17.2 THE MULTIVARIATE NORMAL DISTRIBUTION

17.2.1 The univariate normal

A real random variable $X$ has a normal (or Gaussian) distribution with mean (mode and median) $m$, and variance $V$ if and only if

$$p(X) = (2\pi)^{-1/2} \exp \left[ -\frac{(X - m)^2}{2V} \right], \quad (-\infty < X < \infty).$$

In this case we write $X \sim N[m, V]$.

SUMS OF NORMAL RANDOM VARIABLES.

If $X_i \sim N[m_i, V_i]$, $(i = 1, \ldots, n)$, have covariances $C[X_i, X_j] = c_{ij} = c_{ji}$, then $Y = \sum_{i=1}^{n} a_iX_i + b$ has a normal distribution with

$$E[Y] = \sum_{i=1}^{n} a_im_i + b \quad \text{and} \quad V[Y] = \sum_{i=1}^{n} a_i^2V_i + 2\sum_{i=1}^{n} \sum_{j=1}^{i-1} a_ia_jc_{ij},$$

where $a_i$ and $b$ are constants. In particular, $c_{ij} = 0$ for all $i \neq j$ if and only if the $X_i$ are independent normal, in which case $V[Y] = \sum_{i=1}^{n} a_i^2V_i$.

17.2.2 The multivariate normal

A random $n$-vector $X$ has a multivariate normal distribution in $n$ dimensions if and only if $Y = \sum_{i=1}^{n} a_iX_i$ is normal for all constant, non-zero vectors $a = (a_1, \ldots, a_n)$.

If $X$ is multivariate normal, then $E[X] = m$ and $V[X] = V$ exist, and we use the notation $X \sim N[m, V]$. The moments $m$ and $V$ completely define the distribution whose density is

$$p(X) = \{(2\pi)^{n/2}|V|^{-1/2} \exp \left[ -(X - m)^tV^{-1}(X - m)/2 \right].$$

The subvectors of $X$ are independent if and only if they are uncorrelated. In particular, if $V$ is block diagonal, then the corresponding subvectors of $X$ are mutually independent.

LINEAR TRANSFORMATIONS.

For any constant $A$ and $b$ of suitable dimensions, if $Y = AX + b$, then $Y \sim N[Am + b, AVA']$. If $AVA'$ is diagonal, then the elements of $Y$ are independent normal.

LINEAR FORMS.

Suppose $X_i \sim N[m_i, V_i]$ independently, $i = 1, \ldots, k$, and consider constant matrices and vectors $A_1, \ldots, A_k$ and $b$ of suitable dimensions; $Y = \sum_{i=1}^{k} A_iX_i + b$ is multivariate normal with mean $\sum_{i=1}^{k} A_im_i + b$ and variance matrix $\sum_{i=1}^{k} A_iV_iA_i'$.

MARGINAL DISTRIBUTIONS.

Suppose that we have conformable partitions

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \quad \text{and} \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}.$$

Then $X_i \sim N[m_i, V_{ii}]$, $i = 1, 2$. In particular, if $X_i$ is univariate normal, $X_i \sim N[m_i, V_{ii}]$ for $i = 1, \ldots, n$.

BIVARIATE NORMAL.

Any two elements $X_i$ and $X_j$ of $X$ are bivariate normal with joint density

$$p_{ij}(x_i, x_j) = \left[ (2\pi)\sqrt{V_{ii}V_{jj}(1-\rho_{ij}^2)} \right]^{-1} \exp \left[ -Q(x_1, x_2)/2 \right],$$

where $Q(x_1, x_2)$ is the quadratic form

$$Q(x_1, x_2) = \frac{(x_i - m_i)^2}{V_{ii}} + \frac{(x_j - m_j)^2}{V_{jj}} - 2\rho_{ij} \frac{(x_i - m_i)}{\sqrt{V_{ii}}} \frac{(x_j - m_j)}{\sqrt{V_{jj}}}.$$
with \( \rho_{ij} = \text{corr}(X_i, X_j) = V_{ij}/\sqrt{V_{ii}V_{jj}} \).

**CONDITIONAL DISTRIBUTIONS.**

For the partition of \( X' \) into \( X_1' \) and \( X_2' \), we have

1. \( (X_1 | X_2) \sim N[m_1(X_2), V_{11}(X_2)] \), where
   \[
   m_1(X_2) = m_1 + V_{12}V_{22}^{-1}(X_2 - m_2)
   \]
   and
   \[
   V_{11}(X_2) = V_{11} - V_{12}V_{22}^{-1}V_{21}.
   \]
   The matrix \( A_1 = V_{12}V_{22}^{-1} \) is called the *regression matrix* of \( X_1 \) on \( X_2 \). The conditional moments are given in terms of the regression matrix by
   \[
   m_1(X_2) = m_1 + A_1(X_2 - m_2)
   \]
   and
   \[
   V_{11}(X_2) = V_{11} - A_1V_{22}A_1'.
   \]

2. \( (X_2 | X_1) \sim N[m_2(X_1), V_{22}(X_1)] \), where
   \[
   m_2(X_1) = m_2 + V_{21}V_{11}^{-1}(X_1 - m_1)
   \]
   and
   \[
   V_{22}(X_1) = V_{22} - V_{21}V_{11}^{-1}V_{12}.
   \]
   The matrix \( A_2 = V_{21}V_{11}^{-1} \) is called the *regression matrix* of \( X_2 \) on \( X_1 \). The conditional moments are given in terms of the regression matrix by
   \[
   m_2(X_1) = m_2 + A_2(X_1 - m_1)
   \]
   and
   \[
   V_{22}(X_1) = V_{22} - A_2V_{11}A_2'.
   \]

3. In the special case of the bivariate normal in (1) above, the moments are all scalars, and the correlation between \( X_1 \) and \( X_2 \) is \( \rho = \rho_{12} = V_{12}/\sqrt{V_{11}V_{22}} \). The regressions are determined by regression coefficients \( A_1 \) and \( A_2 \) given by
   \[
   A_1 = \rho\sqrt{V_{11}/V_{22}} \quad \text{and} \quad A_2 = \rho\sqrt{V_{22}/V_{11}}.
   \]
   Also
   \[
   V[X_1 | X_2] = (1 - \rho^2)V_{11} \quad \text{and} \quad V[X_2 | X_1] = (1 - \rho^2)V_{22}.
   \]

17.2.3 *Conditional normals and linear regression*

Many of the important results in this book may be derived directly from the multivariate normal theory reviewed above. A particular regression model is reviewed here to provide the setting for those results.

Suppose that the \( p \)-vector \( Y \) and the \( n \)-vector \( \theta \) are related via the conditional distribution

\[
(Y | \theta) \sim N[F'\theta, V],
\]

where the \((n \times p)\) matrix \( F \) and the \((p \times p)\) positive definite symmetric matrix \( V \) are constant. An equivalent statement is

\[
Y = F'\theta + \nu,
\]
where \( \nu \sim N[0, V] \). Suppose further that the marginal distribution of \( \theta \) is given by

\[
\theta \sim N[a, R],
\]

where both \( a \) and \( R \) are constant, and that \( \theta \) is independent of \( \nu \). Equivalently,

\[
\theta = a + \omega,
\]

where \( \omega \sim N[0, R] \) independently of \( \nu \).

From these distributions it is possible to construct the joint distribution for \( Y \) and \( \theta \) and hence both the marginal for \( Y \) and the conditional for \( (\theta \mid Y) \).

**Multivariate Joint Normal Distribution.**

Since \( \theta = a + \omega \) and \( Y = F^T \theta + \nu = F^T a + F^T \omega + \nu \), then the vector \((Y', \theta')'\) is a linear transformation of \((\nu', \omega')'\). By construction the latter has a multivariate normal distribution, so that \( Y \) and \( \theta \) are jointly normal. Further

1. \( E[\theta] = a \) and \( V[\theta] = R \);
2. \( E[Y] = E[F^T \theta + \nu] = F^T E[\theta] + E[\nu] = F^T a \) and
3. \( C[Y, \theta] = C[F^T \theta + \nu, \theta] = F^T C[\theta, \theta] + C[\nu, \theta] = F^T R F \).

It follows that

\[
\begin{pmatrix} Y \\ \theta \end{pmatrix} \sim N \left[ \begin{pmatrix} F^T a \\ 0 \end{pmatrix}, \begin{pmatrix} F^T R F + V & F^T R \\ R F & R \end{pmatrix} \right].
\]

Therefore, identifying \( Y \) with \( X_1 \) and \( \theta \) with \( X_2 \) in the partition of \( X \) in 17.2.2, we have

4. \( Y \sim N[F^T a, F^T R F + V]; \)
5. \( (\theta \mid Y) \sim N[m, C], \) where

\[
m = a + RF[F^T R F + V]^{-1} [Y - F^T a]
\]

and

\[
C = R - RF[F^T R F + V]^{-1} F R.
\]

By defining \( e = Y - F^T a \), \( Q = F^T R F + V \) and \( A = RFQ^{-1} \), these equations become

\[
m = a + Ae \quad \text{and} \quad C = R - AQA'.
\]

**Multivariate Bayes’ Theorem.**

An alternative derivation of the conditional distribution for \((\theta \mid Y)\) via Bayes’ Theorem provides alternative expressions for \( m \) and \( C \). Note that

\[
p(\theta \mid Y) \propto p(Y \mid \theta)p(\theta)
\]

as a function of \( \theta \), so that

\[
\ln[p(\theta \mid Y)] = k + \ln[p(Y \mid \theta)] + \ln[p(\theta)],
\]

where \( k \) depends on \( Y \) but not on \( \theta \). Therefore

\[
\ln[p(\theta \mid Y)] = k - \frac{1}{2} \left[ (Y - F^T a)' V^{-1} (Y - F^T a) + (\theta - a)' R^{-1} (\theta - a) \right].
\]

The bracketed term here is simply

\[
Y' V^{-1} Y - 2Y' V^{-1} F^T \theta + \theta F^T V F \theta + \theta R^{-1} \theta - 2a' R^{-1} \theta + a' R^{-1} a
\]

\[
= \theta' \left[ F V^{-1} F^T + R^{-1} \right] \theta - 2 \left[ Y' V^{-1} F^T + a' R^{-1} \right] \theta + h
\]
where $h$ depends on $Y$ but not on $\theta$. Completing the quadratic form gives
\[(\theta - m)^\top C^{-1} (\theta - m) + h^* ,\]
where again $h^*$ does not involve $\theta$,
\[C^{-1} = R^{-1} + FV^{-1}F',\]
and
\[m = C \{ FV^{-1} Y + R^{-1} a \} .\]
Hence
\[p(\theta \mid Y) \propto \exp \left[ -\frac{1}{2} (\theta - m)^\top C^{-1} (\theta - m) \right] \]
as a function of $\theta$, so that \((\theta \mid Y) \sim N[\mathbf{m}, C]\), just as derived earlier.

Note that the two derivations give different expressions for $m$ and $C$ that provide, in particular, the matrix identity for $C$ given by
\[C = [R^{-1} + FV^{-1}F']^{-1} = R - RF[RFR + V]^{-1}F'R,\]
that is easily verified once stated.

### 17.3 Joint Normal/Gamma Distributions

#### 17.3.1 The Gamma Distribution

A random variable $\phi > 0$ has a gamma distribution with parameters $n > 0$ and $d > 0$, denoted by $\phi \sim G[n, d]$, if and only if
\[p(\phi) \propto \phi^{n-1} \exp(-\phi d), \quad (\phi > 0).\]
Normalisation leads to $p(\phi) = d^n \Gamma(n)^{-1} \phi^{n-1} \exp(-\phi d)$, where $\Gamma(n)$ is the gamma function. Note that $E[\phi] = n/d$ and $V[\phi] = E[\phi^2]/n$.

Two special cases of interest are
1. $n = 1$, when $\phi$ has a (negative) exponential distribution with mean $1/d$;
2. $\phi \sim G[n/2, d/2]$ when $n$ is a positive integer. In this case $d\phi \sim \chi_n^2$, a chi-squared distribution with $n$ degrees of freedom.

#### 17.3.2 Univariate Normal/Gamma Distribution

Let $\phi \sim G[n/2, d/2]$ for any $n > 0$ and $d > 0$, and suppose that the conditional distribution of a further random variable $X$ given $\phi$ is normal $(X \mid \phi) \sim N[m, C\phi^{-1}]$, for some $m$ and $C$. Note that $E[X] \equiv E[X \mid \phi] = m$ does not depend on $\phi$. However $V[X \mid \phi] = C\phi^{-1}$. The joint distribution of $X$ and $\phi$ is called (univariate) normal/gamma. Note that
\[p(X, \phi) = \left( \frac{\phi}{2\pi C} \right)^{1/2} \exp \left[ -\frac{(X - m)^2}{2C} \right] \times \frac{d^{n/2}}{2^{n/2} \Gamma(n/2)} \phi^{n-1} \exp \left[ -\frac{\phi d}{2} \right] \times \phi^{(n+1)/2} \exp \left[ -\frac{\phi (X - m)^2}{2C} + d \right] ,\]
as a function of $\phi$ and $X$. 

\[ p(\phi \mid X) \propto \phi^{(\frac{n+1}{2})-1} \exp \left[ -\frac{\phi}{2} \left\{ \frac{(X-m)^2}{C} + d \right\} \right]. \]

so that

\[ (\phi \mid X) \sim \mathcal{G} \left[ \frac{n^*}{2}, \frac{d^*}{2} \right], \]

where \( n^* = n + 1 \) and \( d^* = d + (X - m)^2/C \).

(3) \[ p(X) = p(X, \phi)/p(\phi \mid X) \]

\[ \propto [n + (X - m)^2/R]^{-(n+1)/2}, \]

where \( R = C(d/n) = C/E[\phi] \). This is proportional to the density of the Student T distribution with \( n \) degrees of freedom, mode \( m \) and scale \( R \). Hence \( X \sim T_\nu[m, R] \), or \((X - m)/R^{1/2} \sim T_\nu[0,1] \), a standard Student T distribution with \( n \) degrees of freedom.

### 17.3.3 Multivariate normal/gamma distribution

As an important generalisation, suppose that \( \phi \sim \mathcal{G}[n/2, d/2] \) and that the \( p \)-vector \( X \) is normally distributed conditional on \( \phi \), as \( X \sim \mathcal{N}[m, C\phi^{-1}] \). Here the \( p \)-vector \( m \) and the \((p \times p)\) symmetric positive definite matrix \( C \) are known. Thus, each element of \( V[X] \) is scaled by the common factor \( \phi \). The basic results are similar to 17.3.2 in that

1. \((\phi \mid X) \sim \mathcal{G}[n^*/2, d^*/2], \) where \( n^* = n + p \) and \( d^* = d + (X - m)\mathcal{C}^{-1}(X - m)/2 \) (notice the degrees of freedom increases by \( p \)).
2. \( X \) has a (marginal) multivariate T distribution in \( p \) dimensions with \( n \) degrees of freedom, mode \( m \) and scale matrix \( R = C(d/n) = C/E[\phi] \), denoted by \( X \sim T_{\nu}[m, R] \), with density

\[ p(X) \propto [n + (X - m)\mathcal{R}^{-1}(X - m)]^{-(n+p)/2}. \]

In particular, if \( X_i \) is the \( i \)-th element of \( X \), \( m_i \) and \( C_{ii} \) the corresponding mean and diagonal element of \( C \), then

\[ X_i \sim T_{\nu}[m_i, R_{ii}], \]

where \( R_{ii} = C_{ii}(d/n) \).

### 17.3.4 Simple regression model

The normal/gamma distribution plays a key role in providing closed form Bayesian analyses of linear models with unknown scale parameters. Details may be found in De Groot (1971) and Press (1985), for example. A particular regression setting is reviewed here for reference. The details follow from the above joint normal/gamma theory. Suppose that a scalar variable \( Y \) is related to the \( p \)-vector \( \theta \) and the scalar \( \phi \) via

\[ (Y \mid \theta, \phi) \sim \mathcal{N}[F\theta, k\phi^{-1}], \]

where the \( p \)-vector \( F \) and the variance multiple \( k \) are fixed constants. Suppose also that \((\theta, \phi)\) have a joint normal/gamma distribution, namely

\[ (\theta \mid \phi) \sim \mathcal{N}[a, R\phi^{-1}] \]

and

\[ \phi \sim \mathcal{G}[n/2, d/2] \]

for fixed scalars \( n > 0, d > 0 \), \( p \)-vector \( a \) and \((p \times p)\) variance matrix \( R \), and let \( S = d/n = 1/E[\phi] \). Then

1. \((Y \mid \phi) \sim \mathcal{N}[f, Q\phi^{-1}], \) where \( f = Fa \) and \( Q = FRF + k; \)
The probability that 

If this probability is high, then connected regions, or intervals, and these provide interval-based inferences and tests of hypotheses through HPD regions the use of posterior normal, T and F distributions, described below. Fuller theoretical details are provided in the sense that X

\[ \mathbf{X} \sim N[\mathbf{m}, \mathbf{C}] \]

MULTIVARIATE NORMAL POSTERIOR.

(1) Suppose that \( \mathbf{X} = \mathbf{x} \), a scalar, with posterior \( \mathbf{X} \sim N[\mathbf{m}, \mathbf{C}] \). Then, as is always the case with symmetric distributions, HPD regions are intervals symmetrically located about the median (here also the mode and mean) \( \mathbf{m} \). For any \( k > 0 \), the equal-tails interval

\[ m - k\mathbf{C}^{1/2} \leq \mathbf{X} \leq m + k\mathbf{C}^{1/2} \]

is the HPD region with posterior probability

\[ \Pr[|X - m|/\mathbf{C}^{1/2} \leq k] = 2\Phi(k) - 1, \]

where \( \Phi(\cdot) \) is the standard normal cumulative distribution function. With \( k = 1.645 \), so that \( \Phi(k) = 0.95 \), this gives the 90% region \( m \pm 1.645\mathbf{C}^{1/2} \). With \( k = 1.96 \), \( \Phi(k) = 0.975 \) and the 95% region is \( m \pm 1.96\mathbf{C}^{1/2} \). For any \( k > 0 \), the 100\%2\% HPD region for \( \mathbf{X} \) is simply \( m \pm k\mathbf{C}^{1/2} \).

(2) Suppose that \( \mathbf{X} \) is \( n \)-dimensional for some \( n > 1 \),

\[ \mathbf{X} \sim N[\mathbf{m}, \mathbf{C}] \]

denote by elliptical shells centred at the mode \( \mathbf{m} \), defined by the points \( \mathbf{X} \) that lead to common values of the quadratic form in the density, namely

\[ Q(\mathbf{X}) = (\mathbf{X} - \mathbf{m})'\mathbf{C}^{-1}(\mathbf{X} - \mathbf{m}). \]

For any \( k > 0 \), the region

\[ \{\mathbf{X} : Q(\mathbf{X}) \leq k\} \]
is the HPD region with posterior probability
\[ \Pr[Q(X) \leq k] = \Pr[\kappa \leq k], \]
where \( \kappa \) is a gamma distributed random quantity,
\[ \kappa \sim G[n/2,1/2]. \]
When \( n \) is an integer, this gamma distribution is chi-squared with \( n \) degrees of freedom, and so
\[ \Pr[Q(X) \leq k] = \Pr[\chi_n^2 \leq k]. \]

MULTIVARIATE T POSTERIORS.
The results for T distribution parallel those for the normal, T distributions replace normal distributions and F distributions replace gamma.

1. Suppose that \( X = X \), a scalar, with posterior \( X \sim T_r[m,C] \) for some degrees of freedom \( r > 0 \). Again, HPD regions are intervals symmetrically located about the mode \( m \). For any \( k > 0 \), the equal-tails interval
\[ m - kC^{1/2} \leq X \leq m + kC^{1/2} \]
is the HPD region with posterior probability
\[ \Pr[|X - m| / C^{1/2} \leq k] = 2\Psi_r(k) - 1, \]
where \( \Psi_r(.) \) is the cumulative distribution function of the standard Student T distribution on \( r \) degrees of freedom. For any \( k > 0 \), the \( 100(2\Psi_r(k) - 1)\% \) HPD region for \( X \) is simply \( m \pm kC^{1/2} \).

2. Suppose that \( X \) is \( n \)-dimensional for some \( n > 1 \), \( X \sim T_r[m,C] \) for some mean vector \( m \), covariance matrix \( C \), and degrees of freedom \( r > 0 \). HPD regions are again defined by elliptical shells centred at the mode \( m \), identified by values of \( X \) having a common value of the quadratic form
\[ Q(X) = (X - m)^T C^{-1} (X - m). \]
For any \( k > 0 \), the region
\[ \{X : Q(X) \leq k\} \]
is the HPD region with posterior probability
\[ \Pr[Q(X) \leq k] = \Pr[\xi \leq k/n], \]
where \( \xi \) is an F distributed random quantity with \( n \) and \( r \) degrees of freedom,
\[ \xi \sim F_{n,r} \]
(tabulated in Lindley and Scott 1984, pages 50-55). Note that when \( r \) is large, this is approximately a \( \chi_n^2 \) distribution.