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

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

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

SUMS OF NORMAL RANDOM VARIABLES.
If $X_i \sim \mathrm{N}[m_i, V_i]$, $(i = 1, \dots, n)$, have covariances $\mathrm{C}[X_i, X_j] = c_{ij} = c_{ji}$, then $Y = \sum_{i=1}^n a_i X_i + b$ has a normal distribution with

$$\mathrm{E}[Y] = \sum_{i=1}^n a_i m_i + b \qquad \text{and} \qquad \mathrm{V}[Y] = \sum_{i=1}^n a_i^2 V_i + 2\sum_{i=1}^n \sum_{j=1}^{i-1} a_i a_j c_{ij},$$

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 $\mathrm{V}[Y] = \sum_{i=1}^n a_i^2 V_i$.

17.2.2 The multivariate normal

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

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

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

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

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

LINEAR FORMS.
Suppose $\mathbf{X}_i \sim \mathrm{N}[\mathbf{m}_i, \mathbf{V}_i]$ independently, $i = 1, \dots, k$, and consider constant matrices and vectors $\mathbf{A}_1, \dots, \mathbf{A}_k$ and $\mathbf{b}$ of suitable dimensions; $\mathbf{Y} = \sum_{i=1}^k \mathbf{A}_i \mathbf{X}_i + \mathbf{b}$ is multivariate normal with mean $\sum_{i=1}^k \mathbf{A}_i \mathbf{m}_i + \mathbf{b}$ and variance matrix $\sum_{i=1}^k \mathbf{A}_i \mathbf{V}_i \mathbf{A}_i'$.

MARGINAL DISTRIBUTIONS.
Suppose that we have conformable partitions

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \qquad \mathbf{m} = \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}, \quad \text{and} \quad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}.$$

Then $\mathbf{X}_i \sim \mathrm{N}[\mathbf{m}_i, \mathbf{V}_{ii}]$, $i = 1, 2$. In particular, if X_i is univariate normal, $X_i \sim \mathrm{N}[m_i, V_{ii}]$ for $i = 1, \dots, n$.

BIVARIATE NORMAL.
Any two elements X_i and X_j of $\mathbf{X}$ are bivariate normal with joint density

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

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 $\rho_{ij} = \mathrm{corr}(X_i, X_j) = V_{ij}/\sqrt{V_{ii}V_{jj}}$.

CONDITIONAL DISTRIBUTIONS.
For the partition of $\mathbf{X}'$ into $\mathbf{X}_1'$ and $\mathbf{X}_2'$, we have

(1) $(\mathbf{X}_1 \mid \mathbf{X}_2) \sim \mathrm{N}[\mathbf{m}_1(\mathbf{X}_2), \mathbf{V}_{11}(\mathbf{X}_2)]$, where

$$\mathbf{m}_1(\mathbf{X}_2) = \mathbf{m}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{X}_2 - \mathbf{m}_2)$$

and

$$\mathbf{V}_{11}(\mathbf{X}_2) = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}.$$

The matrix $\mathbf{A}_1 = \mathbf{V}_{12}\mathbf{V}_{22}^{-1}$ is called the *regression matrix* of $\mathbf{X}_1$ on $\mathbf{X}_2$. The conditional moments are given in terms of the regression matrix by

$$\mathbf{m}_1(\mathbf{X}_2) = \mathbf{m}_1 + \mathbf{A}_1(\mathbf{X}_2 - \mathbf{m}_2)$$

and

$$\mathbf{V}_{11}(\mathbf{X}_2) = \mathbf{V}_{11} - \mathbf{A}_1\mathbf{V}_{22}\mathbf{A}_1'.$$

(2) $(\mathbf{X}_2 \mid \mathbf{X}_1) \sim \mathrm{N}[\mathbf{m}_2(\mathbf{X}_1), \mathbf{V}_{22}(\mathbf{X}_1)]$, where

$$\mathbf{m}_2(\mathbf{X}_1) = \mathbf{m}_2 + \mathbf{V}_{21}\mathbf{V}_{11}^{-1}(\mathbf{X}_1 - \mathbf{m}_1)$$

and

$$\mathbf{V}_{22}(\mathbf{X}_1) = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}.$$

The matrix $\mathbf{A}_2 = \mathbf{V}_{21}\mathbf{V}_{11}^{-1}$ is called the *regression matrix* of $\mathbf{X}_2$ on $\mathbf{X}_1$. The conditional moments are given in terms of the regression matrix by

$$\mathbf{m}_2(\mathbf{X}_1) = \mathbf{m}_2 + \mathbf{A}_2(\mathbf{X}_1 - \mathbf{m}_1)$$

and

$$\mathbf{V}_{22}(\mathbf{X}_1) = \mathbf{V}_{22} - \mathbf{A}_2\mathbf{V}_{11}\mathbf{A}_2'.$$

(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

$$A_1 = \rho\sqrt{V_{11}/V_{22}} \qquad \text{and} \qquad A_2 = \rho\sqrt{V_{22}/V_{11}}.$$

Also

$$\mathrm{V}[X_1 \mid X_2] = (1-\rho^2)V_{11} \quad \text{and} \quad \mathrm{V}[X_2 \mid X_1] = (1-\rho^2)V_{22}.$$

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 $\mathbf{Y}$ and the n-vector $\boldsymbol{\theta}$ are related via the conditional distribution

$$(\mathbf{Y} \mid \boldsymbol{\theta}) \sim \mathrm{N}[\mathbf{F}'\boldsymbol{\theta}, \mathbf{V}],$$

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

$$\mathbf{Y} = \mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\nu},$$

where $\boldsymbol{\nu} \sim \mathrm{N}[\mathbf{0}, \mathbf{V}]$. Suppose further that the marginal distribution of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta} \sim \mathrm{N}[\mathbf{a}, \mathbf{R}],$$

where both $\mathbf{a}$ and $\mathbf{R}$ are constant, and that θ is independent of ν. Equivalently,

$$\boldsymbol{\theta} = \mathbf{a} + \boldsymbol{\omega},$$

where $\boldsymbol{\omega} \sim \mathrm{N}[\mathbf{0}, \mathbf{R}]$ independently of $\boldsymbol{\nu}$.

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

MULTIVARIATE JOINT NORMAL DISTRIBUTION.
Since $\boldsymbol{\theta} = \mathbf{a} + \boldsymbol{\omega}$ and $\mathbf{Y} = \mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\nu} = \mathbf{F}'\mathbf{a} + \mathbf{F}'\boldsymbol{\omega} + \boldsymbol{\nu}$, then the vector $(\mathbf{Y}', \boldsymbol{\theta}')'$ is a linear transformation of $(\boldsymbol{\nu}', \boldsymbol{\omega}')'$. By construction the latter has a multivariate normal distribution, so that $\mathbf{Y}$ and $\boldsymbol{\theta}$ are jointly normal. Further

(1) $\mathrm{E}[\boldsymbol{\theta}] = \mathbf{a}$ and $\mathrm{V}[\boldsymbol{\theta}] = \mathbf{R}$;
(2) $\mathrm{E}[\mathbf{Y}] = \mathrm{E}[\mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\nu}] = \mathbf{F}'\mathrm{E}[\boldsymbol{\theta}] + \mathrm{E}[\boldsymbol{\nu}] = \mathbf{F}'\mathbf{a}$ and

$$\mathrm{V}[\mathbf{Y}] = \mathrm{V}[\mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\nu}] = \mathbf{F}'\mathrm{V}[\boldsymbol{\theta}]\mathbf{F} + \mathrm{V}[\boldsymbol{\nu}] = \mathbf{F}'\mathbf{R}\mathbf{F} + \mathbf{V};$$

(3) $\mathrm{C}[\mathbf{Y}, \boldsymbol{\theta}] = \mathrm{C}[\mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\nu}, \boldsymbol{\theta}] = \mathbf{F}'\mathrm{C}[\boldsymbol{\theta}, \boldsymbol{\theta}] + \mathrm{C}[\boldsymbol{\nu}, \boldsymbol{\theta}] = \mathbf{F}'\mathbf{R}$.

It follows that

$$\begin{pmatrix} \mathbf{Y} \\ \boldsymbol{\theta} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \mathbf{F}'\mathbf{a} \\ \mathbf{a} \end{pmatrix}, \begin{pmatrix} \mathbf{F}'\mathbf{R}\mathbf{F} + \mathbf{V} & \mathbf{F}'\mathbf{R} \\ \mathbf{R}\mathbf{F} & \mathbf{R} \end{pmatrix}\right].$$

Therefore, identifying $\mathbf{Y}$ with $\mathbf{X}_1$ and $\boldsymbol{\theta}$ with $\mathbf{X}_2$ in the partition of $\mathbf{X}$ in 17.2.2, we have

(4) $\mathbf{Y} \sim \mathrm{N}[\mathbf{F}'\mathbf{a}, \mathbf{F}'\mathbf{R}\mathbf{F} + \mathbf{V}]$;
(5) $(\boldsymbol{\theta} \mid \mathbf{Y}) \sim \mathrm{N}[\mathbf{m}, \mathbf{C}]$, where

$$\mathbf{m} = \mathbf{a} + \mathbf{R}\mathbf{F}[\mathbf{F}'\mathbf{R}\mathbf{F} + \mathbf{V}]^{-1}[\mathbf{Y} - \mathbf{F}'\mathbf{a}]$$

and

$$\mathbf{C} = \mathbf{R} - \mathbf{R}\mathbf{F}[\mathbf{F}'\mathbf{R}\mathbf{F} + \mathbf{V}]^{-1}\mathbf{F}'\mathbf{R}.$$

By defining $\mathbf{e} = \mathbf{Y} - \mathbf{F}'\mathbf{a}$, $\mathbf{Q} = \mathbf{F}'\mathbf{R}\mathbf{F} + \mathbf{V}$ and $\mathbf{A} = \mathbf{R}\mathbf{F}\mathbf{Q}^{-1}$, these equations become

$$\mathbf{m} = \mathbf{a} + \mathbf{A}\mathbf{e} \qquad \text{and} \qquad \mathbf{C} = \mathbf{R} - \mathbf{A}\mathbf{Q}\mathbf{A}'.$$

MULTIVARIATE BAYES' THEOREM.
An alternative derivation of the conditional distribution for $(\boldsymbol{\theta} \mid \mathbf{Y})$ via Bayes' Theorem provides alternative expressions for $\mathbf{m}$ and $\mathbf{C}$. Note that

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})$$

as a function of $\boldsymbol{\theta}$, so that

$$\ln[p(\boldsymbol{\theta} \mid \mathbf{Y})] = k + \ln[p(\mathbf{Y} \mid \boldsymbol{\theta})] + \ln[p(\boldsymbol{\theta})],$$

where k depends on $\mathbf{Y}$ but not on $\boldsymbol{\theta}$. Therefore

$$\ln[p(\boldsymbol{\theta} \mid \mathbf{Y})] = k - \frac{1}{2}\left[(\mathbf{Y} - \mathbf{F}'\boldsymbol{\theta})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{F}'\boldsymbol{\theta}) + (\boldsymbol{\theta} - \mathbf{a})'\mathbf{R}^{-1}(\boldsymbol{\theta} - \mathbf{a})\right].$$

The bracketed term here is simply

$$\begin{aligned} &\mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y} - 2\mathbf{Y}'\mathbf{V}^{-1}\mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbf{F}\mathbf{V}^{-1}\mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbf{R}^{-1}\boldsymbol{\theta} - 2\mathbf{a}'\mathbf{R}^{-1}\boldsymbol{\theta} + \mathbf{a}'\mathbf{R}^{-1}\mathbf{a} \\ &\qquad = \boldsymbol{\theta}'\left\{\mathbf{F}\mathbf{V}^{-1}\mathbf{F}' + \mathbf{R}^{-1}\right\}\boldsymbol{\theta} - 2\left\{\mathbf{Y}'\mathbf{V}^{-1}\mathbf{F}' + \mathbf{a}'\mathbf{R}^{-1}\right\}\boldsymbol{\theta} + h \end{aligned}$$

where h depends on $\mathbf{Y}$ but not on $\boldsymbol{\theta}$. Completing the quadratic form gives

$$(\boldsymbol{\theta}-\mathbf{m})'\,\mathbf{C}^{-1}(\boldsymbol{\theta}-\mathbf{m})+h^*,$$

where again h^* does not involve $\boldsymbol{\theta}$,

$$\mathbf{C}^{-1}=\mathbf{R}^{-1}+\mathbf{F}\mathbf{V}^{-1}\mathbf{F}',$$

and

$$\mathbf{m}=\mathbf{C}\left\{\mathbf{F}\mathbf{V}^{-1}\mathbf{Y}+\mathbf{R}^{-1}\mathbf{a}\right\}.$$

Hence

$$p(\boldsymbol{\theta}\mid\mathbf{Y})\propto\exp\left[-\frac{1}{2}\,(\boldsymbol{\theta}-\mathbf{m})'\,\mathbf{C}^{-1}\,(\boldsymbol{\theta}-\mathbf{m})\right]$$

as a function of $\boldsymbol{\theta}$, so that $(\boldsymbol{\theta}\mid\mathbf{Y})\sim\mathrm{N}[\mathbf{m},\mathbf{C}]$, just as derived earlier.

Note that the two derivations give different expressions for $\mathbf{m}$ and $\mathbf{C}$ that provide, in particular, the matrix identity for $\mathbf{C}$ given by

$$\mathbf{C}=\left[\mathbf{R}^{-1}+\mathbf{F}\mathbf{V}^{-1}\mathbf{F}'\right]^{-1}=\mathbf{R}-\mathbf{R}\mathbf{F}\left[\mathbf{F}'\mathbf{R}\mathbf{F}+\mathbf{V}\right]^{-1}\mathbf{F}'\mathbf{R},$$

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 $\phi\sim\mathrm{G}[n,d]$, if and only if

$$p(\phi)\propto\phi^{n-1}\exp(-\phi d),\qquad(\phi>0).$$

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 $\mathrm{E}[\phi]=n/d$ and $\mathrm{V}[\phi]=\mathrm{E}[\phi]^2/n$.

Two special cases of interest are

(1) $n=1$, when ϕ has a (negative) exponential distribution with mean $1/d$;
(2) $\phi\sim\mathrm{G}[n/2,d/2]$ when n is a positive integer. In this case $d\phi\sim\chi^2_n$, a chi-squared distribution with n degrees of freedom.

17.3.2 Univariate normal/gamma distribution

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

(1)

$$\begin{aligned}p(X,\phi)&=\left(\frac{\phi}{2\pi C}\right)^{1/2}\exp\left[-\frac{\phi(X-m)^2}{2C}\right]\\&\quad\times\frac{d^{n/2}}{2^{n/2}\Gamma(\frac{n}{2})}\phi^{\frac{n}{2}-1}\exp\left[-\frac{\phi d}{2}\right]\\&\propto\phi^{(\frac{n+1}{2})-1}\exp\left[-\frac{\phi}{2}\left\{\frac{(X-m)^2}{C}+d\right\}\right],\end{aligned}$$

as a function of ϕ and X.

(2)

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

so that

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

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

(3)

$$\begin{aligned} p(X) &= p(X,\phi)/p(\phi \mid X) \qquad (-\infty < X < \infty) \\ &\propto [n + (X-m)^2/R]^{-(n+1)/2}, \end{aligned}$$

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

17.3.3 Multivariate normal/gamma distribution

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

(1) $(\phi \mid \mathbf{X}) \sim \mathrm{G}[n^*/2, d^*/2]$, where $n^* = n+p$ and $d^* = d + (\mathbf{X}-\mathbf{m})'\mathbf{C}^{-1}(\mathbf{X}-\mathbf{m})/2$ (notice the degrees of freedom increases by p).

(2) $\mathbf{X}$ has a (marginal) multivariate T distribution in p dimensions with n degrees of freedom, mode $\mathbf{m}$ and scale matrix $\mathbf{R} = \mathbf{C}(d/n) = \mathbf{C}/\mathrm{E}[\phi]$, denoted by $\mathbf{X} \sim \mathrm{T}_n[\mathbf{m}, \mathbf{R}]$, with density

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

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

$$X_i \sim \mathrm{T}_n[m_i, R_{ii}],$$

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 $\boldsymbol{\theta}$ and the scalar ϕ via

$$(Y \mid \boldsymbol{\theta}, \phi) \sim \mathrm{N}[\mathbf{F}'\boldsymbol{\theta}, k\phi^{-1}],$$

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

$$(\boldsymbol{\theta} \mid \phi) \sim \mathrm{N}[\mathbf{a}, \mathbf{R}\phi^{-1}]$$

and

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

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

(1) $(Y \mid \phi) \sim \mathrm{N}[f, Q\phi^{-1}]$, where $f = \mathbf{F}'\mathbf{a}$ and $Q = \mathbf{F}'\mathbf{R}\mathbf{F} + k$;

(2) $Y \sim \mathrm{T}_n[f, QS]$, unconditional on $\boldsymbol{\theta}$ or ϕ;
(3) $(\boldsymbol{\theta}, \phi \mid Y)$ have a joint normal/gamma distribution. Specifically,

$$(\boldsymbol{\theta} \mid \phi, Y) \sim \mathrm{N}[\mathbf{m}, \mathbf{C}\phi^{-1}],$$

where from 17.2.3, $\mathbf{m} = \mathbf{a} + \mathbf{RF}(Y - f)/Q$, $\mathbf{C} = \mathbf{R} - \mathbf{RFF'R}/Q$ and

$$(\phi \mid Y) \sim \mathrm{G}[(n+1)/2, \left\{d + (Y-f)^2/Q\right\}/2];$$

(4) $(\boldsymbol{\theta} \mid Y)$ has a multivariate T distribution in p dimensions with $n+1$ degrees of freedom, mode $\mathbf{m}$ and scale matrix $\mathbf{C}S$, denoted by $(\boldsymbol{\theta} \mid Y) \sim \mathrm{T}_{n+1}[\mathbf{m}, \mathbf{C}S]$. In particular, for the i^{th} element $\boldsymbol{\theta}_i$ of $\boldsymbol{\theta}$,

$$\theta_i \sim \mathrm{T}_{n+1}[m_i, C_{ii}S], \qquad \text{or} \qquad \frac{(\theta_i - m_i)}{\sqrt{C_{ii}S}} \sim \mathrm{T}_{n+1}[0,1].$$

17.3.5 HPD regions

Consider any random vector $\mathbf{X}$ with (posterior) density $p(\mathbf{X})$. Density measures relative support for values of $\mathbf{X}$ in the sense that $\mathbf{X} = \mathbf{x}_0$ is better supported than $\mathbf{X} = \mathbf{x}_1$ if $p(\mathbf{x}_0) > p(\mathbf{x}_1)$. **Highest posterior density** regions (HPD regions) form the bases for inferences about $\mathbf{X}$ based on sets of values with high density. The basic notion is that the posterior probability of the set or region of values of $\mathbf{X}$ that have higher density than any chosen value $\mathbf{x}_0$ measures the support for or against $\mathbf{X} = \mathbf{x}_0$. Formally, the HPD region generated by the value $\mathbf{X} = \mathbf{x}_0$ is

$$\{\mathbf{X} \ : \ p(\mathbf{X}) \geq p(\mathbf{x}_0)\}.$$

The probability that $\mathbf{X}$ lies in this region is simply

$$\Pr[p(\mathbf{X}) \geq p(\mathbf{x}_0)].$$

If this probability is high, then $\mathbf{x}_0$ is poorly supported. In normal linear models, HPD regions are always connected regions, or intervals, and these provide interval-based inferences and tests of hypotheses through the use of posterior normal, T and F distributions, described below. Fuller theoretical details are provided by Box and Tiao (1973) and De Groot (1971).

MULTIVARIATE NORMAL POSTERIOR.

(1) Suppose that $\mathbf{X} = X$, a scalar, with posterior $X \sim \mathrm{N}[m, C]$. Then, as is always the case with symmetric distributions, HPD regions are intervals symmetrically located about the median (here also the mode and mean) 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

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

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

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

for some mean vector $\mathbf{m}$ and covariance matrix $\mathbf{C}$. Then HPD regions are defined 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

$$\Pr[Q(\mathbf{X}) \le k] = \Pr[\kappa \le k],$$

where κ is a gamma distributed random quantity,

$$\kappa \sim \mathrm{G}[n/2, 1/2].$$

When n is an integer, this gamma distribution is chi-squared with n degrees of freedom, and so

$$\Pr[Q(\mathbf{X}) \le k] = \Pr[\chi^2_n \le k].$$

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 $\mathbf{X} = X$, a scalar, with posterior $X \sim \mathrm{T}_r[m, C]$ 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

$$m - kC^{1/2} \le X \le m + kC^{1/2}$$

is the HPD region with posterior probability

$$\Pr[|X - m|/C^{1/2} \le k] = 2\Psi_r(k) - 1,$$

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 $\mathbf{X}$ is n-dimensional for some $n > 1$, $\mathbf{X} \sim \mathrm{T}_r[\mathbf{m}, \mathbf{C}]$ for some mean vector $\mathbf{m}$, covariance matrix $\mathbf{C}$, and degrees of freedom $r > 0$. HPD regions are again defined by elliptical shells centred at the mode $\mathbf{m}$, identified by values of $\mathbf{X}$ having a common value of the quadratic form

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

For any $k > 0$, the region

$$\{\mathbf{X} \ : \ Q(\mathbf{X}) \le k\}$$

is the HPD region with posterior probability

$$\Pr[Q(\mathbf{X}) \le k] = \Pr[\xi \le k/n],$$

where ξ is an F distributed random quantity with n and r degrees of freedom,

$$\xi \sim \mathrm{F}_{n,r}$$

(tabulated in Lindley and Scott 1984, pages 50-55). Note that when r is large, this is approximately a χ^2_n distribution.