Part I: Linear Models

Chapter I.3

Quadratic Forms

Topics

To be discussed...

- Definition of quadratic forms
- Basic properties
- Cochran's theorem
- Applications of Cochran's theorem

I.3.1 Definition

Let $\mathbf{Y}$ be a random vector and $A \in \mathbb{R}^{p \times p}$ be a symmetric matrix. Then $\mathbf{Y}'A\mathbf{Y}$ is called **quadratic form**.

I.3.2 Remark

- For a random vector $\mathbf{Y} \in \mathbb{R}^n$, $\sum_{i=1}^n Y_i^2$ is a quadratic form (for $A = I_n$): $\sum_{i=1}^n Y_i^2 = \mathbf{Y}'\mathbf{Y} = \mathbf{Y}'I_n\mathbf{Y}$.
- Symmetry of A in Definition I.3.1 is not required, since, from $\mathbf{x}'A\mathbf{x} = \mathbf{x}'A'\mathbf{x}$, we have for any $A \in \mathbb{R}^{p \times p}$

$$\mathbf{x}'A\mathbf{x} = \frac{1}{2}(\mathbf{x}'A\mathbf{x} + \mathbf{x}'A'\mathbf{x}) = \mathbf{x}'A_*\mathbf{x}$$

with the symmetric matrix $A_* = \frac{1}{2}(A + A')$. Therefore, without loss of generality, quadratic forms discussed in the following are based on symmetric matrices.

I.3.3 Lemma

Let $\mathbf{Y}$ be a random vector with $E\mathbf{Y} = \boldsymbol{\mu}$ and Cov $(\mathbf{Y}) = \Sigma$ and $A \in \mathbb{R}^{p \times p}$. Then,

$$E\mathbf{Y}'A\mathbf{Y} = \text{trace}(A\Sigma) + \boldsymbol{\mu}'A\boldsymbol{\mu}.$$

Orthogonal projectors & quadratic forms

I.3.4 Lemma

Let $Q \in \mathbb{R}^{p \times p}$ be an orthogonal projector. Then, an eigenvalue λ of Q satisfies $\lambda \in \{0, 1\}$. Furthermore, $\operatorname{rank}(Q) = \operatorname{trace}(Q)$.

I.3.5 Theorem

Let $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, I_p)$ be a random vector and Q be an orthogonal projector. Then,

$$\mathbf{Y}'Q\mathbf{Y} \sim \chi^2\Big(\operatorname{rank}(Q), \frac{1}{2}\boldsymbol{\mu}'Q\boldsymbol{\mu}\Big).$$

Cochran's theorem

I.3.6 Theorem (Cochran)*

Let $\mathbf{X} \sim N_p(0, \sigma^2 I_p)$ and $A_1, \ldots, A_n \in \mathbb{R}^{p \times p}_{\geqslant 0}$ be non-negative definite matrices with $\sum_{j=1}^n A_j = I_p$. Let $r_j = \mathrm{rank}(A_j)$, $1 \leqslant j \leqslant n$. Then, the following conditions are equivalent:

1. $\sum_{j=1}^n r_j = p$.
2. $\frac{1}{\sigma^2}\mathbf{X}' A_j \mathbf{X} \sim \chi^2(r_j)$, $1 \leqslant j \leqslant n$
3. $\mathbf{X}' A_j \mathbf{X}$, $1 \leqslant j \leqslant n$, are mutually independent.

I.3.7 Remark

Theorem I.3.6 may be extended in various directions. In Rencher & Schaalje (2008), Theorem 5.6c, a non-central version is presented, that is, one assumes $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \sigma^2 I_p)$ with $\boldsymbol{\mu} \in \mathbb{R}^p$. Then, the following conditions are equivalent:

1. $\sum_{j=1}^n r_j = p$.
2. $\frac{1}{\sigma^2}\mathbf{X}' A_j \mathbf{X} \sim \chi^2(r_j, \boldsymbol{\mu}' A_j \boldsymbol{\mu}/2)$, $1 \leqslant j \leqslant n$
3. $\mathbf{X}' A_j \mathbf{X}$, $1 \leqslant j \leqslant n$, are mutually independent.

*see, e,g., Gut, A. (2009) An Intermediate Course in Probability. 2nd edn., Springer, New York, Section 5.9.

Application of Cochran's theorem

I.3.8 Corollary (see Theorem I.2.15)

Let $p \geqslant 2$, $\mathbf{Z} = (Z_1, \ldots, Z_p)' \sim N_p(0, \sigma^2 I_p)$, $\overline{Z} = \frac{1}{p}\sum_{j=1}^p Z_j$, $S_Z = \frac{1}{p-1}\sum_{j=1}^p (Z_j - \overline{Z})^2$, and $E_p = I_p - \frac{1}{p}\mathbb{1}_p\mathbb{1}_p'$. Then:

1. $p\dfrac{\overline{Z}^2}{\sigma^2} \sim \chi^2(1)$ and $\dfrac{1}{\sigma^2}\displaystyle\sum_{j=1}^p (Z_j - \overline{Z})^2 = \dfrac{1}{\sigma^2}\mathbf{Z}'E_p\mathbf{Z} \sim \chi^2(p-1)$ are independent.
2. $\dfrac{p\overline{Z}^2}{S_Z} \sim F(1, p-1)$

P3.16 Theorem

1. Let $\mathbf{X} \sim N_p(\mu, \Sigma)$ with $\mu \in \mathbb{R}^p, \Sigma \in \mathbb{R}^{p \times p}_{>0}$. Then, for $A \in \mathbb{R}^{k \times p}, B \in \mathbb{R}^{r \times p}$ with $k, r \in \mathbb{N}$, we get:
   - $A\mathbf{X}$ and $B\mathbf{X}$ are independent if and only if $A\Sigma B' = 0$.
2. Let $p \geqslant 2$, $\mathbf{Z} = (Z_1, \dots, Z_p)' \sim N_p(0, \sigma^2 I_p)$, $\overline{Z} = \frac{1}{p}\sum_{j=1}^p Z_j$, and $E_p = I_p - \frac{1}{p}\mathbb{1}_{p\times p} = I_p - \frac{1}{p}\mathbb{1}_p\mathbb{1}_p'$. Then:
   - $\overline{Z}$ and $\mathbf{Z} - \overline{Z}\mathbb{1}_p = E_p\mathbf{Z}$ are independent.
   - $\overline{Z}$ and $S_Z = \frac{1}{p-1}\sum_{j=1}^p (Z_j - \overline{Z})^2$ are independent.

P3.15 Corollary

1. Let $\mathbf{X} = (X_1, \ldots, X_p)' \sim N_p(\boldsymbol{\mu}, \Sigma)$ with $\Sigma = (\sigma_{ij})_{i,j}$ and $\overline{X} = \frac{1}{p}\sum_{j=1}^{p} X_j$. Then:

$$\overline{X} = \frac{1}{p}\mathbb{1}_p'\mathbf{X} \sim N\Big(\frac{1}{p}\sum_{j=1}^{p}\mu_j, \frac{1}{p^2}\mathbb{1}_p'\Sigma\mathbb{1}_p\Big) = N\Big(\frac{1}{p}\sum_{j=1}^{p}\mu_j, \frac{1}{p^2}\sum_{i,j}\sigma_{ij}\Big).$$

2. If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ then
   - $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
   - $\boxed{\sqrt{n}\,\frac{\overline{X} - \mu}{\sigma} \sim N(0, 1)}$

Two-sample case and F-distribution

I.3.9 Corollary

Let $X_1, \ldots, X_{n_1} \overset{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$, $Y_1, \ldots, Y_{n_2} \overset{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$ be independent samples of random variables with $n_1, n_2 \geqslant 2$ and

$$\widehat{\sigma}_1^2 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_j - \overline{X})^2 = \frac{1}{n_1 - 1} \mathbf{X}' E_{n_1} \mathbf{X} \quad \text{and} \quad \widehat{\sigma}_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (Y_j - \overline{Y})^2 = \frac{1}{n_2 - 1} \mathbf{Y}' E_{n_2} \mathbf{Y}.$$

Then, we get

Z = X - mu, Z^- = X^-mu

1. $\mathbf{X}' E_{n_1} \mathbf{X} / \sigma_1^2 \sim \chi^2(n_1 - 1)$ and $\mathbf{Y}' E_{n_2} \mathbf{Y} / \sigma_2^2 \sim \chi^2(n_2 - 1)$ are independent, and

$$\boxed{\frac{\widehat{\sigma}_1^2 / \sigma_1^2}{\widehat{\sigma}_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)}$$

2. For $\sigma_1 = \sigma_2 = \sigma$:
   - the F-statistic $\boxed{F = \frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2}}$ is F-distributed with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.
   - $\boxed{\frac{(n_1 - 1)\widehat{\sigma}_1^2 + (n_2 - 1)\widehat{\sigma}_2^2}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2).}$

An extension to random vectors

I.3.10 Theorem[†]

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{\text{iid}}{\sim} N_p(\boldsymbol{\mu}, \Sigma)$ with $\boldsymbol{\mu} \in \mathbb{R}^p, \Sigma \in \mathbb{R}^{p \times p}_{>0}$ and $n > p$. Then:

1. $\overline{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\Sigma)$
2. $n\|\overline{\mathbf{X}} - \boldsymbol{\mu}\|^2_{\Sigma} \sim \chi^2(n)$.
3. For the **sample covariance matrix** $\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$, we have
   - $E\hat{\Sigma} = \Sigma$
   - $(n-1)\hat{\Sigma}$ has a so-called **Wishart**-distribution
   - $n\|\overline{\mathbf{X}} - \boldsymbol{\mu}\|^2_{\hat{\Sigma}}$ has an **Hotellings-T^2-distribution** with parameters p and $n-1$ (for short, $n\|\overline{\mathbf{X}} - \boldsymbol{\mu}\|^2_{\hat{\Sigma}} \sim T_{p,n-1}$).
     For $m > p$: if $\mathbf{Y} \sim T_{p,m}$ then $\frac{m-p+1}{mp}\mathbf{Y} \sim F_{p,m-p+1}$

[†] cf. T. W. Anderson (2003) Introduction to Multivariate Statistical Analysis, 3rd ed., New York: Wiley, 2003, Chapters 3 & 5.

$$\boxed{\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)}$$

<proof>

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \sum_{i=1}^{n} Z^2 \sim \sum_{i=1}^{n} \chi^2(1) \sim \chi^2(n)$$

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$$

<proof>

$$\underbrace{\boxed{\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}}}_{\chi^2(n)} = \frac{\sum(X_i-\overline{X})^2}{\sigma^2} + \underbrace{\boxed{\frac{n(\overline{X}_n-\mu)^2}{\sigma^2}}}_{Z^2=\chi^2(1)}$$

$$* \; Z = \frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}$$

$$\chi^2(n) = \alpha + \chi^2(1)$$

$$\alpha = \chi^2(n) - \chi^2(1)$$

$$= \chi^2(n-1)$$

$$\therefore \frac{\sum(X_i-\overline{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

because the sum of n independent chi square distribution with degree of freedom 1 is chi-square(n).

In the matrices in Example 2.6, the diagonal elements are positive. For positive definite matrices, this is true in general.

Theorem 2.6a

(i) If $\mathbf{A}$ is positive definite, then all its diagonal elements a_{ii} are positive.
(ii) If $\mathbf{A}$ is positive semidefinite, then all $a_{ii} \geq 0$.

PROOF

(i) Let $\mathbf{y}' = (0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the ith position and 0's elsewhere. Then $\mathbf{y}'\mathbf{A}\mathbf{y} = a_{ii} > 0$.
(ii) Let $\mathbf{y}' = (0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the ith position and 0's elsewhere. Then $\mathbf{y}'\mathbf{A}\mathbf{y} = a_{ii} \geq 0$. □

Some additional properties of positive definite and positive semidefinite matrices are given in the following theorems.

Theorem 2.6b. Let $\mathbf{P}$ be a nonsingular matrix.

(i) If $\mathbf{A}$ is positive definite, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is positive definite.
(ii) If $\mathbf{A}$ is positive semidefinite, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is positive semidefinite.

PROOF

(i) To show that $\mathbf{y}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{y} > 0$ for $\mathbf{y} \neq \mathbf{0}$, note that $\mathbf{y}'(\mathbf{P}'\mathbf{A}\mathbf{P})\mathbf{y} = (\mathbf{P}\mathbf{y})'\mathbf{A}(\mathbf{P}\mathbf{y})$. Since $\mathbf{A}$ is positive definite, $(\mathbf{P}\mathbf{y})'\mathbf{A}(\mathbf{P}\mathbf{y}) > 0$ provided that $\mathbf{P}\mathbf{y} \neq \mathbf{0}$. By (2.47), $\mathbf{P}\mathbf{y} = \mathbf{0}$ only if $\mathbf{y} = \mathbf{0}$, since $\mathbf{P}^{-1}\mathbf{P}\mathbf{y} = \mathbf{P}^{-1}\mathbf{0} = \mathbf{0}$. Thus $\mathbf{y}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{y} > 0$ if $\mathbf{y} \neq \mathbf{0}$.
(ii) See problem 2.36. □

Corollary 1. Let $\mathbf{A}$ be a $p \times p$ positive definite matrix and let $\mathbf{B}$ be a $k \times p$ matrix of rank $k \leq p$. Then $\mathbf{B}\mathbf{A}\mathbf{B}'$ is positive definite. □

Corollary 2. Let $\mathbf{A}$ be a $p \times p$ positive definite matrix and let $\mathbf{B}$ be a $k \times p$ matrix. If $k > p$ or if rank$(\mathbf{B}) = r$, where $r < k$ and $r < p$, then $\mathbf{B}\mathbf{A}\mathbf{B}'$ is positive semidefinite. □

Theorem 2.6c. A symmetric matrix $\mathbf{A}$ is positive definite if and only if there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}'\mathbf{P}$.

PROOF. We prove the "if" part only. Suppose $\mathbf{A} = \mathbf{P}'\mathbf{P}$ for nonsingular $\mathbf{P}$. Then

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \mathbf{y}'\mathbf{P}'\mathbf{P}\mathbf{y} = (\mathbf{P}\mathbf{y})'(\mathbf{P}\mathbf{y}).$$

This is a sum of squares [see (2.20)] and is positive unless $\mathbf{P}\mathbf{y} = \mathbf{0}$. By (2.47), $\mathbf{P}\mathbf{y} = \mathbf{0}$ only if $\mathbf{y} = \mathbf{0}$. □

Corollary 1. A positive definite matrix is nonsingular. □

One method of factoring a positive definite matrix $\mathbf{A}$ into a product $\mathbf{P}'\mathbf{P}$ as in Theorem 2.6c is provided by the Cholesky decomposition (Seber and Lee 2003, pp. 335–337), by which $\mathbf{A}$ can be factored uniquely into $\mathbf{A} = \mathbf{T}'\mathbf{T}$, where $\mathbf{T}$ is a nonsingular upper triangular matrix.

For any square or rectangular matrix $\mathbf{B}$, the matrix $\mathbf{B}'\mathbf{B}$ is positive definite or positive semidefinite.

Theorem 2.6d. Let $\mathbf{B}$ be an $n \times p$ matrix.

(i) If $\text{rank}(\mathbf{B}) = p$, then $\mathbf{B}'\mathbf{B}$ is positive definite.
(ii) If $\text{rank}(\mathbf{B}) < p$, then $\mathbf{B}'\mathbf{B}$ is positive semidefinite.

PROOF

(i) To show that $\mathbf{y}'\mathbf{B}'\mathbf{B}\mathbf{y} > 0$ for $\mathbf{y} \neq \mathbf{0}$, we note that

$$\mathbf{y}'\mathbf{B}'\mathbf{B}\mathbf{y} = (\mathbf{B}\mathbf{y})'(\mathbf{B}\mathbf{y}),$$

which is a sum of squares and is thereby positive unless $\mathbf{B}\mathbf{y} = \mathbf{0}$. By (2.37), we can express $\mathbf{B}\mathbf{y}$ in the form

$$\mathbf{B}\mathbf{y} = y_1\mathbf{b}_1 + y_2\mathbf{b}_2 + \cdots + y_p\mathbf{b}_p.$$

This linear combination is not $\mathbf{0}$ (for any $\mathbf{y} \neq \mathbf{0}$) because $\text{rank}(\mathbf{B}) = p$, and the columns of $\mathbf{B}$ are therefore linearly independent [see (2.40)].

(ii) If $\text{rank}(\mathbf{B}) < p$, then we can find $\mathbf{y} \neq \mathbf{0}$ such that

$$\mathbf{B}\mathbf{y} = y_1\mathbf{b}_1 + y_2\mathbf{b}_2 + \cdots + y_p\mathbf{b}_p = \mathbf{0}$$

since the columns of $\mathbf{B}$ are linearly dependent [see (2.40)]. Hence $\mathbf{y}'\mathbf{B}'\mathbf{B}\mathbf{y} \geq 0$.

□

can be written in matrix form as

$$\mathbf{Ax} = \mathbf{c}, \tag{2.57}$$

where $\mathbf{A}$ is $n \times p$, $\mathbf{x}$ is $p \times 1$, and $\mathbf{c}$ is $n \times 1$. Note that if $n \neq p$, $\mathbf{x}$ and $\mathbf{c}$ are of different sizes. If $n = p$ and $\mathbf{A}$ is nonsingular, then by (2.47), there exists a unique solution vector $\mathbf{x}$ obtained as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}$. If $n > p$, so that $\mathbf{A}$ has more rows than columns, then $\mathbf{Ax} = \mathbf{c}$ typically has no solution. If $n < p$, so that $\mathbf{A}$ has fewer rows than columns, then $\mathbf{Ax} = \mathbf{c}$ typically has an infinite number of solutions.

If the system of equations $\mathbf{Ax} = \mathbf{c}$ has one or more solution vectors, it is said to be *consistent*. If the system has no solution, it is said to be *inconsistent*.

To illustrate the structure of a consistent system of equations $\mathbf{Ax} = \mathbf{c}$, suppose that $\mathbf{A}$ is $p \times p$ of rank $r < p$. Then the rows of $\mathbf{A}$ are linearly dependent, and there exists some $\mathbf{b}$ such that [see (2.38)]

$$\mathbf{b}'\mathbf{A} = b_1\mathbf{a}'_1 + b_2\mathbf{a}'_2 + \cdots + b_p\mathbf{a}'_p = \mathbf{0}'.$$

Then we must also have $\mathbf{b}'\mathbf{c} = b_1c_1 + b_2c_2 + \cdots + b_pc_p = 0$, since multiplication of $\mathbf{Ax} = \mathbf{c}$ by $\mathbf{b}'$ gives $\mathbf{b}'\mathbf{Ax} = \mathbf{b}'\mathbf{c}$, or $\mathbf{0}'\mathbf{x} = \mathbf{b}'\mathbf{c}$. Otherwise, if $\mathbf{b}'\mathbf{c} \neq 0$, there is no $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{c}$. Hence, in order for $\mathbf{Ax} = \mathbf{c}$ to be consistent, the same linear relationships, if any, that exist among the rows of $\mathbf{A}$ must exist among the elements (rows) of $\mathbf{c}$. This is formalized by comparing the rank of $\mathbf{A}$ with the rank of the *augmented matrix* ($\mathbf{A}$, $\mathbf{c}$). The notation ($\mathbf{A}$, $\mathbf{c}$) indicates that $\mathbf{c}$ has been appended to $\mathbf{A}$ as an additional column.

Theorem 2.7 The system of equations $\mathbf{Ax} = \mathbf{c}$ has at least one solution vector $\mathbf{x}$ if and only if rank($\mathbf{A}$) = rank($\mathbf{A}$, $\mathbf{c}$).

PROOF. Suppose that rank($\mathbf{A}$) = rank($\mathbf{A}$, $\mathbf{c}$), so that appending $\mathbf{c}$ does not change the rank. Then $\mathbf{c}$ is a linear combination of the columns of $\mathbf{A}$; that is, there exists some $\mathbf{x}$ such that

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_p\mathbf{a}_p = \mathbf{c},$$

which, by (2.37), can be written as $\mathbf{Ax} = \mathbf{c}$. Thus $\mathbf{x}$ is a solution.

Conversely, suppose that there exists a solution vector $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{c}$. In general, rank ($\mathbf{A}$) $\leq$ rank($\mathbf{A}, \mathbf{c}$) (Harville 1997, p. 41). But since there exists an $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{c}$, we have

$$\begin{aligned} \text{rank}(\mathbf{A}, \mathbf{c}) &= \text{rank}(\mathbf{A}, \mathbf{Ax}) = \text{rank}[\mathbf{A}(\mathbf{I}, \mathbf{x})] \\ &\leq \text{rank}(\mathbf{A}) \qquad \text{[by Theorem 2.4(i)]}. \end{aligned}$$