6. Expected Value and Covariance Matrices

The main purpose of this section is a discussion of expected value and covariance for random matrices and vectors. These topics are particularly important in multivariate statistical models and the multivariate normal distribution. This section requires some prerequisite knowledge of linear algebra.

Basic Theory

We will let $\mathbb{R}^{m \times n}$ denote the space of all $m \times n$ matrices of real numbers. In particular, we will identify $\mathbb{R}^n$ with $\mathbb{R}^{n \times 1}$, so that an ordered n-tuple can also be thought of as an $n \times 1$ column vector. The transpose of a matrix A is denoted A^T. As usual, our starting point is a random experiment with a probability measure $\mathbb{P}$ on an underlying sample space.

Expected Value of a Random Matrix

Suppose that X is an $m \times n$ matrix of real-valued random variables, whose (i, j) entry is denoted $X_{i,j}$. Equivalently, X can be thought of as a random $m \times n$ matrix. It is natural to define the **expected value** $\mathbb{E}(X)$ to be the $m \times n$ matrix whose (i, j) entry is $\mathbb{E}(X_{i,j})$, the expected value of $X_{i,j}$.

Many of the basic properties of expected value of random variables have analogies for expected value of random matrices, with matrix operation replacing the ordinary ones.

1. Show that $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ if X and Y are random $m \times n$ matrices.

2. Show that $\mathbb{E}(A\,X) = A\,\mathbb{E}(X)$ if A is a non-random $m \times n$ matrix and X is random $n \times k$ matrix.

3. Show that $\mathbb{E}(X\,Y) = \mathbb{E}(X)\,\mathbb{E}(Y)$ if X is a random $m \times n$ matrix, Y is a random $n \times k$ matrix, and X and Y are independent.

Covariance Matrices

Suppose now that X is a random vector in $\mathbb{R}^m$ and Y is a random vector in $\mathbb{R}^n$. The **covariance matrix** of X and Y is the $m \times n$ matrix $\mathrm{cov}(X, Y)$ whose (i, j) entry is $\mathrm{cov}(X_i, Y_j)$ the covariance of X_i and Y_j.

4. Show that $\mathrm{cov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))^T\big)$.

5. Show that $\mathrm{cov}(X, Y) = \mathbb{E}(X\,Y^T) - \mathbb{E}(X)\,\mathbb{E}(Y)^T$

6. Show that $\mathrm{cov}(Y, X) = \mathrm{cov}(X, Y)^T$

7. Show that $\mathrm{cov}(X, Y) = 0$ if and only if each coordinate of X is uncorrelated with each coordinate of Y (in particular, this holds if X and Y are independent).

8. Show that $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$ if X and Y are random vectors in $\mathbb{R}^m$ and Z is a random

vector in $\mathbb{R}^n$.

9. Show that $\operatorname{cov}(X, Y + Z) = \operatorname{cov}(X, Y) + \operatorname{cov}(X, Z)$ if X is a random vector in $\mathbb{R}^m$ and Y and Z are random vectors in $\mathbb{R}^n$.

10. Show that $\operatorname{cov}(A\,X, Y) = A\operatorname{cov}(X, Y)$ if X is a random vector in $\mathbb{R}^m$, Y is a random vector in $\mathbb{R}^n$, and A is a non-random matrix in $\mathbb{R}^{k \times m}$.

11. Show that $\operatorname{cov}(X, A\,Y) = \operatorname{cov}(X, Y)\,A^T$ if X is a random vector in $\mathbb{R}^m$, Y is a random vector in $\mathbb{R}^n$, and A is a non-random matrix in $\mathbb{R}^{k \times n}$.

Variance-Covariance Matrices

Suppose now that $X = (X_1, X_2, ..., X_n)$ is a random vector in $\mathbb{R}^n$. The covariance matrix of X with itself is called the **variance-covariance matrix** of X:

$$\operatorname{VC}(X) = \operatorname{cov}(X, X)$$

12. Show that $\operatorname{VC}(X)$ is a symmetric $n \times n$ matrix with $(\operatorname{var}(X_1), \operatorname{var}(X_2), ..., \operatorname{var}(X_n))$ on the diagonal.

13. Show that $\operatorname{VC}(X + Y) = \operatorname{VC}(X) + \operatorname{cov}(X, Y) + \operatorname{cov}(Y, X) + \operatorname{VC}(Y)$ if X and Y are random vectors in $\mathbb{R}^n$.

14. Show that $\operatorname{VC}(A\,X) = A\operatorname{VC}(X)\,A^T$ if X is a random vector in $\mathbb{R}^n$ and A is a non-random matrix in $\mathbb{R}^{m \times n}$.

If $a \in \mathbb{R}^n$, note that $a^T\,X$ is simply the **inner product** or **dot product** of a with X, and is a linear combination of the coordinates of X:

$$a^T\,X = \sum_{i=1}^n a_i\,X_i$$

15. Show that $\operatorname{var}(a^T\,X) = a^T\operatorname{VC}(X)\,a$ if X is a random vector in $\mathbb{R}^n$ and $a \in \mathbb{R}^n$. Thus conclude that $\operatorname{VC}(X)$ is either positive semi-definite or positive definite. In particular, the eigenvalues and the determinant of $\operatorname{VC}(X)$ are nonnegative.

16. Show that $\operatorname{VC}(X)$ is positive semi-definite (but not positive definite) if and only if there exists $a \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that, with probability 1,
$$a^T\,X = \sum_{i=1}^n a_i\,X_i = c$$

Thus, if $\operatorname{VC}(X)$ is positive semi-definite, then one of the coordinates of X can be written as an affine transformation of the other coordinates (and hence can usually be eliminated in the underlying model). By contrast, if $\operatorname{VC}(X)$ is positive definite, then this cannot happen; $\operatorname{VC}(X)$ has positive eigenvalues and determinant and is invertible.

Best Linear Predictors

Suppose again that $X = (X_1, X_2, ..., X_m)$ is a random vector in $\mathbb{R}^m$ and that $Y = (Y_1, Y_2, ..., Y_n)$ is a random vector in $\mathbb{R}^n$. We are interested in finding the linear (technically affine) function of X,

$$A\,X + b, \quad A \in \mathbb{R}^{n \times m}, \ b \in \mathbb{R}^n$$

that is closest to Y in the mean square sense. This problem is of fundamental importance in statistics when random vector X, the **predictor vector** is observable, but not random vector Y, the **response vector**. Our discussion here generalizes the one-dimensional case, when X and Y are random variables. That problem was solved in the section on Covariance and Correlation. We will assume that $\text{VC}(X)$ is positive definite, so that none of the coordinates of X can be written as an affine function of the other coordinates.

17. Show that $\mathbb{E}\left(\left\|Y - \left(A\,X + b\right)\right\|^2\right)$ is minimized when $A = \text{cov}(Y,\ X)\,\text{VC}(X)^{-1}$ and
$b = \mathbb{E}(Y) - \text{cov}(Y,\ X)\,\text{VC}(X)^{-1}\,\mathbb{E}(X)$

Thus, the linear function of X that is closest to X in the mean square sense is the random vector

$$L(Y|X) = \mathbb{E}(Y) + \text{cov}(Y,\ X)\,\text{VC}(X)^{-1}\,(X - \mathbb{E}(X))$$

.

The function of x given by

$$L(Y|X = x) = \mathbb{E}(Y) + \text{cov}(Y,\ X)\,\text{VC}(X)^{-1}\,(x - \mathbb{E}(X))$$

is known as the **(distribution) linear regression function**. If we observe $X = x$ then $L(Y|X = x)$ is our prediction of Y.

Non-linear regression with a single, real-valued predictor variable can be thought of as a special case of multiple linear regression. Thus, suppose that X is the predictor variable, Y is the response variable, and that $(g_1, g_2, ..., g_n)$ is a sequence of real-valued functions. We can apply the results of Exercise 17 to find the linear function of $(g_1(X), g_2(X), ..., g_n(X))$ that is closest to Y in the mean square sense. We just replace X_i with $g_i(X)$ for each i.

Examples and Applications

18. Suppose that $(X,\ Y)$ has probability density function $f(x,\ y) = x + y$, $0 \le x \le 1,\ 0 \le y \le 1$. Find each of the following:

   a. $\mathbb{E}(X,\ Y)$
   b. $\text{VC}(X,\ Y)$

19. Suppose that $(X,\ Y)$ has probability density function $f(x,\ y) = 2\,(x + y)$, $0 \le x \le y \le 1$. Find each of the following:

   a. $\mathbb{E}(X,\ Y)$
   b. $\text{VC}(X,\ Y)$

20. Suppose that $(X,\ Y)$ has probability density function $f(x,\ y) = 6\,x^2\,y$, $0 \le x \le 1,\ 0 \le y \le 1$. Find each of the following:

   a. $\mathbb{E}(X,\ Y)$

   b. $VC(X, Y)$

21. Suppose that (X, Y) has probability density function $f(x, y) = 15\, x^2\, y,\ 0 \le x \le y \le 1$. Find each of the following:

   a. $\mathbb{E}(X, Y)$
   b. $VC(X, Y)$
   c. $L(Y|X)$
   d. $L\left(Y \middle| X, X^2\right)$
   e. Sketch the regression curves on the same set of axes.

22. Suppose that (X, Y, Z) is uniformly distributed on the region $\left\{(x, y, z) \in \mathbb{R}^3 : 0 \le x \le y \le z \le 1\right\}$ Find each of the following:

   a. $\mathbb{E}(X, Y, Z)$
   b. $VC(X, Y, Z)$
   c. $L(Z|X, Y)$.
   d. $L(Y|X, Z)$.
   e. $L(X|Y, Z)$.

23. Suppose that X is uniformly distributed on $[0, 1]$, and that given X, Y is uniformly distributed on $[0, X]$. Find each of the following:

   a. $\mathbb{E}(X, Y)$
   b. $VC(X, Y)$