/\beta$$
$$\partial^2\ln L/\partial\theta^2 = -n/\theta^2$$
$$\partial^2\ln L/\partial\beta^2 = -\sum_{i=1}^n x_i/\beta^2$$
$$\partial^2\ln L/\partial\beta\partial\theta = 0.$$

Therefore, the maximum likelihood estimators are $\hat{\theta} = 1/\bar{y}$ and $\hat{\beta} = \bar{x}/\bar{y}$ and the asymptotic covariance matrix is the inverse of $E\begin{bmatrix} n/\theta^2 & 0 \\ 0 & \sum_{i=1}^n x_i/\beta^2 \end{bmatrix}$. In order to complete the derivation, we will require the expected value of $\sum_{i=1}^n x_i = nE[x_i]$. In order to obtain $E[x_i]$, it is necessary to obtain the marginal distribution of x_i, which is f(x) = $\int_0^\infty \theta e^{-(\beta+\theta)y}(\beta y)^x / x!\,dy = \beta^x(\theta/x!)\int_0^\infty e^{-(\beta+\theta)y}y^x dy$. This is $\beta^x(\theta/x!)$ times a gamma integral. This is $f(x) = \beta^x(\theta/x!)[\Gamma(x+1)]/(\beta+\theta)^{x+1}$. But, $\Gamma(x+1) = x!$, so the expression reduces to

$$f(x) = [\theta/(\beta+\theta)][\beta/(\beta+\theta)]^x.$$

Thus, x has a geometric distribution with parameter $\pi = \theta/(\beta+\theta)$. (This is the distribution of the number of tries until the first success of independent trials each with success probability 1-π. Finally, we require the expected value of x_i, which is $E[x] = [\theta/(\beta+\theta)]\sum_{x=0}^\infty x[\beta/(\beta+\theta)]^x = \beta/\theta$. Then, the required asymptotic covariance matrix is $\begin{bmatrix} n/\theta^2 & 0 \\ 0 & n(\beta/\theta)/\beta^2 \end{bmatrix}^{-1} = \begin{bmatrix} \theta^2/n & 0 \\ 0 & \beta\theta/n \end{bmatrix}$.

The maximum likelihood estimator of θ/(β+θ) is

$$\widehat{\theta/(\beta+\theta)} = (1/\bar{y})/[\bar{x}/\bar{y} + 1/\bar{y}] = 1/(1+\bar{x}).$$

Its asymptotic variance is obtained using the variance of a nonlinear function

$$V = [\beta/(\beta+\theta)]^2(\theta^2/n) + [-\theta/(\beta+\theta)]^2(\beta\theta/n) = \beta\theta^2/[n(\beta+\theta)^3].$$

The asymptotic variance could also be obtained as $[-1/(1 + E[x])^2]^2$Asy.Var[$\bar{x}$].)

For part (c), we just note that $\gamma = \theta/(\beta+\theta)$. For a sample of observations on x, the log-likelihood would be

$$\ln L = n\ln\gamma + \ln(1-\gamma)\sum_{i=1}^{n} x_i$$

$$\partial \ln L/d\gamma = n/\gamma - \sum_{i=1}^{n} x_i /(1-\gamma).$$

A solution is obtained by first noting that at the solution, $(1-\gamma)/\gamma = \bar{x} = 1/\gamma - 1$. The solution for γ is, thus, $\hat{\gamma} = 1/(1+\bar{x})$.Of course, this is what we found in part b., which makes sense.

For part (d) $f(y|x) = \dfrac{f(x,y)}{f(x)} = \dfrac{\theta e^{-(\beta+\theta)y}(\beta y)^x(\beta+\theta)^x(\beta+\theta)}{x!\quad\theta\quad\beta x}$. Cancelling terms and gathering the remaining like terms leaves $f(y|x) = (\beta+\theta)[(\beta+\theta)y]^x e^{-(\beta+\theta)y}/x!$ so the density has the required form with $\lambda = (\beta+\theta)$. The integral is $\left\{[\lambda^{x+1}]/x!\right\}\int_0^\infty e^{-\lambda y}y^x dy$. This integral is a Gamma integral which equals $\Gamma(x+1)/\lambda^{x+1}$, which is the reciprocal of the leading scalar, so the product is 1. The log-likelihood function is

$$\ln L = n\ln\lambda - \lambda\sum_{i=1}^{n} y_i + \ln\lambda\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln x_i!$$

$$\partial \ln L/\partial\lambda = \left(\sum_{i=1}^{n} x_i + n\right)/\lambda - \sum_{i=1}^{n} y_i .$$

$$\partial^2 \ln L/\partial\lambda^2 = -\left(\sum_{i=1}^{n} x_i + n\right)/\lambda^2.$$

Therefore, the maximum likelihood estimator of λ is $(1+\bar{x})/\bar{y}$ and the asymptotic variance, conditional on the xs is Asy.Var.$\left[\hat{\lambda}\right] = (\lambda^2/n)/(1+\bar{x})$

Part (e.) We can obtain $f(y)$ by summing over x in the joint density. First, we write the joint density as $f(x,y) = \theta e^{-\theta y}e^{-\beta y}(\beta y)^x/x!$. The sum is, therefore, $f(y) = \theta e^{-\theta y}\sum_{x=0}^{\infty} e^{-\beta y}(\beta y)^x/x!$. The sum is that of the probabilities for a Poisson distribution, so it equals 1. This produces the required result. The maximum likelihood estimator of θ and its asymptotic variance are derived from

$$\ln L = n\ln\theta - \theta\sum_{i=1}^{n} y_i$$

$$\partial \ln L/\partial\theta = n/\theta - \sum_{i=1}^{n} y_i$$

$$\partial^2 \ln L/\partial\theta^2 = -n/\theta^2.$$

Therefore, the maximum likelihood estimator is $1/\bar{y}$ and its asymptotic variance is θ^2/n. Since we found $f(y)$ by factoring $f(x,y)$ into $f(y)f(x|y)$ (apparently, given our result), the answer follows immediately. Just divide the expression used in part e. by $f(y)$. This is a Poisson distribution with parameter βy. The log-likelihood function and its first derivative are

$$\ln L = -\beta\sum_{i=1}^{n} y_i + \ln\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i \ln y_i - \sum_{i=1}^{n}\ln x_i!$$

$$\partial \ln L/\partial\beta = -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i /\beta,$$

from which it follows that $\hat{\beta} = \bar{x}/\bar{y}$.

17. Suppose x has the Weibull distribution, $f(x) = \alpha\beta x^{\beta-1}\exp(-\alpha x^\beta)$, $x, \alpha, \beta > 0$.

(a) Obtain the log-likelihood function for a random sample of n observations.

(b) Obtain the likelihood equations for maximum likelihood estimation of α and β. Note that the first provides an explicit solution for α in terms of the data and β. But, after inserting this in the second, we obtain only an implicit solution for β. How would you obtain the maximum likelihood estimators?

(c) Obtain the second derivatives matrix of the log-likelihood with respect to α and β. The exact expectations of the elements involving β involve the derivatives of the Gamma function and are quite messy analytically. Of course, your exact result provides an empirical estimator. How would you estimate the asymptotic covariance matrix for your estimators in part (b)?

(d) Prove that $\alpha\beta\text{Cov}[\ln x, x^\beta] = 1$. (**Hint:** Use the fact that the expected first derivatives of the log-likelihood function are zero.)

The log-likelihood and its two first derivatives are

$$\log L = n\log\alpha + n\log\beta + (\beta-1)\sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n} x_i^\beta$$

$$\partial\log L/\partial\alpha = n/\alpha - \sum_{i=1}^{n} x_i^\beta$$

$$\partial\log L/\partial\beta = n/\beta + \sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n}(\log x_i)x_i^\beta$$

Since the first likelihood equation implies that at the maximum, $\hat{\alpha} = n / \sum_{i=1}^{n} x_i^\beta$, one approach would be to scan over the range of β and compute the implied value of α. Two practical complications are the allowable range of β and the starting values to use for the search.

The second derivatives are

$$\partial^2\ln L/\partial\alpha^2 = -n/\alpha^2$$

$$\partial^2\ln L/\partial\beta^2 = -n/\beta^2 - \alpha\sum_{i=1}^{n}(\log x_i)^2 x_i^\beta$$

$$\partial^2\ln L/\partial\alpha\partial\beta = -\sum_{i=1}^{n}(\log x_i)x_i^\beta .$$

If we had estimates in hand, the simplest way to estimate the expected values of the Hessian would be to evaluate the expressions above at the maximum likelihood estimates, then compute the negative inverse. First, since the expected value of $\partial\ln L/\partial\alpha$ is zero, it follows that $E[x_i^\beta] = 1/\alpha$. Now,

$$E[\partial\ln L/\partial\beta] = n/\beta + E[\sum_{i=1}^{n}\log x_i] - \alpha E[\sum_{i=1}^{n}(\log x_i)x_i^\beta] = 0$$

as well. Divide by n, and use the fact that every term in a sum has the same expectation to obtain

$$1/\beta + E[\ln x_i] - E[(\ln x_i)x_i^\beta]/E[x_i^\beta] = 0.$$

Now, multiply through by $E[x_i^\beta]$ to obtain $E[x_i^\beta] = E[(\ln x_i)x_i^\beta] - E[\ln x_i]E[x_i^\beta]$

or $\qquad 1/(\alpha\beta) = \text{Cov}[\ln x_i, x_i^\beta]$. ~

18. The following data were generated by the Weibull distribution of Exercise 17:

1.3043	.49254	1.2742	1.4019	.32556	.29965	.26423
1.0878	1.9461	.47615	3.6454	.15344	1.2357	.96381
.33453	1.1227	2.0296	1.2797	.96080	2.0070	

(a) Obtain the maximum likelihood estimates of α and β and estimate the asymptotic covariance matrix for the estimates.

(b) Carry out a Wald test of the hypothesis that β = 1.

(c) Obtain the maximum likelihood estimate of α under the hypothesis that β = 1.

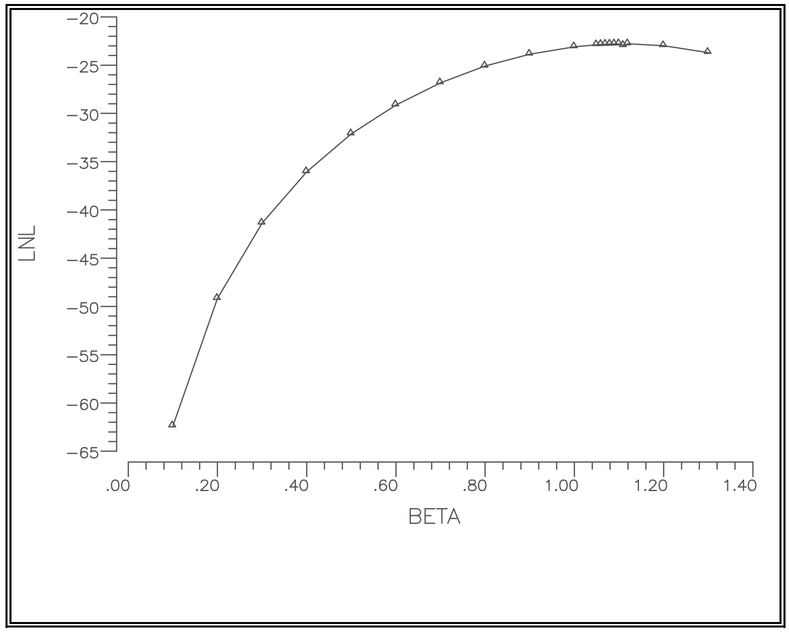
(d) Using the results of a. and c. carry out a likelihood ratio test of the hypothesis that β = 1.

(e) Carry out a Lagrange multiplier test of the hypothesis that β = 1.

As suggested in the previous problem, we can concentrate the log-likelihood over α. From $\partial\log L/\partial\alpha = 0$, we find that at the maximum, $\alpha = 1/[(1/n)\sum_{i=1}^{n} x_i^\beta]$. Thus, we scan over different values of β to seek the value which maximizes log$L$ as given above, where we substitute this expression for each occurrence of α. Values of β and the log-likelihood for a range of values of β are listed and shown in the figure below.

β $\qquad$ log$L$

0.1	-62.386
0.2	-49.175
0.3	-41.381
0.4	-36.051
0.5	-32.122
0.6	-29.127
0.7	-26.829
0.8	-25.098
0.9	-23.866
1.0	-23.101
1.05	-22.891
1.06	-22.863
1.07	-22.841
1.08	-22.823
1.09	-22.809
1.10	-22.800
1.11	-22.796
1.12	-22.797
1.2	-22.984
1.3	-23.693



The maximum occurs at $\beta = 1.11$. The implied value of α is 1.179. The negative of the second derivatives matrix at these values and its inverse are

$$\mathbf{I}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} 25.55 & 9.6506 \\ 9.6506 & 27.7552 \end{bmatrix} \text{ and } \mathbf{I}^{-1}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} .04506 & -.2673 \\ -.2673 & .04148 \end{bmatrix}.$$

The Wald statistic for the hypothesis that $\beta = 1$ is $W = (1.11 - 1)^2/.041477 = .276$. The critical value for a test of size .05 is 3.84, so we would not reject the hypothesis.

If $\beta = 1$, then $\hat{\alpha} = n / \sum_{i=1}^{n} x_i = 0.88496$. The distribution specializes to the geometric distribution if $\beta = 1$, so the restricted log-likelihood would be

$$\log L_r = n\log\alpha - \alpha \sum_{i=1}^{n} x_i = n(\log\alpha - 1) \text{ at the MLE.}$$

$\log L_r$ at $\alpha = .88496$ is -22.44435. The likelihood ratio statistic is $-2\log\lambda = 2(23.10068 - 22.44435) = 1.3126$. Once again, this is a small value. To obtain the Lagrange multiplier statistic, we would compute

$$\begin{bmatrix} \partial \log L / \partial\alpha & \partial \log L / \partial\beta \end{bmatrix} \begin{bmatrix} -\partial^2 \log L / \partial\alpha^2 & -\partial^2 \log L / \partial\alpha\partial\beta \\ -\partial^2 \log L / \partial\alpha\partial\beta & -\partial^2 \log L / \partial\beta^2 \end{bmatrix}^{-1} \begin{bmatrix} \partial \log L / \partial\alpha \\ \partial \log L / \partial\beta \end{bmatrix}$$

at the restricted estimates of $\alpha = .88496$ and $\beta = 1$. Making the substitutions from above, at these values, we would have

$$\partial\log L/\partial\alpha = 0$$

$$\partial\log L/\partial\beta = n + \sum_{i=1}^{n} \log x_i - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i \log x_i = 9.400342$$

$$\partial^2\log L/\partial\alpha^2 = -n\bar{x}^{-2} = -25.54955$$

$$\partial^2\log L/\partial\beta^2 = -n - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i (\log x_i)^2 = -30.79486$$

$$\partial^2\log L/\partial\alpha\partial\beta = -\sum_{i=1}^{n} x_i \log x_i = -8.265.$$

The lower right element in the inverse matrix is .041477. The LM statistic is, therefore, $(9.40032)^2 .041477 = 2.9095$. This is also well under the critical value for the chi-squared distribution, so the hypothesis is not rejected on the basis of any of the three tests.

19. We consider forming a confidence interval for the variance of a normal distribution. As shown in Example 4.29, the interval is formed by finding c_{lower} and c_{upper} such that $\text{Prob}[c_{lower} < \chi^2[n-1] < c_{upper}] = 1 - \alpha$.

The endpoints of the confidence interval are then $(n\text{-}1)s^2/c_{upper}$ and $(n\text{-}1)s^2/c_{lower}$. How do we find the narrowest interval? Consider simply minimizing the width of the interval, $c_{upper} - c_{lower}$ subject to the constraint that the probability contained in the interval is $(1-\alpha)$. Prove that for symmetric and asymmetric distributions alike, the narrowest interval will be such that the density is the same at the two endpoints.

The general problem is to minimize Upper - Lower subject to the constraint F(Upper) - F(Lower) = 1 - α, where $F(.)$ is the appropriate chi-squared distribution. We can set this up as a Lagrangean problem,

$$\min_{L,U} L_* = U - L + \lambda\{(F(U) - F(L)) - (1 - \alpha)\}$$

The necessary conditions are

$$\partial L_*/\partial U = 1 + \lambda f(U) = 0$$

$$\partial L_*/\partial L = -1 - \lambda f(L) = 0$$

$$\partial L_*/\partial \lambda = (F(U) - F(L)) - (1 - \alpha) = 0$$

It is obvious from the first two that at the minimum, $f(U)$ must equal $f(L)$.

20. Using the results in Example 4.26, and Section 4.7.2, estimate the asymptotic covariance matrix of the method of moments estimators of P and λ based on m_{-1}' and m_2'. (**Note:** You will need to use the data in Table 4.1 to estimate **V**.)

Using the income data in Table 4.1, $(1/n)$ times the covariance matrix of $1/x_i$ and x_i^2 is $\mathbf{V} = \begin{bmatrix} .000068456 & -2.811 \\ -2.811 & 228050. \end{bmatrix}$. The moment equations used to estimate P and λ are $E[m_{-1}' - \lambda/(P-1)] = 0$ and $E[m_2' - P(P+1)/\lambda] = 0$. The matrix of derivatives with respect to P and λ is $\mathbf{G} = \begin{bmatrix} \lambda/(P-1)^2 & -\lambda/(P-1) \\ -(2P+1)/\lambda^2 & 2P(P+1)/\lambda^3 \end{bmatrix}$. The estimated asymptotic covariance matrix is

$$[\mathbf{G}\mathbf{V}^{-1}\mathbf{G}']^{-1} = \begin{bmatrix} .17532 & .0073617 \\ .0073617 & .00041871 \end{bmatrix}.$$

Appendix D

Large Sample Distribution Theory

There are no exercises for Appendix D.

Appendix E

Computation and Optimization

1. Show how to maximize the function

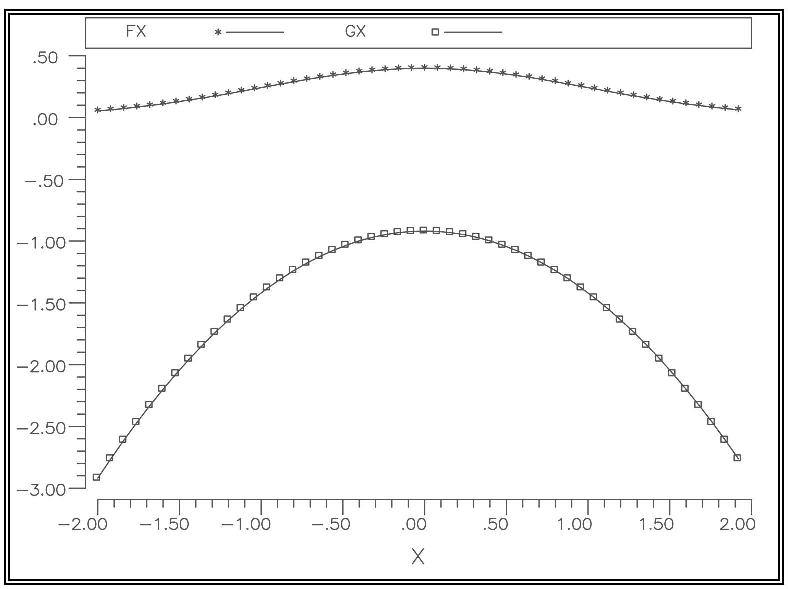
$$f(\beta) \; = \; \frac{1}{\sqrt{2\pi}} e^{-(\beta-c)^2/2}$$

with respect to β for a constant, *c*, using Newton's method. Show that maximizing log*f*(β) leads to the same solution. Plot *f*(β) and log*f*(β).

The necessary condition for maximizing *f*(β) is

$$df(\beta)/d\beta \; = \; \frac{1}{\sqrt{2\pi}} e^{-(\beta-c)^2/2} [-(\beta - c)] \; = \; 0 \; = \; -(\beta - c)\mathrm{f}(\beta).$$

The exponential function can never be zero, so the only solution to the necessary condition is β = *c*. The second derivative is $d^2f(\beta)/d\beta^2 = -(\beta-c)df(\beta)/d\beta - f(\beta) = [(\beta-c)^2 - 1]f(\beta)$. At the stationary value *b* = *c*, the second derivative is negative, so this is a maximum. Consider instead the function *g*(β) = log*f*(β) = $-(1/2)\ln(2\pi) - (1/2)(\beta - c)^2$. The leading constant is obviously irrelevant to the solution, and the quadratic is a negative number everywhere except the point β = *c*. Therefore, it is obvious that this function has the same maximizing value as *f*(β). Formally, $dg(\beta)/d\beta = -(\beta - c) = 0$ at β = *c*, and $d^2g(\beta)/d\beta^2 = -1$, so this is indeed the maximum. A sketch of the two functions appears below.



Note that the transformed function is concave everywhere while the original function has inflection points.

2. Prove that Newton's method for minimizing the sum of squared residuals in the linear regression model will converge to the minimum in one iteration.

The function to be maximized is $f(\beta) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$. The required derivatives are $\partial f(\beta)/\partial\beta = -\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)$ and $\partial^2 f(\beta)/\partial\beta\partial\beta\partial = \mathbf{X}'\mathbf{X}$. Now, consider beginning a Newton iteration at an arbitrary point, β^0. The iteration is defined in (12-17),
$\beta^1 = \beta^0 - (\mathbf{X}'\mathbf{X})^{-1}\{-\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta^0)\} = \beta^0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta^0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{b}$.
Therefore, regardless of the starting value chosen, the next value will be the least squares coefficient vector.

3. For the Poisson regression model, $\text{Prob}[Y_i = y_i|\mathbf{x}_i] = \dfrac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$ where $\lambda_i = e^{\beta'\mathbf{x}_i}$. The log-likelihood function is $\ln L = \sum_{i=1}^n \log\text{Prob}[Y_i = y_i|\mathbf{x}_i]$.

(a) Insert the expression for λ_i to obtain the log-likelihood function in terms of the observed data.
(b) Derive the first order conditions for maximizing this function with respect to β.
(c) Derive the second derivatives matrix of this criterion function with respect to β. Is this matrix negative definite?
(d) Define the computations for using Newton's method to obtain estimates of the unknown parameters.
(e) Write out the full set of steps in an algorithm for obtaining the estimates of the parameters of this model. Include in your algorithm a test for convergence of the estimates based on Belsley's suggested criterion.
(f) How would you obtain starting values for your iterations?
(g) The following data are generated by the Poisson regression model with $\log\lambda = \alpha + \beta x$.

y	6	7	4	10	10	6	4	7	2	3	6	5	3	3	4
x	1.5	1.8	1.8	2.0	1.3	1.6	1.2	1.9	1.8	1.0	1.4	.5	.8	1.1	.7

Use your results from parts (a) - (f) to compute the maximum likelihood estimates of α and β. Also obtain estimates of the asymptotic covariance matrix of your estimates.

The log-likelihood is

$$\log L = \sum_{i=1}^n [-\lambda_i + y_i\ln\lambda_i - \ln y_i!] = -\sum_{i=1}^n e^{\beta'\mathbf{x}_i} + \sum_{i=1}^n y_i(\beta'\mathbf{x}_i) - \sum_{i=1}^n \log y_i!$$
$$= -\sum_{i=1}^n e^{\beta'\mathbf{x}_i} + \beta'\sum_{i=1}^n \mathbf{x}_i y_i - \sum_{i=1}^n \log y_i!$$

The necessary condition is $M\ln L/M\beta = -\sum_{i=1}^n \mathbf{x}_i e^{\beta'\mathbf{x}_i} + \sum_{i=1}^n \mathbf{x}_i y_i = \mathbf{0}$ or $\mathbf{X}N\mathbf{y} = \sum_{i=1}^n \mathbf{x}_i\lambda_i$. It is useful to note, since $E[y_i*\mathbf{x}_i] = \lambda_i = e^{\beta N x_i}$, the first order condition is equivalent to $\sum_{i=1}^n \mathbf{x}_i y_i = \sum_{i=1}^n \mathbf{x}_i E[y_i*\mathbf{x}_i]$ or $\mathbf{X}N\mathbf{y} = \mathbf{X}NE[\mathbf{y}]$, which makes sense. We may write the first order condition as $M\ln L/M\beta = \sum_{i=1}^n \mathbf{x}_i(y_i - \lambda_i) = \mathbf{0}$
which is quite similar to the counterpart for the classical regression if we view $(y_i - \lambda_i) = (y_i - E[y_i*\mathbf{x}_i])$ as a residual. The second derivatives matrix is $\partial\ln L/\partial\beta\partial\beta' = -\sum_{i=1}^n (e^{\beta'\mathbf{x}_i})\mathbf{x}_i\mathbf{x}_i' = -\sum_{i=1}^n \lambda_i\mathbf{x}_i\mathbf{x}_i'$. This is a negative definite matrix. To prove this, note, first, that λ_i must always be positive. Then, let $\mathbf{\Omega}$ be a diagonal matrix whose *i*th diagonal element is $\sqrt{\lambda_i}$ and let $\mathbf{Z} = \mathbf{\Omega X}$. Then, $\partial\ln L/\partial\beta\partial\beta' = -\mathbf{Z}'\mathbf{Z}$ which is clearly negative definite. This implies that the log-likelihood function is globally concave and finding its maximum using NewtonNs method will be straightforward and reliable.

The iteration for NewtonNs method is defined in (5-17). We may apply it directly in this problem. The computations involved in using Newton's method to maximize ln*L* will be as follows:

(1) Obtain starting values for the parameters. Because the log-likelihood function is globally concave, it will usually not matter what values are used. Most applications simply use zero. One suggestion which does appear in the literature is $\beta^0 = \left[\sum_{i=1}^n q_i\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^n q_i\mathbf{x}_i y_i\right]$ where $q_i = \log(\max(1,y_i))$.

(2) The iteration is computed as $\hat{\beta}_{t+1} = \hat{\beta}_t + \left[\sum_{i=1}^n \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \left[\sum_{i=1}^n \mathbf{x}_i (y_i - \hat{\lambda}_i)\right]$.

(3) Each time we compute $\hat{\beta}_{t+1}$, we should check for convergence. Some possibilities are

(a) Gradient: Are the elements of $\partial \ln L/\partial \beta$ small?

(b) Change: Is $\hat{\beta}_{t+1} - \hat{\beta}_t$ small?

(c) Function rate of change: Check the size of

$$\delta_t = \left[\sum_{i=1}^n \mathbf{x}_i (y_i - \hat{\lambda}_i)\right]' \left[\sum_{i=1}^n \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} \left[\sum_{i=1}^n \mathbf{x}_i (y_i - \hat{\lambda}_i)\right]$$

before computing $\hat{\beta}_{t+1}$. This measure describes what will happen to the function at the next value of β. This is Belsley's criterion.

(4) When convergence has been achieved, the asymptotic covariance matrix for the estimates is estimated with the inverse matrix used in the iterations.

Using the data given in the problem, the results of the above computations are

Iter.	α	β	ln*L*	∂ln*L*/∂α	∂ln*L*/∂β	Change
0	0	0	-102.387	65.	95.1	296.261
1	1.37105	2.17816	-1442.38	-1636.25	-2788.5	1526.36
2	.619874	2.05865	-461.989	-581.966	-996.711	516.92
3	.210347	1.77914	-141.022	-195.953	-399.751	197.652
4	.351893	1.26291	-51.2989	-57.9294	-102.847	30.616
5	.824956	.698768	-33.5530	-12.8702	-23.1932	2.75855
6	1.05288	.453352	-32.0824	-1.28785	-2.29289	.032399
7	1.07777	.425239	-32.0660	-.016067	-.028454	.0000051
8	1.07808	.424890	-32.0660	0	0	0

At the final values, the negative inverse of the second derivatives matrix is

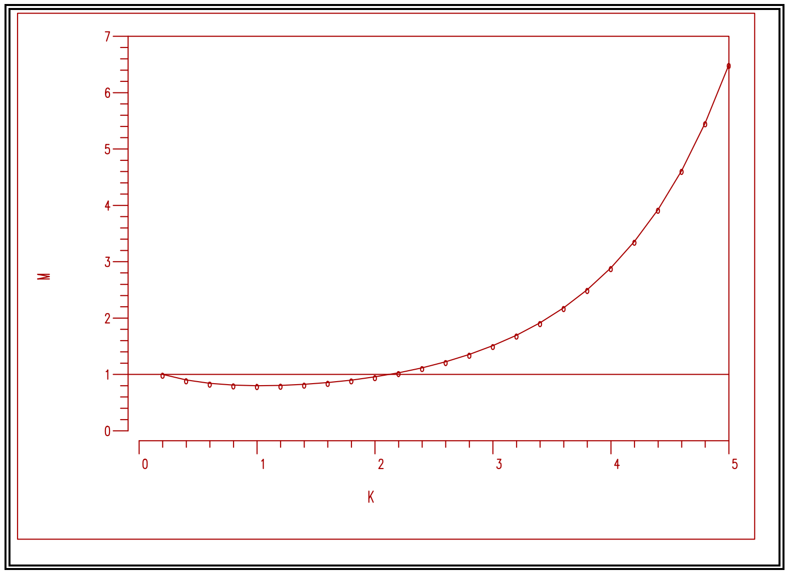
$$\left[\sum_{i=1}^n \hat{\lambda}_i \mathbf{x}_i \mathbf{x}_i'\right]^{-1} = \begin{bmatrix} .151044 & -.095961 \\ -.095961 & .0664665 \end{bmatrix}.$$

4. Use Monte Carlo Integration to plot the function $g(r) = E[x^r * x > 0]$ for the standard normal distribution.

The expected value from the truncated normal distribution is

$$E[x^r | x > 0] = \int_0^\infty x^r f(x|x>0)dx = \frac{\int_0^\infty x^r \phi(x)dx}{\int_0^\infty \phi(x)dx} = \frac{2}{\sqrt{\pi}} \int_0^\infty x^r e^{-\frac{x^2}{2}} dx.$$

To evaluate this expectation, we first sampled 1,000 observations from the truncated standard normal distribution using (5-1). For the standard normal distribution, μ = 0, σ = 1, $P_L = \Phi((0 - 0)/1) = 2$, and $P_U = \Phi((+4 - 0)/1) = 1$. Therefore, the draws are obtained by transforming draws from U(0,1) (denoted F_i) to $x_i = \Phi[2(1 + F_i)]$. Since $0 < F_i < 1$, the argument in brackets must be greater than 2, so $x_i > 0$, which is to be expected. Using the same 1,000 draws each time (so as to obtain smoothness in the figure), we then plot the values of $\bar{x}_r = \frac{1}{1000}\sum_{i=1}^{1000} x_i^r$, r = 0, .2, .4,.6, ..., 5.0. As an additional experiment, we generated a second sample of 1,000 by drawing observations from the standard normal distribution and discarding them and redrawing if they were not positive. The means and standard deviations of the two samples were (0.8097,0.6170) for the first and (0.8059,0.6170) for the second. Drawing the second sample takes approximately twice as long as the second. Why?



5. For the model in Example 5.10, derive the LM statistic for the test of the hypothesis that μ=0.

The derivatives of the log-likelihood with μ = 0 imposed are $g_\mu = n\bar{x}/\sigma^2$ and $g_{\sigma^2} = \dfrac{-n}{2\sigma^2} + \dfrac{\sum_{i=1}^{n} x_i^2}{2\sigma^4}$. The estimator for σ^2 will be obtained by equating the second of these to 0, which will give (of course), $v = \mathbf{x}'\mathbf{x}/n$. The terms in the Hessian are $H_{\mu\mu} = -n/\sigma^2$, $H_{\mu\sigma^2} = -n\bar{x}/\sigma^4$, and $H_{\sigma^2\sigma^2} = n/(2\sigma^4) - \mathbf{x}'\mathbf{x}/\sigma^6$. At the MLE, $g_{\sigma^2} = 0$, exactly. The off diagonal term in the expected Hessian is also zero. Therefore, the LM statistic is $LM = \begin{bmatrix} n\bar{x}/v & 0 \end{bmatrix} \begin{bmatrix} \dfrac{n}{v} & 0 \\ 0 & \dfrac{n}{2v^2} \end{bmatrix}^{-1} \begin{bmatrix} n\bar{x}/v \\ 0 \end{bmatrix} = \left[\dfrac{\bar{x}}{v/\sqrt{n}} \right]^2$.

This resembles the square of the standard *t*-ratio for testing the hypothesis that μ = 0. It would be exactly that save for the absence of a degrees of freedom correction in *v*. However, since we have not estimated μ with $\bar{x}$ in fact, LM *is* exactly the square of a standard normal variate divided by a chi-squared variate over its degrees of freedom. Thus, in this model, LM is exactly an *F* statistic with 1 degree of freedom in the numerator and *n* degrees of freedom in the denominator.

6. In Example 5.10, what is the concentrated over μ log likelihood function?

It is obvious that whatever solution is obtained for σ^2, the MLE for μ will be $\bar{x}$, so the concentrated log-likelihood function is $\log L_c = \dfrac{-n}{2}\left(\log 2\pi + \log\sigma^2\right) - \dfrac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

7. In Example E.13, suppose that $E[y_i] = \mu$, for a nonzero mean.

(a) Extend the model to include this new parameter. What are the new log likelihood, likelihood equation, Hessian, and expected Hessian?

(b) How are the iterations carried out to estimate the full set of parameters?

(c) Show how the *LIMDEP* program should be modified to include estimation of μ.

(d) Using the same data set, estimate the full set of parameters.

If y_i has a nonzero mean, μ, then the log-likelihood is

$$\ln L(\gamma,\mu|\mathbf{Z}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log\sigma_i^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{(y_i-\mu)^2}{\sigma_i^2}\right)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i'\gamma - \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\exp(-\mathbf{z}_i'\gamma).$$

The likelihood equations are

$$\frac{\partial \ln L}{\partial \gamma} = \frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\left(\frac{(y_i-\mu)^2}{\sigma_i^2}-1\right) = -\frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i + \frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\mathbf{z}_i\exp(-\mathbf{z}_i'\gamma)$$

$$= \mathbf{g}_\gamma(\gamma,\mu) = \mathbf{0}$$

and

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n}(y_i-\mu)\exp(-\mathbf{z}_i'\gamma) = g_\mu(\gamma,\mu) = 0.$$

The Hessian is

$$\frac{\partial^2 \ln L}{\partial\gamma\partial\gamma'} = -\frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'\left(\frac{(y_i-\mu)^2}{\sigma_i^2}\right) = -\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\mathbf{z}_i\mathbf{z}_i'\exp(-\mathbf{z}_i'\gamma) = \mathbf{H}_{\gamma\gamma}.$$

$$\frac{\partial^2 \ln L}{\partial\gamma\partial\mu} = -\sum_{i=1}^{n}\mathbf{z}_i(y_i-\mu)\exp(-\mathbf{z}_i'\gamma) = \mathbf{H}_{\gamma\mu}$$

$$\frac{\partial^2 \ln L}{\partial\mu\partial\mu} = -\sum_{i=1}^{n}\exp(-\mathbf{z}_i'\gamma) = \mathbf{H}_{\mu\mu}$$

The expectations in the Hessian are found as follows: Since $E[y_i] = \mu$, $E[\mathbf{H}_{\gamma\mu}] = \mathbf{0}$. There are no stochastic terms in $\mathbf{H}_{\mu\mu}$, so $E[\mathbf{H}_{\mu\mu}] = \mathbf{H}_{\mu\mu} = -\sum_{i=1}^{n}\frac{1}{\sigma_i^2}$. Finally, $E[(y_i - \mu)^2] = \sigma_i^2$, so $E[\mathbf{H}_{\gamma\gamma}] = -1/2(\mathbf{Z}'\mathbf{Z})$.

There is more than one way to estimate the parameters. As in Example 5.13, the method of scoring (using the expected Hessian) will be straightforward in principle - though in our example, it does not work well in practice, so we use Newton's method instead. The iteration, in which we use index 't' to indicate the estimate at iteration t, will be

$$\begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t+1)} = \begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t)} - E[\mathbf{H}(t)]^{-1}\,\mathbf{g}(t).$$

If we insert the expected Hessians and first derivatives in this iteration, we obtain

$$\begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t+1)} = \begin{bmatrix}\mu\\ \gamma\end{bmatrix}_{(t)} + \begin{bmatrix}\sum_{i=1}^{n}\frac{1}{\sigma_i^2(t)} & 0\\ 0 & \frac{1}{2}\mathbf{Z}'\mathbf{Z}\end{bmatrix}^{-1}\begin{bmatrix}\sum_{i=1}^{n}\frac{y_i-\mu(t)}{\sigma_i^2(t)}\\ \frac{1}{2}\sum_{i=1}^{n}\mathbf{z}_i\left(\frac{(y_i-\mu(t))^2}{\sigma_i^2(t)}-1\right)\end{bmatrix}.$$

The zero off diagonal elements in the expected Hessian make this convenient, as the iteration may be broken into two parts. We take the iteration for μ first. With current estimates $\mu(t)$ and $\gamma(t)$, the method of scoring produces this iteration: $\mu(t+1) = \mu(t) + \dfrac{\sum_{i=1}^{n}\frac{y_i-\mu(t)}{\sigma_i^2(t)}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2(t)}}$. As will be explored in Chapters 12 and 13, this is generalized least squares. Let $\mathbf{i}$ denote an $n\times 1$ vector of ones, let $e_i(t) = y_i - \mu(t)$ denote the 'residual' at iteration t and let $\mathbf{e}(t)$ denote the $n\times 1$ vector of residuals. Let $\mathbf{\Omega}(t)$ denote a diagonal matrix which has σ_i^2 on its diagonal (and zeros elsewhere). Then, the iteration for μ is

$\mu(t+1) = \mu(t) + [\mathbf{i}'\mathbf{\Omega}(t)^{-1}\mathbf{i}]^{-1}[\mathbf{i}'\mathbf{\Omega}(t)^{-1}\mathbf{e}(t)]$. This shows how to compute $\mu(t+1)$. The iteration for $\gamma(t+1)$ is exactly as was shown in Example 5.13, save for the single change that in the computation, y_i^2 is changed to $(y_i - \mu(t))^2$. Otherwise, the computation is identical. Thus, we would have

$\gamma(t+1) = \gamma(t) + (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{v}(\gamma(t),\mu(t))$, where $v_i(\gamma(t),\mu(t))$ is the term in parentheses in the iteration shown above. This shows how to compute $\gamma(t+1)$.

```
/*=================================================================
Program Code for Estimation of Harvey's Model
The data set for this model is 100 observations from Greene (1992)
Variables are: Y  = Average monthly credit card expenditure
               Q1 = Age in years+ 12ths of a year
               Q2 = Income, divided by 10,000
               Q3 = OwnRent; individual owns (1) or rents (0) home
               Q4 = Self employed (1=yes, 0=no)
Read     ; Nobs = 200 ; Nvar = 6 ; Names = y,q1,q2,q3,q4
         ; file=d:\DataSets\A5-1.dat$
Namelist ; Z = One,q1,q2,q3,q4 $
=================================================================
Step 1 is to get the starting values and set some values for the
iterations- iter=iteration counter, delta=value for convergence.
*/
Create   ; y0 = y – Xbr(y) ; ui = log(y0^2) $
Matrix   ; gamma0 = <Z'Z> * Z'ui ; EH = 2*<Z'Z> $
Calc     ; c0 = gamma0(1)+1.2704     ? Correction to start value
         ; s20 = y0'y0/n ; delta = 1 ; iter=0 $
Create   ; vi0 = y0^2 / s20 - 1 $  (Used in LM statistic)
? Correct first element in gamma, then set starting vector.
Matrix   ; Gamma0(1) = c0  ; Gamma = Gamma0 $ Start value for gamma
Calc     ; mu0 = Xbr(y); mu = mu0$            Start value for mu
Procedure ----------[This does the iterations]--------------------
Create   ; vari = exp(Z'Gamma)  ;  ei = y-mu  ; varinv=1/vari
         ; hi   = ei^2 / vari
         ; gigamma = .5*(hi - 1); gimu = ei/vari
         ; logli = -.5*(log(2*pi) + log(vari) + hi) $
Matrix   ; ggamma = Z'gigamma ; gmu= 1’gimu
         ; H = 2*<Z’[hi]Z> ; gupdate = H*ggamma
? scoring, update = EH*ggamma
         ; Gamma = Gamma + gupdate $
Calc     ; muupdate = Sum(gimu)/Sum(varinv) ; mu = mu + muupdate $
Matrix   ; update = [gupdate/muupdate] ; g = [ggamma/gmu] $
Calc     ; list ; Iter = Iter+1 ; LogLU = Sum(logli);delta=g'update$
EndProcedure
Execute  ; While  delta > .00001 $ ------------------------------
Matrix   ; Stat (Gamma,H) $
Calc     ; list ; mu ; vmu = 1/Sum(varinv)   ; tmu = mu/Sqr(Vmu) $
Calc     ; list ; Sigmasq = Exp(Gamma(1))    ; K = Col(Z)
                ; SE = Sigmasq * Sqr(H(1,1)) ; TRSE = Sigmasq/SE
                ; LogLR = -n/2*(1 + log(2*pi)+ log(s20))
                ; LRTest = -2*(LogLR - LogLU) $
Matrix          ; Alpha = Gamma(2:K) ; VAlpha = Part(H,2,K,2,K)
         ; list ; WaldTest = Alpha ' <VAlpha> Alpha
                ; LMTest = .5* vi0'Z * <Z'Z> * Z'vi0
                ; EH ; H ; VB = BHHH(Z,gi) ; <VB> $
```

In the Example in the text, μ was constrained to equal $\bar{y}$. In the program, μ is allowed to be a free parameter. The comparison of the two sets of results appears below.

Iteration	(Constrained model, $\mu = \bar{y}$) log likelihood	δ	(Unconstrained model) log-l;ikelihood	δ
1	-698.3888	19.7022	-692.2987	22.8406
2	-692.2986	4.5494	-683.2320	6.9005
3	-689.7029	0.406881	-680.7028	2.7494
4	-689.4980	0.01148798	-679,7461	0.63453
5	-689.4741	0.0000125995	-679.4856	0.27023
6	-689.47407	0.000000000016	-679.4856	0.08124
			-679.4648	0.03079
			-679.4568	0.0101793
			-679.4542	0.00364255
			-679.4533	0.001240906
			-679.4530	0.00043431
			-679.4529	0.0001494193
			-679.4528	0.00005188501
			-679.4528	0.00001790973
			-679.4528	0.00000620193

Estimated Paramaters

Variable	Estimate	Std Error	t-ratio	Estimate	Std Error	t-ratio
Age	0.013042	0.02310	0.565	-0.0134	0.0244	-0.550
Income	0.6432	0.120001	5.360	0.9953	0.1375	7.236
Ownrent	-0.2159	0.3073	-0.703	0.0774	0.3004	0.258
SelfEmployed	-0.4273	0.6677	-0.640	-1.3117	0.6719	-1.952
γ_1	8.465			7.867		
σ^2	4,745.92			2609.72		
μ	189.02	fixed		91.874	15.247	6.026

Tests of the joint hypothesis that all slope coefficients are zero:

LW	40.716	60.759
Wald:	39.024	69.515
LM	35.115	35.115 (same by construction).