

Chapter 3

Random Vectors and Multivariate Normal Distributions

3.1 Random vectors

Definition 3.1.1. Random vector. Random vectors are vectors of random

variables. For instance,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix},$$

where each element represent a random variable, is a random vector.

Definition 3.1.2. Mean and covariance matrix of a random vector.

The mean (expectation) and covariance matrix of a random vector \mathbf{X} is defined as follows:

$$E[\mathbf{X}] = \begin{pmatrix} E[\mathbf{X}_1] \\ E[\mathbf{X}_2] \\ \vdots \\ E[\mathbf{X}_n] \end{pmatrix},$$

and

$$\begin{aligned} cov(\mathbf{X}) &= E\left[\{\mathbf{X} - E(\mathbf{X})\}\{\mathbf{X} - E(\mathbf{X})\}^T\right] \\ &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}, \end{aligned} \tag{3.1.1}$$

where $\sigma_j^2 = var(\mathbf{X}_j)$ and $\sigma_{jk} = cov(\mathbf{X}_j, \mathbf{X}_k)$ for $j, k = 1, 2, \dots, n$.

Properties of Mean and Covariance.

1. If \mathbf{X} and \mathbf{Y} are random vectors and \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{D} are constant matrices, then

$$E[\mathbf{AXB} + \mathbf{CY} + \mathbf{D}] = \mathbf{A}E[\mathbf{X}]\mathbf{B} + \mathbf{C}E[\mathbf{Y}] + \mathbf{D}. \quad (3.1.2)$$

Proof. Left as an exercise. □

2. For any random vector \mathbf{X} , the covariance matrix $cov(\mathbf{X})$ is symmetric.

Proof. Left as an exercise. □

3. If $X_j, j = 1, 2, \dots, n$ are independent random variables, then $cov(\mathbf{X}) = diag(\sigma_j^2, j = 1, 2, \dots, n)$.

Proof. Left as an exercise. □

4. $cov(\mathbf{X} + \mathbf{a}) = cov(\mathbf{X})$ for a constant vector \mathbf{a} .

Proof. Left as an exercise. □

Properties of Mean and Covariance (cont.)

5. $\text{cov}(\mathbf{AX}) = \mathbf{A}\text{cov}(\mathbf{X})\mathbf{A}^T$ for a constant matrix \mathbf{A} .

Proof. Left as an exercise. □

6. $\text{cov}(\mathbf{X})$ is positive semi-definite. **Meaningful** covariance matrices are positive definite.

Proof. Left as an exercise. □

7. $\text{cov}(\mathbf{X}) = E[\mathbf{XX}^T] - E[\mathbf{X}]\{E[\mathbf{X}]\}^T$.

Proof. Left as an exercise. □

Definition 3.1.3. Correlation Matrix.

A correlation matrix of a vector of random variable \mathbf{X} is defined as the matrix of pairwise correlations between the elements of \mathbf{X} . Explicitly,

$$\text{corr}(\mathbf{X}) = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{bmatrix}, \quad (3.1.3)$$

where $\rho_{jk} = \text{corr}(\mathbf{X}_j, \mathbf{X}_k) = \sigma_{jk}/(\sigma_j\sigma_k)$, $j, k = 1, 2, \dots, n$.

Example 3.1.1. If only successive random variables in the random vector \mathbf{X} are correlated and have the same correlation ρ , then the correlation matrix $\text{corr}(\mathbf{X})$ is given by

$$\text{corr}(\mathbf{X}) = \begin{bmatrix} 1 & \rho & 0 & \dots & 0 \\ \rho & 1 & \rho & \dots & 0 \\ 0 & \rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad (3.1.4)$$

Example 3.1.2. If every pair of random variables in the random vector \mathbf{X} have the same correlation ρ , then the correlation matrix $corr(\mathbf{X})$ is given by

$$corr(\mathbf{X}) = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}, \quad (3.1.5)$$

and the random variables are said to be exchangeable.

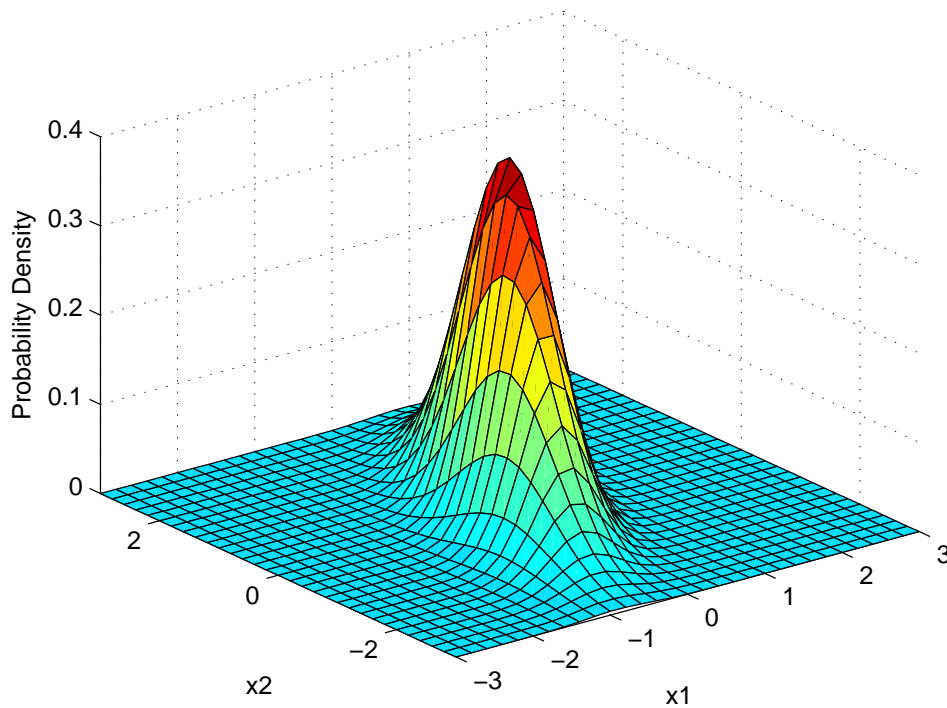
3.2 Multivariate Normal Distribution

Definition 3.2.1. Multivariate Normal Distribution. A random vector $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^T$ is said to follow a multivariate normal distribution with mean μ and covariance matrix Σ if \mathbf{X} can be expressed as

$$\mathbf{X} = \mathbf{AZ} + \mu,$$

where $\Sigma = \mathbf{AA}^T$ and $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_n)$ with $\mathbf{Z}_i, i = 1, 2, \dots, n$ iid $N(0, 1)$ variables.

Bivariate normal distribution with mean $(0, 0)^T$ and covariance matrix $\begin{bmatrix} 0.25 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}$



Definition 3.2.2. Multivariate Normal Distribution. A random vector $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^T$ is said to follow a multivariate normal distribution with mean μ and a positive definite covariance matrix Σ if \mathbf{X} has the density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right] \quad (3.2.1)$$

.

Properties

1. Moment generating function of a $N(\mu, \Sigma)$ random variable \mathbf{X} is given by

$$M_{\mathbf{X}}(\mathbf{t}) = \exp \left\{ \mu^T \mathbf{t} + \frac{1}{2} \mathbf{t}^T \Sigma \mathbf{t} \right\}. \quad (3.2.2)$$

2. $E(\mathbf{X}) = \mu$ and $cov(\mathbf{X}) = \Sigma$.
3. If $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are i.i.d $N(0, 1)$ random variables, then their joint distribution can be characterized by $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)^T \sim N(0, \mathbf{I}_n)$.
4. $\mathbf{X} \sim N_n(\mu, \Sigma)$ if and only if all non-zero linear combinations of the components of \mathbf{X} are normally distributed.

Linear transformation

5. If $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{A}_{m \times n}$ is a constant matrix of rank m , then $\mathbf{Y} = \mathbf{A}\mathbf{x} \sim N_p(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$.

Proof. Use definition 3.2.1 or property 1 above. □

Orthogonal linear transformation

6. If $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \mathbf{I}_n)$ and $\mathbf{A}_{n \times n}$ is an orthogonal matrix and $\boldsymbol{\Sigma} = \mathbf{I}_n$, then $\mathbf{Y} = \mathbf{A}\mathbf{x} \sim N_n(\mathbf{A}\boldsymbol{\mu}, \mathbf{I}_n)$.

Marginal and Conditional distributions

Suppose \mathbf{X} is $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and \mathbf{X} is partitioned as follows,

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix},$$

where \mathbf{X}_1 is of dimension $p \times 1$ and \mathbf{X}_2 is of dimension $n - p \times 1$. Suppose the corresponding partitions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

respectively. Then,

7. **Marginal distribution.** \mathbf{X}_1 is multivariate normal - $N_p(\mu_1, \boldsymbol{\Sigma}_{11})$.

Proof. Use the result from property 5 above. □

8. **Conditional distribution.** The distribution of $\mathbf{X}_1|\mathbf{X}_2$ is p-variate normal - $N_p(\mu_{1|2}, \boldsymbol{\Sigma}_{1|2})$, where,

$$\mu_{1|2} = \mu_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \mu_2),$$

and

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21},$$

provided $\boldsymbol{\Sigma}$ is positive definite.

Proof. See Result 5.2.10, page 156 (Ravishanker and Dey). □

Uncorrelated implies independence for multivariate normal random variables

9. If \mathbf{X} , μ , and Σ are partitioned as above, then \mathbf{X}_1 and \mathbf{X}_2 are independent if and only if $\Sigma_{12} = 0 = \Sigma_{21}^T$.

Proof. We will use m.g.f to prove this result. Two random vectors \mathbf{X}_1 and \mathbf{X}_2 are independent iff

$$M_{(\mathbf{X}_1, \mathbf{X}_2)}(t_1, t_2) = M_{\mathbf{X}_1}(t_1)M_{\mathbf{X}_2}(t_2).$$

□

3.3 Non-central distributions

We will start with the standard chi-square distribution.

Definition 3.3.1. Chi-square distribution. If X_1, X_2, \dots, X_n be n independent $N(0, 1)$ variables, then the distribution of $\sum_{i=1}^n X_i^2$ is χ_n^2 (ch-square with degrees of freedom n).

χ_n^2 -distribution is a special case of gamma distribution when the scale parameter is set to $1/2$ and the shape parameter is set to be $n/2$. That is, the density of χ_n^2 is given by

$$f_{\chi_n^2}(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} e^{-x/2} x^{n/2-1}, \quad x \geq 0; \quad n = 1, 2, \dots, \quad (3.3.1)$$

Example 3.3.1. The distribution of $(n-1)S^2/\sigma^2$, where $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ is the sample variance of a random sample of size n from a normal distribution with mean μ and variance σ^2 , follows a χ_{n-1}^2 .

The moment generating function of a chi-square distribution with n d.f. is given by

$$M_{\chi_n^2}(t) = (1 - 2t)^{-n/2}, \quad t < 1/2. \quad (3.3.2)$$

The m.g.f (3.3.2) shows that the sum of two independent ch-square random variables is also a ch-square. Therefore, differences of sequential sums of squares of independent normal random variables will be distributed independently as chi-squares.

Theorem 3.3.2. *If $\mathbf{X} \sim N_n(\mu, \Sigma)$ and Σ is positive definite, then*

$$(\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu) \sim \chi_n^2. \quad (3.3.3)$$

Proof. Since Σ is positive definite, there exists a non-singular $\mathbf{A}_{n \times n}$ such that $\Sigma = \mathbf{A}\mathbf{A}^T$ (Cholesky decomposition). Then, by definition of multivariate normal distribution,

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mu,$$

where \mathbf{Z} is a random sample from a $N(0, 1)$ distribution. Now, □

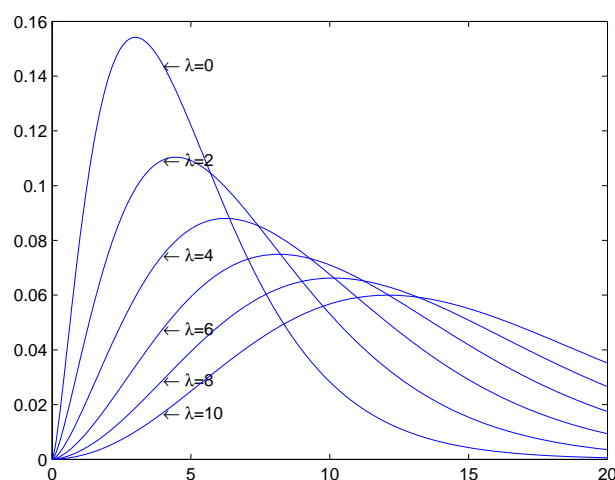


Figure 3.1: Non-central chi-square densities with df 5 and non-centrality parameter λ .

Definition 3.3.2. Non-central chi-square distribution. Suppose X 's are as in Definition (3.3.1) except that each X_i has mean μ_i , $i = 1, 2, \dots, n$. Equivalently, suppose, $\mathbf{X} = (X_1, \dots, X_n)^T$ be a random vector distributed as $N_n(\mu, \mathbf{I}_n)$, where $\mu = (\mu_1, \dots, \mu_n)^T$. Then the distribution of $\sum_{i=1}^n X_i^2 = \mathbf{X}^T \mathbf{X}$ is referred to as non-central chi-square with d.f. n and non-centrality parameter $\lambda = \sum_{i=1}^n \mu_i^2/2 = \frac{1}{2} \mu^T \mu$. The density of such a non-central chi-square variable $\chi_n^2(\lambda)$ can be written as a infinite poisson mixture of central chi-square densities as follows:

$$f_{\chi_n^2(\lambda)}(x) = \sum_{j=1}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \frac{(1/2)^{(n+2j)/2}}{\Gamma((n+2j)/2)} e^{-x/2} x^{(n+2j)/2-1}. \quad (3.3.4)$$

Properties

1. The moment generating function of a non-central chi-square variable

$\chi_n^2(\lambda)$ is given by

$$M_{\chi_n^2(n,\lambda)}(t) = (1 - 2t)^{-n/2} \exp\left\{\frac{2\lambda t}{1 - 2t}\right\}, \quad t < 1/2. \quad (3.3.5)$$

2. $E[\chi_n^2(\lambda)] = n + 2\lambda.$

3. $Var[\chi_n^2(\lambda)] = 2(n + 4\lambda).$

4. $\chi_n^2(0) \equiv \chi_n^2.$

5. For a given constant c ,

(a) $P(\chi_n^2(\lambda) > c)$ is an increasing function of λ .

(b) $P(\chi_n^2(\lambda) > c) \geq P(\chi_n^2 > c).$

Theorem 3.3.3. *If $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma}$ is positive definite, then*

$$\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \sim \chi_n^2(\lambda = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} / 2). \quad (3.3.6)$$

Proof. Since $\boldsymbol{\Sigma}$ is positive definite, there exists a non-singular matrix $\mathbf{A}_{n \times n}$ such that $\boldsymbol{\Sigma} = \mathbf{A} \mathbf{A}^T$ (Cholesky decomposition). Define,

$$\mathbf{Y} = \mathbf{A}^{-1} \mathbf{X}.$$

Then,

$$\mathbf{X} = \mathbf{A} \mathbf{Y},$$

and

$$\begin{aligned} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} &= \mathbf{Y}^T \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{A} \mathbf{Y} \\ &= \mathbf{Y}^T \mathbf{A}^T \{ \mathbf{A} \mathbf{A}^T \}^{-1} \mathbf{A} \mathbf{Y} \\ &= \mathbf{Y}^T \mathbf{Y} = \sum_{i=1}^n Y_i^2. \end{aligned}$$

But notice that \mathbf{Y} is a linear combination of \mathbf{X} and hence is distributed as $N_n(\mathbf{A}^{-1} \boldsymbol{\mu}, \mathbf{A}^{-1} \boldsymbol{\Sigma} \{ \mathbf{A}^{-1} \}^T = I_n)$. Therefore components of \mathbf{Y} are independently distributed. Specifically, $Y_i \sim N(\tau_i, 1)$, where $\boldsymbol{\tau} = \mathbf{A}^{-1} \boldsymbol{\mu}$. Thus, by definition of non-central chi-square, $\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} = \sum_{i=1}^n Y_i^2 \sim \chi_n^2(\lambda)$, where $\lambda = \sum_{i=1}^n \tau_i^2 / 2 = \boldsymbol{\tau}^T \boldsymbol{\tau} / 2 = \boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} / 2$. \square

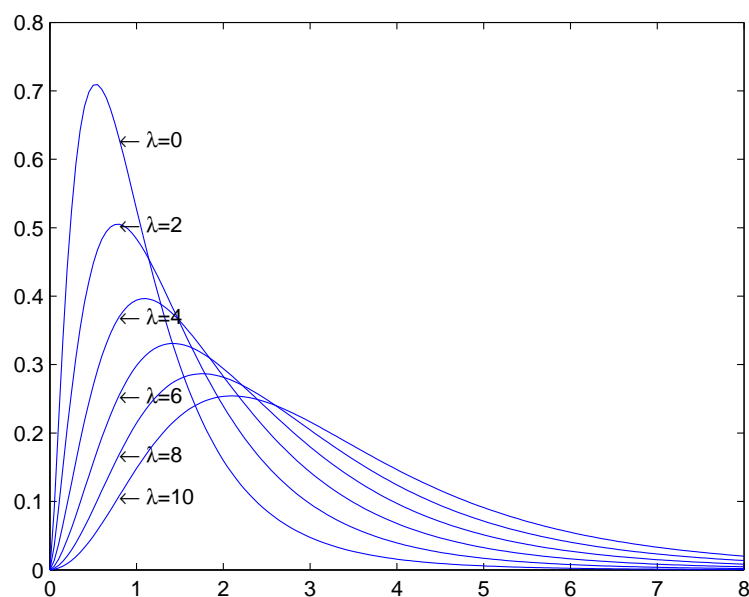


Figure 3.2: Non-central F -densities with df 5 and 15 and non-centrality parameter λ .

Definition 3.3.3. Non-central F -distribution. If $U_1 \sim \chi_{n_1}^2(\lambda)$ and $U_2 \sim \chi_{n_2}^2$ and U_1 and U_2 are independent, then, the distribution of

$$F = \frac{U_1/n_1}{U_2/n_2} \quad (3.3.7)$$

is referred to as non-central F -distribution with df n_1 and n_2 , and non-centrality parameter λ .

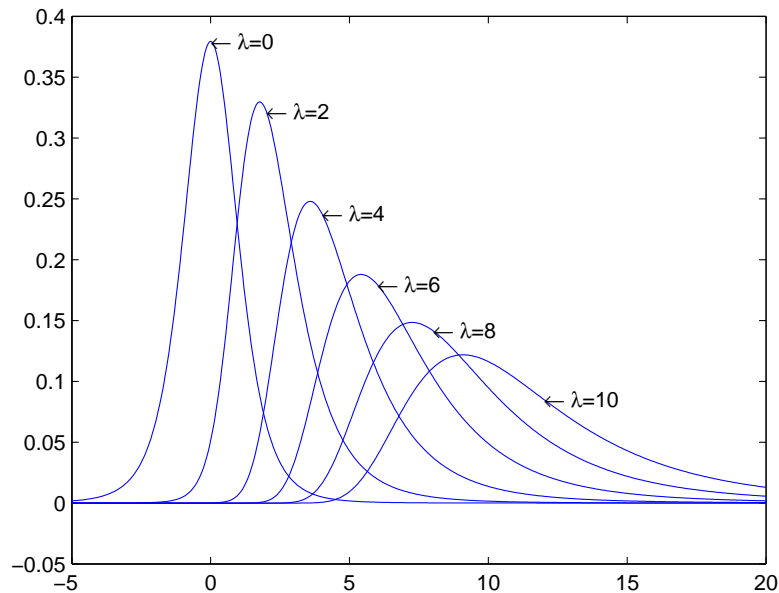


Figure 3.3: Non-central t -densities with df 5 and non-centrality parameter λ .

Definition 3.3.4. Non-central t -distribution. If $U_1 \sim N(\lambda, 1)$ and $U_2 \sim \chi_n^2$ and U_1 and U_2 are independent, then, the distribution of

$$T = \frac{U_1}{\sqrt{U_2/n}} \quad (3.3.8)$$

is referred to as non-central t -distribution with df n and non-centrality parameter λ .

3.4 Distribution of quadratic forms

Caution: We assume that our matrix of quadratic form is symmetric.

Lemma 3.4.1. *If $\mathbf{A}_{n \times n}$ is symmetric and idempotent with rank r , then r of its eigenvalues are exactly equal to 1 and $n - r$ are equal to zero.*

Proof. Use spectral decomposition theorem. (See Result 2.3.10 on page 51 of Ravishanker and Dey). \square

Theorem 3.4.2. *Let $\mathbf{X} \sim N_n(0, \mathbf{I}_n)$. The quadratic form $\mathbf{X}^T \mathbf{A} \mathbf{X} \sim \chi_r^2$ iff \mathbf{A} is idempotent with $\text{rank}(\mathbf{A}) = r$.*

Proof. Let \mathbf{A} be (symmetric) idempotent matrix of rank r . Then, by spectral decomposition theorem, there exists an orthogonal matrix \mathbf{P} such that

$$\mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{\Lambda} = \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix}. \quad (3.4.1)$$

Define $\mathbf{Y} = \mathbf{P}^T \mathbf{X} = \begin{bmatrix} \mathbf{P}_1^T \mathbf{X} \\ \mathbf{P}_2^T \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}$, so that $\mathbf{P}_1^T \mathbf{P}_1 = \mathbf{I}_r$. Thus, $\mathbf{X} =$

\mathbf{PY} and $\mathbf{Y}_1 \sim N_r(0, \mathbf{I}_r)$. Now,

$$\begin{aligned} \mathbf{X}^T \mathbf{A} \mathbf{x} &= (\mathbf{PY})^T \mathbf{A} \mathbf{PY} \\ &= \mathbf{Y}^T \begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix} \mathbf{Y} \\ &= \mathbf{Y}_1^T \mathbf{Y}_1 \sim \chi_r^2. \end{aligned} \quad (3.4.2)$$

Now suppose $\mathbf{X}^T \mathbf{A} \mathbf{X} \sim \chi_r^2$. This means that the moment generating function of $\mathbf{X}^T \mathbf{A} \mathbf{X}$ is given by

$$M_{\mathbf{X}^T \mathbf{A} \mathbf{X}}(t) = (1 - 2t)^{-r/2}. \quad (3.4.3)$$

But, one can calculate the m.g.f. of $\mathbf{X}^T \mathbf{A} \mathbf{X}$ directly using the multivariate normal density as

$$\begin{aligned} M_{\mathbf{X}^T \mathbf{A} \mathbf{X}}(t) &= E [\exp \{(\mathbf{X}^T \mathbf{A} \mathbf{X})t\}] \\ &= \int \exp \{(\mathbf{X}^T \mathbf{A} \mathbf{X})t\} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &= \int \exp \{(\mathbf{X}^T \mathbf{A} \mathbf{X})t\} \frac{1}{(2\pi)^{n/2}} \exp \left[-\frac{1}{2} \mathbf{x}^T \mathbf{x} \right] d\mathbf{x} \\ &= \int \frac{1}{(2\pi)^{n/2}} \exp \left[-\frac{1}{2} \mathbf{x}^T (\mathbf{I}_n - 2t\mathbf{A}) \mathbf{x} \right] d\mathbf{x} \\ &= |\mathbf{I}_n - 2t\mathbf{A}|^{-1/2} \\ &= \prod_{i=1}^n (1 - 2t\lambda_i)^{-1/2}. \end{aligned} \quad (3.4.4)$$

□

Equate (3.4.3) and (3.4.4) to obtain the desired result.

Theorem 3.4.3. *Let $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is positive definite. The quadratic form $\mathbf{X}^T \mathbf{A} \mathbf{X} \sim \chi_r^2(\lambda)$ where $\lambda = \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} / 2$, iff $\mathbf{A} \boldsymbol{\Sigma}$ is idempotent with $\text{rank}(\mathbf{A} \boldsymbol{\Sigma}) = r$.*

Proof. Omitted. □

Theorem 3.4.4. Independence of two quadratic forms. *Let $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is positive definite. The two quadratic forms $\mathbf{X}^T \mathbf{A} \mathbf{X}$ and $\mathbf{X}^T \mathbf{B} \mathbf{X}$ are independent if and only if*

$$\mathbf{A} \boldsymbol{\Sigma} \mathbf{B} = \mathbf{0} = \mathbf{B} \boldsymbol{\Sigma} \mathbf{A}. \quad (3.4.5)$$

Proof. Omitted. □

Remark 3.4.1. Note that in the above theorem, the two quadratic forms need not have a chi-square distribution. When they are, the theorem is referred to as **Craig's theorem**.

Theorem 3.4.5. Independence of linear and quadratic forms. *Let $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is positive definite. The quadratic form $\mathbf{X}^T \mathbf{A} \mathbf{X}$ and the linear form $\mathbf{B} \mathbf{X}$ are independently distributed if and only if*

$$\mathbf{B} \boldsymbol{\Sigma} \mathbf{A} = \mathbf{0}. \quad (3.4.6)$$

Proof. Omitted. □

Remark 3.4.2. Note that in the above theorem, the quadratic form need not have a chi-square distribution.

Example 3.4.6. Independence of sample mean and sample variance. Suppose $\mathbf{X} \sim N_n(0, \mathbf{I}_n)$. Then $\bar{X} = \sum_{i=1}^n X_i/n = \mathbf{1}^T \mathbf{X}/n$ and $S_{\bar{X}}^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ are independently distributed.

Proof.

□

Theorem 3.4.7. *Let $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then*

$$E[\mathbf{X}^T \mathbf{A} \mathbf{X}] = \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + \text{trace}(\mathbf{A} \boldsymbol{\Sigma}). \quad (3.4.7)$$

Remark 3.4.3. Note that in the above theorem, the quadratic form need not have a chi-square distribution.

Proof.

□

Theorem 3.4.8. Fisher-Cochran theorem. Suppose $\mathbf{X} \sim N_n(\mu, \mathbf{I}_n)$. Let $\mathbf{Q}_j = \mathbf{X}^T \mathbf{A}_j \mathbf{X}$, $j = 1, 2, \dots, k$ be k quadratic forms with $\text{rank}(\mathbf{A}_j) = r_j$ such that $\mathbf{X}^T \mathbf{X} = \sum_{j=1}^k \mathbf{Q}_j$. Then, Q_j 's are independently distributed as $\chi_{r_j}^2(\lambda_j)$ where $\lambda_j = \mu^T \mathbf{A}_j \mu / 2$ if and only if $\sum_{j=1}^k r_j = n$.

Proof. Omitted. □

Theorem 3.4.9. Generalization of Fisher-Cochran theorem. Suppose $\mathbf{X} \sim N_n(\mu, \mathbf{I}_n)$. Let \mathbf{A}_j , $j = 1, 2, \dots, k$ be k $n \times n$ symmetric matrices with $\text{rank}(\mathbf{A}_j) = r_j$ such that $\mathbf{A} = \sum_{j=1}^k \mathbf{A}_j$ with $\text{rank}(\mathbf{A}) = r$. Then,

1. $\mathbf{X}^T \mathbf{A}_j \mathbf{X}$'s are independently distributed as $\chi_{r_j}^2(\lambda_j)$ where $\lambda_j = \mu^T \mathbf{A}_j \mu / 2$,
and
2. $\mathbf{X}^T \mathbf{A} \mathbf{X} \sim \chi_r^2(\lambda)$ where $\lambda = \sum_{j=1}^k \lambda_j$

if and only if ANY ONE of the following conditions is satisfied.

C1. $\mathbf{A}_j \boldsymbol{\Sigma}$ is idempotent for all j and $\mathbf{A}_j \boldsymbol{\Sigma} \mathbf{A}_k = 0$ for all $j < k$.

C2. $\mathbf{A}_j \boldsymbol{\Sigma}$ is idempotent for all j and $\mathbf{A} \boldsymbol{\Sigma}$ is idempotent.

C3. $\mathbf{A}_j \boldsymbol{\Sigma} \mathbf{A}_k = 0$ for all $j < k$ and $\mathbf{A} \boldsymbol{\Sigma}$ is idempotent.

C4. $r = \sum_{j=1}^k r_j$ and $\mathbf{A} \boldsymbol{\Sigma}$ is idempotent.

C5. the matrices $\mathbf{A} \boldsymbol{\Sigma}$, $\mathbf{A}_j \boldsymbol{\Sigma}$, $j = 1, 2, \dots, k-1$ are idempotent and $\mathbf{A}_k \boldsymbol{\Sigma}$ is non-negative definite.