Chapter 26: Exercises and Complements

Exercise 1: Covariance Matrix

If x is an n-dimensional column vector of random variables with mean vector μ, prove that the covariance matrix of x, $\sum = E[(x-\mu)(x-\mu)^T]$ is positive semidefinite, by using the fact that the scalar $c^T x$ has a non-negative variance. If $\sum$ is not positive definite, what does this imply about x?

Solution 1:

Let c be any non-zero $n \times 1$ vector. Define $Y = c^T x$. The mean of Y is:

$$E[Y] = E[c^T x] = c^T E[x] = c^T \mu.$$

The variance of Y is:

$$\mathrm{var}(Y) = E[(Y - E[Y])^2] = E[(c^T x - c^T \mu)(c^T x - c^T \mu)^T].$$

Since Y is a scalar, $(Y - E[Y])^2 = (Y - E[Y])(Y - E[Y]) = (Y - E[Y])(Y - E[Y])^T$. Thus,

$$\mathrm{var}(Y) = E[(c^T x - c^T \mu)(c^T x - c^T \mu)^T] \qquad = E[c^T (x-\mu)(x-\mu)^T c] \qquad = c^T E[(x-\mu)(x-\mu)^T] c \qquad = c^T \sum c.$$

Since the variance of any random variable is non-negative, $\mathrm{var}(Y) \geq 0$. Therefore, $c^T \sum c \geq 0$ for any non-zero vector c. This means that $\sum$ is **positive semidefinite**.

If $\sum$ is not **positive definite**, it means that $\sum$ is **positive semidefinite** but not **positive definite**. This occurs when there exists some non-zero vector c such that $c^T \sum c = 0$. This implies that $\mathrm{var}(Y) = \mathrm{var}(c^T x) = 0$. If a random variable has zero variance it must be a constant. In this instance, $c^T x = c^T \mu$, which is a constant. In turn, $c^T x$ is a linear combination of elements of x that is constant, indicating that there exists **linear dependence** among the elements of the random vector x.

Exercise 2: Matrix/vector notation

Write the following expressions in matrix/vector notation:

a. $\sum_{i=1}^n a_i b_i$;

b. $a_i b_j, i = 1, \ldots, n; j = 1, \ldots, J$

c. $\sum_{j=1}^J a_{ij} x_j, i = 1, \ldots, n$;

d. $\sum_{j=1}^n a_{ij} b_{jk}, i = 1, \ldots, n; k = 1, \ldots, K$;

e. $\sum_{j=1}^K \sum_{k=1}^K a_{ij} b_{kj} c_{km}, i = 1, \ldots, n; m = 1, \ldots, M$;

f. $\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$.

Solution 2:

a. Let $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$. Then $\sum_{i=1}^n a_i b_i = a^T b$.

b. Let $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ \vdots \\ b_J \end{bmatrix}$. Then $ab^T = \begin{bmatrix} a_1 b_1 & \ldots & a_1 b_J \\ \vdots & \ddots & \vdots \\ a_n b_1 & \ldots & a_n b_J \end{bmatrix}$, which gives $a_i b_j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, J$.

c. Let $A = [a_{ij}]$ be an $n \times J$ matrix and $x = \begin{bmatrix} x_1 \\ \vdots \\ x_J \end{bmatrix}$. Then Ax is an $n \times 1$ vector. The i-th element of Ax is $\sum_{j=1}^J a_{ij} x_j$.

d. Let $A = [a_{ij}]$ be an $n \times n$ matrix and $B = [b_{jk}]$ be an $n \times K$ matrix. Then, the matrix product AB is an $n \times K$ matrix whose (i,k)-th element is $\sum_{j=1}^n a_{ij} b_{jk}$.

e. Let $A = [a_{ij}]$ be an $n \times K$ matrix, $B = [b_{jk}]$ be a $K \times K$ matrix, and $C = [c_{km}]$ be a $K \times M$ matrix. The product ABC is an $n \times M$ matrix. The (i,m)-th element of ABC is given by $\sum_{j=1}^K \sum_{k=1}^K a_{ij} b_{jk} c_{km}$. Specifically, let us calculate the matrix multiplication step by step. Let $D = AB$, D

will be an $n \times K$ matrix. $d_{ik} = \sum_{j=1}^{K} a_{ij} b_{jk}$. Now let us calculate DC, and call this matrix E. $E = DC$ is an $n \times M$ matrix. $e_{im} = \sum_{k=1}^{K} d_{ik} c_{km} = \sum_{k=1}^{K} (\sum_{j=1}^{K} a_{ij} b_{jk}) c_{km} = \sum_{j=1}^{K} \sum_{k=1}^{K} a_{ij} b_{jk} c_{km}$.

f. Let $A = [a_{ij}]$ be an $n \times n$ matrix and $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$. The given expression can be written as $x^T A x$.

Exercise 3: Trace of matrix products

Prove that if A and B are matrices such that AB and BA both exist, then AB and BA have the same sum of diagonal elements, i.e. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. Extend the result to show that

$$\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA),$$

provided the matrices are conformable. Show that, however, $\operatorname{tr}(BAC)$ and $\operatorname{tr}(ACB)$ may be different.

Solution 3:

Suppose A is $m \times n$ and B is $n \times m$, so that AB is $m \times m$ and BA is $n \times n$. Then,

$$\operatorname{tr}(AB) = \sum_{i=1}^{m} (AB)_{ii} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji} \qquad \operatorname{tr}(BA) = \sum_{j=1}^{n} (BA)_{jj} = \sum_{j=1}^{n} \sum_{i=1}^{m} b_{ji} a_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ji}.$$

Thus, $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Now suppose that A is $n \times p$, B is $p \times q$, and C is $q \times n$, so that ABC is $n \times n$. Then we have

$$\operatorname{tr}(ABC) = \operatorname{tr}((AB)C) = \operatorname{tr}(C(AB)) = \operatorname{tr}((CAB)),$$

and,

$$\operatorname{tr}(ABC) = \operatorname{tr}(A(BC)) = \operatorname{tr}((BC)A) = \operatorname{tr}(BCA).$$

For the second part, let

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Then

$$BAC = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \operatorname{tr}(BAC) = 1 \qquad ACB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \operatorname{tr}(ACB) = 0.$$

Exercise 4: Linear System

Suppose that A is $n \times n$, B is $K \times K$, and C is $n \times K$ are given matrices. Consider the system of equations in $n \times K$ matrices X:

$$AXB = C. \text{ (26.1)}$$

Show that this system of equations is linear and solve for X.

Solution 4:

This equation is linear in X, because for any X, X^* of same dimensions

$$AXB + AX^*B = A(X + X^*)B$$

If A, B are invertible, then we have a unique solution

$$X = A^{-1} C B^{-1}.$$

Example: Special Case

Special case $K = 1$. We have $axb = c$, which implies that $x = c/ab$. Question: Suppose that $K = n$.

$$AX^{-1}B = C$$

then solve for X.

Solution: Special Case

We pre-multiply by A^{-1} and post-multiply by B^{-1} to get $X^{-1} = A^{-1}CB^{-1}$. Taking the inverse again yields, $X = (A^{-1}CB^{-1})^{-1} = BC^{-1}A$ (assuming all matrices are square and invertible).

Exercise 5: Properties of X^TX

If X is a non-zero matrix of order $T \times K$, prove that X^TX is:

a. symmetric and

b. positive semi-definite;

c. Under what conditions on X is X^TX positive definite.

Solution 5:

a. $(X^TX)^T = X^T(X^T)^T = X^TX$. Therefore, X^TX is **symmetric**.

b. Let z be any non-zero $K \times 1$ vector. Then $z^TX^TXz = (Xz)^TXz = y^Ty$, where $y = Xz$. Since y^Ty is a sum of squares, $y^Ty \geq 0$. Therefore, $z^TX^TXz \geq 0$ for any non-zero vector z, implying that X^TX is **positive semi-definite**.

c. For X^TX to be **positive definite**, we need $z^TX^TXz > 0$ for any non-zero $K \times 1$ vector z. This condition is equivalent to $Xz \neq 0$ for any non-zero z. In turn, this is true if and only if X has **full column rank**, meaning the columns of X are linearly independent. If the columns of X are linearly independent, then $Xz \neq 0$, so X^TX is positive definite.

Exercise 6: Properties of Symmetric Matrix

For a real symmetric $n \times n$ matrix A, prove:

a. $\det(A) = \prod_{j=1}^n \lambda_j$, where λ_j are the eigenvalues of A;

b. $\mathrm{tr}(A) = \sum_{j=1}^n \lambda_j$, where λ_j are the eigenvalues of A;

c. A is positive definite, if and only if its eigenvalues are positive;

d. A is positive definite, if and only if A^{-1} is positive definite;

e. If A is positive definite, then there exists an $n \times n$ matrix P such that $A = PP^T$.

Solution 6:

Since A is a real symmetric matrix, it can be diagonalized as $A = Q\Lambda Q^T$, where Q is an orthogonal matrix ($Q^TQ = QQ^T = I$) containing the eigenvectors of A, and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix containing the corresponding eigenvalues.

a. $\det(A) = \det(Q\Lambda Q^T) = \det(Q)\det(\Lambda)\det(Q^T) = \det(QQ^T)\det(\Lambda) = \det(I)\det(\Lambda) = 1 \cdot \prod_{j=1}^n \lambda_j = \prod_{j=1}^n \lambda_j$.

b. $\mathrm{tr}(A) = \mathrm{tr}(Q\Lambda Q^T) = \mathrm{tr}(\Lambda Q^TQ) = \mathrm{tr}(\Lambda I) = \mathrm{tr}(\Lambda) = \sum_{j=1}^n \lambda_j$.

c. If A is positive definite, then for any non-zero vector x, $x^TAx > 0$. Let x be an eigenvector of A, say q_i. Then $Aq_i = \lambda_i q_i$, and $q_i^TAq_i = q_i^T\lambda_i q_i = \lambda_i q_i^Tq_i = \lambda_i > 0$, since $q_i^Tq_i = 1$. Therefore, all eigenvalues are positive. Conversely, if all eigenvalues are positive, then for any non-zero x, we can write $x = Qz$ for some $z \neq 0$, since Q is invertible, and so $x^TAx = z^TQ^TQ\Lambda Q^TQz = z^T\Lambda z = \sum_i \lambda_i z_i^2 > 0$. Hence A is positive definite.

d. If A is positive definite, then its eigenvalues are all positive. The eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A, i.e. $1/\lambda_i$. Thus, if $\lambda_i > 0$, then $1/\lambda_i > 0$, so all eigenvalues of A^{-1} are positive, meaning A^{-1} is positive definite. Conversely, if A^{-1} is positive definite, all its eigenvalues are positive. The eigenvalues of A are the reciprocals of the eigenvalues of A^{-1}. Thus, the eigenvalues of A are all positive, and A is positive definite.

e. Since A is positive definite, all of its eigenvalues are positive. We can define $\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$. Then we can write $A = Q\Lambda Q^T = Q\Lambda^{1/2}\Lambda^{1/2}Q^T = Q\Lambda^{1/2}(Q\Lambda^{1/2})^T$. Let $P = Q\Lambda^{1/2}$. Then, $A = PP^T$.

Exercise 7: Idempotent Matrix

A square matrix A is **idempotent** if $A = A^2$. Prove that:

a. The eigenvalues of A are either zero or one;

b. $\mathrm{rank}(A) = \mathrm{tr}(A)$.

Solution 7:

a. Let λ be an eigenvalue of A, and x be the corresponding eigenvector. Then, $Ax = \lambda x$. Since A is idempotent, $A^2 = A$. Multiplying both sides by x, $A^2 x = Ax$. Substituting $Ax = \lambda x$ gives, $A(Ax) = Ax$, $A(\lambda x) = \lambda x$, $\lambda Ax = \lambda x$, $\lambda(\lambda x) = \lambda x$, $\lambda^2 x = \lambda x$. Since x is an eigenvector it cannot be the 0 vector, so we can divide both sides by x to obtain, $(\lambda^2 - \lambda)x = 0 \implies \lambda^2 - \lambda = 0$. This means that $\lambda(\lambda - 1) = 0$, so $\lambda = 0$ or $\lambda = 1$.

b. Since A is idempotent, and therefore symmetric, it can be diagonalized as $A = Q\Lambda Q^T$, where Q is an orthogonal matrix and Λ is a diagonal matrix of eigenvalues. Each eigenvalue is either 0 or 1. Then $\text{rank}(A) = \text{rank}(\Lambda)$, and the rank of Λ is equal to the number of non-zero eigenvalues, which is the number of eigenvalues equal to 1. Also, $\text{tr}(A) = \text{tr}(Q\Lambda Q^T) = \text{tr}(\Lambda Q^T Q) = \text{tr}(\Lambda)$. The trace of Λ is equal to the sum of its diagonal elements, which are the eigenvalues. Since each eigenvalue is either 0 or 1, the trace is the number of eigenvalues equal to 1. Therefore, $\text{rank}(A) = \text{tr}(A)$.

Exercise 8: Quadratic Equation with Matrices

Suppose that X is a real symmetric matrix. Calculate the eigenvectors and eigenvalues of the matrix X^2 in terms of the eigenvalues and eigenvectors of the matrix X. Now consider the matrix quadratic equation

$$2X^2 - 3X + I_n = 0_n,$$

where $X, I_n, 0_n$ are $n \times n$ matrices. Find real valued matrix solutions to this equation, i.e., find the X that solves this equation.

Solution 8:

If X is symmetric, $X = U\Lambda U^T$, where U is an orthogonal matrix whose columns are eigenvectors of X, and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ contains corresponding eigenvalues. Then $X^2 = U\Lambda U^T U\Lambda U^T = U\Lambda^2 U^T$, where $\Lambda^2 = \text{diag}(\lambda_1^2, \dots, \lambda_n^2)$. Thus, the eigenvalues of X^2 are the squares of the eigenvalues of X, and the eigenvectors are the same.

If X is symmetric

$$X = U\Lambda U^T,$$

where $UU^T = I$. It follows that

$$2U\Lambda^2 U^T - 3U\Lambda U^T + I_n = 0_n.$$

Collecting terms we have

$$U \begin{bmatrix} 2\lambda_1^2 - 3\lambda_1 + 1 & & 0 \\ & \ddots & \\ 0 & & 2\lambda_n^2 - 3\lambda_n + 1 \end{bmatrix} U^T = 0_n.$$

Consider the quadratic equation

$$2x^2 - 3x + 1 = 0$$

This can be factorized as

$$(2x - 1)(x - 1) = 0$$

so that the set of solutions includes all matrices of the form

$$\{X = U\Lambda U^T : \lambda_i \in \{1, 1/2\}, i = 1, \dots, n, \text{where } U \text{ is any orthonormal matrix}\}.$$

Exercise 9: Characteristic Polynomial

Suppose that the matrix A has characteristic polynomial

$$\lambda^2 - a\lambda - b = 0.$$

Using the Cayley-Hamilton theorem, show that

$$A^{-1} = \tfrac{1}{b}(A - aI).$$

Solution 9:

The Cayley-Hamilton theorem states that a matrix satisfies its own characteristic equation. Therefore, $A^2 - aA - bI = 0$. Rearranging gives $A^2 - aA = bI$, which can be factored as $A(A - aI) = bI$. Dividing both sides by b (assuming $b \neq 0$), we get $A(\frac{1}{b}(A - aI)) = I$. Thus, $A^{-1} = \frac{1}{b}(A - aI)$. Note that for A^{-1} to exist, the characteristic polynomial must have a non-zero constant term.

Exercise 10: Idempotent Matrix

Show that the matrix $I - Z(Z^T Z)^{-1} Z^T$, where Z is any $n \times K$ matrix of full rank K, is idempotent and find its rank.

Solution 10:

Let $M = I - Z(Z^TZ)^{-1}Z^T$. Then

$$M^2 = (I - Z(Z^TZ)^{-1}Z^T)(I - Z(Z^TZ)^{-1}Z^T) \qquad = I - 2Z(Z^TZ)^{-1}Z^T + Z(Z^TZ)^{-1}Z^TZ(Z^TZ)^{-1}Z^T$$
$$= I - 2Z(Z^TZ)^{-1}Z^T + Z(Z^TZ)^{-1}Z^T \qquad = I - Z(Z^TZ)^{-1}Z^T \qquad = M.$$

Therefore, M is idempotent. The rank of an idempotent matrix is equal to its trace. So

$$\text{rank}(M) = \text{tr}(M) = \text{tr}(I_n - Z(Z^TZ)^{-1}Z^T) \qquad = \text{tr}(I_n) - \text{tr}(Z(Z^TZ)^{-1}Z^T) \qquad = \text{tr}(I_n) - \text{tr}((Z^TZ)^{-1}Z^TZ) \qquad = \text{tr}(I_n) - \text{tr}(I_K)$$
$$= n - K.$$

Exercise 11: Cross-product Matrix

An $n \times K$ matrix X is partitioned column-wise

$$X = (X_1, X_2),$$

where X_1 is $n \times K_1$ and X_2 is $n \times K_2$ where $K_1 + K_2 = K$. Write the cross-product matrix X^TX in terms of X_1 and X_2. An $n \times K$ matrix W is partitioned column-wise

$$W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix},$$

where W_1 is $n_1 \times K$ and W_2 is $n_2 \times K$ where $n_1 + n_2 = n$. Write the cross-product matrix W^TW in terms of W_1 and W_2.

Solution 11:

The cross-product matrix X^TX is

$$X^TX = \begin{pmatrix} X_1^T \\ X_2^T \end{pmatrix} \begin{pmatrix} X_1 & X_2 \end{pmatrix} = \begin{pmatrix} X_1^TX_1 & X_1^TX_2 \\ X_2^TX_1 & X_2^TX_2 \end{pmatrix}.$$

The cross-product matrix W^TW is

$$W^TW = \begin{pmatrix} W_1^T & W_2^T \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = W_1^TW_1 + W_2^TW_2.$$

Exercise 12: Distribution of Quadratic Form

If u is an n-dimensional vector of random variables distributed as $N(0, \sigma^2 I_n)$, and A is a symmetric, idempotent matrix of order n and rank p, show that u^TAu/σ^2 is distributed as χ^2 with p degrees of freedom.

Solution 12:

Since A is symmetric and idempotent, and $u \sim N(0, \sigma^2 I_n)$, the quadratic form u^TAu/σ^2 is distributed as a chi-squared distribution with degrees of freedom equal to the rank of A. Given that A has rank p, we have $u^TAu/\sigma^2 \sim \chi^2(p)$.

To see why, we know that $A = A^T$ and $A = A^2$ so we can diagonalize $A = Q\Lambda Q^T$, where Q is orthonormal and Λ is diagonal of eigenvalues. Then,

$$u^TAu = u^TQ\Lambda Q^Tu = (Q^Tu)^T\Lambda(Q^Tu)$$

and since Q^T is also orthonormal,

$$Q^Tu \sim N(0, \sigma^2 I).$$

If we call $v = Q^Tu$, and knowing that the eigenvalues of A are just 0 and 1, the quadratic form becomes

$$u^TAu = v^T\Lambda v = \sum_{i=1}^n \lambda_i v_i^2 = \sum_{i=1}^p v_i^2$$

if A has p eigenvalues equal to one, and the rest equal to zero. If u follows a normal distribution, so does v. Hence each v_i will be a standard normal. Hence we have the sum of squares of p standard normal random variables. If we divide by σ^2 we obtain a χ^2 distribution with p degrees of freedom.

Exercise 13: Distribution of Quadratic Form

If u is an n-dimensional vector of random variables distributed as $N(0, \sum)$, where $\sum$ is non-singular, show that $u^T \sum^{-1} u$ is distributed as χ^2 with n degrees of freedom.

Solution 13:

Since $\sum$ is a symmetric, positive definite matrix, we can write $\sum = PP^T$, where P is non-singular. Then $\sum^{-1} = (P^{-1})^T P^{-1}$. Let $v = P^{-1}u$. Then $v \sim N(0, P^{-1}\sum(P^{-1})^T) = N(0, P^{-1}PP^T(P^{-1})^T) = N(0, I_n)$. Thus,

$$u^T \sum^{-1} u = u^T (P^{-1})^T P^{-1} u = (P^{-1}u)^T (P^{-1}u) = v^T v = \sum_{i=1}^n v_i^2.$$

Since v_i are i.i.d. standard normal random variables, $v^T v$ follows a χ^2 distribution with n degrees of freedom.

Exercise 14: Inverse of Matrix

Consider the matrix

$$A = \text{diag}(s) - ss^T,$$

where $s = (s_1, \ldots, s_n)^T$ is such that $i^T s \neq 0$ and $s_i \neq 0$ for $i = 1, \ldots, n$. Is this matrix symmetric? Show that its inverse is

$$A^{-1} = \text{diag}(1/s_1, \ldots, 1/s_n) + \frac{1}{1 - i^T s} ii^T.$$

Solution 14:

Yes, the matrix is symmetric because $A^T = (\text{diag}(s) - ss^T)^T = \text{diag}(s)^T - (ss^T)^T = \text{diag}(s) - ss^T = A$. Let $B = \text{diag}(1/s_1, \ldots, 1/s_n) + \frac{1}{1-i^T s} ii^T = \text{diag}(1/s) + \frac{1}{1-i^T s} ii^T$, $A = \text{diag}(s) - ss^T$, and $i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

We want to show that $AB = I$.

$$AB = (\text{diag}(s) - ss^T)(\text{diag}(1/s) + \frac{1}{1-i^T s} ii^T) \quad = \text{diag}(s)\text{diag}(1/s) + \frac{1}{1-i^T s}\text{diag}(s) ii^T - ss^T \text{diag}(1/s) - \frac{1}{1-i^T s} ss^T ii^T$$
$$= I + \frac{1}{1-i^T s} si^T - si^T - \frac{i^T s}{1-i^T s} ss^T \quad = I + \frac{1}{1-i^T s}(si^T - si^T(1 - i^T s) - (i^T s) ss^T) \quad = I + \frac{1}{1-i^T s}(si^T - si^T + s(i^T s) i^T - (i^T s) ss^T)$$
$$= I + \frac{1}{1-i^T s}(s(i^T s) i^T - (i^T s) ss^T) \quad = I + \frac{i^T s}{1-i^T s}(si^T - ss^T) = I.$$

Exercise 15: Orthonormal Matrix

Verify that the matrix

$$U = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

is orthonormal for any $\theta \in \mathbb{R}$.

Solution 15:

U is orthonormal if $U^T U = UU^T = I$.

$$U^T U = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & -\cos(\theta)\sin(\theta) + \sin(\theta)\cos(\theta) \\ -\sin(\theta)\cos(\theta) + \cos(\theta)\sin(\theta) & \sin^2(\theta) + \cos^2(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

$$UU^T = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & \cos(\theta)\sin(\theta) - \sin(\theta)\cos(\theta) \\ \sin(\theta)\cos(\theta) - \cos(\theta)\sin(\theta) & \sin^2(\theta) + \cos^2(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

Exercise 16: Skew Symmetric Matrix

For any skew symmetric matrix A (such that $A^T = -A$), let

$$U = (I + A)(I - A)^{-1}.$$

Show that U is orthonormal.

Solution 16:

We need to show that $U^T U = I$. Since $A^T = -A$,

$$U^T = ((I - A)^{-1})^T (I + A)^T = ((I - A)^T)^{-1}(I + A)^T = (I - A^T)^{-1}(I + A^T) = (I + A)^{-1}(I - A).$$ Then,

$$U^T U = (I + A)^{-1}(I - A)(I + A)(I - A)^{-1}.$$

Note that $(I - A)(I + A) = I - A^2$ and $(I + A)(I - A) = I - A^2$, so $(I - A)(I + A) = (I + A)(I - A)$.

Therefore,

$$U^T U = (I+A)^{-1}(I+A)(I-A)(I-A)^{-1} = I.$$

Exercise 17: Conditional Distribution

Suppose that $Y \in \mathbb{R}^{d_y}, X \in \mathbb{R}^{d_x}$ with

$$X \sim N(\mu_x, \sum_x) \qquad Y|X = x \sim N(a + Bx, \sum_y),$$

where B is $d_y \times d_x$. Prove that

$$X|Y = y \sim N(\mu_{x|y}, \sum_{x|y}) \ (26.2)$$

$$\mu_{x|y} = \sum_{x|y}(B^T \sum_y^{-1}(y-a) + \sum_x^{-1} \mu_x) \qquad \sum_{x|y} = (\sum_x^{-1} + B^T \sum_y^{-1} B)^{-1}.$$

Solution 17:

This result can be derived from the properties of multivariate normal distributions. The joint distribution of X and Y can be written as:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ a + B\mu_x \end{pmatrix}, \begin{pmatrix} \sum_x & \sum_{xy} \\ \sum_{yx} & \sum_y + B \sum_x B^T \end{pmatrix}\right),$$

where $\sum_{xy} = \text{Cov}(X,Y) = \sum_x B^T$ and $\sum_{yx} = \text{Cov}(Y,X) = B\sum_x$. From properties of the multivariate normal distribution, the conditional distribution of $X|Y = y$ is also normal, with

$$\mu_{x|y} = \mu_x + \sum_{xy} \sum_{yy}^{-1}(y - \mu_y) = \mu_x + \sum_x B^T(\sum_y + B\sum_x B^T)^{-1}(y - a - B\mu_x),$$

and

$$\sum_{x|y} = \sum_x - \sum_{xy} \sum_{yy}^{-1} \sum_{yx} = \sum_x - \sum_x B^T(\sum_y + B\sum_x B^T)^{-1} B \sum_x.$$

We use the matrix identity $(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1}$ with $A = \sum_y, U = B, B = \sum_x, V = B^T$, which yields,

$$(\sum_y + B\sum_x B^T)^{-1} = \sum_y^{-1} - \sum_y^{-1} B(\sum_x^{-1} + B^T \sum_y^{-1} B)^{-1} B^T \sum_y^{-1}.$$

Substituting into the expression of $\sum_{x|y}$,

$$\sum_{x|y} = \sum_x - \sum_x B^T(\sum_y^{-1} - \sum_y^{-1} B(\sum_x^{-1} + B^T \sum_y^{-1} B)^{-1} B^T \sum_y^{-1}) B \sum_x$$
$$= \sum_x - \sum_x B^T \sum_y^{-1} B \sum_x + \sum_x B^T \sum_y^{-1} B(\sum_x^{-1} + B^T \sum_y^{-1} B)^{-1} B^T \sum_y^{-1} B \sum_x.$$

Let $D = (\sum_x^{-1} + B^T \sum_y^{-1} B)$. Thus,

$$\sum_{x|y} = \sum_x - \sum_x B^T \sum_y^{-1} B \sum_x + \sum_x B^T \sum_y^{-1} BD^{-1} B^T \sum_y^{-1} B \sum_x. \qquad = \sum_x - \sum_x B^T \sum_y^{-1} B(\sum_x - D^{-1} B^T \sum_y^{-1} B \sum_x)$$
$$= \sum_x - \sum_x B^T \sum_y^{-1} B(\sum_x - (\sum_x^{-1} + B^T \sum_y^{-1} B)^{-1} B^T \sum_y^{-1} B \sum_x)$$
$$= \sum_x - \sum_x B^T \sum_y^{-1} B(\sum_x - (I + \sum_x B^T \sum_y^{-1} B)^{-1} \sum_x B^T \sum_y^{-1} B \sum_x)$$ After many manipulations one can establish that

$$\sum_{x|y} = (\sum_x^{-1} + B^T \sum_y^{-1} B)^{-1}.$$

Substituting in the expression of $\mu_{x|y}$, and using the simplified version of $\sum_{x|y}$ leads to the stated result.

Exercise 18: Regression

Suppose that

$$y_i = \beta x_i + \epsilon_i,$$

where (x_i, ϵ_i) are i.i.d. with ϵ_i independent of x_i with mean zero but with unknown distribution. Suppose that $x_i \sim N(0,1)$. Then claim that

$$t = \frac{\hat{\beta}}{s(\sum_{i=1}^n x_i^2)^{-1/2}} \sim t_{n-1},$$

where s is the usual standard error. That is, the exact t-test is valid as before even though the distribution of ϵ could be anything.

Solution 18:

We have

$$\hat{\beta} - \beta_0 = \frac{\sum_{i=1}^n x_i \epsilon_i}{\sum_{i=1}^n x_i^2} \qquad t = \frac{\sum_{i=1}^n x_i \epsilon_i}{\sqrt{\sum_{i=1}^n \epsilon_i^2}} \times \frac{\sqrt{\frac{\sum_{i=1}^n \epsilon_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n x_i^2 \frac{\sum_{i=1}^n \epsilon_i^2}{\sum_{i=1}^n \hat{\epsilon}_i^2}}} = \frac{\sum_{i=1}^n x_i \epsilon_i}{\sqrt{\sum_{i=1}^n x_i^2}} / \sqrt{\frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n-1}}.$$

Note that conditional on $\epsilon_1, \ldots, \epsilon_n$

$$\sum_{i=1}^n x_i \epsilon_i \sim N(0, \sum_{i=1}^n \epsilon_i^2)$$

so that conditionally and unconditionally

$$\frac{\sum_{i=1}^n x_i \epsilon_i}{\sqrt{\sum_{i=1}^n \epsilon_i^2}} \sim N(0, 1).$$

Furthermore, we can write

$$(n-1) s_*^2 \frac{(\hat{\epsilon}^T \epsilon)}{(\epsilon^T \epsilon)} = \frac{(\epsilon^T \epsilon)}{\epsilon^T \epsilon} \sum_{i=1}^n \epsilon_i^2 \qquad = \frac{(x^T x)}{(\epsilon^T \epsilon)} \epsilon^T \epsilon - \frac{(x^T x)}{(\epsilon^T \epsilon)} \epsilon^T x (x^T x)^{-1} x^T \epsilon \qquad = \frac{(x^T x)}{(\epsilon^T \epsilon)} (\epsilon^T \epsilon) - \epsilon^T x (\epsilon^T \epsilon)^{-1} x^T \epsilon \qquad = x^T M_\epsilon x,$$

which is distributed as a chi-squared random variables with $n-1$ degrees of freedom conditional on $\epsilon_1, \ldots, \epsilon_n$. Moreover, they two terms are independent. Voila!

Exercise 19: Regression Statistics

Derive the equation of the regression line, the R^2 and the t-statistics for the slope coefficient for each of the following datasets $(X_i, Y_{ji}), i = 1, \ldots, 11$ and $j = 1, 2, 3$. Graph the data and comment on your findings.

X	Y_1	Y_2	Y_3
10.0	8.04	9.14	7.46
8.0	6.95	8.14	6.77
13.0	7.58	8.74	12.74
9.0	8.81	8.77	7.11
11.0	8.33	9.26	7.81
14.0	9.96	8.10	8.84
6.0	7.24	6.13	6.08
4.0	4.26	3.10	5.39
12.0	10.84	9.13	8.15
7.0	4.82	7.26	6.42
5.0	5.68	4.74	5.73

Solution 19:

This exercise illustrates Anscombe's quartet, a set of four datasets that have nearly identical simple descriptive statistics, yet appear very different when graphed. It highlights the importance of visualizing data before analyzing it. Let's denote each dataset as (X, Y_j) for $j = 1, 2, 3$. The linear regression model is $Y_{ji} = \alpha_j + \beta_j X_i + \epsilon_{ji}$. The OLS estimators are:

$$\hat{\beta}_j = \frac{\sum_{i=1}^{11} (X_i - \bar{X})(Y_{ji} - \bar{Y}_j)}{\sum_{i=1}^{11} (X_i - \bar{X})^2}$$

$$\hat{\alpha}_j = \bar{Y}_j - \hat{\beta}_j \bar{X}$$

The R^2 is:

$$R_j^2 = \frac{\sum_{i=1}^{11} (\hat{Y}_{ji} - \bar{Y}_j)^2}{\sum_{i=1}^{11} (Y_{ji} - \bar{Y}_j)^2} = \hat{\beta}_j^2 \frac{\sum_{i=1}^{11} (X_i - \bar{X})^2}{\sum_{i=1}^{11} (Y_{ji} - \bar{Y}_j)^2}$$

The t-statistic for the slope coefficient β_j is:

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{\frac{s_j^2}{\sum_{i=1}^{11} (X_i - \bar{X})^2}}}, \text{ where } s_j^2 = \frac{1}{11-2} \sum_{i=1}^{11} \hat{\epsilon}_{ji}^2$$

Using software (e.g., R, Python, or even a calculator) to compute these values:

For all three datasets, we have approximately:

- $\bar{X} = 9$
- $\bar{Y}_j = 7.5$
- $\hat{\beta}_j = 0.5$
- $\hat{\alpha}_j = 3$

- Regression Equation: $Y_{ji} = 3 + 0.5X_i$
- $R_j^2 = 0.667$
- $t_j \approx 4.24$

The findings highlight that despite having the same regression line, R^2, and t-statistic, the datasets are visually very different. Dataset 1 is a standard linear relationship. Dataset 2 demonstrates a clear non-linear relationship. Dataset 3 is linear, but has an influential outlier. The summary statistics and regression results are identical, but the underlying data and appropriateness of the linear model vary drastically across the datasets. This emphasizes the importance of graphing data and not relying solely on summary statistics.

Exercise 20: Hypothesis testing in Regression Model

Consider the regression model

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

where $\epsilon_i \sim N(0, \sigma^2)$ and X_i are fixed numbers. Let $\theta = (\alpha, \beta, \sigma^2)$. Provide the Lagrange Multiplier, Wald, and Likelihood Ratio statistics [and give their asymptotic distribution] for testing the null hypothesis that $\sigma^2 = 1$ versus (a) the two sided alternative $\sigma^2 \neq 1$; (b) the one-sided alternative $\sigma^2 < 1$.

Solution 20:

The t-test is straightforward. The log likelihood

$$\log L(\alpha, \beta, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$$

$$\log L(\alpha, \beta, 1) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$$

In this case, the estimates of α, β are unaffected by the restriction. Therefore, since

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2$$

$$LR = 2\{\log L(\hat{\theta}) - \log L(\tilde{\theta})\} = -n\log\tilde{\sigma}^2 + n(\tilde{\sigma}^2 - 1).$$

Now this looks complicated. But write $\delta = \tilde{\sigma}^2 - 1$ and then

$$LR = -n(\log(1+\delta) - \delta).$$

We now apply the fact that $\ln(1+\epsilon) - \epsilon \sim \epsilon^2/2$ for small ϵ so that

$$LR \approx n(\tilde{\sigma}^2 - 1)^2.$$

Formally, apply the delta method. Then further approximate by

$$n\left(\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2 - 1\right)^2 \xrightarrow{D} \chi^2(1).$$

In this case, we reject the null that $\sigma^2 = 1$ against the two-sided alternative if $LR > \chi^2_\alpha(1)$. To test the one sided alternative we consider the signed likelihood ratio test with

$$LR_+ = -n(\log(1+\delta) - \delta) \times 1(\delta > 0).$$

Exercise 21: Regression with Fourth Power

Consider the regression model $Y_i = \beta X_i + \epsilon_i$, where X_i is i.i.d., while ϵ_i are i.i.d. with mean zero and variance σ^2. Discuss the consistency and asymptotic normality of $\hat{\beta}$, where $\hat{\beta}$ minimizes

$$\sum_{i=1}^n (Y_i - \beta X_i)^4.$$

Discuss the case $E(\epsilon_i^3) = 0$ and the case $E(\epsilon_i^3) \neq 0$.

Solution 21:

Let

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^n (Y_i - \beta X_i)^4.$$

We can write

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^n (\epsilon_i + (\beta_0 - \beta)X_i)^4$$
$$= \frac{1}{n}\sum_{i=1}^n \epsilon_i^4 + 4(\beta_0 - \beta)\frac{1}{n}\sum_{i=1}^n \epsilon_i^3 X_i + 6(\beta_0 - \beta)^2\frac{1}{n}\sum_{i=1}^n \epsilon_i^2 X_i^2 + 4(\beta_0 - \beta)^3\frac{1}{n}\sum_{i=1}^n \epsilon_i X_i^3 + (\beta_0 - \beta)^4\frac{1}{n}\sum_{i=1}^n X_i^4.$$

Then, provided fourth moments exist, we have

$$Q_n(\beta) \xrightarrow{P} Q(\beta) = a_0 + a_1(\beta_0 - \beta) + a_2(\beta_0 - \beta)^2 + a_3(\beta_0 - \beta)^3 + a_4(\beta_0 - \beta)^4.$$

This convergence is uniform in β over a compact set because of the separation of the random variables from the parameters. When $E(\epsilon_i^3) = 0$, we have

$$Q(\beta) = a_0 + a_2(\beta_0 - \beta)^2 + a_4(\beta_0 - \beta)^4.$$

This function is uniquely minimized at $\beta = \beta_0$ because $a_2, a_4 > 0$. In general a_1, a_3 may not be zero, and the analysis is more complicated.

Exercise 22: Identification in GMM

Suppose that

$$y_i = \sin \beta x_i + \epsilon_i,$$

where $E(\epsilon_i|x_i) = 0$. Suppose also that x_i is uniformly distributed on $[a, b]$ for some a, b. The conditional moment restriction implies that $E\epsilon_i h(x_i) = 0$ for any measurable function h. Consider two choices $h(x) = 1$ and $h(x) = x \cos \beta x$. Check the identification issue for the GMM method with these two choices of instruments in the case that $\beta = 0$.

Solution 22:

We have to check that

$$Eh(x_i) \sin \beta x_i = 0$$

if and only if $\beta = 0$. The *if* part always works by assumption. We compute the above expectation for different h and a, b. First, suppose that $[a, b] = [0, \pi]$. Then

$$\frac{1}{\pi} \int_0^\pi \sin \beta x dx = \frac{1 - \cos \pi\beta}{\beta}$$

$$\frac{1}{\pi} \int_0^\pi x \cos \beta x \sin \beta x dx = \frac{-2\pi\beta \cos^2 \pi\beta + \cos \pi\beta \sin \pi\beta + \pi\beta}{\beta^2}.$$

These two functions are graphed below on the range $-4 \leq \beta \leq 4$:

They both have very similar behavior, namely many isolated zeros. However, in the first graph, the zeros are at points where the derivative is zero, which suggests that these points do not correspond to a minimum (of whatever criterion function). In the second graph there are some zeros where the derivative is also positive, which suggests that these points correspond to local minima. It turns out that the zero at $\beta = 0$ corresponds to the global minimum of the nonlinear least squares criterion function, which can be checked by looking at the quantity

$$\frac{1}{\pi} \int_0^\pi (\sin(\beta x))^2 dx = \frac{-\cos \pi\beta \sin \pi\beta + \pi\beta}{2\pi},$$

which is plotted below

Finally, we repeat the exercise for the case where $[a, b] = [-\pi/2, \pi/2]$. In this case,

$$\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \sin(\beta x) dx = 0$$

$$\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \sin(\beta x) x \cos(\beta x) dx = \frac{2\pi\beta \cos^2 \frac{1}{2}\pi\beta - 2 \cos \frac{1}{2}\pi\beta \sin \frac{1}{2}\pi\beta - \pi\beta}{4\pi}$$

$$\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} (\sin(\beta x))^2 dx = -\frac{-2 \cos \frac{1}{2}\pi\beta \sin \frac{1}{2}\pi\beta + \pi\beta}{2\pi}.$$

Now of course, the instrument $h = 1$ is totally useless, while the nonlinear least squares instruments still work, as can be seen from the graph below of the first order condition and the least squares criterion

Exercise 23: Coastline of Britain

How long is the coastline of Britain? Lewis Fry Richardson conjectured that the length of the coastline L was related to the length of the ruler G by

$$L = MG^{1-D},$$

where M, D are constants with $D \geq 1$. Suppose that you have a sample of observations $\{L_i, G_i\}_{i=1}^n$, how would you test the hypothesis that the length is finite against the alternative that it is infinite (i.e., as $G \to 0$)?

Solution 23:

Taking logs, we have

$$\ln L_i = \ln M + (1 - D) \ln G_i.$$

Let $y_i = \ln L_i$, $x_i = \ln G_i$, $\alpha = \ln M$, and $\beta = 1 - D$. Then we have a linear regression model

$$y_i = \alpha + \beta x_i + \epsilon_i.$$

We assume that ϵ_i are i.i.d. errors with mean zero. We can estimate α and β by OLS. The hypothesis that the length is finite corresponds to $\beta = 0$ ($D = 1$). The alternative hypothesis that the length is infinite is $\beta < 0$ ($D > 1$). Thus, we can conduct a one-sided t-test of $H_0 : \beta = 0$ versus $H_1 : \beta < 0$.

Exercise 24: Measuring a Table

Suppose that you want to measure the sides of a table, i.e., length (L) and width (W). However, your research assistant reports to you only the area (A). Luckily, she was trained at Oxford and so makes an error in each measurement. Specifically,

$$L_i = L + \epsilon_i; \quad W_i = W + \eta_i,$$

where ϵ_i, η_i are mutually independent standard normal random variables. The RA reports $\{A_i\}_{i=1}^n$, where $A_i = L_i W_i$. Suggest Method of Moments estimators of L, W based on the sample information. Now suppose that ϵ_i and η_i are both $N(0, \sigma^2)$ for some unknown σ^2. Show how to estimate L, W, and σ^2 from the sample data $\{A_i\}_{i=1}^n$.

Solution 24:

We have

$$E(A_i) = E[(L + \epsilon_i)(W + \eta_i)] = LW$$
$$E(A_i^2) = E[(L + \epsilon_i)^2(W + \eta_i)^2] = E[(L^2 + 2L\epsilon_i + \epsilon_i^2)(W^2 + 2W\eta_i + \eta_i^2)] = (L^2 + 1)(W^2 + 1) = L^2W^2 + L^2 + W^2 + 1.$$
$$\text{var}(A_i) = L^2 + W^2 + 1$$

These are two moment conditions in two unknowns. In fact by substitution $L = E(A)/W$ we get the quadratic equation

$$\theta^2 + (1 - E(A))\theta + E^2A = 0$$

where $\theta = W^2$. Note that the coefficient on θ is negative and the intercept is positive. We have

$$\theta = \frac{E(A) - 1 \pm \sqrt{(1 - E(A))^2 - 4E^2A}}{2}$$

assuming this is well defined. We need

$$(1 - E(A))^2 - 4E^2A \geq 0$$

which will be satisfied because we can expand out and find this quantity is $(L^2 - W^2)^2$. We should take the positive root. Therefore, we estimate using sample mean and sample variance of the area measurements to obtain $\hat{\theta}$ and then estimate $\hat{W} = \sqrt{\hat{\theta}}$ and $\hat{L} = E(A)/\sqrt{\hat{\theta}}$.

Exercise 25: Regression with Error in Variables

Consider the following regression model

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon,$$

where X_j is an $n \times K_j$ matrix of non-random regressors for $j = 1, 2$. The disturbance term ϵ satisfies

$$E(\epsilon) = X_1\gamma$$

for some non-zero vector γ, and further suppose that

$$E[(\epsilon - E(\epsilon))(\epsilon - E(\epsilon))^T] = \sigma^2 I_n.$$

Calculate $E(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_1)$.

Solution 25:

The OLS estimator of β_1 is obtained by regressing y on X_1 and X_2. Let $X = (X_1 \quad X_2)$. Then the OLS estimator is $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X^TX)^{-1}X^Ty$. Let $M_1 = I - X_1(X_1^TX_1)^{-1}X_1^T$. Then, the OLS estimator of β_1 is
$\hat{\beta}_1 = (X_1^TX_1)^{-1}X_1^Ty - (X_1^TX_1)^{-1}X_1^TX_2(X_2^TM_1X_2)^{-1}X_2^TM_1y = (X_1^TM_2X_1)^{-1}X_1^TM_2y$, where $M_2 = (I - X_2(X_2^TX_2)^{-1}X_2^T)$.

$$E[\hat{\beta}_1] = E[(X_1^TM_2X_1)^{-1}X_1^TM_2(X_1\beta_1 + X_2\beta_2 + \epsilon)] \qquad = \beta_1 + (X_1^TM_2X_1)^{-1}X_1^TM_2X_1\gamma \qquad = \beta_1 + \gamma. \qquad E(\epsilon|X) = X_1\gamma$$

$$\text{var}(\hat{\beta}_1) = (X_1^TM_2X_1)^{-1}X_1^TM_2\text{Var}(\epsilon)M_2X_1(X_1^TM_2X_1)^{-1} \qquad \text{var}(\hat{\beta}_1) = \sigma^2(X_1^TM_2X_1)^{-1}.$$

Exercise 26: True or False (a, b, c)

True, False, or Indeterminate, and Explain. (a) Let X and Y be two mean zero random variables. Then, $E(X|Y)=0$ and $E(Y|X)=0$ if and only if X and Y are uncorrelated.

b. Two events A and B with $P(A)=1$ and $P(B)=1$ must be mutually independent.

c. Whenever the cumulative distribution function is discontinuous, the median is not well-defined.

Solution 26:

a. **False**. If $E(Y|X)=0$ or $E(X|Y)$, then X and Y are uncorrelated, but not vice versa. Uncorrelatedness means $E[XY]=E[X]E[Y]$, while $E(Y|X)=0$ is a stronger condition, implying $E[XY]=E[XE[Y|X]]=E[X\cdot 0]=0$. However, two uncorrelated variables with means of zeros, may still have dependence between them that is missed by the measure of correlation.

b. **True**. We have

$$P(A\cap B)=P(A)+P(B)-P(A\cup B)\geq P(A)+P(B)-1=1$$

so $P(A\cap B)=P(A)P(B)$.

c. **False**. The cumulative distribution function could be discontinuous at some points but be continuous at the median.

Exercise 27: True or False (a, b, c)

True, False, or Indeterminate, and Explain.

a. The Cramer-Rao Theorem says that with normally distributed errors the maximum likelihood estimator is Best Unbiased.

b. Suppose that for some estimator $\hat{\theta}$, we have as $n\to\infty$

$$n(\hat{\theta}-2\pi)\xrightarrow{D}N(0,1),$$

where n is the sample size. Then

$$n\sin\hat{\theta}\xrightarrow{D}N(0,1).$$

c. Suppose that A,B are symmetric real matrices. The eigenvalues of $A+B$ are the same as the eigenvalues of $B+A$.

Solution 27:

a. **False**. The CR theorem says that any unbiased estimator has variance greater than or equal to the inverse information. In the normal case, the MLE is unbiased and achieves the CR lower bound. The statement should say "… the MLE is Best Linear Unbiased"

b. **True**. This is true from a simple application of the delta method with $f(x)=\sin(x)$ has derivative $\cos(x)$, which is equal to 1 when $x=2\pi$;

c. **True** because $A+B=B+A$.

Exercise 28: Reverse Regression

Consider the linear regression model

$$Y=X\beta+\epsilon,$$

where $Y=(y_1,\ldots,y_n)^T$ and X is the $n\times K$ matrix containing the regressors. Suppose that the classical assumptions hold, i.e., ϵ_i are i.i.d. with $\epsilon_i\sim N(0,1)$ and independent of X. Consider the reverse regression estimator $\hat{\beta^R}=(\hat{\beta_1^R},\ldots,\hat{\beta_K^R})^T$, where

$$\hat{\beta_j^R}=(Y^TY)^{-1}Y^TX_j,$$

where X_j are $n\times 1$ vectors containing the observations on the jth regressor. Let x_i be the $K\times 1$ vector containing the ith observations on the regressors, and let $\hat{x_i}=\hat{\beta^R}y_i$ and $\hat{u_i}=x_i-\hat{x_i}$. Which, if any, of the following hold and why:

i. $n^{-1}\sum_{i=1}^n x_i\hat{u_i}=0$;

ii. $n^{-1}\sum_{i=1}^n \hat{u_i}=0$;

iii. $n^{-1}\sum_{i=1}^n y_i\hat{u_i}=0$. This is the only necessarily true answer, because the OLS procedure is constructed to make the right hand side variable, in this case y, orthogonal to the residual $\hat{u}$.

iv. $X^Ty=X^TX\hat{\beta^R}$

Derive the probability limit of $\hat{\beta^R}$ as n gets large. State clearly any additional assumptions you need to make. Consider the scalar case ($K=1$). Then

$$\hat{\beta}^R = \frac{Y^TX}{Y^TY} = \frac{\frac{1}{n}X^TY}{\frac{1}{n}Y^TY}.$$

Solution 28:

We have

$$\frac{1}{n}X^TY = \beta\frac{1}{n}X^TX + \frac{1}{n}X^T\epsilon \qquad \frac{1}{n}Y^TY = \beta^2\frac{1}{n}X^TX + 2\beta\frac{1}{n}X^T\epsilon + \frac{1}{n}\epsilon^T\epsilon.$$

We apply law of large numbers to show that

$$\frac{1}{n}X^TX \xrightarrow{P} Ex_i^2 \qquad \frac{1}{n}X^T\epsilon \xrightarrow{P} 0 \qquad \frac{1}{n}\epsilon^T\epsilon \xrightarrow{P} 1,$$

provided x_i are i.i.d. with finite variance. We then apply Slutsky's theorem to establish

$$\hat{\beta}^R \xrightarrow{P} \frac{\beta Ex_i^2}{\beta^2 Ex_i^2+1}.$$

Exercise 29: Hypothesis Testing

Consider the regression model

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

where $\epsilon_i \sim N(0,\sigma^2)$ and are independent of X_i, which you may treat as fixed numbers. Let $\{Y_i, X_i, i = 1,\ldots,n\}$ be the observed sample, and let $\theta = (\alpha, \beta, \sigma^2)$ be unknown parameters. Provide the Likelihood Ratio statistic for testing the null hypothesis that $\sigma^2 = 1$ versus the two sided alternative $\sigma^2 \neq 1$. Derive its large sample distribution under the null hypothesis. You may use the expansion $\log(1+x) = x - x^2/2 + x^3/3 + \ldots$ for $x \in (-1,1)$.

Solution 29:

The t-test is straightforward. The log likelihood

$$\log L(\alpha,\beta,\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$$

$$\log L(\alpha,\beta,1) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2.$$

In this case, the estimates of α, β are unaffected by the restriction. Therefore, since

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2$$

$$LR = 2\{\log L(\hat{\theta}) - \log L(\tilde{\theta})\} = -n\log\hat{\sigma}^2 + n(\hat{\sigma}^2 - 1)$$

Now this looks nasty. But write $\delta = \hat{\sigma}^2 - 1$ and then

$LR = -n(\log(1+\delta) - \delta)$. We now apply the fact that $\ln(1+\epsilon) - \epsilon \approx \epsilon^2/2$ for small ϵ, so that

$$LR \approx n(\hat{\sigma}^2 - 1)^2$$

and then approximate by

$$n\left(\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2 - 1\right)^2 \xrightarrow{d} \chi^2(1).$$

Exercise 30: Rogoff-Satchell Estimator

Suppose that $y_i = \beta x_i + \epsilon_i$, where ϵ_i, x_i are i.i.d. and mutually independent with $\epsilon_i, x_i \sim N(0,1)$. Consider the Rogoff-Satchell estimator. The idea is to remove observations in such a way as to change the sign of the estimator. One version of this is defined as follows:

$$\hat{\beta}_{RS} = \frac{\sum_{i=1}^n x_i y_i 1(x_i y_i \hat{\beta} < 0)}{\sum_{i=1}^n x_i^2 1(x_i y_i \hat{\beta} < 0)}.$$

Note that $x_i y_i \hat{\beta} < 0$ if either $\hat{\beta} > 0$ and $x_i y_i < 0$ or $\hat{\beta} < 0$ and $x_i y_i > 0$. We have $x_i y_i < 0$ if either $x_i < 0$ and $y_i > 0$ or $x_i > 0$ and $y_i < 0$. Essentially this keeps only the data in quadrants opposite to the least squares slope and then fits least squares to them. This estimator is nonlinear and it is biased. Derive the properties of this estimator in two cases:

a. $\beta \neq 0$;

b. $\beta = 0$.

Solution 30:

When $\beta \neq 0$ we have

$$\hat{\beta}^*_{R-S} = \hat{\beta}_{R-S} + o_p(n^{-1/2})$$

$$\hat{\beta}^*_{R-S} = \frac{\sum_{i=1}^n x_i y_i 1(x_i y_i \beta < 0)}{\sum_{i=1}^n x_i^2 1(x_i y_i \beta < 0)}$$

and this can be treated by standard arguments. When $\beta = 0$, we argue as follows. Firstly, we may show that

$$\sqrt{n}\hat{\beta} \xrightarrow{D} Z,$$

where Z is standard normal. Then we have

$$\hat{\beta}_{R-S} = \hat{\beta}^*_{R-S} + o_p(n^{-1/2}) \qquad \hat{\beta}^*_{R-S} = \frac{\sum_{i=1}^n x_i y_i 1(x_i y_i Z < 0)}{\sum_{i=1}^n x_i^2 1(x_i y_i Z < 0)}.$$

Since Z is independent of the data we may condition on it. Furthermore, note that if $Z > 0$ then $x_i y_i < 0$ to ensure 1 and vice versa. We have

$$\frac{1}{n}\sum_{i=1}^n x_i y_i 1(x_i y_i Z < 0) \xrightarrow{P}_{Z>0} E[x_i y_i 1(x_i y_i < 0)] \qquad \frac{1}{n}\sum_{i=1}^n x_i y_i 1(x_i y_i Z < 0) \xrightarrow{P}_{Z<0} E[x_i y_i 1(x_i y_i > 0)].$$

Similarly

$$\frac{1}{n}\sum_{i=1}^n x_i^2 1(x_i y_i Z < 0) \xrightarrow{P}_{Z>0} E[x_i^2 1(x_i y_i < 0)] \; \frac{1}{n}\sum_{i=1}^n x_i^2 1(x_i y_i Z < 0) \xrightarrow{P}_{Z<0} E[x_i^2 1(x_i y_i > 0)].$$

Therefore, when $Z > 0$

$$\hat{\beta}_{R-S} \xrightarrow{P}_{Z>0} \frac{E[x_i y_i 1(x_i y_i < 0)]}{E[x_i^2 1(x_i y_i < 0)]}$$

but when $Z < 0$

$$\hat{\beta}_{R-S} \xrightarrow{P}_{Z<0} \frac{E[x_i y_i 1(x_i y_i > 0)]}{E[x_i^2 1(x_i y_i > 0)]}.$$

This shows that

$$\hat{\beta}_{R-S} \xrightarrow{D} W = \begin{cases} \frac{E[x_i y_i 1(x_i y_i < 0)]}{E[x_i^2 1(x_i y_i < 0)]} & \text{if } Z > 0 \\ \frac{E[x_i y_i 1(x_i y_i > 0)]}{E[x_i^2 1(x_i y_i > 0)]} & \text{if } Z < 0, \end{cases}$$

which is a two point distribution.

Exercise 31: True or False

Establish whether the following statements are True, False, or Indeterminate. Explain your reasoning.

i. If X_n is a discrete random variable for each n, and if $X_n \xrightarrow{P} 0$, then $Pr(X_n = 0) \to 1$.

ii. The Gauss-Markov theorem says that with normally distributed errors the ordinary least squares estimator is Best Linear Unbiased.

iii. Suppose that A, B are symmetric real matrices. Let λ be a nonzero eigenvalue of AB, then λ^2 is an eigenvalue of $(AB)^2$.

Solution 31:

i. **False**. For example, $X_n = \pm 1/n$ with probability $1/2$. However, if support of X_n is fixed then this is true.

ii. **True**, but don't need normality, unnecessary condition

iii. **True**. Suppose that

$ABx = \lambda x.$

Then

$ABABx = \lambda ABx = \lambda^2 x.$

Exercise 32: Hypothesis Testing in Regression

Suppose that you observe data $\{Y_i, X_i, i = 1, \ldots, n\}$ with $X_i \in \mathbb{R}^K$. You believe that the linear regression model holds, so that

$Y_i = \alpha + \beta^T X_i + \epsilon_i$

where $\beta \in \mathbb{R}^K$ is an unknown parameter vector and $\alpha \in \mathbb{R}$ is an unknown scalar parameter. Explain how you would test the following hypotheses. In your answer, be clear about what additional assumptions you make and what role the assumptions have in making the test valid. Also give the test statistics and the critical values you would use (no table work).

i. The null hypothesis that $\alpha = 0$ versus the alternative that $\alpha > 0$.

ii. The null hypothesis that $\beta = 0$ versus the alternative that $\beta \neq 0$.

iii. The null hypothesis that $\alpha = 0$ and $\beta = 0$ versus the alternative that $\alpha \neq 0$ or $\beta \neq 0$.

Solution 32:

We assume i.i.d. observations, with $E[\epsilon_i|X_i] = 0$ and $E[\epsilon_i^2|X_i] = \sigma^2$. We also need that second moments exist ($E||X_i||^2 < \infty$).

i. Under the null hypothesis $H_0 : \alpha = 0$, we can use a one-sided t-test. The test statistic is:

$t = \frac{\hat{\alpha}-0}{SE(\hat{\alpha})}$

where $\hat{\alpha}$ is the OLS estimator of α and $SE(\hat{\alpha})$ is its standard error. Under the null hypothesis and the assumptions given, this statistic follows a t_{n-K-1} distribution (or asymptotically a standard normal distribution). We reject the null if $t > t_{n-K-1,1-\gamma}$, where γ is the significance level.

ii. Under the null hypothesis $H_0 : \beta = 0$, we use an F-test. The test statistic is based on comparing the restricted and unrestricted sum of squared residuals (RSS). Let $RSS_R = \sum_{i=1}^n (Y_i - \hat{\alpha})^2$ be the restricted RSS (under the null that $\beta = 0$) Let $RSS_U = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}^T X_i)^2$ be the unrestricted RSS. The F-statistic is $F = \frac{(RSS_R - RSS_U)/K}{RSS_U/(n-K-1)}$. Under the null hypothesis and our assumptions, this statistic follows an $F_{K,n-K-1}$ distribution. We reject the null if $F > F_{K,n-K-1,1-\gamma}$.

iii. Under the null hypothesis $H_0 : \alpha = 0$ and $\beta = 0$, we again use an F-test. This is a joint test. Let $RSS_R = \sum_{i=1}^n Y_i^2$ be the restricted RSS (under the null that $\alpha = \beta = 0$). Let $RSS_U = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta}^T X_i)^2$ be the unrestricted RSS. The F-statistic is: $F = \frac{(RSS_R - RSS_U)/(K+1)}{RSS_U/(n-K-1)}$. Under the null hypothesis, and assumptions given, $F \sim F_{K+1,n-K-1}$. Reject the null if $F > F_{K+1,n-K-1,1-\gamma}$.

Author: Peter Fuleky

This book was built with Quarto