Ch 4. The Multivariate Normal Distribution

- Univariate normal density $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2} \qquad -\infty < x < \infty$$

- The term

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

measures the square of the distance from x to μ in standard deviation units.

- This can be generalized for a $p\times 1$ vector x of observations on several variables as

$$(x-\mu)'\Sigma^{-1}(x-\mu),$$

where μ represents the $p\times 1$ vector of the expected value of the random vector X,

and Σ represents the $p\times p$ variance-covariance matrix of X.

- Assume that the symmetric matrix Σ is positive definite.

- Multivariate normal density $N_p(\mu, \Sigma)$

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-(x-\mu)'\Sigma^{-1}(x-\mu)/2} \qquad -\infty < x_i < \infty, i = 1, \cdots, p.$$

4.2. The Multivariate Normal Density and its Properties

- Example 4.1. Bivariate Normal Density

- $\mu_1 = E(X_1)$, $\mu_2 = E(X_2)$, $\sigma_{11} = Var(X_1)$, $\sigma_{22} = Var(X_2)$,
 and $\rho_{12} = \sigma_{12} \Big/ \left(\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}\right) = Corr(X_1, X_2)$.

- Since $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$, $\quad \Sigma^{-1} = \dfrac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix}$.

 Since the correlation coefficient, ρ_{12}, is expressed as $\rho_{12} = \sigma_{12} \Big/ \sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$,

$$\sigma_{11}\sigma_{22} - \sigma_{12}^2 = \sigma_{11}\sigma_{22}\left(1 - \rho_{12}^2\right).$$

 Then, the squared distance becomes

$$(x-\mu)'\Sigma^{-1}(x-\mu) = [x_1-\mu_1, x_2-\mu_2]\frac{1}{\sigma_{11}\sigma_{22}\left(1-\rho_{12}^2\right)}\begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}} & \sigma_{11} \end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$= \frac{\sigma_{22}(x_1-\mu_1)^2 + \sigma_{11}(x_2-\mu_2)^2 - 2\rho_{12}\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}(x_1-\mu_1)(x_2-\mu_2)}{\sigma_{11}\sigma_{22}\left(1-\rho_{12}^2\right)}$$

$$= \frac{1}{1-\rho_{12}^2}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right].$$

Therefore, the bivariate (p = 2) normal density is

$$f(x_1,x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}\left(1-\rho_{12}^2\right)}}\exp\left\{-\frac{1}{2\left(1-\rho_{12}^2\right)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}.$$

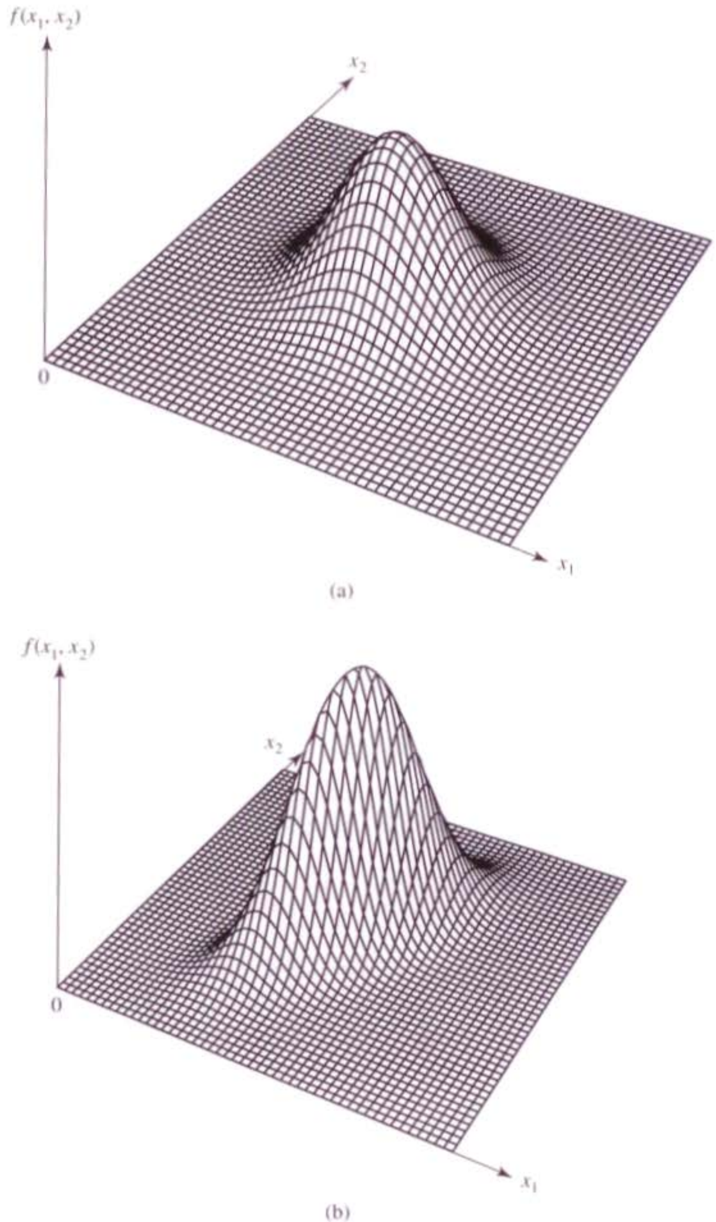

Figure 4.2 Two bivariate normal distributions. (a) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = 0$. (b) $\sigma_{11} = \sigma_{22}$ and $\rho_{12} = .75$.

4.2. The Multivariate Normal Density and its Properties

- From the multivariate normal density,

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)},$$

it is clear that the paths of x values yielding a constant height for the density are ellipsoids.

- The multivariate normal density is constant on surfaces where the square of the distance $(x-\mu)'\Sigma^{-1}(x-\mu)$ is constant. These paths are called contours:

Constant probability density contour

= {all x such that $(x-\mu)'\Sigma^{-1}(x-\mu) = c^2$ }

= surface of an ellipsoid centered at μ.

- The axes of each ellipsoid of constant density are in the direction of the eigenvectors of Σ^{-1}, and their lengths are proportional to the reciprocals of the square roots of the eigenvalues of Σ^{-1}.

- We can avoid the calculation of Σ^{-1} when determining the axes, since these ellipsoids are also determined by the eigenvalues and eigenvectors of Σ.

4.2. The Multivariate Normal Density and its Properties

- Result 4.1. If Σ is positive definite, so that Σ^{-1} exists, then

$$\Sigma e = \lambda e \quad \text{implies} \quad \Sigma^{-1} e = (1/\lambda) e$$

so (λ, e) is an eigenvalue-eigenvector pair for Σ corresponding to the pair $(1/\lambda, e)$ for Σ^{-1}. Also, Σ^{-1} is positive definite.

4.2. The Multivariate Normal Density and its Properties

- The spectral decomposition of Σ^{-1}

$$\Sigma^{-1} = \frac{1}{\lambda_1} e_1 e_1' + \frac{1}{\lambda_2} e_2 e_2' + \cdots + \frac{1}{\lambda_p} e_p e_p'$$

- $$x'\Sigma^{-1}x = x'\left(\frac{1}{\lambda_1} e_1 e_1' + \frac{1}{\lambda_2} e_2 e_2' + \cdots + \frac{1}{\lambda_p} e_p e_p' \right) x$$

$$= x' \frac{1}{\lambda_1} e_1 e_1' x + x' \frac{1}{\lambda_2} e_2 e_2' x + \cdots + x' \frac{1}{\lambda_p} e_p e_p' x$$

$$= \frac{1}{\lambda_1} x' e_1 e_1' x + \frac{1}{\lambda_2} x' e_2 e_2' x + \cdots + \frac{1}{\lambda_p} x' e_p e_p' x$$

$$= \frac{1}{\lambda_1} (e_1' x)^2 + \frac{1}{\lambda_2} (e_2' x)^2 + \cdots + \frac{1}{\lambda_p} (e_p' x)^2$$

$$= \frac{1}{\lambda_1} y_1^2 + \frac{1}{\lambda_2} y_2^2 + \cdots + \frac{1}{\lambda_p} y_p^2$$

- $\frac{1}{\lambda_1} y_1^2 + \frac{1}{\lambda_2} y_2^2 + \cdots + \frac{1}{\lambda_p} y_p^2 = c^2$ indicates an ellipsoid in $y_1 = e_1' x,\ y_2 = e_2' x, \ldots, y_p = e_p' x.$

4.2. The Multivariate Normal Density and its Properties

- With $x = c\sqrt{\lambda_1} e_1$,

$$x'\Sigma^{-1}x = \frac{1}{\lambda_1}\left(c\sqrt{\lambda_1}e_1'e_1\right)^2 + \frac{1}{\lambda_2}\left(c\sqrt{\lambda_1}e_2'e_1\right)^2 + \cdots + \frac{1}{\lambda_p}\left(c\sqrt{\lambda_1}e_p'e_1\right)^2 = c^2$$

gives the appropriate distance in the e_1 direction.

- With $x = c\sqrt{\lambda_2} e_2$,

$$x'\Sigma^{-1}x = \frac{1}{\lambda_1}\left(c\sqrt{\lambda_2}e_1'e_2\right)^2 + \frac{1}{\lambda_2}\left(c\sqrt{\lambda_2}e_2'e_2\right)^2 + \cdots + \frac{1}{\lambda_p}\left(c\sqrt{\lambda_2}e_p'e_2\right)^2 = c^2$$

gives the appropriate distance in the e_2 direction.

$$\vdots$$

- The points at distance c lie on an ellipsoid whose axes are given by the eigenvectors of Σ with lengths proportional to the square roots of the eigenvalues.

4.2. The Multivariate Normal Density and its Properties

Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by x such that

$$(x-\mu)'\Sigma^{-1}(x-\mu)=c^2.$$

These ellipsoids are centered at μ and have axes $\pm c\sqrt{\lambda_i}e_i$, where $\Sigma e_i = \lambda_i e_i$ for $i = 1,2,\ldots, p$.

- Example 4.2. Contours of the bivariate normal density

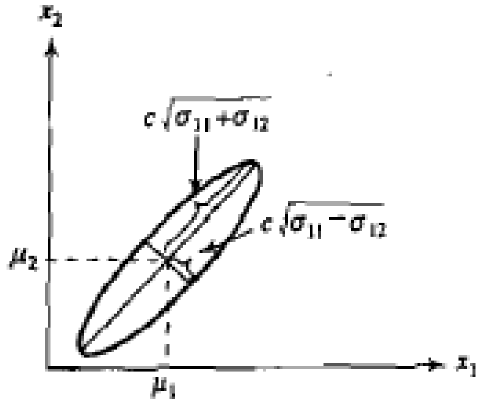

Figure 4.3 A constant-density contour for a bivariate normal distribution with $\sigma_{11} = \sigma_{22}$ and $\sigma_{12} > 0$ (or $\rho_{12} > 0$).

4.2. The Multivariate Normal Density and its Properties

- The choice $c^2 = \chi^2_p(\alpha)$, where $\chi^2_p(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with p degrees of freedom, leads to contours that contain $(1 - \alpha)\times 100\%$ of the probability.

For a p-dimensional normal distribution, the solid ellipsoid of x values satisfying

$$(x-\mu)'\Sigma^{-1}(x-\mu) \le \chi^2_p(\alpha)$$

has probability $1- \alpha$.

- See Figure 4.4 (p. 155).

- The p-variate normal density has a maximum value when the squared distance is zero – that is, when $x = \mu$.

- μ is the point of maximum density, or **mode**, as well as the expected value of X, **mean**.

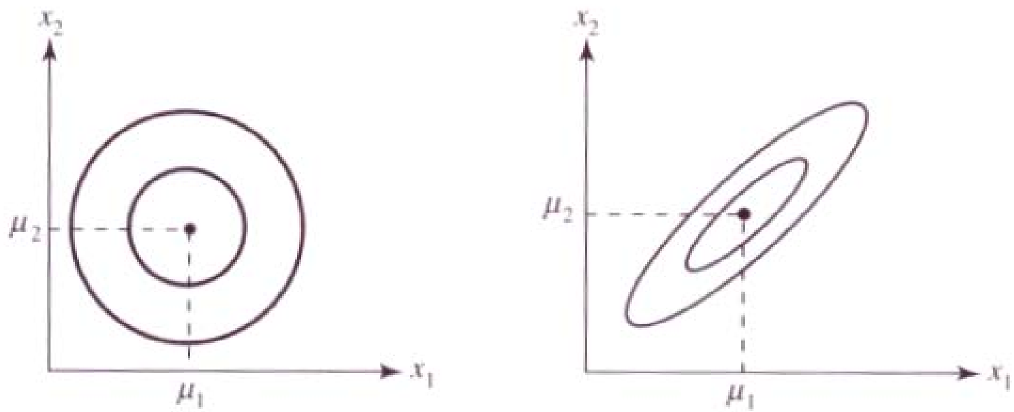

Figure 4.4 The 50% and 90% contours for the bivariate normal distributions in Figure 4.2.

Additional Properties of the Multivariate Normal Distribution

- The following are true for a random vector X having a multivariate normal distribution:

1. Linear combinations of the components of X are normally distributed.
2. All subsets of the components of X have a (multivariate) normal distribution.
3. Zero covariance implies that the corresponding components are independently distributed.
4. The conditional distributions of the components are (multivariate) normal.

Additional Properties of the Multivariate Normal Distribution

- Result 4.2. If X is distributed as $N_p(\mu, \Sigma)$, then any linear combination of variables $a'X = a_1X_1 + a_2X_2 + \cdots + a_pX_p$ is distributed as $N(a'\mu, a'\Sigma a)$. Also, if $a'X$ is distributed as $N(a'\mu, a'\Sigma a)$ for every a, then X must be $N_p(\mu, \Sigma)$.

- Example 4.3. The distribution of a linear combination of the components of a normal random vector

  Consider the linear combination $a'X$ of a multivariate normal random vector determined by the choice $a' = [1,0,\ldots,0]$.

Additional Properties of the Multivariate Normal Distribution

- Result 4.3. If X is distributed as $N_p(\mu, \Sigma)$, the q linear combinations

$$A_{(q\times p)}X_{(p\times 1)} = \begin{bmatrix} a_{11}X_1 + \cdots + a_{1p}X_p \\ a_{21}X_1 + \cdots + a_{2p}X_p \\ \vdots \\ a_{q1}X_1 + \cdots + a_{qp}X_p \end{bmatrix}$$

are distributed as $N_q(A\mu, A\Sigma A')$. Also, $X_{(p\times 1)} + d_{(p\times 1)}$, where d is a vector of constants, is distributed as $N_p(\mu + d, \Sigma)$.

- Result 4.4. All subsets of X are normally distributed. If we respectively partition X, its mean vector μ, and its covariance matrix Σ as

$$X_{(p\times 1)} = \begin{bmatrix} \underset{(q\times 1)}{X_1} \\ \hline \underset{((p-q)\times 1)}{X_2} \end{bmatrix}, \quad \mu = \begin{bmatrix} \underset{(q\times 1)}{\mu_1} \\ \hline \underset{((p-q)\times 1)}{\mu_2} \end{bmatrix}, \quad \text{and } \Sigma = \left[\begin{array}{c|c} \underset{(q\times q)}{\Sigma_{11}} & \underset{(q\times(p-q))}{\Sigma_{21}} \\ \hline \underset{((p-q)\times q)}{\Sigma_{21}} & \underset{((p-q)\times(p-q))}{\Sigma_{22}} \end{array}\right]$$

then X_1 is distributed as $N_q(\mu_1, \Sigma_{11})$.

Additional Properties of the Multivariate Normal Distribution

- Example 4.5. The distribution of a subset of a normal random vector

  - If X is distributed as $N_5(\mu, \Sigma)$, find the distribution of $\begin{bmatrix} X_2 \\ X_4 \end{bmatrix}$.

Set $X_1 = \begin{bmatrix} X_2 \\ X_4 \end{bmatrix}, \quad \mu_1 = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}.$

Then,

$$X = \begin{bmatrix} X_2 \\ X_4 \\ \hline X_1 \\ X_3 \\ X_5 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_2 \\ \mu_4 \\ \hline \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix} \quad \Sigma = \left[\begin{array}{cc|ccc} \sigma_{22} & \sigma_{24} & \sigma_{12} & \sigma_{23} & \sigma_{25} \\ \sigma_{24} & \sigma_{44} & \sigma_{14} & \sigma_{34} & \sigma_{45} \\ \hline \sigma_{12} & \sigma_{14} & \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{23} & \sigma_{34} & \sigma_{13} & \sigma_{33} & \sigma_{35} \\ \sigma_{25} & \sigma_{45} & \sigma_{15} & \sigma_{35} & \sigma_{55} \end{array}\right],$$

so, we have the distribution

$$N_2(\mu_1, \Sigma_{11}) = N_2\left(\begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix} \right).$$

Additional Properties of the Multivariate Normal Distribution

- <u>Result 4.5</u>.

  (a) If X_1 and X_2 are independent, then $Cov(X_1, X_2) = \mathbf{0}$, a $q_1 \times q_2$ matrix of zeros.

  (b) If $\left[\begin{array}{c} X_1 \\ \hline X_2 \end{array}\right]$ is $N_{q_1+q_2}\left(\left[\begin{array}{c} \mu_1 \\ \hline \mu_2 \end{array}\right], \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]\right)$, then X_1 and X_2 are independent if and only if $\Sigma_{12} = 0$.

  (c) If X_1 and X_2 are independent and are distributed as $N_{q_1}(\mu_1, \Sigma_{11})$ and $N_{q_2}(\mu_2, \Sigma_{22})$, respectively, then $\left[\begin{array}{c} X_1 \\ \hline X_2 \end{array}\right]$ has the multivariate normal distribution $N_{q_1+q_2}\left(\left[\begin{array}{c} \mu_1 \\ \hline \mu_2 \end{array}\right], \left[\begin{array}{c|c} \Sigma_{11} & 0 \\ \hline 0 & \Sigma_{22} \end{array}\right]\right)$.

- <u>Example 4.6</u>. The equivalence of zero covariance and independence for normal variables

  - Let X be $N_3(\mu, \Sigma)$ with $\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

  Are X_1 and X_2 independent? What about (X_1, X_2) and X_3?

Additional Properties of the Multivariate Normal Distribution

- Result 4.6.

Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be distributed as $N_p(\mu, \Sigma)$ with $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$,

and $|\Sigma_{22}| > 0$. Then the conditional distribution of X_1, given that $X_2 = x_2$, is normal and has

$$\text{Mean} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(x_2 - \mu_2\right)$$

and

$$\text{Covariance} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Note that the covariance does not depend on the value of x_2 of the conditioning variable.

Additional Properties of the Multivariate Normal Distribution

- Example 4.7. The conditional density of a bivariate normal distribution

  The conditional density of X_1, given that $X_2 = x_2$, is defined by

$$f(x_1 | x_2) = \{\text{conditional density of } X_1 \text{ given that } X_2 = x_2\} = \frac{f(x_1, x_2)}{f(x_2)},$$

where $f(x_2)$ is the marginal distribution of X_2. If $f(x_1, x_2)$ is the bivariate normal density, show that $f(x_1|x_2)$ is

$$N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right).$$

Proof

Note that $\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} = \sigma_{11}(1 - \rho_{12}^2)$. The two terms involving X_1 - μ_1 in the exponent of the bivariate normal density becomes, apart from the multiplicative constant $-\frac{1}{2}(1 - \rho_{12}^2)$,

$$\frac{(x_1 - \mu_1)^2}{\sigma_{11}} - 2\rho_{12}\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} = \frac{1}{\sigma_{11}}\left[x_1 - \mu_1 - \rho_{12}\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}(x_2 - \mu_2)\right]^2 - \frac{\rho_{12}^2}{\sigma_{22}}(x_2 - \mu_2)^2.$$

Additional Properties of the Multivariate Normal Distribution

- Example 4.7. The conditional density of a bivariate normal distribution

  Proof (continued)

  Because $\rho_{12} = \dfrac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}}$, or $\rho_{12}\dfrac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}} = \dfrac{\sigma_{12}}{\sigma_{22}}$, the complete exponent is

$$-\frac{1}{2\left(1-\rho_{12}^2\right)}\left(\frac{\left(x_1-\mu_1\right)^2}{\sigma_{11}} - 2\rho_{12}\frac{\left(x_1-\mu_1\right)\left(x_2-\mu_2\right)}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} + \frac{\left(x_2-\mu_2\right)^2}{\sigma_{22}}\right)$$

$$= -\frac{1}{2\sigma_{11}\left(1-\rho_{12}^2\right)}\left[x_1-\mu_1-\rho_{12}\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}\left(x_2-\mu_2\right)\right]^2 - \frac{1}{2\left(1-\rho_{12}^2\right)}\left(\frac{1}{\sigma_{22}} - \frac{\rho_{12}^2}{\sigma_{22}}\right)\left(x_2-\mu_2\right)^2$$

$$= -\frac{1}{2\sigma_{11}\left(1-\rho_{12}^2\right)}\left[x_1-\mu_1-\frac{\sigma_{12}}{\sigma_{22}}\left(x_2-\mu_2\right)\right]^2 - \frac{1}{2}\frac{\left(x_2-\mu_2\right)^2}{\sigma_{22}}$$

  The constant term $2\pi\sqrt{\sigma_{11}\sigma_{22}\left(1-\rho_{12}^2\right)}$ also factors as

$$\sqrt{2\pi}\sqrt{\sigma_{22}}\sqrt{2\pi}\sqrt{\sigma_{11}\left(1-\rho_{12}^2\right)}.$$

  Dividing the joint density of X_1 and X_2 by the marginal density

$$f\left(x_2\right) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{22}}}e^{-\frac{\left(x_2-\mu_2\right)^2}{2\sigma_{22}}},$$

  and canceling terms yields the conditional density

$$f\left(x_1 \middle| x_2\right) = \frac{f\left(x_1,x_2\right)}{f\left(x_2\right)} = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}\left(1-\rho_{12}^2\right)}}e^{-\frac{\left[x_1-\mu_1-\frac{\sigma_{12}}{\sigma_{22}}\left(x_2-\mu_2\right)\right]^2}{2\sigma_{11}\left(1-\rho_{12}^2\right)}} \qquad -\infty < x_1 < \infty.$$

Additional Properties of the Multivariate Normal Distribution

- Example 4.7. The conditional density of a bivariate normal distribution

  Proof (continued)

Thus the conditional distribution of X_1, given that $X_2 = x_2$, is

$$N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}\right).$$

Now,

$$\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} = \sigma_{11}\left(1 - \rho_{12}^2\right) \text{and } \Sigma_{12}\Sigma_{22}^{-1} = \frac{\sigma_{12}}{\sigma_{22}},$$ agreeing with Result 4.6.

Additional Properties of the Multivariate Normal Distribution

- Result 4.7. Let X be distributed as $N_p(\mu, \Sigma)$ with $|\Sigma| > 0$. Then

  (a) $(X-\mu)'\Sigma^{-1}(X-\mu)$ is distributed as χ^2_p, where χ^2_p denotes the chi-square distribution with p degrees of freedom.

  (b) The $N_p(\mu, \Sigma)$ distribution assigns probability 1 - α to the solid ellipsoid $\left\{x : (x-\mu)'\Sigma^{-1}(x-\mu) \le \chi^2_p(\alpha)\right\}$, where χ^2_p denotes the upper (100α)th percentile of the χ^2_p distribution.

  Proof of (a).

  χ^2_p is defined as the distribution of the sum $Z_1^2 + Z_2^2 + \cdots + Z_p^2$, where $Z_1, Z_2, \cdots, Z_p$ are independent $N(0,1)$ random variables.

  By the spectral decomposition, $\Sigma^{-1} = \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)e_i e_i'$, where $\Sigma e_i = \lambda_i e_i$, so $\Sigma^{-1} e_i = \left(\frac{1}{\lambda_i}\right)e_i$.

  Consequently, $(X-\mu)'\Sigma^{-1}(X-\mu) = \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)(X-\mu)' e_i e_i'(X-\mu) = \sum_{i=1}^{p}\left(\frac{1}{\lambda_i}\right)\left(e_i'(X-\mu)\right)^2$

  $$= \sum_{i=1}^{p}\left[\left(\frac{1}{\sqrt{\lambda_i}}\right)e_i'(X-\mu)\right]^2 = \sum_{i=1}^{p} Z_i^2, \text{ for instance.}$$

Additional Properties of the Multivariate Normal Distribution

- Result 4.7.

  Proof. (continued)

  Now, $Z = A(X - \mu)$, where $Z_{(p\times 1)} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix}$, $A_{(p\times p)} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} e_1' \\ \frac{1}{\sqrt{\lambda_2}} e_2' \\ \vdots \\ \frac{1}{\sqrt{\lambda_p}} e_p' \end{bmatrix}$

  and $X - \mu$ is distributed as $N_p(0, \Sigma)$. Therefore, by Result 4.3, $Z = A(X - \mu)$ is distributed as $N_p(0, A\Sigma A')$, where

$$\underset{(p\times p)}{A}\underset{(p\times p)}{\Sigma}\underset{(p\times p)}{A'} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} e_1' \\ \frac{1}{\sqrt{\lambda_2}} e_2' \\ \vdots \\ \frac{1}{\sqrt{\lambda_p}} e_p' \end{bmatrix} \left[\sum_{i=1}^{p} \lambda_i e_i e_i' \right] \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} e_1 & \frac{1}{\sqrt{\lambda_2}} e_2 & \cdots & \frac{1}{\sqrt{\lambda_p}} e_p \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{\lambda_1} e_1' \\ \sqrt{\lambda_2} e_2' \\ \vdots \\ \sqrt{\lambda_p} e_p' \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} e_1 & \frac{1}{\sqrt{\lambda_2}} e_2 & \cdots & \frac{1}{\sqrt{\lambda_p}} e_p \end{bmatrix} = I.$$

Additional Properties of the Multivariate Normal Distribution

- Result 4.7.

  Proof. (continued)

  By Result 4.5, $Z_1, Z_2, \cdots, Z_p$ are *independent* standard normal variables. Therefore, $(X-\mu)'\Sigma^{-1}(X-\mu)$ has a χ^2_p distribution.

  Proof of (b)

  Note that $P\left[(X-\mu)'\Sigma^{-1}(X-\mu)\le c^2\right]$ is the probability assigned to the ellipsoid $(X-\mu)'\Sigma^{-1}(X-\mu)\le c^2$ by the density $N_p(\mu, \Sigma)$.

  From (a), since $P\left[(X-\mu)'\Sigma^{-1}(X-\mu)\le \chi^2_p(\alpha)\right]=1-\alpha$, and (b) holds.

Additional Properties of the Multivariate Normal Distribution

- Interpretation of statistical distance
  - When X is distributed as $N_p(\mu, \Sigma)$, $(X-\mu)'\Sigma^{-1}(X-\mu)$ is the squared statistical distance from X to the population mean vector μ.
  - If one component has a much larger variance than another, it will contribute less to the squared distance.
  - Two highly correlated random variables will contribute less than two variables that are nearly uncorrelated.
  - The use of the inverse of the covariance matrix (1) standardizes all of the variables and (2) eliminates the effects of correlation.
  - From the proof of Result 4.7, $(X-\mu)'\Sigma^{-1}(X-\mu) = Z_1^2 + Z_2^2 + \cdots + Z_p^2$. In terms of $\Sigma^{-1/2}$, $Z = \Sigma^{-1/2}(X-\mu)$ has a $N_p(0, I_p)$ distribution, and $(X-\mu)'\Sigma^{-1}(X-\mu) = (X-\mu)'\Sigma^{-1/2}\Sigma^{-1/2}(X-\mu) = Z'Z = Z_1^2 + Z_2^2 + \cdots + Z_p^2$.
  - The squared statistical distance is calculated as if the random vector X were transformed to p independent standard normal random variables and then the usual squared distance, the sum of the squares of the variables, were applied.

Additional Properties of the Multivariate Normal Distribution

- Result 4.8. Let $X_1, X_2, \ldots, X_n$ be mutually independent with X_j distributed as $N_p(\mu_j, \Sigma)$. (Note that each X_j has the same covariance matrix Σ). Then

$$V_1 = c_1X_1 + c_2X_2 + \ldots + c_nX_n$$

is distributed as $N_p\left(\sum_{j=1}^{n} c_j\mu_j, \left(\sum_{j=1}^{n} c_j^2\right)\Sigma\right)$.

Moreover, V_1 and $V_2 = b_1X_1 + b_2X_2 + \ldots + b_nX_n$ are jointly multivariate normal with covariance matrix

$$\begin{bmatrix} \left(\sum_{j=1}^{n} c_j^2\right)\Sigma & (b'c)\Sigma \\ (b'c)\Sigma & \left(\sum_{j=1}^{n} b_j^2\right)\Sigma \end{bmatrix}.$$

Consequently, V_1 and V_2 are independent if $b'c = \sum_{j=1}^{n} c_j b_j = 0$.

4.3. Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation

- The $p \times 1$ vectors $X_1, X_2, \ldots, X_n$ represent a random sample from a multivariate normal population with mean vector μ and covariance matrix Σ.

  - The joint density function of all the observations is the product of the marginal normal densities:

$$\left\{ \begin{array}{c} \text{Joint density} \\ \text{of } X_1, X_2, \cdots, X_n \end{array} \right\} = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x_j - \mu)' \Sigma^{-1} (x_j - \mu)} \right\}$$

$$= \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{j=1}^{n}(x_j - \mu)' \Sigma^{-1} (x_j - \mu)}.$$

  - This function of μ and Σ for the fixed set of observations $x_1, x_2, \ldots, x_n$ is called the **likelihood**.

- **Maximum likelihood estimation** selects a value of population parameters that maximize the joint density evaluated at the observations.

  - The maximizing parameter values are called **maximum likelihood estimates**.

4.3. Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation

- Result 4.9. Let A be a $k \times k$ symmetric matrix and x be a $k \times 1$ vector. Then
  (a) $x'Ax = \mathrm{tr}(x'Ax) = \mathrm{tr}(Axx')$
  (b) $\mathrm{tr}(A) = \sum_{i=1}^{k} \lambda_i$, where the λ_i are the eigenvalues of A.

- The exponent in the joint normal density can be simplified as

$$\sum_{j=1}^{n}(x_j-\mu)'\Sigma^{-1}(x_j-\mu) = \sum_{j=1}^{n}\mathrm{tr}\left[(x_j-\mu)'\Sigma^{-1}(x_j-\mu)\right] = \sum_{j=1}^{n}\mathrm{tr}\left[\Sigma^{-1}(x_j-\mu)(x_j-\mu)'\right] = \mathrm{tr}\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\mu)(x_j-\mu)'\right)\right]$$

$$= \mathrm{tr}\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x}+\bar{x}-\mu)(x_j-\bar{x}+\bar{x}-\mu)'\right)\right] = \mathrm{tr}\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'+n(\bar{x}-\mu)(\bar{x}-\mu)'\right)\right].$$

- The likelihood function

$$L(\mu,\Sigma) = (2\pi)^{-\frac{np}{2}}|\Sigma|^{-\frac{n}{2}} e^{\left\{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'+n(\bar{x}-\mu)(\bar{x}-\mu)'\right)\right]\right\}}$$

Maximum Likelihood Estimation of μ and Σ

- Result 4.10. Given a $p\times p$ symmetric positive definite matrix B and a scalar $b > 0$, it follows that

$$\frac{1}{|\Sigma|^b} e^{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}B\right)} \le \frac{1}{|B|^b}(2b)^{pb} e^{-bp}$$

for all positive definite Σ, with equality holding only for $\Sigma = (1/2b)B$.

Proof.

Let $B^{1/2}$ be the symmetric square root of B, so $B^{1/2}B^{1/2} = B$, $B^{1/2}B^{-1/2} = I$, and $B^{-1/2}B^{-1/2} = B^{-1}$. Then, $\mathrm{tr}(\Sigma^{-1}B) = \mathrm{tr}(\Sigma^{-1}B^{1/2}B^{1/2}) = \mathrm{tr}[(\Sigma^{-1}B^{1/2})B^{1/2}] = \mathrm{tr}[B^{1/2}(\Sigma^{-1}B^{1/2})]$.

Let η be an eigenvalue of $B^{1/2}\Sigma^{-1}B^{1/2}$. This matrix is positive definite because $y'B^{1/2}\Sigma^{-1}B^{1/2}y = (B^{1/2}y)'\,\Sigma^{-1}(B^{1/2}y) > 0$ if $B^{1/2}y \neq 0$ or, equivalently, $y \neq 0$. Thus, the eigenvalues η_i of $B^{1/2}\Sigma^{-1}B^{1/2}$ are positive.

Result 4.9(b) then gives

$$tr\left(\Sigma^{-1}B\right) = tr\left(B^{1/2}\Sigma^{-1}B^{1/2}\right) = \sum_{i=1}^{p}\eta_i \quad \text{and} \quad \left|B^{1/2}\Sigma^{-1}B^{1/2}\right| = \prod_{i=1}^{p}\eta_i .$$

Then $\left|B^{1/2}\Sigma^{-1}B^{1/2}\right| = \left|B^{1/2}\right|\left|\Sigma^{-1}\right|\left|B^{1/2}\right| = \left|\Sigma^{-1}\right|\left|B^{1/2}\right|\left|B^{1/2}\right| = \left|\Sigma^{-1}\right|\left|B\right| = \frac{1}{|\Sigma|}|B|$

or $\displaystyle \frac{1}{|\Sigma|} = \frac{\left|B^{1/2}\Sigma^{-1}B^{1/2}\right|}{|B|} = \frac{\prod_{i=1}^{p}\eta_i}{|B|}.$

Maximum Likelihood Estimation of μ and Σ

- Result 4.10. *Proof.* (continued)

Combining the results for the trace and the determinant yields

$$\frac{1}{|\Sigma|^b} e^{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}B\right)} = \frac{\left(\prod_{i=1}^{p} \eta_i\right)^b}{|B|^b} e^{-\frac{1}{2}\sum_{i=1}^{p}\eta_i} = \frac{1}{|B|^b} \prod_{i=1}^{p} \eta_i^b e^{-\frac{\eta_i}{2}}.$$

The function $\eta^b e^{-\eta/2}$ has a maximum, with respect to η, of $(2b)^b e^{-b}$, occurring at $\eta = 2b$. The choice $\eta_i = 2b$, for each i, therefore gives

$$\frac{1}{|\Sigma|^b} e^{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}B\right)} \le \frac{1}{|B|^b} (2b)^{pb} e^{-bp}.$$

The upper bound is uniquely attained when $\Sigma = (1/2b)B$, since, for this choice,

$$B^{1/2}\Sigma^{-1}B^{1/2} = B^{1/2}(2b)B^{-1}B^{1/2} = (2b)I_{(p\times p)}$$

and $tr\left(\Sigma^{-1}B\right) = tr\left(B^{1/2}\Sigma^{-1}B^{1/2}\right) = tr\left((2b)I\right) = 2bp.$

Moreover,

$$\frac{1}{|\Sigma|} = \frac{\left|B^{1/2}\Sigma^{-1}B^{1/2}\right|}{|B|} = \frac{|(2b)I|}{|B|} = \frac{(2b)^p}{|B|}.$$

Straightforward substitution for $\mathrm{tr}[\Sigma^{-1}B]$ and $1/|\Sigma|^b$ yields the bound asserted.

Maximum Likelihood Estimation of μ and Σ

- Result 4.11. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and covariance Σ. Then

$$\hat{\mu} = \bar{X} \text{ and } \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})' = \frac{(n-1)}{n}S$$

are the **maximum likelihood estimators** of μ and Σ, respectively. Their observed values, $\bar{x}$ and $\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'$, are called the **maximum likelihood estimates** of μ and Σ.

Proof.

The exponent in the likelihood function, apart from the multiplicative factor -1/2, is

$$tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'\right)\right] + n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu).$$

By Result 4.1, Σ^{-1} is positive definite, so the distance $(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) > 0$ unless $\mu = \bar{x}$. Thus, the likelihood is maximized with respect to μ at $\hat{\mu} = \bar{x}$. It remains to maximize

$$L(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'\right)\right]}$$

over Σ.

Maximum Likelihood Estimation of μ and Σ

- Result 4.11. *Proof.* (continued)

  By Result 4.10 with $b = n/2$ and $B = \sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'$, the maximum occurs at $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'$, as stated.

  The maximum likelihood estimators are random quantities. They are obtained by replacing the observations $x_1, x_2, \ldots, x_n$ in the expressions for $\hat{\mu}$ and $\hat{\Sigma}$ with the corresponding random vectors, $X_1, X_2, \ldots, X_n$.

Maximum Likelihood Estimation of μ and Σ

- The maximum likelihood estimator $\overline{X}$ is a random vector and the maximum likelihood estimator $\hat{\Sigma}$ is a random matrix.

  - The maximum likelihood estimates are their particular values for the given data set.

  - The maximum of the likelihood is

$$L\left(\hat{\mu},\hat{\Sigma}\right)=\frac{1}{(2\pi)^{np/2}}e^{-np/2}\frac{1}{\left|\hat{\Sigma}\right|^{n/2}}$$

or since $\left|\hat{\Sigma}\right|=\left[(n-1)/n\right]^{p}\left|S\right|$,

$$L\left(\hat{\mu},\hat{\Sigma}\right) = \text{constant} \times (\text{generalized variance})^{-n/2}$$

  - The generalized variance determines the “peakedness” of the likelihood function, and consequently, is a natural measure of variability when the parent population is multivariate normal.

Maximum Likelihood Estimation of μ and Σ

- **Invariance property**

  - $\hat{\theta}$: the maximum likelihood estimator of θ

  - The *maximum likelihood estimate* of $h(\theta)$, a function of θ, is given by $h\left(\hat{\theta}\right)$.

  - Example.

  1. The maximum likelihood estimator of $\mu'\Sigma^{-1}\mu$ is $\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu}$, where $\hat{\mu} = \overline{X}$ and $\hat{\Sigma} = \left((n-1)/n\right)S$ are the maximum likelihood estimators of μ and Σ, respectively.

  2. The maximum likelihood estimator of $\sqrt{\sigma_{ii}}$ is $\sqrt{\hat{\sigma}_{ii}}$, where

$$\hat{\sigma}_{ii} = \frac{1}{n}\sum_{j=1}^{n}\left(X_{ij} - \overline{X}_i\right)^2$$

  is the maximum likelihood estimator of $\sigma_{ii} = Var(X_i)$.

Sufficient Statistics

- For a random sample, X_1, X_2,…, X_n, from a multivariate normal population with mean μ and covariance Σ, $\overline{X}$ and S are *sufficient statistics*.

 - The joint density depends on the whole set of observations x_1, x_2,…, x_n only through the sample mean $\overline{x}$ and the sum-of-squares-and-cross-product matrix $\sum_{j=1}^{n}\left(X_j-\overline{X}\right)\left(X_j-\overline{X}\right)'=(n-1)S$.

 Then, $\overline{x}$ and $(n$ - 1$)S$ (or S) are **sufficient statistics**.

 - This is generally not true for nonnormal populations. If the data cannot be regarded as multivariate normal, techniques that depend solely on $\overline{x}$ and S may be ignoring other useful sample information.

4.4. The Sampling Distribution of $\overline{X}$ and S

- In the univariate case (p = 1), $\overline{X}$ is normal with mean μ = (population mean) and variance $\frac{1}{n}\sigma^2 = \frac{\text{population variance}}{\text{sample size}}$.
- For the multivariate case ($p \geq 2$), $\overline{X}$ has a normal distribution with mean vector μ and covariance matrix $(1/n)\Sigma$.

- In the univariate case, recall $(n-1)s^2 = \sum_{j=1}^{n}(X_j - \overline{X})^2$ is distributed as σ^2 times a chi-square variable having n – 1 degrees of freedom (d.f.).

- This chi-square is the distribution of a sum of squares of independent standard normal random variables.

- $(n - 1)s^2$ is distributed as $\sigma^2\left(Z_1^2 + \cdots + Z_{n-1}^2\right) = (\sigma Z_1)^2 + \cdots + (\sigma Z_{n-1})^2$.

- The individual terms σZ_i are independently distributed as $N(0, \sigma^2)$.

4.4. The Sampling Distribution of $\bar{X}$ and S

- The sampling distribution of the sample covariance matrix is called the **Wishart distribution**:

  $W_m(\cdot|\Sigma)$ = Wishart distribution with m d.f.

  $$= \text{distribution of } \sum_{j=1}^{m} Z_j Z_j',$$

  where the Z_j are each independently distributed as $N_p(0, \Sigma)$.

- Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a p-variate normal distribution with mean μ and covariance matrix Σ. Then

1. $\bar{X}$ is distributed as $N_p(\mu, (1/n)\Sigma)$.
2. $(n - 1)S$ is distributed as a Wishart random matrix with $n - 1$ d.f.
3. $\bar{X}$ and S are independent.

- Because Σ is unknown, the distribution of $\bar{X}$ cannot be used directly to make inference about μ. However, S provides independent information about Σ, and the distribution of S does not depend on μ.

4.4. The Sampling Distribution of $\overline{X}$ and S

- Properties of the Wishart Distribution

1. If A_1 is distributed as $W_{m_1}(A_1|\Sigma)$ independently of A_2, which is distributed as $W_{m_2}(A_2|\Sigma)$, then $A_1 + A_2$ is distributed as $W_{m_1+m_2}(A_1 + A_2|\Sigma)$. That is, the degrees of freedom add.
2. If A is distributed as $W_m(A|\Sigma)$, then CAC' is distributed as $W_m(CAC'|C\Sigma C')$.

- The probability density function of the Wishart distribution at the positive definite matrix A is

$$w_{n-1}(A|\Sigma)=\frac{|A|^{(n-p-2)/2}e^{-\mathrm{tr}[\mathrm{A}\Sigma^{-1}]/2}}{2^{p(n-1)/2}\pi^{p(p-1)/4}|\Sigma|^{(n-1)/2}\prod_{i=1}^{p}\Gamma\left(\frac{1}{2}(n-i)\right)},$$

where $\Gamma(\cdot)$ is the gamma function.

4.5. Large-Sample Behavior of $\bar{X}$ and S

- Suppose the quantity X is determined by a large number of independent causes $V_1, V_2, \ldots, V_n$, where the random variable V_i representing the causes have approximately the same variability. If X is the sum

$$X = V_1 + V_2 + \cdots + V_n,$$

then the central limit theorem applies, and conclude that X has a distribution that is nearly normal. This is true for virtually any parent distribution of the V_i's, provided that n is large enough.

- The univariate central limit theorem also tells us that the sampling distribution of the sample mean, $\bar{X}$, for a large sample size is nearly normal, whatever the form of the underlying population distribution is.

- Certain multivariate statistics, like $\bar{X}$ and S, have large sample properties analogous to their univariate counterparts.

4.5. Large-Sample Behavior of $\overline{X}$ and S

- Result 4.12 (**Law of large numbers**). Let $Y_1, Y_2, \ldots, Y_n$ be independent observations from a population with mean $E(Y_i) = \mu$. Then

$$\overline{Y} = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$$

converges in probability to μ as n increases without bound. That is, for any prescribed accuracy $\varepsilon > 0$, $P\left[-\varepsilon < \overline{Y} - \mu < \varepsilon\right]$ approaches unity as $n \to \infty$.

- The law of large numbers says that each $\overline{X}_i$ converges in probability to μ_i, $i = 1,2,\ldots, p$. Therefore, $\overline{X}$ converges in probability to μ.

- Similarly, since each sample covarianace s_{ik} converges in probability to σ_{ik}, $i, k = 1,2,\ldots, p$, S (or $\hat{\Sigma} = S_n$) converges in probability to Σ, since

$$(n-1)s_{ik} = \sum_{j=1}^{n}\left(X_{ji} - \overline{X}_i\right)\left(X_{jk} - \overline{X}_k\right) = \sum_{j=1}^{n}\left(X_{ji} - \mu_i + \mu_i - \overline{X}_i\right)\left(X_{jk} - \mu_k + \mu_k - \overline{X}_k\right)$$

$$= \sum_{j=1}^{n}\left(X_{ji} - \mu_i\right)\left(X_{jk} - \mu_k\right) + n\left(\overline{X}_i - \mu_i\right)\left(\overline{X}_k - \mu_k\right)$$

and the first term converges to σ_{ik} and the second term converges to zero, by applying the law of large numbers.

4.5. Large-Sample Behavior of $\bar{X}$ and S

- Result 4.13. (**The central limit theorem**). Let $X_1, X_2, \ldots, X_n$ be independent observations from any population with mean vector μ and finite covariance Σ. Then $\sqrt{n}\left(\bar{X}-\mu\right)$ has an approximate $N_p(0, \Sigma)$ distribution for large sample sizes. Here n should also be large relative to p.

- The approximation provided by the central limit theorem applies to discrete, as well as continuous, multivariate populations.
  - $\bar{X}$ is exactly normally distributed when the underlying population is normal. Therefore, the central limit theorem approximation would be quite good for moderate n when the parent population is nearly normal.

- $n\left(\bar{X}-\mu\right)'\Sigma^{-1}\left(\bar{X}-\mu\right)$ has a χ^2_p distribution when $\bar{X}$ is distributed as $N_p(\mu, (1/n)\Sigma)$, or, equivalently, when $\sqrt{n}\left(\bar{X}-\mu\right)$ has an $N_p(0, \Sigma)$ distribution. The χ^2_p distribution is *approximately* the sampling distribution of $n\left(\bar{X}-\mu\right)'\Sigma^{-1}\left(\bar{X}-\mu\right)$ when $\bar{X}$ is approximately normally distributed. Replacing Σ^{-1} by S^{-1} does not seriously affect this approximation for n large and much greater than p.

4.5. Large-Sample Behavior of $\overline{X}$ and S

- Let $X_1, X_2, \ldots, X_n$ be independent observations from a population with mean μ and finite (nonsingular) covariance Σ. Then

  $\sqrt{n}\left(\overline{X} - \mu\right)$ is approximately $N_p(0, \Sigma)$

  and

  $n\left(\overline{X} - \mu\right)' S^{-1}\left(\overline{X} - \mu\right)$ is approximately χ^2_p

  for n - p large.

4.6 Assessing the Assumption of Normality

- Many statistical techniques assume that each vector observation X_j comes from a multivariate normal distribution.

- **Evaluating the Normality of the Univariate Marginal Distributions**
  - Dot diagram for smaller n and histograms for $n > 25$ help reveal situations where one tail of a univariate distribution is much longer than the other.
  - Can check further by counting the number of observations in certain interval based on a univariate normal distribution.
  - Q-Q plots can be used to assess the assumption of normality.

- The Q-Q plot

1. Order the original observations to get $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ and their corresponding probability values $(1 - 0.5)/n$, $(2 - 0.5)/n$, …, $(n - 0.5)/n$;
2. Calculate the standard normal quantiles $q_{(1)}, q_{(2)}, \ldots, q_{(n)}$,

$$P\left[Z \le q_{(j)}\right] = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \frac{j - 0.5}{n};$$

3. Plot the pairs of observations $(q_{(1)}, x_{(1)})$, $(q_{(2)}, x_{(2)})$, …, $(q_{(n)}, x_{(n)})$, and examine the "straightness" of the outcome.

- If data arise from a normal population, the pairs $(q_{(j)}, x_{(j)})$ will be approximately linearly related.

4.6 Assessing the Assumption of Normality

- The straightness of the *Q*-*Q* plot can be measured by calculating the correlation coefficient of the points in the plot:

$$r_Q = \frac{\sum_{j=1}^{n}\left(x_{(j)} - \bar{x}\right)\left(q_{(j)} - \bar{q}\right)}{\sqrt{\sum_{j=1}^{n}\left(x_{(j)} - \bar{x}\right)^2}\sqrt{\sum_{j=1}^{n}\left(q_{(j)} - \bar{q}\right)^2}}.$$

4.6 Assessing the Assumption of Normality

- **Evaluating the Multivariate Normality**

- A somewhat more formal method for judging the joint normality of a data set is based on the squared generalized distances

$$d_j^2 = (x_j - \bar{x})' S^{-1} (x_j - \bar{x}), \quad j = 1, 2, \ldots, n,$$

where $x_1, x_2, \ldots, x_n$ are the sample observations.

- When the parent population is multivariate normal and both n and $n - p$ are greater than 25 or 30, each of the squared distances $d_1^2, d_2^2, \ldots, d_n^2$ should behave like a chi-square random variable.

- The chi-square plot

1. Order the squared distances $d_1^2, d_2^2, \ldots, d_n^2$ from smallest to largest as $d_{(1)}^2, d_{(2)}^2, \ldots, d_{(n)}^2$.
2. Graph the pairs $(q_{c,p}((j - 0.5)/n), d_{(j)}^2)$, where $q_{c,p}((j - 0.5)/n)$ is the $100(j - 0.5)/n$ quantile of the chi-square distribution with p degrees of freedom.

- The plot should resemble a straight line through the origin having slope 1.

4.7 Detecting Outliers and Cleaning Data

- Steps for detecting outliers

1. Make a dot plot for each variable.
2. Make a scatter plot for each pair of variables.
3. Calculate the standardized values

$$z_{jk} = \frac{(x_{jk} - \bar{x}_k)}{\sqrt{s_{kk}}}$$

for j = 1, 2, …, n and for each column k = 1, 2, …, p. Examine these standardized values for large or small values.

4. Calculate the generalized squared distances $(x_j - \bar{x})' S^{-1} (x_j - \bar{x})$. Examine these distances for unusually large values. In a chi-square plot, these would be the points farthest from the origin.

4.8 Transformations to Near Normality

- If normality is not a viable assumption, one alternative is to make nonnormal data more "normal looking" by considering **transformations** of the data.

  - Normal-theory analysis can then be carried out with the suitably transformed data.

- Helpful Transformations To Near Normality

Original Scale	Transformed Scale
Counts, y	$\sqrt{y}$
Proportions, $\hat{p}$	$\text{logit}(\hat{p}) = \frac{1}{2}\log\left(\frac{\hat{p}}{1-\hat{p}}\right)$
Correlations, r	Fisher's $z(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$

- Box-Cox Transformations

$$x^{(\lambda)} = \begin{cases} \dfrac{x^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log x & \lambda = 0 \end{cases}$$

  - With multivariate observations, a power transformation must be selected for each of the variables.