CHAPTER 8

SOME ASPECTS OF MULTIVARIATE ANALYSIS

8.1 INTRODUCTION

In all the problems we have so far considered, observations are made of a single unidimensional response or output y. The inference problems that result are called univariate problems. In this and the next chapter, we shall consider problems which arise when the output is multidimensional. Thus, in the study of a chemical process, at each experimental setting one might observe yield y_1, density y_2, and color y_3 of the product. Similarly, in a study of consumer behavior, for each household one might record spending on food y_1, spending on durables y_2, and spending on travel and entertainment y_3. We would then say that a three-dimensional output or response is observed. Inference problems which arise in the analysis of such data are called *multivariate*.

In this chapter, we shall begin by reviewing some *univariate* problems in a general setting which can be easily extended to the multivariate case.

8.1.1 A General Univariate Model

It is often desired to make inferences about *parameters* $\theta_1, \ldots, \theta_k$ contained in the relationship between a single observed *output variable* or *response* y subject to error and p *input invariables* $\xi_1, \ldots, \xi_p$ whose values are assumed exactly known. It should be understood that the inputs could include qualitative as well as quantitative variables. For example, ξ_i might take values of 0 or 1 depending on whether some particular quality was absent or present in which case ξ_i is called an *indicator variable* or less appropriately a *dummy* variable.

The Design Matrix

Suppose, in an investigation, n experimental "runs" are made, and the uth run consists of making an observation y_u at some fixed set of input conditions $\boldsymbol{\xi}'_u = (\xi_{u1}, \xi_{u2}, \ldots, \xi_{up})$. The $n \times p$ *design matrix*

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}'_1 \\ \vdots \\ \boldsymbol{\xi}'_u \\ \vdots \\ \boldsymbol{\xi}'_n \end{bmatrix} = \begin{bmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1p} \\ \vdots & & & \\ \xi_{u1} & \xi_{u2} & \cdots & \xi_{up} \\ \vdots & & & \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{np} \end{bmatrix} \tag{8.1.1}$$

lists the p input conditions to be used in each of the n projected runs and the uth row of $\boldsymbol{\xi}$ is the vector $\boldsymbol{\xi}_u'$. The phraseology "experimental run", "experimental design" is most natural in a situation in which a scientific experiment is being conducted and in which the levels of the inputs are at our choice. In some applications, however, and particularly in economic studies, it is often impossible to choose the experimental conditions. We have only historical data generated for us in circumstances beyond our control and often in a manner we would not choose. It is convenient here to extend the terminologies "experimental run" and "experimental design" to include experiments designed by nature, but we must, of course, bear in mind the limitations of such historical data.

To obtain a mathematical model for our set-up we need to link the n observations $\mathbf{y}' = (y_1, ..., y_n)$ with the inputs $\boldsymbol{\xi}$. This we do by defining two functions called respectively the *expectation function* and the *error function*.

The Expectation Function

The expected value $E(y_u)$ of the output from the uth run is assumed to be a known function η_u of the p fixed inputs $\boldsymbol{\xi}_u$ employed during that run, involving k unknown parameters $\boldsymbol{\theta}' = (\theta_1, ..., \theta_k)$,

$$E(y_u) = \eta_u = \eta(\boldsymbol{\xi}_u, \boldsymbol{\theta}). \tag{8.1.2}$$

The vector valued function $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\xi}, \boldsymbol{\theta}), \boldsymbol{\eta}' = (\eta_1, \ldots, \eta_u, \ldots, \eta_n)$, is called the *expectation function.*

The Error Function

The expectation function links $E(y_u)$ to $\boldsymbol{\xi}_u$ and $\boldsymbol{\theta}$. We now have to link y_u to $E(y_u) = \eta_u$. This is done by means of an *error distribution function* in $\boldsymbol{\varepsilon}' = (\varepsilon_1, ..., \varepsilon_n)$. The n experimental errors $\boldsymbol{\varepsilon} = \mathbf{y} - \boldsymbol{\eta}$ which occur in making the runs are assumed to be random variables having zero means but in general not necessarily independently or Normally distributed. We denote the density function of the n errors by $p(\boldsymbol{\varepsilon} \mid \boldsymbol{\pi})$ where $\boldsymbol{\pi}$ is a set of error distribution parameters whose values are in general unknown.

Finally, then, the output in the form of the n observations $\mathbf{y}$ and the input in the form of the n sets of conditions $\boldsymbol{\xi}$ are linked together by a *mathematical model* containing the error function and the expectation function as follows

$$\boldsymbol{\xi} \xrightarrow{\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\xi}, \boldsymbol{\theta})} \boldsymbol{\eta} \xrightarrow{p(\mathbf{y} - \boldsymbol{\eta} \mid \boldsymbol{\pi})} \mathbf{y}. \tag{8.1.3}$$

This model involves a function

$$f(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\xi}) \tag{8.1.4}$$

of the observations $\mathbf{y}$, the parameters $\boldsymbol{\theta}$ of the expectation function, the parameters $\boldsymbol{\pi}$ of the error distribution and the design $\boldsymbol{\xi}$.

Data Generation Model

If we knew $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ and the design $\boldsymbol{\xi}$, we could use the function (8.1.4) to calculate the probability density associated with any particular set of data $\mathbf{y}$. This data generation model (which might, for example, be directly useful for simulation and Monte-Carlo studies) is the function $f(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\xi})$ with $\boldsymbol{\theta}, \boldsymbol{\pi}$ and $\boldsymbol{\xi}$ held fixed and we denote it by

$$p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\xi}) = f(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\xi}), \tag{8.1.5}$$

which emphasizes that the density is a function of $\mathbf{y}$ alone for *fixed* $\boldsymbol{\theta}$, $\boldsymbol{\pi}$, and $\boldsymbol{\xi}$.

The Likelihood Function and the Posterior Distribution

In ordinary statistical practice, we are not directly interested in probabilities associated with *various* sets of data, given *fixed* values of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$. On the contrary, we are concerned with the probabilities associated with *various* sets of parameter values, given a *fixed* set of data which is known to have occurred.

After an experiment has been performed, $\mathbf{y}$ is known and fixed (as is $\boldsymbol{\xi}$) but $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are unknown. The likelihood function has the same form as (8.1.4), but in it $\mathbf{y}$ and $\boldsymbol{\xi}$ are fixed and $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ are not to be regarded as variables. Thus, the likelihood may be written

$$l(\boldsymbol{\theta}, \boldsymbol{\pi} \mid \mathbf{y}, \boldsymbol{\xi}) = f(\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\xi}). \tag{8.1.6}$$

In what follows we usually omit specific note of dependence on $\boldsymbol{\xi}$ and write $l(\boldsymbol{\theta}, \boldsymbol{\pi} \mid \mathbf{y}, \boldsymbol{\xi})$ as $l(\boldsymbol{\theta}, \boldsymbol{\pi} \mid \mathbf{y})$.

In the Bayesian framework, inferences about $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ can be made by suitable study of the posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\pi} \mid \mathbf{y})$ of $\boldsymbol{\theta}$ and $\boldsymbol{\pi}$ obtained by combining the likelihood with the appropriate prior distribution $p(\boldsymbol{\theta}, \boldsymbol{\pi})$,

$$p(\boldsymbol{\theta}, \boldsymbol{\pi} \mid \mathbf{y}) \propto l(\boldsymbol{\theta}, \boldsymbol{\pi} \mid \mathbf{y})\, p(\boldsymbol{\theta}, \boldsymbol{\pi}). \tag{8.1.7}$$

An example in which the expectation function is nonlinear and the error distribution is non-Normal was given in Section 3.5. In this chapter, we shall from now on assume Normality but will extend our general model to cover multivariate problems.

8.2 A GENERAL MULTIVARIATE NORMAL MODEL

Suppose now that a number of output responses are measured in each experimental run. Thus, in a chemical experiment, at each setting of the process conditions ξ_1 = temperature and ξ_2 = concentration, observations might be made on the output responses y_1 = yield of product A, y_2 = yield of product B, and y_3 = yield of product C. In general, then, from each experimental run the

m-variate observation

$$\mathbf{y}'_{(u)} = (y_{u1}, \ldots, y_{ui}, \ldots, y_{um})$$

would be available. There would now be m expectation functions

$$E(\mathbf{y}_{(u)}) = \boldsymbol{\eta}_{(u)} = (\eta_{u1}, \ldots, \eta_{um})'$$

where

$$\begin{aligned}
E(y_{u1}) &= \eta_{u1} = \eta_1(\boldsymbol{\xi}_{u1}, \boldsymbol{\theta}_1)\\
&\vdots\\
E(y_{ui}) &= \eta_{ui} = \eta_i(\boldsymbol{\xi}_{ui}, \boldsymbol{\theta}_i)\\
&\vdots\\
E(y_{uj}) &= \eta_{uj} = \eta_j(\boldsymbol{\xi}_{uj}, \boldsymbol{\theta}_j)\\
&\vdots\\
E(y_{um}) &= \eta_{um} = \eta_m(\boldsymbol{\xi}_{um}, \boldsymbol{\theta}_m)
\end{aligned} \tag{8.2.1}$$

where $\boldsymbol{\xi}_{ui}$ would contain p_i elements $(\xi_{u1i}, \ldots, \xi_{usi}, \ldots, \xi_{up_ii})$ and $\boldsymbol{\theta}_i$ would contain k_i elements $(\theta_{1i}, \ldots, \theta_{gi}, \ldots, \theta_{k_ii})$. The expectation functions η_{ui} might be linear or nonlinear both in the parameters $\boldsymbol{\theta}_i$ and the inputs $\boldsymbol{\xi}_{ui}$. Also, depending on the problem, some or all of the p_i elements of $\boldsymbol{\xi}_{ui}$ might be the same as those of $\boldsymbol{\xi}_{uj}$ and some or all of the elements of $\boldsymbol{\theta}_i$ might be the same as those of $\boldsymbol{\theta}_j$. That is to say, a given output would involve certain inputs and certain parameters which might or might not be shared by other outputs.

8.2.1 The Likelihood Function

Let us now consider the problem of making inferences about the $\boldsymbol{\theta}_i$ for a set of n m-variate observations. We assume that the error vector

$$\boldsymbol{\varepsilon}_{(u)} = \mathbf{y}_{(u)} - \boldsymbol{\eta}_{(u)} = (\varepsilon_{u1}, \ldots, \varepsilon_{um})', \qquad u = 1, \ldots, n, \tag{8.2.2}$$

is, for given $\boldsymbol{\theta}$ and Σ, distributed as the m-variate Normal $M_m(\mathbf{0}, \Sigma)$, and that the runs are made in such a way that it can be assumed that from run to run the observations are independent. Thus, in terms of the general framework of (8.1.4), $\boldsymbol{\Sigma} = \boldsymbol{\pi}$ are the parameters of an error distribution which is multivariate Normal. We first derive some very general results which apply to any model of this type, and then consider in detail the various important special cases that emerge if the expectation functions are supposed linear in the parameters $\boldsymbol{\theta}_i$.

The joint distribution of the n vectors of errors $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_{(1)}, \ldots, \boldsymbol{\varepsilon}_{(u)}, \ldots, \boldsymbol{\varepsilon}_{(n)})'$ is

$$\begin{aligned}
p(\boldsymbol{\varepsilon} \mid \boldsymbol{\Sigma}, \boldsymbol{\theta}) &= \prod_{u=1}^{n} p(\boldsymbol{\varepsilon}_{(u)} \mid \boldsymbol{\Sigma}, \boldsymbol{\theta})\\
&= (2\pi)^{-mn/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2}\sum_{u=1}^{n} \boldsymbol{\varepsilon}'_{(u)} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_{(u)}\right)\\
&-\infty < \varepsilon_{ui} < \infty, \quad i = 1, \ldots, m, \quad u = 1, \ldots, n,
\end{aligned} \tag{8.2.3}$$

where $\Sigma = \{\sigma_{ij}\}$ is the $m \times m$ covariance matrix, $\Sigma_{-1} = \{\sigma^{ij}\}$ its inverse and θ refers to the complete set of all the $(k_1 + \cdots + k_m)$ parameters $\theta_1, \ldots, \theta_m$.

Denoting $\mathbf{S}(\theta)$ to be the $m \times m$ symmetric matrix

$$\mathbf{S}(\theta) = \{S_{ij}(\theta_i, \theta_j)\}$$

with

$$S_{ij}(\theta_i, \theta_j) = \sum_{u=1}^{n} \varepsilon_{ui}\,\varepsilon_{uj} = \sum_{u=1}^{n} [y_{ui} - \eta_i(\xi_{ui}, \theta_i)]\,[y_{uj} - \eta_j(\xi_{uj}, \theta_j)], \qquad i, j = 1, \ldots, m, \tag{8.2.4}$$

then the exponent in (8.2.3) can be expressed as

$$\sum_{u=1}^{n} \varepsilon'_{(u)}\,\Sigma^{-1}\varepsilon_{(u)} = \operatorname{tr} \mathbf{S}(\theta)\Sigma^{-1} = \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma^{ij} S_{ij}(\theta_i, \theta_j) \tag{8.2.5}$$

where $\operatorname{tr}\mathbf{A}$ means the trace of the matrix $\mathbf{A}$. Given the observations, the likelihood function can thus be written

$$\begin{aligned} l(\theta, \Sigma \mid \mathbf{y}) &\propto p(\varepsilon \mid \Sigma, \theta) \\ &\propto |\Sigma|^{-n/2} \exp\left[-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}\mathbf{S}(\theta)\right]. \end{aligned} \tag{8.2.6}$$

To clarify the notation, we emphasize that $\mathbf{y}$ refers to the $n \times m$ matrix of observations

$$\mathbf{y} = \begin{bmatrix} y_{11} & \cdots & y_{1i} & \cdots & y_{1m} \\ \vdots & & & & \\ y_{u1} & \cdots & y_{ui} & \cdots & y_{um} \\ \vdots & & & & \\ y_{n1} & \cdots & y_{ni} & \cdots & y_{nm} \end{bmatrix} = [\mathbf{y}_1, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_m] = \begin{bmatrix} \mathbf{y}'_{(1)} \\ \vdots \\ \mathbf{y}'_{(u)} \\ \vdots \\ \mathbf{y}'_{(n)} \end{bmatrix},$$

where $\mathbf{y}_i = (y_{1i}, \ldots, y_{ni})'$ is the vector of n observations corresponding to the ith response and $\mathbf{y}_{(u)} = (y_{u1}, \ldots, y_{um})'$ is the vector of m observations of the uth experimental run. Similarly, ε refers to the $n \times m$ matrix of errors

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1i} & \cdots & \varepsilon_{1m} \\ \vdots & & & & \\ \varepsilon_{u1} & \cdots & \varepsilon_{ui} & \cdots & \varepsilon_{um} \\ \vdots & & & & \\ \varepsilon_{n1} & \cdots & \varepsilon_{ni} & \cdots & \varepsilon_{nm} \end{bmatrix} = [\varepsilon_1, \ldots, \varepsilon_i, \ldots, \varepsilon_m] = \begin{bmatrix} \varepsilon'_{(1)} \\ \vdots \\ \varepsilon'_{(u)} \\ \vdots \\ \varepsilon'_{(n)} \end{bmatrix}.$$

8.2.2 Prior Distribution of (θ, Σ)

For the prior distribution of the parameters (θ, Σ), we shall first of all assume that θ and Σ are approximately independent so that

$$p(\theta, \Sigma) \doteq p(\theta)\,p(\Sigma). \tag{8.2.7}$$

We shall further suppose that the parameterization in terms of $\boldsymbol{\theta}$ is so chosen such that it is appropriate to take $\boldsymbol{\theta}$ as locally uniform,†

$$p(\boldsymbol{\theta}) \propto \text{constant}. \tag{8.2.8}$$

For the prior distribution of the $\frac{1}{2}m(m+1)$ distinct elements of $\boldsymbol{\Sigma}$, application of the argument in Section 1.3 for the multiparameter situation leads to the non informative reference prior

$$p(\boldsymbol{\Sigma}) \propto |\mathscr{I}(\boldsymbol{\Sigma})|^{1/2}. \tag{8.2.9}$$

Now,

$$|\mathscr{I}(\boldsymbol{\Sigma})| = |\mathscr{I}(\boldsymbol{\Sigma}^{-1})| \left| \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\Sigma}^{-1}} \right|^{-2}, \tag{8.2.10}$$

where

$$\left| \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\Sigma}^{-1}} \right| = \left| \frac{\partial(\sigma_{11}, \sigma_{12}, \ldots, \sigma_{mm})}{\partial(\sigma^{11}, \sigma^{12}, \ldots, \sigma^{mm})} \right| \tag{8.2.11}$$

is the Jacobian of the transformation from the elements σ_{ij} of $\boldsymbol{\Sigma}$ to the elements σ^{ij} of $\boldsymbol{\Sigma}^{-1}$. It is shown in Appendix A8.2 that

$$|\mathscr{I}(\boldsymbol{\Sigma}^{-1})| \propto \left| \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\Sigma}^{-1}} \right| \tag{8.2.12}$$

and that

$$\left| \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\Sigma}^{-1}} \right| = |\boldsymbol{\Sigma}|^{m+1}. \tag{8.2.13}$$

Thus,

$$p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)}. \tag{8.2.14}$$

In this special case $m = 1$, (8.2.14) reduces to

$$p(\sigma_{11}) \propto \frac{1}{\sigma_{11}} \tag{8.2.15}$$

which coincides with the usual assumption concerning a noninformative prior distribution for a single variance. Another special case of interest is when the errors $(\varepsilon_{u1}, \ldots, \varepsilon_{um})$ are uncorrelated, that is, $\sigma_{ij} = 0$ if $i \neq j$. In this case, the same argument leads to

$$p(\boldsymbol{\Sigma} \mid \sigma_{ij} = 0, i \neq j) = p(\sigma_{11}, \ldots, \sigma_{mm}) \propto \prod_{i=1}^{m} \sigma_{ii}^{-1}. \tag{8.2.16}$$

† As we have mentioned earlier, when the parameter space is of high dimension, the use of the locally uniform prior may be inappropriate and more careful considerations should be given to the structure of the model in selecting a noninformative prior.

8.2.3 Posterior Distribution of $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$

Using (8.2.6), (8.2.8), and (8.2.14), the joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ is

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(n+m+1)} \exp\left[-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \mathbf{S}(\boldsymbol{\theta})\right], \qquad -\infty < \boldsymbol{\theta} < \infty, \quad \boldsymbol{\Sigma} > \mathbf{0}, \tag{8.2.17}$$

where the notation $-\infty < \boldsymbol{\theta} < \infty$ means that each element of the set of parameters $\boldsymbol{\theta}$ can vary from $-\infty$ to ∞, and the notation $\boldsymbol{\Sigma} > \mathbf{0}$ means that the $\frac{1}{2}m(m+1)$ elements σ_{ij} are such that the random matrix $\boldsymbol{\Sigma}$ is positive definite.

It is sometimes convenient to work with the elements of $\boldsymbol{\Sigma}^{-1} = \{\sigma^{ij}\}$ rather than the elements of $\boldsymbol{\Sigma}$. Since

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{y}) = p(\boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{y}) \left| \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{\Sigma}^{-1}} \right|, \tag{8.2.18}$$

it follows from (8.2.13) that the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\Sigma}^{-1})$ is

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}(n-m-1)} \exp\left[-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} \mathbf{S}(\boldsymbol{\theta})\right], \qquad -\infty < \boldsymbol{\theta} < \infty, \quad \boldsymbol{\Sigma}^{-1} > \mathbf{0}. \tag{8.2.19}$$

8.2.4 The Wishart Distribution

We now introduce a distribution which is basic in Normal theory multivariate problems. Let $\mathbf{Z}$ be a $m \times m$ positive definite symmetric random matrix which consists of $\frac{1}{2}m(m+1)$ distinct random variables z_{ij} $(i, j = 1, \ldots, m; i \geqslant j)$. Let $q > 0$, and $\mathbf{B}$ be a $m \times m$ positive definite symmetric matrix of fixed constants. The distribution of z_{ij},

$$p(\mathbf{Z}) \propto |\mathbf{Z}|^{\frac{1}{2}q-1} \exp\left(-\tfrac{1}{2} \operatorname{tr} \mathbf{Z}\mathbf{B}\right), \qquad \mathbf{Z} > \mathbf{0} \tag{8.2.20}$$

obtained by Wishart (1928), is a multivariate generalization of the χ^2 distribution. It can be shown that

$$\int_{\mathbf{Z}>\mathbf{0}} |\mathbf{Z}|^{\frac{1}{2}q-1} \exp\left(-\tfrac{1}{2} \operatorname{tr} \mathbf{Z}\mathbf{B}\right) d\mathbf{Z} = |\mathbf{B}|^{-\frac{1}{2}(q+m-1)}\, 2^{\frac{1}{2}m(q+m-1)}\, \Gamma_m\left(\frac{q+m-1}{2}\right) \tag{8.2.21}$$

where $\Gamma_p(b)$ is the generalized gamma function, Siegel (1935)

$$\Gamma_p(b) = \left[\Gamma(\tfrac{1}{2})\right]^{\frac{1}{2}p(p-1)} \prod_{\alpha=1}^{p} \Gamma\left(b + \frac{\alpha - p}{2}\right), \qquad b > \frac{p-1}{2}. \tag{8.2.22}$$

We shall denote the distribution (8.2.20) by $W_m(\mathbf{B}^{-1}, q)$ and say that $\mathbf{Z}$ is distributed as Wishart with q degrees of freedom and parameter matrix $\mathbf{B}^{-1}$. For a discussion of the properties of the Wishart distribution, see for example Anderson (1958). Note carefully that the parameterization used in (8.2.20) is different from the one used in Anderson in one respect. In his notation, the distribution in (8.2.20) is denoted as $W(\mathbf{B}^{-1}, \nu)$ where $\nu = q + m - 1$ is said to be the degrees of freedom.

As an application of the Wishart distribution, we see in (8.2.19) that, given $\boldsymbol{\theta}, \boldsymbol{\Sigma}^{-1}$ is distributed as $W_m[\mathbf{S}^{-1}(\boldsymbol{\theta}), n - m + 1]$ provided $n \geqslant m$.

8.2.5 Posterior Distribution of $\boldsymbol{\theta}$

Using the identity (8.2.21), we immediately obtain from (8.2.19) the marginal posterior distribution of $\boldsymbol{\theta}$ as

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto |\mathbf{S}(\boldsymbol{\theta})|^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.2.23}$$

provided $n \geqslant m$.

This extremely simple result is remarkable because of its generality. It will be noted that to reach it we have *not* had to assume either:

a) that any of the input variables $\boldsymbol{\xi}_{ui}$ were or were not common to more than one output, or
b) that the parameters $\boldsymbol{\theta}_i$ were or were not common to more than one output, or
c) that the expectation functions were linear or were nonlinear in the parameters.

This generality may be contrasted with the specification needed to obtain "nice" sampling theory results. For example, a common formulation assumes that the $\boldsymbol{\xi}_{ui}$ are common, that the $\boldsymbol{\theta}_i$ are *not*, and that the expectation functions are all linear in the parameters.

In the special case in which there is only one output response y, (8.2.23) reduces to

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto [S(\boldsymbol{\theta})]^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.2.24}$$

with $S(\boldsymbol{\theta}) = \sum_{u=1}^{n} [y_u - \eta(\boldsymbol{\xi}_u, \boldsymbol{\theta})]^2$. As we have seen, this result can be regarded as supplying a Bayesian justification of least squares, since the modal values of $\boldsymbol{\theta}$ (those associated with maximum posterior density) are those which minimize S. The general result (8.2.23) supplies then, among other things, an appropriate Bayesian multivariate generalization of least squares. The "most probable" values of $\boldsymbol{\theta}$ being simply those which minimize the determinant $|\mathbf{S}(\boldsymbol{\theta})|$.

Finally, in the special case $\sigma_{ij} = 0, i \neq j$, combining (8.2.16) with (8.2.6) and integrating out $\sigma_{11}, \ldots, \sigma_{mm}$ yields

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \prod_{i=1}^{m} [S_{ii}(\boldsymbol{\theta}_i)]^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty. \tag{8.2.25}$$

8.2.6 Estimation of Common Parameters in a Nonlinear Multivariate Model

We now illustrate the general applicability of the result (8.2.23) by considering an example in which:

a) certain of the θ's are common to more than one output, and

b) the expectation functions are nonlinear in the parameters.

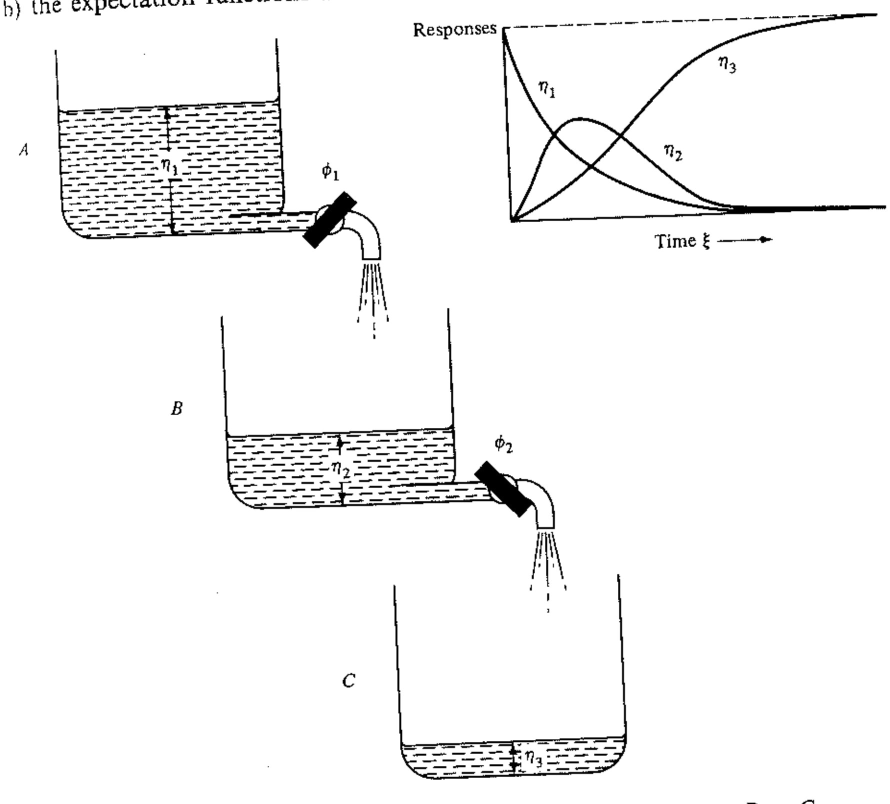

Fig. 8.2.1 Diagrammatic representation of a system $A \to B \to C$.

Suppose we have the consecutive system indicated in Fig. 8.2.1, which shows water running from a tank A via a tap opened an amount ϕ_1 into a tank B which then runs into a tank C via a tap opened an amount ϕ_2.

If η_1, η_2 and η_3 are the proportions of A, B, and C present at time ξ, with initial conditions ($\eta_1 = 1, \eta_2 = 0, \eta_3 = 0$), the system can be described by the differential equations

$$\frac{d\eta_1}{d\xi} = -\phi_1\eta_1,$$

$$\frac{d\eta_2}{d\xi} = \phi_1\eta_1 - \phi_2\eta_2, \tag{8.2.26}$$

$$\frac{d\eta_3}{d\xi} = \phi_2\eta_2.$$

Systems of this kind have many applications in engineering and in the physical and biological sciences. In particular, the equation (8.2.26) could represent a consecutive first-order chemical reaction in which a substance A decomposed to form B, which in turn decomposed to form C. The responses η_1, η_2, η_3 would then be the mole fractions of A, B, and C present at time ξ and the quantities ϕ_1 and ϕ_2 would then be rate constants associated with the first and second decompositions and would normally have to be estimated from data.

If we denote by y_1, y_2, and y_3 the *observed* values of η_1, η_2, and η_3, then, on integration of (8.2.26), we have the expectation functions

$$E(y_1) = \eta_1 = e^{-\phi_1\xi}, \tag{8.2.27a}$$

$$E(y_2) = \eta_2 = (e^{-\phi_1\xi} - e^{-\phi_2\xi})\,\phi_1/(\phi_2 - \phi_1), \tag{8.2.27b}$$

$$E(y_3) = \eta_3 = 1 + (-\,\phi_2\,e^{-\phi_1\xi} + \phi_1\,e^{-\phi_2\xi})/(\phi_2 - \phi_1), \tag{8.2.27c}$$

and it is to be noted that for all ξ,

$$\eta_1 + \eta_2 + \eta_3 \equiv 1. \tag{8.2.27d}$$

Observations on y_1 could yield information only on ϕ_1, but observations on y_2 and y_3 could each provide information on both ϕ_1 and ϕ_2. If measurements of more than one of the quantities (y_1, y_2, y_3) were available, we should certainly expect to be able to estimate the parameters more precisely. The Bayesian approach allows us to pool the information from (y_1, y_2, y_3) and makes it easy

Table 8.2.1

Observations on the yield of three substances in a chemical reaction

Time $=\xi_u$	Yield of A y_{1u}	Yield of B y_{2u}	Yield of C y_{3u}
$\frac{1}{2}$	0.959	0.025	0.028
$\frac{1}{2}$	0.914	0.061	0.000
1	0.855	0.152	0.068
1	0.785	0.197	0.096
2	0.628	0.130	0.090
2	0.617	0.249	0.118
4	0.480	0.184	0.374
4	0.423	0.298	0.358
8†	0.166	0.147	0.651
8†	0.205	0.050	0.684
16†	0.034	0.000	0.899
16†	0.054	0.047	0.991

† These four runs are omitted in the second analysis.

to appreciate the contribution from each of the three responses. In this example ξ is the only input variable and is the elapsed time since the start of the reaction. We denote by $\mathbf{y}'_{(u)} = (y_{u1}, y_{u2}, y_{u3})$ a set of $m = 3$ observations made on $\eta_{1u}, \eta_{2u}, \eta_{3u}$ at time ξ_u. A typical set of such observations is shown in Table 8.2.1.

In some cases observations may not be available on all three of the outputs. Thus only the concentration y_2 of the product B might be observable, or y_2 and y_3 might be known, but there might be no independently measured observation y_1 of the concentration of A.†

We suppose that the observations of Table 8.2.1 may be treated as having arisen from 12 *independent* experimental runs, as might be appropriate if the runs were carried out in random order in sealed tubes, each reaction being terminated at the appropriate time by sudden cooling. Furthermore, we suppose that (y_1, y_2, y_3) are functionally independent so that the 3×3 matrix $\boldsymbol{\Sigma}$ may be assumed to be positive definite and contains three variances and three covariances, all unknown. It is perhaps most natural for the experimenter to think in terms of the logarithms $\theta_1 = \log \phi_1$ and $\theta_2 = \log \phi_2$ of the rate constants and to regard these as locally uniformally distributed *a priori*.‡ We shall, therefore, choose as our reference priors for (θ_1, θ_2) and $\boldsymbol{\Sigma}$ the distributions in (8.2.8) and (8.2.14), respectively.

---

† When the chemist has difficulty in determining one of the products he sometimes makes use of relations like (8.2.27d) to "obtain it by calculation." Thus he might "obtain" y_1 from the relation $y_1 = 1 - y_2 - y_3$. For the resulting data set, the 3×3 covariance matrix $\boldsymbol{\Sigma}$ will of course not be positive definite, and the analysis in terms of three-dimensional responses will be inappropriate. In particular, the determinant of the sums of squares and products which appears in (8.2.23) will be zero *whatever* the values of the parameters. The difficulty is of course overcome very simply. The quantity y_1 is not an observation and the *data* has two dimensions, not three. The analysis should be carried through with y_2 and y_3 which *have* actually been measured. For a fuller treatment of problems of this kind arising because of data dependence or near dependence, see Box, Erjavec, Hunter and MacGregor (1972).

‡ Suppose that (a) the expectation functions were *linear* in $\theta_1(\boldsymbol{\phi})$ and $\theta_2(\boldsymbol{\phi})$ where $\boldsymbol{\phi} = (\phi_1, \phi_2)$, (b) little was known *a priori* about either parameter compared with the information supplied by the data, and (c) any prior information about one parameter would supply essentially none about the other.

Then, arguing as in Section 1.3, a noninformative reference prior to $\boldsymbol{\theta}$ should be locally uniform.

Conditions (b) and (c) are likely to be applicable to this problem at least as approximations, but condition (a) is not, because the expectation functions are non-linear in ϕ_1 and ϕ_2 and no general linearizing transformation exists. However, [see for example Beale (1960), and Guttman and Meeter (1965)] the expectation functions are more "nearly linear" in $\theta_1 = \log \phi_1$ and $\theta_2 = \log \phi_2$. Thus, the assumption that θ_1 and θ_2 are locally uniform provides a better approximation to a noninformative prior for the rate constants. For reasons we have discussed earlier, the assumption is not critical and, if for example we assume ϕ_1 and ϕ_2 themselves to be locally uniform, the posterior distribution is not altered appreciably.

Expression (8.2.23) makes it possible to compute the posterior density for the parameters assuming observations are available on some or all of the products A, B, and C. Thus, we may consider the posterior distribution of $\boldsymbol{\theta} = (\theta_1, \theta_2)'$

a) if only yields $\mathbf{y}_2$ of product B are available

$$p(\boldsymbol{\theta} \mid \mathbf{y}_2) \propto [S_{22}(\boldsymbol{\theta})]^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.2.28a}$$

b) if only yields $\mathbf{y}_3$ of product C are available,

$$p(\boldsymbol{\theta} \mid \mathbf{y}_3) \propto [S_{33}(\boldsymbol{\theta})]^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.2.28b}$$

c) if only yields $\mathbf{y}_2$ and $\mathbf{y}_3$ of B and C are available

$$p(\boldsymbol{\theta} \mid \mathbf{y}_2, \mathbf{y}_3) \propto \begin{vmatrix} S_{22}(\boldsymbol{\theta}) & S_{23}(\boldsymbol{\theta}) \\ S_{23}(\boldsymbol{\theta}) & S_{33}(\boldsymbol{\theta}) \end{vmatrix}^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.2.28c}$$

and

d) if yields $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$ of the products A, B and C are all available

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto |\mathbf{S}(\boldsymbol{\theta})|^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.2.28d}$$

where $\mathbf{S}(\boldsymbol{\theta}) = \{S_{ij}(\boldsymbol{\theta})\}$, $i, j = 1, 2, 3$.

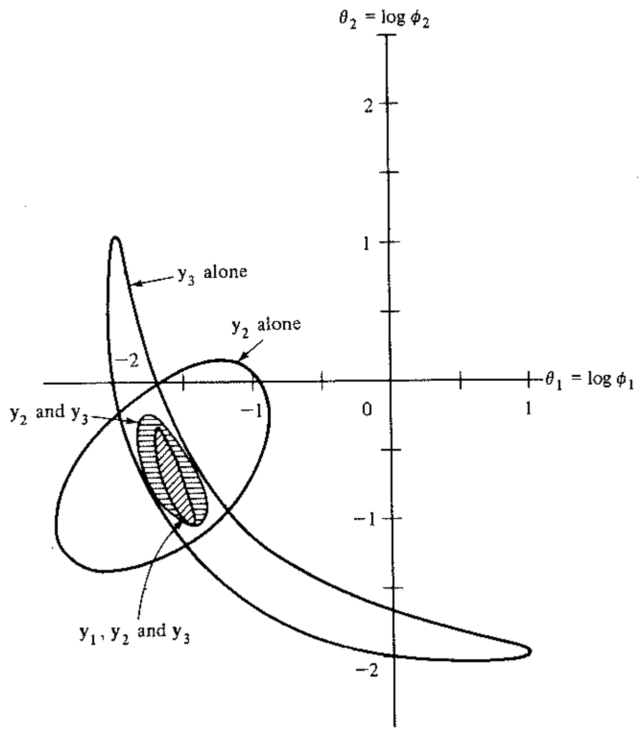

Fig. 8.2.2 99.75% H.P.D. regions for θ_1 and θ_2 for the chemical reaction data.

Since there are only two parameters θ_1 and θ_2, the posterior distributions can be represented by contour diagrams which may be superimposed to show the contributions made by the various output responses. Single contours are shown in Fig. 8.2.2 of the posterior distributions of θ_1 and θ_2 for (a) $\mathbf{y}_2$ alone, (b) $\mathbf{y}_3$ alone, (c) $\mathbf{y}_2$ and $\mathbf{y}_3$ jointly, and (d) $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ jointly. The contours actually shown are those which should correspond to an H.P.D. region containing approximately 99.75% of the probability mass calculated from

$$\log p(\hat{\boldsymbol{\theta}} \mid \cdot) - \log p(\boldsymbol{\theta} \mid \cdot) = \tfrac{1}{2}\chi^2(2, \alpha), \qquad \alpha = 0.0025$$

where $p(\boldsymbol{\theta} \mid \cdot)$ refers to the appropriate distributions in (8.2.28a–d) and $\hat{\boldsymbol{\theta}}$, the corresponding modal values of $\boldsymbol{\theta}$. In this example, it is apparent, particularly for $\mathbf{y}_3$, that the posterior distributions are non-Normal. Nevertheless, the above very crude approximation will suffice for the purpose of the present discussion.

In studying Figure 8.2.2, we first consider the moon-shaped contour obtained from observations $\mathbf{y}_3$ on the end product C alone. In any sequential reaction $A \to B \to C \to \ldots$ etc., we should expect that observation of *only* the end product (C in this case) could provide little or no information about the *individual* parameters but only about some aggregate of these rate constants. A diagonally attenuated ridge-like surface is therefore to be expected. However, it should be further noted that since in this specific instance η_3 is *symmetric* in θ_1 and θ_2 [see expression (8.2.27c)], the posterior surface is *completely symmetric* about the line $\theta_1 = \theta_2$. In particular, if $(\hat{\theta}_1, \hat{\theta}_2)$ is a point of maximum density the point $(\hat{\theta}_2, \hat{\theta}_1)$ will also give the same maximum density. In general the surface will be bimodal and have two peaks of equal height symmetrically situated about the equi-angular line. Marginal distributions will thus display precisely the kind of behaviour shown in Fig. A5.6.1.

Figure 8.2.2 shows, how, for this data, the inevitable ambiguity arising when only observations $\mathbf{y}_3$ on product C are utilized, is resolved as soon as the additional information supplied by values $\mathbf{y}_2$ on the intermediate product B is considered. As can be expected, the nature of the evidence that the intermediate product $\mathbf{y}_2$ contributes, is preferentially concerned with the difference of the parameters. This is evidenced by the tendency of the region to be obliquely oriented approximately at right angles to that for $\mathbf{y}_3$. By combining information from the two sources we obtain a much smaller region contained within the intersection of the individual regions. Finally, information from $\mathbf{y}_1$ which casts further light on the value of θ_1, causes the region to be further reduced.

Data of this kind sometimes occur in which available observations trace only part of the reaction. To demonstrate the effect of this kind of inadequacy in the experimental design, the analysis is repeated omitting the last four observations in Table 8.2.1. As shown in Fig. 8.2.3, over the ranges studied, the contours for $\mathbf{y}_2$ alone and $\mathbf{y}_3$ alone do not now close. Nevertheless, quite precise estimation is possible using $\mathbf{y}_2$ and $\mathbf{y}_3$ together and the addition of $\mathbf{y}_1$ improves the estimation further.

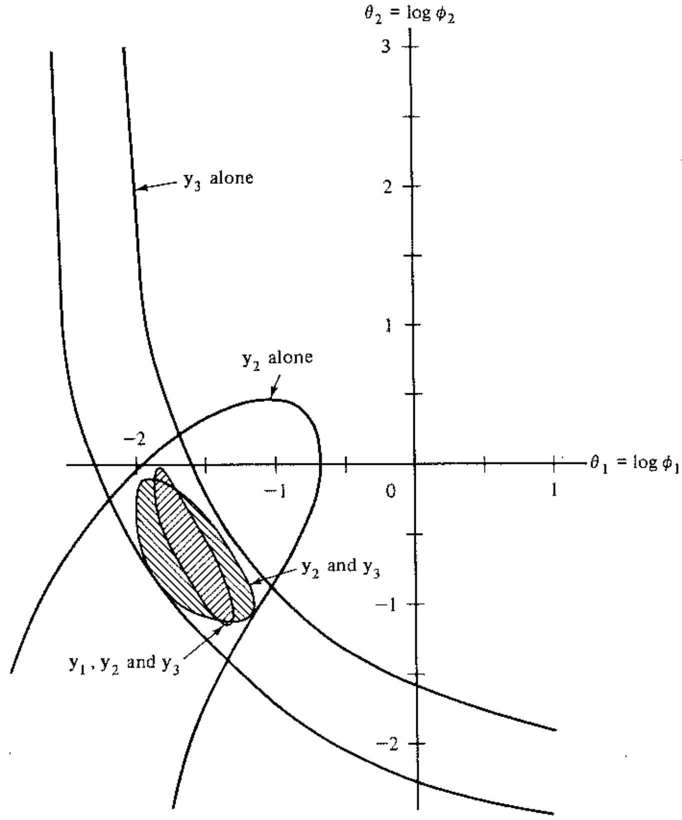

Fig. 8.2.3 99.75% H.P.D. regions for θ_1 and θ_2, excluding the last four observations.

Precautions in the Estimation of Common Parameters

Even in cases where only a single response is being considered, caution is needed in the fitting of functions. As explained in Section 1.1.4, fitting should be regarded as merely one element in the iterative model building process. The appropriate attitude is that when the model is initially fitted it is tentatively entertained rather than assumed. Careful checks on residuals are applied in a process of model criticism to see whether there is reason to doubt its applicability to the situation under consideration.

The importance of such precaution is even greater when several responses are considered. In multivariate problems, not only should each response model be checked individually but they must also be checked for overall consistency. The investigator should in practice *not* revert immediately to a joint analysis of responses. He should:

1) check the individual fit of each response,
2) compare posterior distributions to appraise the consistency of the information from the various responses (an aspect discussed in more detail in Chapter 9).

Only in those cases where he is satisfied with the individual fit and with the consistency shall he revert to the joint analysis.

8.3 LINEAR MULTIVARIATE MODELS

In discussing the general m-variate Normal model above, we have not needed to assume anything specific about the form of the m expectation functions $\boldsymbol{\eta}$. In particular, they need not be linear in the parameters† nor does it matter whether or not some parameters appear in more than one of the expectation functions. Many interesting and informative special cases arise if we suppose the expectation functions to be linear in the θ's. Moreover, as will be seen, the linear results can sometimes supply adequate local approximations for models non-linear in the parameters. From now on then we assume that

$$E(y_{ui}) = \eta_i(\boldsymbol{\xi}_{ui}, \boldsymbol{\theta}_i) = \mathbf{x}'_{(ui)}\boldsymbol{\theta}_i, \qquad i = 1, \ldots, m, \quad u = 1, \ldots, n, \tag{8.3.1}$$

where

$$\boldsymbol{\theta}'_i = (\theta_{1i}, \ldots, \theta_{gi}, \ldots, \theta_{k_i i})$$

and

$$\mathbf{x}'_{(ui)} = (x_{u1i}, \ldots, x_{ugi}, \ldots, x_{uk_i i})$$

with

$$x_{ugi} = \frac{\partial \eta_i(\boldsymbol{\xi}_{ui}, \boldsymbol{\theta}_i)}{\partial \theta_{gi}}$$

independent of all the θ's.

The $n \times k_i$ matrix $\mathbf{X}_i$ whose uth row is $\mathbf{x}'_{(ui)}$ will be called the *derivative* matrix for the ith response.

Our linear m-variate model may now be written as

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\theta}_1 + \boldsymbol{\varepsilon}_1 \\ &\vdots \\ \mathbf{y}_i &= \mathbf{X}_i\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i \\ &\vdots \\ \mathbf{y}_m &= \mathbf{X}_m\boldsymbol{\theta}_m + \boldsymbol{\varepsilon}_m. \end{aligned} \tag{8.3.2}$$

Certain characteristics of this mode of writing the model should be noted. In particular, it is clear that while the elements of $\mathbf{x}_{(ui)}$ will be functions of the

† Although, so that a uniform density can represent an approximately noninformative prior and also to assist local linear approximation, parameter transformations in terms of which the expectation function is more nearly linear, will often be employed.

elements of the vector input variables $\boldsymbol{\xi}_{ui}$, they will in general not be proportiona to the elements of $\boldsymbol{\xi}_{ui}$ themselves. Thus if

$$E(y_{ui}) = \frac{\theta_1 \log \xi_{u1i} + \theta_2\, \xi_{u1i}\, \xi_{u3i}}{\xi_{u2i}},$$

then

$$x_{u1i} = \frac{\log \xi_{u1i}}{\xi_{u2i}} \qquad \text{and} \qquad x_{u2i} = \frac{\xi_{u1i}\, \xi_{u3i}}{\xi_{u2i}}.$$

8.3.1 The Use of Linear Theory Approximations when the Expectation is Nonlinea in the Parameters

The specific form of posterior distributions which we shall obtain for the linea case will often provide reasonably close approximations even when th expectation functions $\boldsymbol{\eta}_{(u)}$ is *nonlinear* in $\boldsymbol{\theta}$. This is because we need only tha the expectation functions are approximately linear *in the region of the paramete space covered by most of the posterior distribution*, say within the 95% H.P.D region.† For moderate n, this can happen with functions that are highl nonlinear in the parameters when considered over their whole range. Then, in th region where the posterior probability mass is concentrated (say the 95% H.P.D. region), we may expand the expectation function around the mode $\hat{\boldsymbol{\theta}}_i$

$$E(y_{ui}) = \eta_{ui} \doteq \eta_i(\boldsymbol{\xi}_{ui}, \hat{\boldsymbol{\theta}}_i) + \sum_{g=1}^{k_i} x_{ugi}\,(\theta_{gi} - \hat{\theta}_{gi}), \qquad (8.3.3)$$

where

$$x_{ugi} = \left.\frac{\partial \eta_i(\boldsymbol{\xi}_{ui}, \boldsymbol{\theta}_i)}{\partial \theta_{gi}}\right|_{\boldsymbol{\theta}_i = \hat{\boldsymbol{\theta}}_i},$$

which is, approximately, in the form of a linear model. Thus, the posterio distributions found from linear theory can, in many cases, provide clos approximations to the true distributions. For example, in the univariate cas ($m = 1$) with a single parameter θ, the posteriror distribution in (8.2.24) would b approximately

$$p(\theta \mid \mathbf{y}) \mathrel{\dot{\propto}} [\nu s^2 + (\Sigma\, x_u^2)\,(\theta - \hat{\theta})^2]^{-n/2} \qquad (8.3.4)$$

where

$$\nu = n - 1, \qquad s^2 = \frac{1}{\nu}\,\Sigma[y_u - \eta(\xi_u, \hat{\theta})]^2 \qquad \text{and} \qquad x_u = \left.\frac{\partial \eta(\xi_u, \theta)}{\partial \theta}\right|_{\theta = \hat{\theta}},$$

so that the quantity

$$\frac{\sqrt{\Sigma x_u^2}\,(\theta - \hat{\theta})}{s} \qquad (8.3.5)$$

would be approximately distributed as $t(0, 1, \nu)$.

---

† A possibility that can be checked a posteriori for any specific case.

When, as in the case in which multivariate or univariate least squares is appropriate, a convenient method of calculating the $\hat{\theta}$'s is available for the linear case but not for the corresponding nonlinear situation, the linearization may be used iteratively to find the $\hat{\theta}$'s for the nonlinear situation.

For example, in the univariate model containing a single parameter θ, with a first guess θ_0, we can write approximately

$$E(z_{u0}) \doteq (\theta - \theta_0)x_u, \tag{8.3.6}$$

where

$$z_{u0} = y_u - \eta(\xi_u, \theta_0) \quad \text{and} \quad x_u = \left.\frac{\partial \eta(\xi_u, \theta)}{\partial \theta}\right|_{\theta=\theta_0}$$

Applying ordinary least squares to the model, we obtain an estimate of the correction $\theta - \theta_0$ and hence hopefully an improved "guess" θ_1 from

$$\theta_1 = \theta_0 + \frac{\Sigma z_{u0}\, x_u}{\Sigma x_u^2}. \tag{8.3.7}$$

This is the well-known Newton–Gauss method of iteration for nonlinear least squares, Box (1957, 1960), Hartley (1961), Marquardt (1963), and under favorable conditions the successive iterants will converge to $\hat{\theta}$.

8.3.2 Special Cases of the General Linear Multivariate Model

In general, the joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ is given by (8.2.17) and the marginal distribution of $\boldsymbol{\theta}$ is that in (8.2.23) quite independently of whether $\eta_i(\boldsymbol{\xi}_{ui}, \boldsymbol{\theta}_i)$ is linear in $\boldsymbol{\theta}_i$ or not. For practical purposes, however, it is of interest to consider a number of special cases.

For orientation we reconsider for a moment the linear *univariate* situation discussed earlier in Section 2.7,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{8.3.8}$$

where $\mathbf{y}$ is a $n \times 1$ vector of observations, $\mathbf{X}$ a $n \times k$ matrix of fixed elements, $\boldsymbol{\theta}$ a $k \times 1$ vector of parameters and $\boldsymbol{\varepsilon}$ a $n \times 1$ vector of errors. In this case,

$$p(\boldsymbol{\theta}; \sigma^2 \mid \mathbf{y}) \propto (\sigma^2)^{-(\frac{1}{2}n+1)} \exp\left[-\frac{S(\boldsymbol{\theta})}{2\sigma^2}\right], \qquad \sigma^2 > 0, \quad -\infty < \boldsymbol{\theta} < \infty, \tag{8.3.9}$$

and

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto [S(\boldsymbol{\theta})]^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.3.10}$$

The determinant $|\mathbf{S}(\boldsymbol{\theta})|$ in (8.2.23) becomes the single sum of squares

$$S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}). \tag{8.3.11}$$

In this linear case, we may write

$$S(\boldsymbol{\theta}) = (n-k)s^2 + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \tag{8.3.12}$$

where

$$(n-k)s^2 = (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\theta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\theta}})$$

and, assuming $\mathbf{X}$ is of rank k,

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

so that, writing $\nu = n - k$,

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \left[1 + \frac{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})}{\nu s^2}\right]^{-\frac{1}{2}(\nu+k)}, \qquad -\infty < \boldsymbol{\theta} < \infty. \tag{8.3.13}$$

The posterior distribution of $\boldsymbol{\theta}$ is thus the k dimensional $t_k[\hat{\boldsymbol{\theta}}, s^2(\mathbf{X}'\mathbf{X})^{-1}, \nu]$ distribution. Further, integrating out $\boldsymbol{\theta}$ from (8.3.9) yields the distribution of σ^2,

$$p(\sigma^2 \mid \mathbf{y}) \propto (\sigma^2)^{-(\frac{1}{2}\nu+1)} \exp\left(-\frac{\nu s^2}{2\sigma^2}\right), \qquad \sigma^2 > 0, \tag{8.3.14}$$

so that $\sigma^2/(\nu s^2)$ has the χ_ν^{-2} distribution. All the above results have, of course, been already obtained earlier in Section 2.7.

It is clear that the general linear model (8.3.2) which can be regarded as the multivariate generalization of (8.3.8) need not be particularized in any way. The matrices $\mathbf{X}_1, \ldots, \mathbf{X}_m$ *may or may not* have elements in common; furthermore, the vectors of parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ *may or may not* have elements in common. Using sampling theory, Zellner (1962, 1963) attempted to study the situation in which the $\mathbf{X}_i$ were not assumed to be identical. The main difficulty with his approach was that the minimum variance estimator for $\boldsymbol{\theta}$ involves the unknown $\boldsymbol{\Sigma}$, and the estimators proposed are "optimal" only in the asymptotic sense.

Cases of special interest which are associated with practical problems of importance and which relate to known results include:

a) when the derivative matrices $\mathbf{X}_1 = \ldots = \mathbf{X}_m = \mathbf{X}$ are common but the parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ are not,
b) when $\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_m$ but the matrices $\mathbf{X}_1, \ldots, \mathbf{X}_m$ are not, and
c) when $\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_m$ *and* $\mathbf{X}_1 = \ldots = \mathbf{X}_m$.

In the remaining part of this chapter, we shall discuss case (a). The problem of estimating common parameters which includes (b) and (c) will be treated in the next chapter.

8.4 INFERENCES ABOUT θ FOR THE CASE OF A COMMON DERIVATIVE MATRIX X

The model for which $\mathbf{X}_1 = \ldots = \mathbf{X}_m = \mathbf{X}$ (so that $k_1 = \ldots = k_m = k$) and $\boldsymbol{\theta}_1 \neq \ldots \neq \boldsymbol{\theta}_m$ has received most attention in the sampling theory framework —see, for example, Anderson (1958). From the Bayesian point of view, the problem has been studied by Savage (1961a), Geisser and Cornfield (1963), Geisser (1965a), Ando and Kaufman (1965), and others. In general, the multivariate model in (8.3.2) can now be written

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{8.4.1}$$

$$\underset{n\times m}{[\,\mathbf{y}\,]} = \underset{n\times k}{[\,\mathbf{X}\,]}\ \underset{k\times m}{[\,\boldsymbol{\theta}\,]} + \underset{n\times m}{[\,\boldsymbol{\varepsilon}\,]}$$

where the notation beneath the matrices indicates that $\mathbf{y}$ is an $n\times m$ matrix of m-variate observations, $\boldsymbol{\theta}$ is a $k\times m$ matrix of parameters and $\boldsymbol{\varepsilon}$ an $n\times m$ matrix of errors.

The model would be appropriate for example if say a 2^p factorial experiment had been conducted on a chemical process and the output y_1 = product yield, y_2 = product purity, y_3 = product density had been measured. The elements of each column of the common matrix $\mathbf{X}$ would then be an appropriate sequence of $+1$'s and -1's corresponding to the experimental conditions and the "effect" parameters $\boldsymbol{\theta}_i$ would be different for each output. In econometrics, the model (8.4.1) is frequently encountered in the analysis of the reduced form of simultaneous equation systems.

We note that the $k\times m$ matrix of parameters

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_{11} & \cdots & \theta_{1i} & \cdots & \theta_{1m} \\ \vdots & & \vdots & & \vdots \\ \theta_{g1} & \cdots & \theta_{gi} & \cdots & \theta_{gm} \\ \vdots & & \vdots & & \vdots \\ \theta_{k1} & \cdots & \theta_{ki} & \cdots & \theta_{km} \end{bmatrix} \tag{8.4.2}$$

can be written in the two alternative forms

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i, \ldots, \boldsymbol{\theta}_m] = \begin{bmatrix} \boldsymbol{\theta}'_{(1)} \\ \vdots \\ \boldsymbol{\theta}'_{(g)} \\ \vdots \\ \boldsymbol{\theta}'_{(k)} \end{bmatrix} \tag{8.4.3}$$

where $\boldsymbol{\theta}_i$ is the ith column vector and $\boldsymbol{\theta}'_{(g)}$ is the gth row vector of $\boldsymbol{\theta}$. For simplicity, we shall assume throughout the chapter that the rank of $\mathbf{X}$ is k.

8.4.1 Distribution of θ

Consider the elements of the $m\times m$ matrix $\mathbf{S}(\boldsymbol{\theta}) = \{S_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)\}$ of (8.2.4). When $\mathbf{X}_1 = \ldots = \mathbf{X}_m = \mathbf{X}$, we can write

$$\begin{aligned} S_{ij}(\boldsymbol{\theta}_i,\boldsymbol{\theta}_j) &= (\mathbf{y}_i - \mathbf{X}\boldsymbol{\theta}_i)'(\mathbf{y}_j - \mathbf{X}\boldsymbol{\theta}_j) \\ &= (\mathbf{y}_i - \mathbf{X}\hat{\boldsymbol{\theta}}_i)'(\mathbf{y}_j - \mathbf{X}\hat{\boldsymbol{\theta}}_j) + (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_j - \hat{\boldsymbol{\theta}}_j) \end{aligned} \tag{8.4.4}$$

where $\hat{\boldsymbol{\theta}}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i$ is the least squares estimates of $\boldsymbol{\theta}_i, i = 1, \ldots, m$.

Consequently,

$$\mathbf{S}(\boldsymbol{\theta}) = \mathbf{A} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}), \tag{8.4.5}$$

where $\hat{\boldsymbol{\theta}}$ is the $k \times m$ matrix of least squares estimates

$$\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_i, \ldots, \hat{\boldsymbol{\theta}}_m] = \begin{bmatrix} \hat{\boldsymbol{\theta}}'_{(1)} \\ \vdots \\ \hat{\boldsymbol{\theta}}'_{(g)} \\ \vdots \\ \hat{\boldsymbol{\theta}}'_{(k)} \end{bmatrix}, \tag{8.4.6}$$

and $\mathbf{A}$ is the $m \times m$ matrix

$$\mathbf{A} = \{a_{ij}\}$$

with

$$a_{ij} = (\mathbf{y}_i - \mathbf{X}\hat{\boldsymbol{\theta}}_i)'(\mathbf{y}_j - \mathbf{X}\hat{\boldsymbol{\theta}}_j), \qquad i,j = 1, \ldots, m, \tag{8.4.7}$$

that is, $\mathbf{A}$ is proportional to the sample covariance matrix. For simplicity, we shall assume that $\mathbf{A}$ is positive definite. From the general result in (8.2.23), the posterior distribution of $\boldsymbol{\theta}$ is then

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto |\mathbf{A} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty. \tag{8.4.8}$$

As mentioned earlier, when there is a single output ($m = 1$), (8.4.8) is in the form of a k-dimensional multivariate t distribution. The distribution in (8.4.8) is a matric-variate generalization of the t distribution. It was first obtained by Kahirsagar (1960). A comprehensive discussion of its properties has been given by Dickey (1967b).

8.4.2 Posterior Distribution of the Means from a m-dimensional Normal Distribution

In the case $k = 1$ where each $\boldsymbol{\theta}_i$ consists of a single element and $\mathbf{X}$ is a $n \times 1$ vector of ones, expression (8.4.8) is the joint posterior distribution of the m means when sampling from an m-dimensional multivariate Normal distribution $N_m(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. In this case

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_i, \ldots, \theta_m), \quad \hat{\boldsymbol{\theta}} = (\bar{y}_1, \ldots, \bar{y}_i, \ldots, \bar{y}_m),$$

$$\mathbf{X}'\mathbf{X} = n \quad \text{and} \quad a_{ij} = \sum_{u=1}^{n} (y_{ui} - \bar{y}_i)(y_{uj} - \bar{y}_j), \tag{8.4.9}$$

where

$$\bar{y}_i = \frac{1}{n} \sum_{u=1}^{n} y_{ui}.$$

The posterior distribution of $\boldsymbol{\theta}$ can be written

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto |\mathbf{A} + n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-n/2}$$
$$\propto |\mathbf{I} + n\,\mathbf{A}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-n/2}, \qquad -\infty < \boldsymbol{\theta} < \infty. \tag{8.4.10}$$

We now make use of the fundamental identity

$$|\mathbf{I}_k - \mathbf{P}\,\mathbf{Q}| = |\mathbf{I}_l - \mathbf{Q}\,\mathbf{P}|, \tag{8.4.11}$$

where $\mathbf{I}_k$ and $\mathbf{I}_l$ are, respectively, $k \times k$ and a $l \times l$ identity matrices, $\mathbf{P}$ is a $k \times l$ matrix and $\mathbf{Q}$ is a $l \times k$ matrix. Noting that $(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ is a $1 \times m$ vector, we immediately obtain

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto [1 + n(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\,\mathbf{A}^{-1}\,(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})']^{-n/2}, \qquad -\infty < \theta_i < \infty, \quad i = 1, \ldots, m, \tag{8.4.12}$$

which is a m-dimensional $t_m[\hat{\boldsymbol{\theta}}', n^{-1}(n-m)^{-1}\mathbf{A}, n-m]$ distribution, a result first published by Geisser and Cornfield (1963). Thus, by comparing (8.4.12) with (8.3.13), we see that both when $m = 1$ and when $k = 1$, the distribution in (8.4.8) can be put in the multivariate t form.

8.4.3 Some Properties of the Posterior Matric-variate t Distribution of $\boldsymbol{\theta}$

When neither m nor k is equal to one, it is not possible to express the distribution of $\boldsymbol{\theta}$ as a multivariate t distribution. As we have mentioned, the distribution in (8.4.8) can be thought of as a matric-variate extension of the t distribution. We now discuss some properties of this distribution.

Two equivalent Representations of the Distribution of $\boldsymbol{\theta}$

It is shown in Appendix A8.3 that, for $\nu > 0$,

$$\int_{-\infty < \boldsymbol{\theta} < \infty} |\mathbf{I}_m + \mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-\frac{1}{2}(\nu+k+m-1)}\,d\boldsymbol{\theta}$$
$$= c(m, k, \nu)\,|\mathbf{X}'\mathbf{X}|^{-m/2}\,|\mathbf{A}^{-1}|^{-k/2}, \tag{8.4.13}$$

where

$$\nu = n - (k + m) + 1,$$
$$c(m, k, \nu) = [\Gamma(\tfrac{1}{2})]^{mk}\,\frac{\Gamma_m[\frac{1}{2}(\nu + m - 1)]}{\Gamma_m[\frac{1}{2}(\nu + k + m - 1)]} \tag{8.4.14}$$

and $\Gamma_p(b)$ is the generalized Gamma function defined in (8.2.22). Thus,

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = [c(m,k,\nu)]^{-1}\,|\mathbf{X}'\mathbf{X}|^{m/2}\,|\mathbf{A}^{-1}|^{k/2}\,|\mathbf{I}_m + \mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-\frac{1}{2}(\nu+k+m-1)},$$
$$-\infty < \boldsymbol{\theta} < \infty. \tag{8.4.15}$$

and we shall say that the $k \times m$ matrix of parameters $\boldsymbol{\theta}$ is distributed as $t_{km}[\hat{\boldsymbol{\theta}}, (\mathbf{X}'\mathbf{X})^{-1}, \mathbf{A}, \nu]$. Note that by applying the identity (8.4.11), we can write

$$|\mathbf{I}_m + \mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})| = |\mathbf{I}_k + (\mathbf{X}'\mathbf{X})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'| \quad (8.4.16)$$

so that in terms of the $m \times k$ matrix $\boldsymbol{\theta}'$ the roles of m and k on the one hand and of the matrices $(\mathbf{X}'\mathbf{X})^{-1}$ and $\mathbf{A}$ on the other are simultaneously interchanged. Thus, we may conclude that if

$$\boldsymbol{\theta} \sim t_{km}[\hat{\boldsymbol{\theta}}, (\mathbf{X}'\mathbf{X})^{-1}, \mathbf{A}, \nu],$$

then

$$\boldsymbol{\theta}' \sim t_{mk}[\hat{\boldsymbol{\theta}}', \mathbf{A}, (\mathbf{X}'\mathbf{X})^{-1}, \nu]. \quad (8.4.17)$$

It follows from these two equivalent representations that

$$c(k, m, \nu) \equiv c(m, k, \nu),$$

that is,

$$\frac{\Gamma_k[\frac{1}{2}(\nu + k - 1)]}{\Gamma_k[\frac{1}{2}(\nu + k + m - 1)]} \equiv \frac{\Gamma_m[\frac{1}{2}(\nu + m - 1)]}{\Gamma_m[\frac{1}{2}(\nu + k + m - 1)]}. \quad (8.4.18)$$

Marginal and Conditional Distributions of Subsets of Columns of $\boldsymbol{\theta}$

We now show that the marginal and conditional distributions of subsets of the m columns of $\boldsymbol{\theta}$ are also matric-variate t distributions. Let $m = m_1 + m_2$ and partition the matrices $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, and $\mathbf{A}$ into

$$\boldsymbol{\theta} = [\overset{m_1}{\boldsymbol{\theta}_{1*}} \;\vdots\; \overset{m_2}{\boldsymbol{\theta}_{2*}}]_k, \qquad \hat{\boldsymbol{\theta}} = [\overset{m_1}{\hat{\boldsymbol{\theta}}_{1*}} \;\vdots\; \overset{m_2}{\hat{\boldsymbol{\theta}}_{2*}}]_k,$$

$$\mathbf{A} = \begin{array}{cc} \begin{array}{cc} m_1 & m_2 \end{array} & \\ \left[\begin{array}{c:c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hdashline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] & \begin{array}{c} m_1 \\ m_2 \end{array} \end{array}. \quad (8.4.19)$$

Then:

a) conditional on $\boldsymbol{\theta}_{1*}$, the subset $\boldsymbol{\theta}_{2*}$ is distributed as

$$\boldsymbol{\theta}_{2*} \sim t_{km_2}(\tilde{\boldsymbol{\theta}}_{2*}, \mathbf{H}^{-1}, \mathbf{A}_{22\cdot 1}, \nu + m_1), \quad (8.4.20)$$

where

$$\mathbf{H}^{-1} = (\mathbf{X}'\mathbf{X})^{-1} + (\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})\mathbf{A}_{11}^{-1}(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})',$$

$$\tilde{\boldsymbol{\theta}}_{2*} = \hat{\boldsymbol{\theta}}_{2*} + (\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})\mathbf{A}_{11}^{-1}\mathbf{A}_{12},$$

$$\mathbf{A}_{22\cdot 1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}.$$

b) $\boldsymbol{\theta}_{1*}$ is distributed as

$$\boldsymbol{\theta}_{1*} \sim t_{km_1}[\hat{\boldsymbol{\theta}}_{1*}, (\mathbf{X}'\mathbf{X})^{-1}, \mathbf{A}_{11}, \nu]. \quad (8.4.21)$$

To prove these results, we can write

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' = (\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})\mathbf{A}_{11}^{-1}(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})'$$
$$+ (\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})\mathbf{A}_{22\cdot 1}^{-1}(\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})'. \quad (8.4.22)$$

The determinant on the right-hand side of (8.4.16) can now be written

$$|\mathbf{I}_k + (\mathbf{X}'\mathbf{X})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'| = |\mathbf{I}_k + (\mathbf{X}'\mathbf{X})(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})\mathbf{A}_{11}^{-1}(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})'| \times |\mathbf{I}_k + \mathbf{H}(\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})\mathbf{A}_{22\cdot 1}^{-1}(\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})'|. \qquad (8.4.23)$$

Substituting (8.4.23) into (8.4.15), we see that, given $\boldsymbol{\theta}_{1*}$, the conditional distribution of $\boldsymbol{\theta}_{2*}$ is

$$p(\boldsymbol{\theta}_{2*} | \boldsymbol{\theta}_{1*}, \mathbf{y}) \propto |\mathbf{I}_k + \mathbf{H}(\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})\mathbf{A}_{22\cdot 1}^{-1}(\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})'|^{-\frac{1}{2}(\nu+k+m-1)}$$
$$\propto |\mathbf{I}_{m_2} + \mathbf{A}_{22\cdot 1}^{-1}(\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})'\mathbf{H}(\boldsymbol{\theta}_{2*} - \tilde{\boldsymbol{\theta}}_{2*})|^{-\frac{1}{2}[(\nu+m_1)+k+m_2-1]}$$
$$-\infty < \boldsymbol{\theta}_{2*} < \infty. \qquad (8.4.24)$$

From (8.4.13), the normalizing constant is

$$[c(m_2, k, \nu + m_1)]^{-1}\, |\mathbf{H}|^{m_2/2}\, |\mathbf{A}_{22\cdot 1}^{-1}|^{k/2}. \qquad (8.4.25)$$

Thus, given $\boldsymbol{\theta}_{1*}$, the $k \times m_2$ matrix of parameters $\boldsymbol{\theta}_{2*}$ is distributed as $t_{km_2}(\tilde{\boldsymbol{\theta}}_{2*}, \mathbf{H}^{-1}, \mathbf{A}_{22\cdot 1}, \nu + m_1)$. For the marginal distribution of $\boldsymbol{\theta}_{1*}$, since

$$p(\boldsymbol{\theta}_{1*} | \mathbf{y}) = \frac{p(\boldsymbol{\theta} | \mathbf{y})}{p(\boldsymbol{\theta}_{2*} | \boldsymbol{\theta}_{1*}, \mathbf{y})},$$

use of (8.4.23) through (8.4.25) yields

$$p(\boldsymbol{\theta}_{1*} | \mathbf{y}) \propto |\mathbf{H}|^{-m_2/2}\, |\mathbf{I}_k + (\mathbf{X}'\mathbf{X})(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})\mathbf{A}_{11}^{-1}(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})'|^{-\frac{1}{2}(\nu+k+m-1)}$$
$$\propto |\mathbf{I}_{m_1} + \mathbf{A}_{11}^{-1}(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_{1*} - \hat{\boldsymbol{\theta}}_{1*})|^{-\frac{1}{2}(\nu+k+m_1-1)},$$
$$-\infty < \boldsymbol{\theta}_{1*} < \infty. \qquad (8.4.26)$$

That is, the $k \times m_1$ matrix of parameters $\boldsymbol{\theta}_{1*}$ is distributed as

$$t_{km_1}[\hat{\boldsymbol{\theta}}_{1*}, (\mathbf{X}'\mathbf{X})^{-1}, \mathbf{A}_{11}, \nu].$$

Marginal Distribution of a Particular Column of $\boldsymbol{\theta}$

In particular, by setting $m_1 = 1$ in (8.4.21), the marginal distribution of $\boldsymbol{\theta}_{1*} = \boldsymbol{\theta}_1$ is the k-dimensional multivariate t distribution

$$p(\boldsymbol{\theta}_1 | \mathbf{y}) \propto [1 + a_{11}^{-1}(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)]^{-\frac{1}{2}(\nu+k)}, \qquad -\infty < \boldsymbol{\theta}_1 < \infty, \qquad (8.4.27)$$

where $a_{11} = \mathbf{A}_{11}$ is now a scalar, that is, $\boldsymbol{\theta}_1 \sim t_k[\hat{\boldsymbol{\theta}}_1, \nu^{-1} a_{11}(\mathbf{X}'\mathbf{X})^{-1}, \nu]$.

By mere relabeling we may conclude that the marginal distribution of the ith column of $\boldsymbol{\theta}$ in (8.4.2) is

$$p(\boldsymbol{\theta}_i | \mathbf{y}) \propto [1 + a_{ii}^{-1}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)]^{-\frac{1}{2}(\nu+k)}, \qquad -\infty < \boldsymbol{\theta}_i < \infty. \qquad (8.4.28)$$

It will be noted that this distribution is identical to that obtained in (8.3.13) when only a single output was considered, except that in (8.3.13)

$\nu = n - k$, but in (8.4.28) $\nu = n - k - (m - 1)$. In a certain sense, the reduction of the degrees of freedom by $m - 1$ is not surprising. In adopting the multivariate framework, $m(m-1)/2$ additional parameters σ_{ij} $(i \neq j)$ are introduced. A part of the information from the sample is therefore utilized to estimate these parameters and $(m - 1)$ of them $(\sigma_{i1}, \ldots, \sigma_{i(i-1)}, \sigma_{i(i+1)}, \ldots, \sigma_{im})$ are connected with $\mathbf{y}_i$. We may say that 'one degree of freedom is lost' for each of the $(m - 1)$ additional parameters.

On the other hand, it is somewhat puzzling that if we ignored the multivariate structure of the problem and treated $\mathbf{y}_i$ as the output of a *univariate* response, then on the basis of the noninformative prior $p(\boldsymbol{\theta}_i, \sigma_{ii}) \propto \sigma_{ii}^{-1}$, we would obtain a posterior multivariate t distribution for $\boldsymbol{\theta}_i$ with $(m - 1)$ additional degrees of freedom. This would seem to imply that, by ignoring the information from the other $(m - 1)$ responses $\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_m$, more precise inference about $\boldsymbol{\theta}_i$ could be made than when all the m responses were considered jointly. This phenomenon is related to the "paradox" pointed out by Dempster (1963) and the criticisms of the prior in (8.2.14) by Stone (1964).

The above implication is admittedly perplexing, and further research is needed to clarify the situation. We feel, however, that the multivariate results presented in this chapter are of considerable interest and certainly provide a sensible basis for inference in the common practical situation when $(n - k)$ is large relative to m.

Distribution of $\boldsymbol{\theta}$ *Expressed as a Product of Multivariate t Distributions*

We note that for the partition in (8.4.19), if we set $m_1 = m - 1$ and $m_2 = 1$, then from (8.4.20) the conditional distribution of $\boldsymbol{\theta}_{2*} = \boldsymbol{\theta}_m$, given $\boldsymbol{\theta}_{1*} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m-1}]$, is

$$\boldsymbol{\theta}_m \sim t_k[\hat{\boldsymbol{\theta}}_m, (\nu + m - 1)^{-1} a_{m\cdot 12\ldots(m-1)} \mathbf{H}^{-1}, \nu + m - 1], \tag{8.4.29}$$

where

$$a_{m\cdot 12\ldots(m-1)} = \mathbf{A}_{22\cdot 1}.$$

From the marginal distribution of $\boldsymbol{\theta}_{1*} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m-1}]$ in (8.4.21) if we partition $\boldsymbol{\theta}_{1*}$ into $[\boldsymbol{\theta}_{s*} \vdots \boldsymbol{\theta}_{m-1}]$ where $\boldsymbol{\theta}_{s*} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m-2}]$, it is clear that the conditional distribution of $\boldsymbol{\theta}_{m-1}$, given $\boldsymbol{\theta}_{s*}$, is again a k-dimensional multivariate t distribution. It follows by repeating the process $m - 1$ times that, if we express $p(\boldsymbol{\theta} \mid \mathbf{y})$ as the product

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = p(\boldsymbol{\theta}_1 \mid \mathbf{y}) p(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \mathbf{y}) \cdots p(\boldsymbol{\theta}_m \mid \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m-1}, \mathbf{y}), \tag{8.4.30}$$

then each factor on the right-hand side is a k-dimensional multivariate t distribution.

Marginal and Conditional Distributions of Rows of $\boldsymbol{\theta}$

Results very similar to those given above for column decomposition of $\boldsymbol{\theta}$ can now be obtained for the rows of $\boldsymbol{\theta}$. Consider the partitions

$$\boldsymbol{\theta}' = [\overset{k_1}{\boldsymbol{\theta}_{(1)*}} \,\vdots\, \overset{k_2}{\boldsymbol{\theta}_{(2)*}}]\, m, \qquad \hat{\boldsymbol{\theta}}' = [\overset{k_1}{\hat{\boldsymbol{\theta}}_{(1)*}} \,\vdots\, \overset{k_2}{\hat{\boldsymbol{\theta}}_{(2)*}}]\, m \tag{8.4.31}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{C} = \begin{array}{c} \begin{array}{cc} k_1 & k_2 \end{array} \\ \left[\begin{array}{c|c} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \hline \mathbf{C}_{21} & \mathbf{C}_{22} \end{array}\right] \end{array} \begin{array}{c} k_1 \\ k_2 \end{array}, \qquad k_1 + k_2 = k,$$

where it is to be remembered that $\boldsymbol{\theta}'_{(1)*}$ are the first k_1 rows and $\boldsymbol{\theta}'_{(2)*}$ the remaining k_2 rows of $\boldsymbol{\theta}$. Since the $m \times k$ matrix $\boldsymbol{\theta}'$ is distributed as $t_{mk}[\hat{\boldsymbol{\theta}}', \mathbf{A}, (\mathbf{X}'\mathbf{X})^{-1}, \nu]$, it can be readily shown that

a) given $\boldsymbol{\theta}_{(1)*}$,

$$\boldsymbol{\theta}_{(2)*} \sim t_{mk_2}(\tilde{\boldsymbol{\theta}}_{(2)*}, \mathbf{G}, \mathbf{C}_{22\cdot 1}, \nu + k_1), \tag{8.4.32}$$

where

$$\mathbf{C}_{22\cdot 1} = \mathbf{C}_{22} - \mathbf{C}_{21}\, \mathbf{C}_{11}^{-1}\, \mathbf{C}_{12},$$

$$\tilde{\boldsymbol{\theta}}_{(2)*} = \hat{\boldsymbol{\theta}}_{(2)*} + (\boldsymbol{\theta}_{(1)*} - \hat{\boldsymbol{\theta}}_{(1)*})\, \mathbf{C}_{11}^{-1}\, \mathbf{C}_{12},$$

and

$$\mathbf{G} = \mathbf{A} + (\boldsymbol{\theta}_{(1)*} - \hat{\boldsymbol{\theta}}_{(1)*})\, \mathbf{C}_{11}^{-1}\, (\boldsymbol{\theta}_{(1)*} - \hat{\boldsymbol{\theta}}_{(1)*})'.$$

b) Marginally,

$$\boldsymbol{\theta}_{(1)*} \sim t_{mk_1}[\hat{\boldsymbol{\theta}}_{(1)*}, \mathbf{A}, \mathbf{C}_{11}, \nu] \tag{8.4.33}$$

or equivalently,

$$\boldsymbol{\theta}'_{(1)*} \sim t_{k_1 m}[\hat{\boldsymbol{\theta}}'_{(1)*}, \mathbf{C}_{11}, \mathbf{A}, \nu].$$

c) The gth row of $\boldsymbol{\theta}$, $\boldsymbol{\theta}'_{(g)}$, is distributed as

$$\boldsymbol{\theta}_{(g)} \sim t_m[\hat{\boldsymbol{\theta}}_{(g)}, \nu^{-1}\, c_{gg}\, \mathbf{A}, \nu] \tag{8.4.34}$$

where c_{gg} is the (gg)th element of $\mathbf{C}$.

d) The distribution of $\boldsymbol{\theta}$ can alternatively be expressed as the product

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = p(\boldsymbol{\theta}_{(1)} \mid \mathbf{y})\, p(\boldsymbol{\theta}_{(2)} \mid \boldsymbol{\theta}_{(1)}, \mathbf{y}) \cdots p(\boldsymbol{\theta}_{(k)} \mid \boldsymbol{\theta}_{(1)}, \ldots, \boldsymbol{\theta}_{(k-1)}, \mathbf{y}), \tag{8.4.35}$$

where, parallel to (8.4.30), each factor on the right-hand side is an m-dimensional multivariate t distribution.

Comparing expression (8.4.34) with the result in (8.4.12), we see that, as was the case with a column vector of $\boldsymbol{\theta}$, the two distributions are of the same form except for the difference in the "degrees of freedom." They now differ by $(k - 1)$, simply because an additional $(k - 1)$ "input variables" are included in the model.

Marginal Distribution of a Block of Elements of $\boldsymbol{\theta}$

Finally consider now the partitions

$$\boldsymbol{\theta} = \begin{array}{c} \begin{array}{cc} m_1 & m_2 \end{array} \\ \left[\begin{array}{c|c} \boldsymbol{\theta}_{11} & \boldsymbol{\theta}_{12} \\ \hline \boldsymbol{\theta}_{21} & \boldsymbol{\theta}_{22} \end{array}\right] \end{array} \begin{array}{c} k_1 \\ k_2 \end{array}, \qquad \hat{\boldsymbol{\theta}} = \begin{array}{c} \begin{array}{cc} m_1 & m_2 \end{array} \\ \left[\begin{array}{c|c} \hat{\boldsymbol{\theta}}_{11} & \hat{\boldsymbol{\theta}}_{12} \\ \hline \hat{\boldsymbol{\theta}}_{21} & \hat{\boldsymbol{\theta}}_{22} \end{array}\right] \end{array} \begin{array}{c} k_1 \\ k_2 \end{array}. \tag{8.4.36}$$

It follows from (8.4.21) and (8.4.33) that the $k_1 \times m_1$ matrix of parameters
is distributed as

$$\boldsymbol{\theta}_{11} \sim t_{k_1 m_1}(\hat{\boldsymbol{\theta}}_{11}, \mathbf{C}_{11}, \mathbf{A}_{11}, \nu). \qquad (8.4.$$

The marginal distributions of $\boldsymbol{\theta}_{12}$, $\boldsymbol{\theta}_{21}$, $\boldsymbol{\theta}_{22}$, and indeed that of any *block*
elements of $\boldsymbol{\theta}$ can be similarly obtained, and are left to the reader.

In the above we have shown that the marginal and the conditio
distributions of certain subsets of $\boldsymbol{\theta}$ are of the matric-variate t form. This
however, not true in general. For example, one can show that neither the margi
distribution of $(\boldsymbol{\theta}_{11}, \boldsymbol{\theta}_{22})$ nor the conditional distribution of $(\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_{21})$ gi
$(\boldsymbol{\theta}_{11}, \boldsymbol{\theta}_{22})$ is a matric-variate t distribution. The problem of obtaining expl
expressions for the marginal and the conditional distributions in general is qu
complex, and certain special cases have recently been considered by Dr
and Morales (1970) and Tiao, Tan, and Chang (1970).

Means and Covariance Matrix of $\boldsymbol{\theta}$

From (8.4.28), the matrix of means of the posterior distribution of $\boldsymbol{\theta}$ is

$$E(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}} \qquad (8.4.$$

and the covariance matrix of $\boldsymbol{\theta}_i$ is

$$\mathrm{Cov}\,(\boldsymbol{\theta}_i) = \frac{a_{ii}}{\nu - 2}\,(\mathbf{X}'\mathbf{X})^{-1}, \qquad i = 1, \ldots, m. \qquad (8.4.$$

For the covariance matrix of $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$, with no loss in generality we consi
the case $i = 1$ and $j = 2$. Now

$$E(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2)' = \underset{\boldsymbol{\theta}_1}{E}\,(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1) \underset{\boldsymbol{\theta}_2|\boldsymbol{\theta}_1}{E}\,(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2)'.$$

If we set $m_1 = 2$ in (8.4.21) and perform a column decomposition of the k
matrix $\boldsymbol{\theta}_{1*} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$, it is then clear from (8.4.20) that

$$\underset{\boldsymbol{\theta}_2|\boldsymbol{\theta}_1}{E}\,(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2) = a_{11}^{-1}\,a_{12}\,(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)$$

so that, as might be expected,

$$E(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2)' = \frac{a_{12}}{\nu - 2}\,(\mathbf{X}'\mathbf{X})^{-1}. \qquad (8.4.$$

Thus,

$$E\begin{bmatrix} \boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1 \\ \boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2 \end{bmatrix} [(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)', (\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2)'] = \frac{1}{\nu - 2}\begin{bmatrix} a_{11}\,(\mathbf{X}'\mathbf{X})^{-1} & a_{12}\,(\mathbf{X}'\mathbf{X})^{-1} \\ a_{12}\,(\mathbf{X}'\mathbf{X})^{-1} & a_{22}\,(\mathbf{X}'\mathbf{X})^{-1} \end{bmatrix}$$

$$= \frac{1}{\nu - 2}\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \otimes (\mathbf{X}'\mathbf{X})^{-1} \qquad (8.4.$$

where ⊗ denotes the Kronecker product—see Appendix A8.4. In general, if we write the elements of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ as

$$\boldsymbol{\Theta}' = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_m'), \qquad \hat{\boldsymbol{\Theta}}' = (\hat{\boldsymbol{\theta}}_1', \ldots, \hat{\boldsymbol{\theta}}_m'), \tag{8.4.42a}$$

where $\boldsymbol{\Theta}$ and $\hat{\boldsymbol{\Theta}}$ are $km \times 1$ vectors, then

$$\mathrm{Cov}(\boldsymbol{\Theta}) = E(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}})(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}})' = \frac{1}{\nu - 2}\mathbf{A} \otimes (\mathbf{X}'\mathbf{X})^{-1}. \tag{8.4.42b}$$

By a similar argument, if we write

$$\boldsymbol{\Theta}_*' = (\boldsymbol{\theta}_{(1)}', \ldots, \boldsymbol{\theta}_{(k)}'), \qquad \hat{\boldsymbol{\Theta}}_* = (\hat{\boldsymbol{\theta}}_{(1)}', \ldots, \hat{\boldsymbol{\theta}}_{(k)}'), \tag{8.4.43a}$$

then

$$\mathrm{Cov}(\boldsymbol{\Theta}_*) = \frac{1}{\nu - 2}(\mathbf{X}'\mathbf{X})^{-1} \otimes \mathbf{A}. \tag{8.4.43b}$$

Linear Transformation of $\boldsymbol{\theta}$

Let $\mathbf{P}$ be a $k_1 \times k$ $(k_1 \leqslant k)$ matrix of rank k_1 and $\mathbf{Q}$ be a $m \times m_1$ $(m_1 \leqslant m)$ matrix of rank m_1. Suppose $\boldsymbol{\phi}$ is the $k_1 \times m_1$ matrix of random variables obtained from the linear transformation

$$\boldsymbol{\phi} = \mathbf{P}\,\boldsymbol{\theta}\,\mathbf{Q}. \tag{8.4.44}$$

Then $\boldsymbol{\phi}$ is distributed as $t_{k_1 m_1}[\mathbf{P}\,\hat{\boldsymbol{\theta}}\,\mathbf{Q}, \mathbf{P}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{P}', \mathbf{Q}'\mathbf{A}\mathbf{Q}, \nu]$. The proof is left as an exercise for the reader.

Asymptotic Distribution of $\boldsymbol{\theta}$

When ν tends to infinity, the distribution of $\boldsymbol{\theta}$ approaches a km dimensional multivariate Normal distribution,

$$\lim_{\nu\to\infty} p(\boldsymbol{\theta} \mid \mathbf{y}) = (\sqrt{2\pi})^{-mk}|\hat{\boldsymbol{\Sigma}}^{-1}|^{k/2}\,|\mathbf{X}'\mathbf{X}|^{m/2}$$
$$\times \exp\left[-\tfrac{1}{2}\,\mathrm{tr}\,\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\,\mathbf{X}'\mathbf{X}\,(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right], \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{8.4.45}$$

where

$$\hat{\boldsymbol{\Sigma}} = \nu^{-1}\mathbf{A},$$

and we shall say that, asymptotically, $\boldsymbol{\theta} \sim N_{mk}[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}} \otimes (\mathbf{X}'\mathbf{X})^{-1}]$.

To see this, in (8.4.15) let

$$\mathbf{Q} = \nu\,\mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$
$$= \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Then, we may write

$$|\mathbf{I}_m + \nu^{-1}\mathbf{Q}| = \prod_{i=1}^{m}(1 + \nu^{-1}\lambda_i) \tag{8.4.46}$$

where $(\lambda_1, \ldots, \lambda_m)$ are the latent roots of $\mathbf{Q}$. Thus, as $\nu \to \infty$

$$\lim_{\nu \to \infty} |\mathbf{I}_m + \nu^{-1}\mathbf{Q}|^{-\frac{1}{2}(\nu+k+m-1)} = \exp\left(-\frac{1}{2}\sum_{i=1}^{m} \lambda_i\right)$$
$$= \exp(-\tfrac{1}{2}\operatorname{tr}\mathbf{Q}).$$

Since

$$\operatorname{tr}\mathbf{Q} = (\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}})' \hat{\boldsymbol{\Sigma}}^{-1} \otimes (\mathbf{X}'\mathbf{X})(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}) \qquad (8.4.47)$$

where $\boldsymbol{\Theta}$ and $\hat{\boldsymbol{\Theta}}$ are the $km \times 1$ vectors defined in (8.4.42a), and noting that

$$|\hat{\boldsymbol{\Sigma}}^{-1} \otimes (\mathbf{X}'\mathbf{X})| = |\hat{\boldsymbol{\Sigma}}^{-1}|^{k/2} |\mathbf{X}'\mathbf{X}|^{m/2},$$

the desired result follows at once.

It follows that

$$E(\boldsymbol{\Theta}) = \hat{\boldsymbol{\Theta}}, \qquad \operatorname{Cov}(\boldsymbol{\Theta}) = \hat{\boldsymbol{\Sigma}} \otimes (\mathbf{X}'\mathbf{X})^{-1} \qquad (8.4.48a)$$

or, alternatively,

$$E(\boldsymbol{\Theta}_*) = \hat{\boldsymbol{\Theta}}_*, \qquad \operatorname{Cov}(\boldsymbol{\Theta}_*) = (\mathbf{X}'\mathbf{X})^{-1} \otimes \hat{\boldsymbol{\Sigma}} \qquad (8.4.48b)$$

where $(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}})$ and $(\boldsymbol{\Theta}_*, \hat{\boldsymbol{\Theta}}_*)$ are defined in (8.4.42a) and (8.4.43a), respectively

8.4.4 H.P.D. Regions of $\boldsymbol{\theta}$

Expressions (8.4.28) and (8.4.34) allow us to make inferences about a specific column or row of $\boldsymbol{\theta}$. Using properties of the multivariate t distribution, H.P.D regions of the elements of a row or a column can be easily determined.

We now discuss a procedure for the complete set of parameters $\boldsymbol{\theta}$, which makes it possible to decide whether a general point $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is or is not included in an H.P.D. region of approximate content $(1 - \alpha)$.

It is seen in expression (8.4.15) that the posterior distribution of $\boldsymbol{\theta}$ is a monotonic increasing function of the quantity $U(\boldsymbol{\theta})$, where

$$U(\boldsymbol{\theta}) = \frac{|\mathbf{A}|}{|\mathbf{A} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|} = |\mathbf{I}_m \mathbf{A} +^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-1} \qquad (8.4.49)$$

Consequently, the parameter point $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ lies inside the $(1 - \alpha)$ H.P.D. region if and only if

$$\Pr\{U(\boldsymbol{\theta}) > U(\boldsymbol{\theta}_0) \mid \mathbf{y}\} \leqslant (1 - \alpha), \qquad (8.4.50)$$

8.4.5 Distribution of $U(\boldsymbol{\theta})$

To obtain the posterior distribution of $U(\boldsymbol{\theta})$ so that (8.4.50) can be calculated we first derive the moments of $U(\boldsymbol{\theta})$. Applying the integral identity (8.4.13) th

hth moment of $U(\boldsymbol{\theta})$ is found to be

$$E[U(\boldsymbol{\theta})^h \mid \mathbf{y}] = \frac{c(m, k, \nu + 2h)}{c(m, k, \nu)}$$

$$= \prod_{s=1}^{m} \frac{\Gamma[\frac{1}{2}(\nu - 1 + s) + h]\,\Gamma[\frac{1}{2}(\nu - 1 + k + s)]}{\Gamma[\frac{1}{2}(\nu - 1 + s)]\,\Gamma[\frac{1}{2}(\nu - 1 + k + s) + h]} \qquad (8.4.51)$$

From (8.4.46) and (8.4.49) it follows that $U = U(\boldsymbol{\theta})$ is a random variable defined on the interval $(0, 1)$ so that the distribution of U is uniquely determined by its moments. Further, expression (8.4.51) shows that distribution of U is a function of (m, k, ν). Adopting the symbol $U_{(m,k,\nu)}$ to mean a random variable whose probability distribution is that implied by the moments in (8.4.51), we have the following general result.

Theorem 8.4.1 Let $\hat{\boldsymbol{\theta}}$ be a $k \times m$ matrix of constants, $\mathbf{X}'\mathbf{X}$ and $\mathbf{A}$ be, respectively, a $k \times k$ and a $m \times m$ positive definite symmetric matrix of constants, and $\nu > 0$. If the $k \times m$ matrix of random variables

$$\boldsymbol{\theta} \sim t_{km}[\hat{\boldsymbol{\theta}}, (\mathbf{X}'\mathbf{X})^{-1}, \mathbf{A}, \nu],$$

then

$$U(\boldsymbol{\theta}) \sim U_{(m,k,\nu)}$$

where

$$U(\boldsymbol{\theta}) = |\mathbf{I}_m + \mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\,\mathbf{X}'\mathbf{X}\,(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-1}.$$

As noted by Geisser (1965a), expression (8.4.51) correspond exactly to that for the hth sampling moment of $U(\boldsymbol{\theta}_0)$ in the sampling theory framework when $\boldsymbol{\theta}$ are regarded as fixed and $\mathbf{y}$ random variables. Thus, the Bayesian probability $\Pr\{U(\boldsymbol{\theta}) > U(\boldsymbol{\theta}_0) \mid \mathbf{y}\}$ is numerically equivalent to the significance level associated with the null hypothesis $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

Some Distributional Properties of $U_{(m,k,\nu)}$.

Following the development, for example, in Anderson (1958), we now discuss some properties of the distribution of $U_{(m,k,\nu)}$. It will be noted that the notation $U_{(m,k,\nu)}$ here is slightly different from the one used in Anderson. Specifically, in his notation, ν is replaced by $\nu + m - 1$.

a) Since from (8.4.18) $c(k, m, \nu) = c(k, m, \nu)$, the hth moment in (8.4.51) can be alternatively expressed as

$$E[U(\boldsymbol{\theta})^h \mid \mathbf{y}] = \prod_{t=1}^{k} \frac{\Gamma[\frac{1}{2}(\nu - 1 + t) + h]\,\Gamma[\frac{1}{2}(\nu - 1 + m + t)]}{\Gamma[\frac{1}{2}(\nu - 1 + t)]\,\Gamma[\frac{1}{2}(\nu - 1 + m + t) + h]}. \qquad (8.4.52)$$

By comparing (8.4.51) with (8.4.52), we see that the roles played by m and k can be interchanged. That is, the distribution of $U_{(m,k,\nu)}$ is identical to that of $U_{(k,m,\nu)}$. In other words, the distribution of $U = U(\boldsymbol{\theta})$ arising from a multivariate model with m output variables and k regression coefficients for each output is identical to that from a multivariate model with k output variables and m regression coefficients for each output. With no loss in generality, we shall proceed with the m-output model, i.e., the $U_{(m,k,\nu)}$ distribution.

b) Now (8.4.51) can be written

$$E(U^h \mid \mathbf{y}) = \prod_{s=1}^{m} B\left(\frac{\nu - 1 + s}{2} + h, \frac{k}{2}\right) \Big/ B\left(\frac{\nu - 1 + s}{2}, \frac{k}{2}\right), \tag{8.4.53}$$

where $B(p,q)$ is the complete beta function. The right-hand side is the hth moment of the product of m independent variables $x_1, \ldots, x_m$ having beta distributions with parameters

$$\left(\frac{\nu - 1 + s}{2}, \frac{k}{2}\right), \quad s = 1, \ldots, m.$$

It follows that U is distributed as the product $x_1 \ldots x_m$.

c) Suppose m is even. Then we can write (8.4.53) as

$$E(U^h \mid \mathbf{y}) = \prod_{t=1}^{m/2} \frac{B\left[\dfrac{\nu + 2(t-1)}{2} + h, \dfrac{k}{2}\right] B\left(\dfrac{\nu - 1 + 2t}{2} + h, \dfrac{k}{2}\right)}{B\left[\dfrac{\nu + 2(t-1)}{2}, \dfrac{k}{2}\right] B\left(\dfrac{\nu - 1 + 2t}{2}, \dfrac{k}{2}\right)}. \tag{8.4.54}$$

Using the duplication formula

$$\Gamma(p + \tfrac{1}{2})\Gamma(p) = \frac{\sqrt{\pi}\,\Gamma(2p)}{2^{2p-1}} \tag{8.4.55}$$

so that

$$B(p + \tfrac{1}{2}, q)\, B(p,q) = 2^{2q}\, B(2p, 2q)\, B(q,q), \tag{8.4.56}$$

we obtain

$$\begin{aligned} E(U^h \mid \mathbf{y}) &= \prod_{t=1}^{m/2} \frac{B[\nu + 2(t - 1 + h), k]}{B[\nu + 2(t-1), k]} \\ &= E(z_1^2 \cdots z_{m/2}^2)^h, \end{aligned} \tag{8.4.57}$$

where $z_1, \ldots, z_{m/2}$ are $m/2$ independent random variables having beta distributions with parameters $[\nu + 2(t-1), k]$, $t = 1, \ldots, m/2$, respectively. Thus, U is distributed as the product $z_1^2 \cdots z_{m/2}^2$.

The Case $m = 1$.

When $m = 1$ so that $\nu = n - k$, it follows from (8.4.53) that, U has the beta distribution with parameters $[(n-k)/2, k/2]$ so that the quantity $(n-k)\,(1-U)/(kU)$ is distributed as F with $(k, n-k)$ degrees of freedom. This result is of course to be expected, since for $m = 1$, we have $|\mathbf{A}| = (n-k)s^2$, where

$$s^2 = (n-k)^{-1}(\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}),$$

so that

$$\left(\frac{1-U}{U}\right)\left(\frac{n-k}{k}\right) = \frac{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})}{ks^2} \tag{8.4.58}$$

which, from (2.7.21), has the F distribution with $(k, n-k)$ degrees of freedom.

The Case $m = 2$.

When $m = 2, v = n - k - 1$ and from (8.4.57) $U^{1/2}$ is distributed as a beta variable with parameters $(n-k-1, k)$. Thus, the quantity

$$\left(\frac{1-U^{1/2}}{U^{1/2}}\right)\left(\frac{n-k-1}{k}\right) = (|\mathbf{I}_2 + \mathbf{A}^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})|^{1/2}) - 1)\left(\frac{n-k-1}{k}\right) \tag{8.4.59}$$

has the F distribution with $[2k, 2(n-k-1)]$ degrees of freedom.

8.4.6 An Approximation to the Distribution of U for General m

For $m \geqslant 3$, the exact distribution of U is complicated, see e.g. Schatzoff (1966) and Pillai and Gupta (1969). We now give an approximation method following Bartlett (1938) and Box (1949). In expression (8.4.51), we make the substitutions

$$M = -\phi v \log U, \qquad t = -h/(\phi v),$$
$$x = \tfrac{1}{2}v, \qquad a_s = \tfrac{1}{2}(s+k-1) \quad \text{and} \quad b_s = \tfrac{1}{2}(s-1), \tag{8.4.60}$$

where ϕ is some arbitrary positive number, so that

$$\begin{aligned} E(U^h \mid \mathbf{y}) &= E(e^{tM} \mid \mathbf{y}) \\ &= \prod_{s=1}^{m} \frac{\Gamma(x+a_s)}{\Gamma(x+b_s)} \frac{\Gamma[\phi x(1-2t) + x(1-\phi) + b_s]}{\Gamma[\phi x(1-2t) + x(1-\phi) + a_s]}. \end{aligned} \tag{8.4.61}$$

In terms of the random variable M, (8.4.61) is then its moment-generating function. Taking logarithms and employing Stirling's series (see Appendix A2.2), we obtain the cumulant generating function of M as

$$\kappa_M(t) = -\frac{mk}{2}\log(1-2t) - \sum_{r=1}^{\infty} \omega_r[(1-2t)^{-r} - 1], \tag{8.4.62}$$

where

$$\omega_r = \frac{(-1)^r}{r(r+1)(\phi x)^r} \sum_{s=1}^{m} \{B_{r+1}[x(1-\phi)+b_s] - B_{r+1}[x(1-\phi)+a_s]\}$$

and $B_r(z)$ is the Bernoulli polynomial of degree r and order one. Th asymptotic expansion in (8.4.62) is valid provided ϕ is so chosen that $x(1-$ is bounded. In this case, ω_r is of order $O[(\phi x)^{-r}]$ in magnitude.

The series in (8.4.62) is of the same type as the one obtains in (2.12.1 for the comparison of the spread of k Normal populations. In particular, t distribution of $M = -\phi\nu \log U$ can be expressed as a weighted series of χ^2 densitie the leading term having mk degrees of freedom.

The Choice of ϕ

It follows that, if we take the leading term alone, then to order $O[(\phi x)^{-1}]$ $O[(\phi\frac{1}{2}\nu)^{-1}]$ the quantity

$$M = -\phi\nu \log U \tag{8.4.6}$$

is distributed as χ^2_{mk}, provided $\frac{1}{2}\nu(1-\phi)$ is bounded. In particular, if w take $\phi = 1$, we then have that $M = -\nu \log U$ is distributed approximate as χ^2_{mk}.

For moderate values of ν, the accuracy of the χ^2 approximation can b improved by suitably choosing ϕ so that $\omega_1 = 0$. This is because when $\omega_1 =$ the quantity M will be distributed as χ^2_{mk} to order $O[(\phi\frac{1}{2}\nu)^{-2}]$. Using the fa that

$$B_2(z) = z^2 - z + \tfrac{1}{6} \tag{8.4.6}$$

it is straight forward to verify that for $\omega_1 = 0$, we require

$$\phi = 1 + \frac{1}{2\nu}(m + k - 3). \tag{8.4.6}$$

This choice of ϕ gives very close approximations in practice. An examp with $\nu = 9$, $m = k = 2$ will be given later in Section 8.4.8 to compare th approximation with the exact distribution.

It follows from the above discussion that to order $O[(\phi\frac{1}{2}\nu)^{-2}]$,

$$\Pr\{U(\boldsymbol{\theta}) > U(\boldsymbol{\theta}_0) \mid \mathbf{y}\} \doteq \Pr\{\chi^2_{mk} < -\phi\nu \log U(\boldsymbol{\theta}_0)\} \tag{8.4.6}$$

with

$$\phi = 1 + \frac{1}{2\nu}(m + k - 3),$$

$$\log U(\boldsymbol{\theta}_0) = -\log|\mathbf{I}_m + \mathbf{A}^{-1}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})|$$

which can be employed to decide whether the parameter point $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ lie approximately inside or outside the $(1-\alpha)$ H.P.D. region.

8.4.7 Inferences about a General Parameter Point of a Block Submatrix of $\boldsymbol{\theta}$

In the above, we have discussed inference procedures for (a) a specific colum of $\boldsymbol{\theta}$, (b) a specific row of $\boldsymbol{\theta}$, and (c) a parameter point $\boldsymbol{\theta}_0$ for the complete set of n some problems, we may be interested in making inferences about the arameters belonging to a certain block submatrix of $\boldsymbol{\theta}$. Without loss of enerality we consider only the problem for the $k_1 \times m_1$ matrix $\boldsymbol{\theta}_{11}$ defined n (8.4.36). From (8.4.37) and Theorem 8.4.1 (on page 449), it follows that the quantity

$$U(\boldsymbol{\theta}_{11}) = |\mathbf{I}_{m_1} + \mathbf{A}_{11}^{-1}(\boldsymbol{\theta}_{11} - \hat{\boldsymbol{\theta}}_{11})' \mathbf{C}_{11}^{-1}(\boldsymbol{\theta}_{11} - \hat{\boldsymbol{\theta}}_{11})|^{-1} \tag{8.4.67}$$

is distributed as $U_{(m_1, k_1, \nu)}$. This distribution would then allow us to decide whether a particular value of $\boldsymbol{\theta}_{11}$ lay inside or outside a desired H.P.D. region. In particular, for $m_1 > 2$ and $k_1 > 2$, we may then make use of the approximation

$$-\phi_1 \nu \log U(\boldsymbol{\theta}_{11}) \sim \chi^2_{m_1 k_1}, \tag{8.4.68}$$

where

$$\phi_1 = 1 + \frac{1}{2\nu}(m_1 + k_1 - 3),$$

so that the parameter point $\boldsymbol{\theta}_{11,0}$ lies inside the $(1-\alpha)$ H.P.D. region if and only if

$$\Pr\{U(\boldsymbol{\theta}_{11}) > U(\boldsymbol{\theta}_{11,0}) \mid \mathbf{y}\} \doteq \Pr\{\chi^2_{m_1 k_1} < -\phi_1 \nu \log U(\boldsymbol{\theta}_{11,0})\} < (1-\alpha). \tag{8.4.69}$$

8.4.8 An Illustrative Example

An experiment was conducted to study the effect of temperature on the yield of the product y_1 and the by-product y_2 of a chemical process. Twelve runs were made at different temperature settings ranging from 161.3°F to 195.7°F. The data are given in Table 8.4.1.

The average temperature employed is $\bar{T} = 177.86$. We suppose a model to be entertained whereby, over the range of temperature explored, the relationships between product yield and temperature and by-product yield and temperature were nearly linear so that to an adequate approximation

$$E(y_{u1}) = \theta_{11} + \theta_{21} x_u,$$
$$E(y_{u2}) = \theta_{12} + \theta_{22} x_u, \qquad u = 1, \ldots, 12, \tag{8.4.70}$$

where $x_u = (T_u - \bar{T})/100$, the divisor 100 being introduced for convenience in calculation. The parameters θ_{11} and θ_{12} will thus determine the locations of the yield-temperature lines at the average temperature $\bar{T}$ while θ_{21} and θ_{22} will represent the slopes of these lines. The experimental runs were set up independently and we should therefore expect experimental errors to be independent from *run to run*. However, in any particular run, we should expect the error in y_1 to be correlated with that in y_2 since slight aberations in reaction

Table 8.4.1
Yield of product and by-product of a chemical process

Temp. °F	Product y_1	By-product y_2
161.3	63.7	20.3
164.0	59.5	24.2
165.7	67.9	18.0
170.1	68.8	20.5
173.9	66.1	20.1
176.2	70.4	17.5
177.6	70.0	18.2
181.7	73.7	15.4
185.6	74.1	17.8
189.0	79.6	13.3
193.5	77.1	16.7
195.7	82.8	14.8

conditions or in analytical procedures could simultaneously affect observat of both product and by-product yields. Finally, then, the tentative model wa

$$y_{u1} = \theta_{11}x_{u1} + \theta_{21}x_{u2} + \varepsilon_{u1}$$

$$y_{u2} = \theta_{12}x_{u1} + \theta_{22}x_{u2} + \varepsilon_{u2},$$ (8.4

where $x_{u2} = x_u$, and $x_{u1} = 1$ is a dummy variable introduced to "carry" parameters θ_{11} and θ_{12}. It was supposed that $(\varepsilon_{u1}, \varepsilon_{u2})$ followed the bivar Normal distribution $N_2(\mathbf{0}, \boldsymbol{\Sigma})$.

Given this setup, we apply the results arrived at earlier in this section to m inferences about the regression coefficients

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} = [\boldsymbol{\theta}_1\ ,\ \boldsymbol{\theta}_2] = \begin{bmatrix} \boldsymbol{\theta}'_{(1)} \\ \boldsymbol{\theta}'_{(2)} \end{bmatrix}$$ (8.4

against the background of a noninformative reference prior distribution $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$.

The relevant sample quantities are summarized below:

$$n = 12 \qquad m = k = 2$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 12 & 0 \\ 0 & 0.14546 \end{bmatrix} \qquad \mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.0833 & 0 \\ 0 & 6.8750 \end{bmatrix}$$ (8.4

$$\mathbf{A} = \begin{bmatrix} 61.4084 & -38.4823 \\ -38.4823 & 36.7369 \end{bmatrix} \qquad \mathbf{A}^{-1} = \begin{bmatrix} 0.0474 & 0.0496 \\ 0.0496 & 0.0792 \end{bmatrix}$$

and

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} 71.1417 & 18.0666 \\ 54.4355 & -20.0933 \end{bmatrix} = [\hat{\boldsymbol{\theta}}_1 \;,\; \hat{\boldsymbol{\theta}}_2] = \begin{bmatrix} \hat{\boldsymbol{\theta}}'_{(1)} \\ \hat{\boldsymbol{\theta}}'_{(2)} \end{bmatrix}.$$

The fitted lines

$$\hat{y}_1 = \hat{\theta}_{11} + \hat{\theta}_{21}(T - \bar{T}) \times 10^{-2}$$
$$\hat{y}_2 = \hat{\theta}_{12} + \hat{\theta}_{22}(T - \bar{T}) \times 10^{-2}$$

together with the data are shown in Fig. 8.4.1. As explained earlier, in a real data analysis, we should pause at this point to critically examine the conditional inference by study of residuals. We shall here proceed with further analysis supposing that such checks have proved satisfactory.

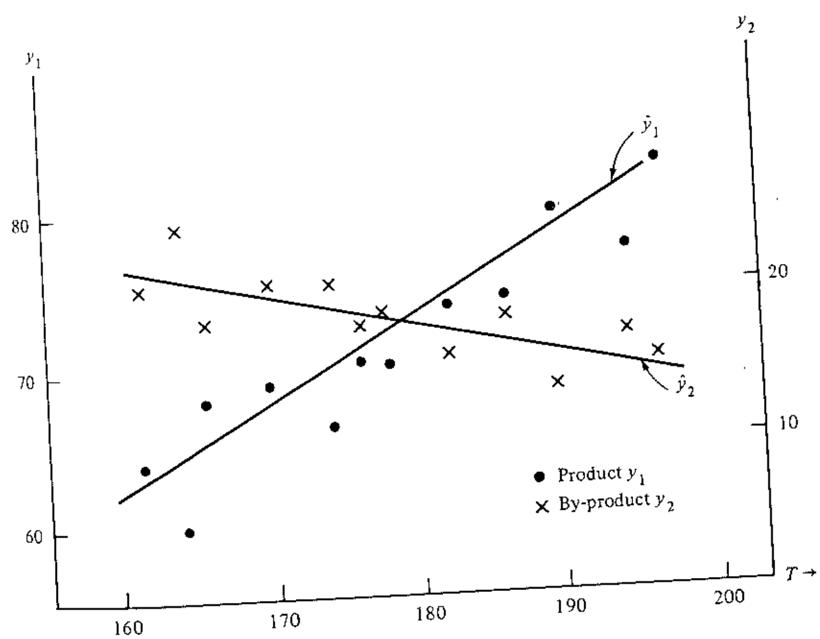

Fig. 8.4.1 Scatter diagram of the product y_1 and the by-product y_2 together with the best fitting lines.

Inferences about $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{21})'$

When interest centers primarily on the parameters $\boldsymbol{\theta}_1$ for the product y_1, we have from (8.4.28) that

$$p(\boldsymbol{\theta}_1 \mid \mathbf{y}) \propto [61.4084 + (\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)' \mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1)]^{-11/2} \qquad (8.4.74)$$

that is, a bivariate $t_2[\hat{\boldsymbol{\theta}}_1, 6.825(\mathbf{X}'\mathbf{X})^{-1}, 9]$ distribution. Since the matrix $\mathbf{X}'\mathbf{X}$ is diagonal, θ_{11} and θ_{21} are uncorrelated (but of course not independent).

Figure 8.4.2a shows contours of the 50, 75 and 95 per cent H.P.D. regio together with the mode $\hat{\boldsymbol{\theta}}_1$, from which overall conclusions about $\boldsymbol{\theta}_1$ may be draw

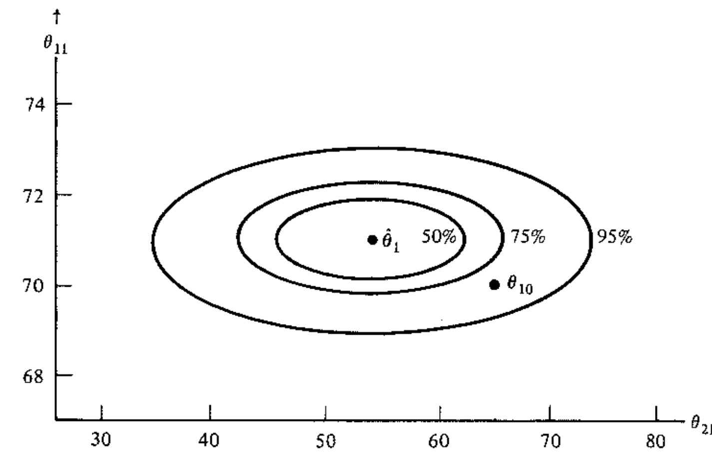

Fig. 8.4.2a Contours of the posterior distribution of $\boldsymbol{\theta}_1$, the parameters of the produ straight line.

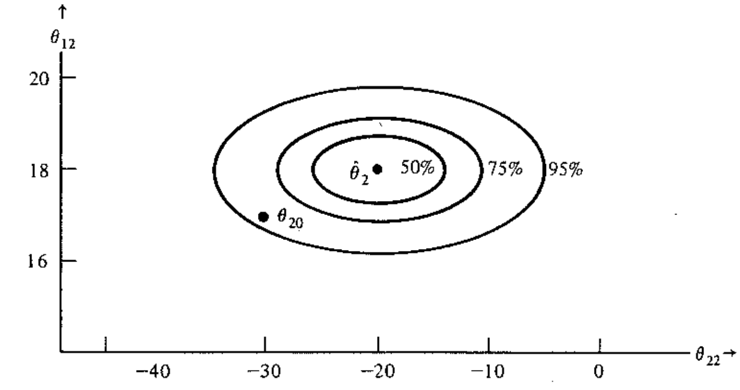

Fig. 8.4.2b Contours of the posterior distribution of $\boldsymbol{\theta}_2$, the parameters of the by-produc straight line.

Inferences about $\boldsymbol{\theta}_2 = (\theta_{12}, \theta_{22})'$

Similarly, from (8.4.28), the posterior distribution of $\boldsymbol{\theta}_2$ for the by-product y_2 is

$$p(\boldsymbol{\theta}_2 \mid \mathbf{y}) \propto [36.7369 + (\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2)' \mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2)]^{-11/2}, \qquad (8.4.75)$$

which is a $t_2[\hat{\boldsymbol{\theta}}_2, 4.08(\mathbf{X}'\mathbf{X})^{-1}, 9]$ distribution. Again, the parameters θ_{12} and θ_{22} are uncorrelated. The 50, 75 and 95 per cent H.P.D. contours together with the mode $\hat{\boldsymbol{\theta}}_2$ for this distribution are shown in Fig. 8.4.2b. The contours have exactly the same shape and orientation as those in Fig. 8.4.2a because the same $\mathbf{X}'\mathbf{X}$ matrix is employed; the spread for $\boldsymbol{\theta}_2$ is however smaller than that for $\boldsymbol{\theta}_1$ since the sample variance from y_2 is less than that from y_1.

Inferences about $\boldsymbol{\theta}_{(2)} = (\theta_{21}, \theta_{22})'$

In problems of the kind considered, interest often centers on $\boldsymbol{\theta}'_{(2)} = (\theta_{21}, \theta_{22})$ which measure respectively the slopes of the yield/temperature lines for the product and by-product. From (8.4.34), the posterior distribution of $\boldsymbol{\theta}_{(2)}$ is

$$p(\boldsymbol{\theta}_{(2)} \mid \mathbf{y}) \propto [6.875 + (\boldsymbol{\theta}_{(2)} - \hat{\boldsymbol{\theta}}_{(2)})'\mathbf{A}^{-1}(\boldsymbol{\theta}_{(2)} - \hat{\boldsymbol{\theta}}_{(2)})]^{-11/2}, \qquad (8.4.76)$$

that is, a $t_2[\hat{\boldsymbol{\theta}}_{(2)}, 0.764\mathbf{A}, 9]$ distribution. Figure 8.4.3 shows the 50, 75 and 95 per cent contours together with the mode $\hat{\boldsymbol{\theta}}_{(2)}$. Also shown in the same figure are the marginal distributions $t(\hat{\theta}_{21}, 46.90, 9)$ and $t(\hat{\theta}_{22}, 28.05, 9)$ and the

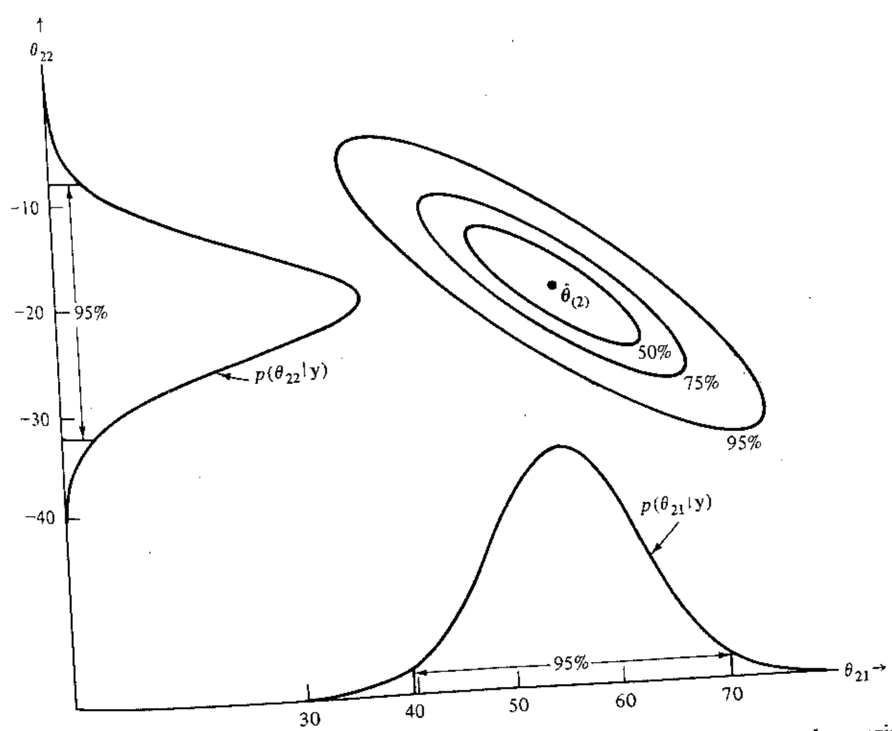

Fig. 8.4.3 Contours of the posterior distribution of $\boldsymbol{\theta}_{(2)}$ and the associated marginal distributions for the product and by-product data.

corresponding 95 per cent H.P.D. intervals for θ_{21} and θ_{22}. Figure 8.4.3 summarizes, then, the information about the slopes $(\theta_{21}, \theta_{22})$ coming from the data on the basis of a noninformative reference prior. The parameters are negatively correlated; the correlation between $(\theta_{21}, \theta_{22})$ is in fact the sample correlation between the errors $(\varepsilon_{u1}, \varepsilon_{u2})$

$$r_{12} = \frac{a_{12}}{(a_{11}\, a_{22})^{1/2}} = -\,0.81. \tag{8.4.77}$$

It is clear from the figure that care must be taken to distinguish between individual and joint inferences about $(\theta_{21}, \theta_{22})$. It could be exceedingly misleading to make inferences from the individual H.P.D. intervals about the parameters jointly (see the discussion in Section 2.7).

Joint Inferences about $\boldsymbol{\theta}$

To make overall inferences about the parameters $[\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$, we need to calculate the distribution of the quantity $U(\boldsymbol{\theta})$ defined in (8.4.49). For instance, suppose we wish to decide whether or not the parameter point

$$\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{10}\ ,\ \boldsymbol{\theta}_{20}] = \begin{bmatrix} 70 & 17 \\ 65 & -30 \end{bmatrix} \tag{8.4.78}$$

lies inside the 95 per cent H.P.D. region for $\boldsymbol{\theta}$. We have

$$(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) = \begin{bmatrix} 31.8761 & -0.6108 \\ -0.6108 & 27.9272 \end{bmatrix} \tag{8.4.79}$$

so that

$$U(\boldsymbol{\theta}_0) = \frac{|\mathbf{A}|}{|\mathbf{A} + (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})|} = \frac{775.0668}{4503.8878} = 0.1720, \tag{8.4.80}$$

Since $m = 2$, we may use the result in (8.4.59) to calculate the exact probability that $U(\boldsymbol{\theta})$ exceeds $U(\boldsymbol{\theta}_0)$. We obtain

$$\Pr\{U(\boldsymbol{\theta}) > U(\boldsymbol{\theta}_0) \mid \mathbf{y}\} = 1 - I_{\sqrt{0.1720}}(9, 2) = 1 - 0.0022 = 0.9978. \tag{8.4.81}$$

From (8.4.50), we conclude that the point $\boldsymbol{\theta}_0$ lies outside the 95 per cent H.P.D region of $\boldsymbol{\theta}$. Note that while the point $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})$ is excluded from the 95 per cent region, Figs. 8.4.2a,b show that both the points $\boldsymbol{\theta}_{10}$ and $\boldsymbol{\theta}_{20}$ are included in the corresponding marginal 95 per cent H.P.D. regions. This serves to illustrate once more the distinction between joint inferences and marginal inferences.

Approximating Distribution of $U = U(\boldsymbol{\theta})$

It is informative to compare the exact distribution of $U(\boldsymbol{\theta})$ with the approximation in (8.4.63) using the present example ($\nu = 9$, $m = k = 2$). From (8.4.59), the exact distribution of U is found to be

$$p(U) = \frac{1}{2B(2, 9)} U^{3.5} (1 - U^{1/2}), \qquad 0 < U < 1. \tag{8.4.82}$$

Using the approximation given in (8.4.63) to (8.4.65), we find

$$\phi = \tfrac{19}{18}, \qquad \phi\nu = 9.5, \qquad M = -9.5 \log U \tag{8.4.83}$$

and

$$p(M) \doteq \tfrac{1}{4}M \exp\left(-\tfrac{1}{2}M\right), \qquad 0 < M < \infty.$$

This implies that the distribution of U is approximately

$$p(U) \doteq (22.5625)(-\log U)\, U^{3.75}, \qquad 0 < U < 1. \tag{8.4.84}$$

Table 8.4.2 gives a specimen of the exact and the approximate densities of U calculated from (8.4.82) and (8.4.84). Although the sample size is only 10, the agreement is very close.

Table 8.4.2

Comparison of the exact and the approximate distributions of U for $n = 12$ and $m = k = 2$

	$p(U)$	
U	Exact	Approximate
0.05	0.00098	0.00089
0.10	0.00973	0.00924
0.20	0.08900	0.08688
0.30	0.30098	0.29731
0.40	0.66947	0.66548
0.50	1.16498	1.16239
0.60	1.69708	1.69718
0.70	2.10935	2.11244
0.80	2.17561	2.18045
0.90	1.59706	1.60130
0.95	0.95220	0.95478

8.5 SOME ASPECTS OF THE DISTRIBUTION OF Σ FOR THE CASE OF A COMMON DERIVATIVE MATRIX X

We discuss in this section certain results pertaining to the posterior distribution of the elements of the covariance matrix $\Sigma = \{\sigma_{ij}\}$.†

† For the important problem of making inferences about the latent roots and vectors of Σ which is not discussed in this book, see Geisser (1965a) and Tiao and Fienberg (1969).

8.5.1 Joint Distribution of $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$

When the $\mathbf{X}_i$'s are common, the joint posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ in (8.2.1 can be written

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(v+k+2m)} \exp\{-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} [\mathbf{A} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})]\},$$
$$-\infty < \boldsymbol{\theta} < \infty, \quad \boldsymbol{\Sigma} > \mathbf{0}, \qquad (8.5.$$

where $v = n - (k + m) + 1$ and use is made of (8.4.4) and (8.4.5). T individual and joint inferences about $(\theta_{21}, \theta_{22})$. It could be exceeding distribution can be written as the product $p(\boldsymbol{\theta}, \boldsymbol{\Sigma} \mid \mathbf{y}) = p(\boldsymbol{\theta} \mid \boldsymbol{\Sigma}, \mathbf{y})\, p(\boldsymbol{\Sigma} \mid \mathbf{y})$.

Conditional Distribution of $\boldsymbol{\theta}$ *given* $\boldsymbol{\Sigma}$

Given $\boldsymbol{\Sigma}$, we have that

$$p(\boldsymbol{\theta} \mid \boldsymbol{\Sigma}, \mathbf{y}) \propto \exp[-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})], \qquad -\infty < \boldsymbol{\theta} < \infty, \qquad (8.5.$$

which by comparison with (8.4.45), is the km-dimensional Normal distributi

$$N_{mk}[\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}]. \qquad (8.5.$$

Marginal Distribution of $\boldsymbol{\Sigma}$

Thus, the marginal posterior distribution of $\boldsymbol{\Sigma}$ is

$$p(\boldsymbol{\Sigma} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-(\frac{1}{2}v+m)} \exp(-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\mathbf{A}), \qquad \boldsymbol{\Sigma} > \mathbf{0}, \qquad (8.5.$$

From the Jacobian in (8.2.13), the distribution of $\boldsymbol{\Sigma}^{-1}$ is

$$p(\boldsymbol{\Sigma}^{-1} \mid \mathbf{y}) \propto |\boldsymbol{\Sigma}^{-1}|^{\frac{1}{2}v-1} \exp(-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\mathbf{A}), \qquad \boldsymbol{\Sigma}^{-1} > \mathbf{0}, \qquad (8.5.$$

which, by comparing with (8.2.20), is recognized as the Wishart distributi $W_m(\mathbf{A}^{-1}, v)$ provided $v > 0$. The distribution of $\boldsymbol{\Sigma}$ in (8.5.4) may thus be call an m-dimensional "inverted" Wishart distribution with v degrees of freedo and be denoted by $W_m^{-1}(\mathbf{A}, v)$. From (8.2.21) the normalizing constant for t distributions of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ is

$$|\mathbf{A}|^{\frac{1}{2}(v+m-1)}\, 2^{-\frac{1}{2}m(v+m-1)} \left[\Gamma_m\!\left(\frac{v+m-1}{2}\right)\right]^{-1}. \qquad (8.5.$$

Note that when $m = 1$ and $v = n - k$ the distribution in (8.5.4) reduces an inverted χ^2 distribution,

$$p(\sigma_{11} \mid \mathbf{y}) \propto \sigma_{11}^{-\frac{1}{2}(n-k+2)} \exp\left(-\frac{a_{11}}{2\sigma_{11}}\right), \qquad \sigma_{11} > 0, \qquad (8.5.$$

which is the posterior distribution of $\sigma^2 = \sigma_{11}$ with data from a univaria multiple regression problem (see Section 2.7).

From the results in (8.5.1), (8.5.3), (8.5.4) and (8.4.15), we have t following useful theorem.

5

heorem 8.5.1 Let $\hat{\theta}$ be a $k \times m$ matrix of constants, $X'X$ and A be, respectively, $k \times k$ and a $m \times m$ positive definite symmetric matrix of constants, and $\nu > 0$. f the joint distribution of the elements of the $k \times m$ matrix θ and the $m \times m$ ositive definite symmetric matrix Σ is

$$p(\theta, \Sigma \mid y) \propto |\Sigma|^{-\frac{1}{2}(\nu+k+2m)} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}[A + (\theta - \hat{\theta})' X'X(\theta - \hat{\theta})]\},$$

hen, (a) given Σ, $\theta \sim N_{mk}[\hat{\theta}, \Sigma \otimes (X'X)^{-1}]$, (b) $\Sigma \sim W_m^{-1}(A, \nu)$ and (c) $\theta \sim t_{mk}[\hat{\theta}, (X'X)^{-1}, A, \nu]$.

8.5.2 Some Properties of the Distribution of Σ

Consider the partition

$$\Sigma = \begin{array}{c} \begin{array}{cc} m_1 & m_2 \end{array} \\ \left[\begin{array}{c:c} \Sigma_{11} & \Sigma_{12} \\ \hdashline \Sigma_{21} & \Sigma_{22} \end{array}\right] \end{array} \begin{array}{l} \\ m_1 \\ m_2 \end{array}, \qquad m_1 + m_2 = m. \tag{8.5.8}$$

We now proceed to obtain the posterior distribution of (Σ_{11}, Ω, T), where

$$\Omega = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{8.5.9}$$

and

$$T = \Sigma_{11}^{-1}\Sigma_{12}.$$

It is to be remembered that Σ_{11} is the covariance matrix of the m_1 errors $(\varepsilon_{u1}, \ldots, \varepsilon_{um_1})$, Ω and T are, respectively, the covariance matrix and the $m_1 \times m_2$ matrix of "regression coefficients" for the conditional distribution of the remaining m_2 errors $(\varepsilon_{u(m_1+1)}, \ldots, \varepsilon_{um})$ given $(\varepsilon_{u1}, \ldots, \varepsilon_{um_1})$. Denoting

$$A = \begin{array}{c} \begin{array}{cc} m_1 & m_2 \end{array} \\ \left[\begin{array}{c:c} A_{11} & A_{12} \\ \hdashline A_{21} & A_{22} \end{array}\right] \end{array} \begin{array}{l} \\ m_2 \\ m_1 \end{array}, \qquad A_{22\cdot 1} = A_{22} - A_{21} A_{11}^{-1} A_{12} \tag{8.5.10}$$

and

$$\hat{T} = A_{11}^{-1} A_{12},$$

it can be readily shown that

a) Σ_{11} is distributed independently of (T, Ω), (8.5.11)

b) $$\Sigma_{11} \sim W_{m_1}^{-1}(A_{11}, \nu), \tag{8.5.12}$$

c) $$\Omega \sim W_{m_2}^{-1}(A_{22\cdot 1}, \nu + m_1), \tag{8.5.13}$$

d) $$T \sim t_{m_1 m_2}(\hat{T}, A_{11}^{-1}, A_{22\cdot 1}, \nu + m_1),$$

$$U(T) = |I_{m_2} + A_{22\cdot 1}^{-1}(T - \hat{T})' A_{11}(T - \hat{T})|^{-1} \sim U_{(m_2, m_1, \nu + m_1)}. \tag{8.5.14}$$

The above results may be verified as follows. Since Σ is positive definite, w
express the determinant and the inverse of Σ as

$$|\Sigma| = |\Sigma_{11}|\,|\Omega|$$

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \mathbf{M}, \qquad (8.$$

where

$$\mathbf{M} = \begin{bmatrix} \mathbf{T}\Omega^{-1}\mathbf{T}' & -\mathbf{T}\Omega^{-1} \\ -\Omega^{-1}\mathbf{T}' & \Omega^{-1} \end{bmatrix}.$$

Expression (8.5.15) may be verified by showing $\Sigma^{-1}\Sigma = \mathbf{I}$. Thus, the distributi
(8.5.4) can be written

$$p(\Sigma \mid \mathbf{y}) \propto [|\Sigma_{11}|\,|\Omega|]^{-(\frac{1}{2}\nu+m)} \exp\left(-\tfrac{1}{2}\operatorname{tr}\Sigma_{11}^{-1}\mathbf{A}_{11} - \tfrac{1}{2}\operatorname{tr}\mathbf{MA}\right), \qquad \Sigma > 0. \qquad (8$$

For fixed Σ_{11}, it is readily seen by making use of (A8.1.1) in Appendix A8.1 th
Jacobian of the transformation from $(\Sigma_{12}, \Sigma_{22})$ to $(\mathbf{T}, \Omega)$ is

$$J = \left|\frac{\partial(\Sigma_{12}, \Sigma_{22})}{\partial(\mathbf{T}, \Omega)}\right| = |\Sigma_{11}|^{m_2}. \qquad (8$$

Noting that $\mathbf{M}$ does not depend on Σ_{11}, it follows from (8.5.16) and (8.5.17) th
is independent of $(\mathbf{T}, \Omega)$ so that

$$p(\Sigma_{11}, \mathbf{T}, \Omega \mid \mathbf{y}) = p(\mathbf{T}, \Omega \mid \mathbf{y})\, p(\Sigma_{11} \mid \mathbf{y}) \qquad (8$$

where

$$p(\mathbf{T}, \Omega \mid \mathbf{y}) \propto |\Omega|^{-(\frac{1}{2}\nu+m)} \exp\left(-\tfrac{1}{2}\operatorname{tr}\mathbf{MA}\right), \qquad \Omega > 0, \quad -\infty < \mathbf{T} < \infty \qquad (8$$

and

$$P(\Sigma_{11} \mid \mathbf{y}) \propto |\Sigma_{11}|^{-(\frac{1}{2}\nu+m_1)} \exp\left(-\tfrac{1}{2}\operatorname{tr}\Sigma_{11}^{-1}\mathbf{A}_{11}\right), \qquad \Sigma_{11} > 0. \qquad (8$$

Thus, Σ_{11} is distributed as the inverted Wishart given in (8.5.11).

From (8.5.10) and (8.5.15), we may express the exponent of the distributi
(8.5.19) as

$$\begin{aligned} \operatorname{tr}\mathbf{MA} &= \operatorname{tr}(\mathbf{T}\Omega^{-1}\mathbf{T}'\mathbf{A}_{11} - \mathbf{T}\Omega^{-1}\mathbf{A}_{21} - \Omega^{-1}\mathbf{T}'\mathbf{A}_{12} + \Omega^{-1}\mathbf{A}_{22}) \\ &= \operatorname{tr}\Omega^{-1}[\mathbf{A}_{22\cdot 1} + (\mathbf{T} - \hat{\mathbf{T}})'\mathbf{A}_{11}(\mathbf{T} - \hat{\mathbf{T}})] \qquad (8 \end{aligned}$$

In obtaining the extreme right of (8.5.21), repeated use is made of the fac
$\operatorname{tr}\mathbf{AB} = \operatorname{tr}\mathbf{BA}$.
Thus,

$$p(\mathbf{T}, \Omega \mid \mathbf{y}) \propto |\Omega|^{-\frac{1}{2}(\nu'+m_1+2m_2)} \exp\left\{-\tfrac{1}{2}\operatorname{tr}\Omega^{-1}[\mathbf{A}_{22\cdot 1} + (\mathbf{T} - \hat{\mathbf{T}})'\mathbf{A}_{11}(\mathbf{T} - \hat{\mathbf{T}})]\right\}$$
$$-\infty < \mathbf{T} < \infty, \quad \Omega$$

where $\nu' = \nu + m_1$, which is in exactly the same form as the joint distributi
(Σ, θ) in (8.5.1). The results in (8.5.12) and (8.5.13) follow by application of Th
8.5.1 (on p. 461). Further, by using Theorem 8.4.1 (on p. 449), we obtain (8.5.1

Distribution of σ_{11}

By setting $m_1 = 1$ in (8.5.11), the distribution of σ_{11} is

$$p(\sigma_{11} \mid \mathbf{y}) \propto \sigma_{11}^{-(\frac{1}{2}\nu+1)} \exp\left(-\frac{a_{11}}{2\sigma_{11}}\right), \qquad \sigma_{11} > 0, \tag{8.5.22}$$

a χ^{-2} distribution with $\nu = n - k - (m-1)$ degrees of freedom. Comparing with the univariate case in (8.5.7), we see that the two distributions are identical except, as expected, that the degrees of freedom differ by $(m-1)$. The difference is of minor importance when m is small relative to $n - k$ but can be appreciable otherwise—see the discussion about the posterior distribution of $\boldsymbol{\theta}_i$ in (8.4.28).

It is clear that the distribution of σ_{ii}, the ith diagonal element of Σ, is given by simply replacing the subscripts (1,1) in (8.5.22) by (i, i).

The Two Regression Matrices $\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ *and* $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$

The $m_1 \times m_2$ matrix of "regression coefficients" $\mathbf{T} = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ measures the dependence of the conditional expectation of the errors $(\varepsilon_{u(m_1+1)}, \ldots, \varepsilon_{um})$ on $(\varepsilon_{u1}, \ldots, \varepsilon_{um_1})$. From the posterior distribution of $\mathbf{T}$ in (8.5.13) the plausibility of different values of the measure $\mathbf{T}$ may be compared in terms of e.g. the H.P.D. regions. In deciding whether a parameter point $\mathbf{T} = \mathbf{T}_0$ lies inside or outside a desired H.P.D. region, from (8.5.14) and (8.4.66) one may then calculate

$$\Pr\{U(\mathbf{T}) > U(\mathbf{T}_0) \mid \mathbf{y}\} \doteq \Pr\{\chi^2_{m_1 m_2} < -\phi'(\nu + m_1)\log U(\mathbf{T}_0)\}, \tag{8.5.23}$$

where

$$\phi' = 1 + \frac{m_1 + m_2 - 3}{2(\nu + m_1)}.$$

In particular, if $\mathbf{T}_0 = \mathbf{0}$ which corresponds to $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, that is, $(\varepsilon_{u1}, \ldots, \varepsilon_{um_1})$ are independent of $(\varepsilon_{u(m_1+1)}, \ldots, \varepsilon_{um})$, then

$$U(\mathbf{T}_0 = \mathbf{0}) = \frac{|\mathbf{A}|}{|\mathbf{A}_{11}|\,|\mathbf{A}_{22}|}. \tag{8.5.24}$$

Consider now the $m_2 \times m_1$ matrix of "regression coefficients"

$$\mathbf{Z} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{8.5.25}$$

It is clear from the development leading to (8.5.13) and (8.5.14) that by interchanging the roles of m_1 and m_2, we have

$$\mathbf{Z} \sim t_{m_2 m_1}(\hat{\mathbf{Z}}, \mathbf{A}_{22}^{-1}, \mathbf{A}_{11\cdot 2}, \nu + m_2) \tag{8.5.26}$$

where

$$\hat{\mathbf{Z}} = \mathbf{A}_{22}^{-1}\mathbf{A}_{21}, \qquad \mathbf{A}_{11\cdot 2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21},$$

and

$$U(\mathbf{Z}) = |\mathbf{I}_{m_1} + \mathbf{A}_{11\cdot 2}^{-1}(\mathbf{Z} - \hat{\mathbf{Z}})'\mathbf{A}_{22}(\mathbf{Z} - \hat{\mathbf{Z}})|^{-1}$$
$$\sim U_{(m_1, m_2, \nu+m_2)}. \tag{8.5.27}$$

Thus, in deciding whether the parameter point $\mathbf{Z}_0$ is included in a desired H.P.D. region, we calculate

$$\Pr\{U(\mathbf{Z}) > U(\mathbf{Z}_0) \mid \mathbf{y}\} \doteq \Pr\{\chi^2_{m_2 m_1} < -\phi''(\nu + m_2)\log U(\mathbf{Z}_0)\} \qquad (8.5.28)$$

where

$$\phi'' = 1 + \frac{m_1 + m_2 - 3}{2(\nu + m_2)}.$$

In particular, if $\mathbf{Z}_0 = \mathbf{0}$ which again corresponds to $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, then

$$U(\mathbf{Z}_0 = \mathbf{0}) = \frac{|\mathbf{A}|}{|\mathbf{A}_{11}|\,|\mathbf{A}_{22}|}. \qquad (8.5.29)$$

Although $U(\mathbf{Z}_0 = \mathbf{0}) = U(\mathbf{T}_0 = \mathbf{0})$ the probabilities on the right-hand sides of (8.5.23) and (8.5.28) will be different whenever $m_1 \neq m_2$. This is not a surprising result. For, it can be seen from (8.5.18) that the distribution of $\mathbf{T}$ is in fact proportional to the conditional distribution of $\boldsymbol{\Sigma}_{12}$, given $\boldsymbol{\Sigma}_{11}$. Inferences about the parameter point $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ in terms of the probability $\Pr\{\chi^2_{m_1 m_2} < -\phi'(\nu + m_1)\log U(\mathbf{T}_0 = \mathbf{0})\}$ can thus be interpreted as made conditionally for fixed $\boldsymbol{\Sigma}_{11}$. That is to say that we are comparing the plausibility of $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ with other values of $\boldsymbol{\Sigma}_{12}$ in relation to a fixed $\boldsymbol{\Sigma}_{11}$. On the other hand, in terms of $\Pr\{\chi^2_{m_1 m_2} < -\phi''(\nu + m_2)\log U(\mathbf{Z}_0 = \mathbf{0})\}$, inferences about $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ can be regarded as conditional on fixed $\boldsymbol{\Sigma}_{22}$. Thus, one would certainly not expect that, in general, the two types of conditional inferences about $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ will be identical.

8.5.3 An Example

For illustration, consider again the product and by-product data in Table 8.4.1. The relevant sample quantities are given in (8.4.73).

When interest centers on the variance σ_{11} of the error ε_{u1} corresponding to the product y_1, we have from (8.5.22)

$$\sigma_{11} \sim 61.4084\,\chi_9^{-2} \qquad (8.5.30)$$

Using Table II (at the end of this book), limits of the 95 per cent H.P.D. interval of $\log\sigma_{11}$ in terms of σ_{11} are (3.02, 20.79).

Similarly, the posterior distribution of σ_{22} is such that

$$\sigma_{22} \sim 36.7369\,\chi_9^{-2} \qquad (8.5.31)$$

and the limits of the corresponding 95 per cent H.P.D. interval are (1.81, 12.44).

From (8.5.13), and since $m_1 = m_2 = 1$, the posterior distribution of $T = \sigma_{11}^{-1}\sigma_{12}$ is the univariate $t(\hat{T}, s_1^2, \nu_1)$ distribution where

$$\hat{T} = a_{11}^{-1}a_{12} = -0.627, \qquad \nu_1 = 10,$$

and

$$s_1^2 = \nu_1^{-1} a_{11}^{-1} a_{22\cdot 1} = \frac{a_{22} - a_{12}^2/a_{11}}{\nu_1 \times a_{11}} = 0.0206.$$

Thus,

$$\frac{T + 0.627}{0.143} \sim t_{10} \tag{8.5.32}$$

so that from Table IV at the end of this book, limits of the 95 per cent H.P.D. interval are $(-0.95, -0.31)$. In particular, the parameter point $\sigma_{11}^{-1}\sigma_{12} = 0$ (corresponds to $\sigma_{12} = 0$) is excluded from the 95 per cent interval.

Finally, from (8.5.26) $Z = \sigma_{22}^{-1}\sigma_{12}$ is distributed as $t(\hat{Z}, s_2^2, \nu_2)$ where

$$\hat{Z} = a_{22}^{-1} a_{12} = -1.05, \qquad \nu_2 = 10$$

and

$$s_2^2 = \nu_2^{-1} a_{22}^{-1} a_{11\cdot 2} = \frac{a_{11} - a_{12}^2/a_{22}}{\nu_2 \times a_{22}} = 0.0574.$$

Thus,

$$\frac{Z + 1.05}{0.240} \sim t_{10}. \tag{8.5.33}$$

Limits of the 95 per cent H.P.D. interval are $(-1.58, -0.52)$ and the point $\sigma_{22}^{-1}\sigma_{12} = 0$ is again excluded. Further, from (8.5.14), (8.5.27), (8.5.32) and (8.5.33)

$$\Pr\{U(T) > U(0) \mid \mathbf{y}\} = \Pr\{U(Z) > U(0) \mid \mathbf{y}\} = \Pr\{|t_{10}| > 4.37\} \tag{8.5.34}$$

so that inferences about $\sigma_{12} = 0$ in terms of either T or Z are identical. This is of course to be expected since, for this example, $m_1 = m_2 = 1$.

8.5.4 Distribution of the Correlation Coefficient ρ_{12}

The two regression matrices $\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ are measures of the dependence (or association) between the two set of responses $(y_{u1}, \ldots, y_{um_1})$ and $(y_{u(m_1+1)}, \ldots, y_{um})$. When interest centers at the association between two specific responses y_{ui} and y_{uj}, the most natural measure of association is the correlation coefficient $\rho_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$. Without loss of generality, we now consider how inferences may be made about ρ_{12}.

By setting $m_1 = 2$ in the distribution of $\boldsymbol{\Sigma}_{11}$ in (8.5.11), we can follow the development in Jeffreys (1961, p. 174) to obtain the posterior distribution of the correlation coefficient ρ_{12} as

$$p(\rho \mid \mathbf{y}) \propto (1 - \rho^2)^{\frac{1}{2}(\nu-2)} \int_0^\infty \omega^{-1}\left(\omega + \frac{1}{\omega} - 2\rho r\right)^{-(\nu+1)} d\omega, \qquad -1 < \rho < 1, \tag{8.5.35}$$

where $\rho = \rho_{12}$,

$$r = r_{12} = \frac{a_{12}}{(a_{11}\, a_{22})^{1/2}}$$

is the sample correlation coefficient, and the normalizing constant is

$$2(1 - r^2)^{(\nu+1)/2}\, \Gamma(\nu + 1) \Big/ \left[\pi^{1/2}\, \Gamma\!\left(\frac{\nu}{2}\right) \Gamma\!\left(\frac{\nu+1}{2}\right)\right].$$

It is noted that this distribution depends only upon the sample correlati coefficient r.

To see this, for $m_1 = 2$, the posterior distribution of the elements $(\sigma_{11}, \sigma_{22}, \sigma_{12})$ $\boldsymbol{\Sigma}_{11}$ in (8.5.11) is

$$p(\sigma_{11}, \sigma_{22}, \sigma_{12} \mid \mathbf{y}) \propto [\sigma_{11}\, \sigma_{22}\, (1 - \rho^2)]^{-(\frac{1}{2}\nu+2)}$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{a_{11}}{\sigma_{11}} + \frac{a_{22}}{\sigma_{22}} - \frac{2\rho\, a_{12}}{(\sigma_{11}\,\sigma_{22})^{1/2}}\right]\right\},$$
$$\sigma_{11} > 0, \quad \sigma_{22} > 0, \quad \sigma_{11}\sigma_{22} > \sigma_{12}^2, \qquad (8.5.36)$$

where from (8.5.6), the normalizing constant is,

$$(a_{11}a_{22} - a_{12}^2)^{(\nu+1)/2} \Big/ \left\{2^{(\nu+1)}\, \pi^{1/2} \prod_{i=1}^{2} \Gamma[\tfrac{1}{2}(\nu + 2 - i)]\right\}$$

We now make the transformation, due to Fisher (1915),

$$x = \left(\frac{\sigma_{11}\,\sigma_{22}}{a_{11}\,a_{22}}\right)^{1/2}, \qquad \omega = \left(\frac{\sigma_{11}\,a_{22}}{\sigma_{22}\,a_{11}}\right)^{1/2}, \qquad \rho = \frac{\sigma_{12}}{(\sigma_{11}\,\sigma_{22})^{1/2}}. \qquad (8.5.37)$$

The Jacobian is

$$\left|\frac{\partial(\sigma_{11}\,\sigma_{22}\,\sigma_{12})}{\partial(x, \omega, \rho)}\right| = 2a_{11}\,\sigma_{22}(\sigma_{11}\,\sigma_{22})^{1/2} \qquad (8.5.38)$$

so that the distribution of (x, ω, ρ) is

$$p(x, \omega, \rho \mid \mathbf{y}) \propto (1 - \rho^2)^{\frac{1}{2}(\nu+4)}\, \omega^{-1}\, x^{-(\nu+2)} \exp\left[-\frac{1}{2(1-\rho^2)x}\left(\omega + \frac{1}{\omega} - 2\rho r\right)\right],$$
$$\omega > 0, \quad x > 0, \quad -1 < \rho < 1. \qquad (8.5.39)$$

Upon integrating out x,

$$p(\omega, \rho \mid \mathbf{y}) \propto (1 - \rho^2)^{\frac{1}{2}(\nu-2)}\, \omega^{-1}\left(\omega + \frac{1}{\omega} - 2\,\rho r\right)^{-(\nu+1)} \qquad \omega > 0, \quad -1 < \rho < 1,$$
$$(8.5.40)$$

from which we obtain the distribution of ρ given in (8.5.35).

The Special Case When $r = 0$

When $r = 0$, the distribution in (8.5.35) reduces to

$$p(\rho \mid r = 0) \propto (1 - \rho^2)^{\frac{1}{2}(\nu-2)}, \qquad -1 < \rho < 1, \tag{8.5.41}$$

which is symmetric at $\rho = 0$, and is identical in form to the sampling distribution of r on the null hypothesis that $\rho = 0$. In this case, if we make the transformation

$$\rho = \frac{t}{(\nu + t^2)^{1/2}}, \tag{8.5.42}$$

then the distribution of t is

$$p(t) \propto (1 + t^2/\nu)^{-\frac{1}{2}(\nu+1)}, \qquad -\infty < t < \infty,$$

so that the quantity t is distributed as $t(0, 1, \nu)$.

The General Case When $r \neq 0$

In general, the density function (8.5.35) cannot be expressed in terms of simple functions of r. With the availability of a computer, it can always be evaluated by numerical integration, however. To illustrate, consider again the bivariate product and by-product data introduced in Table 8.4.1. Figure 8.5.1 shows the posterior distribution of ρ calculated from (8.5.35). For this example $\nu = 9$ and $r = -0.81$. The distribution is skewed to the right and concentrated rather sharply about its mode at $\rho \doteq -0.87$; it practically rules out values of ρ exceeding -0.3.

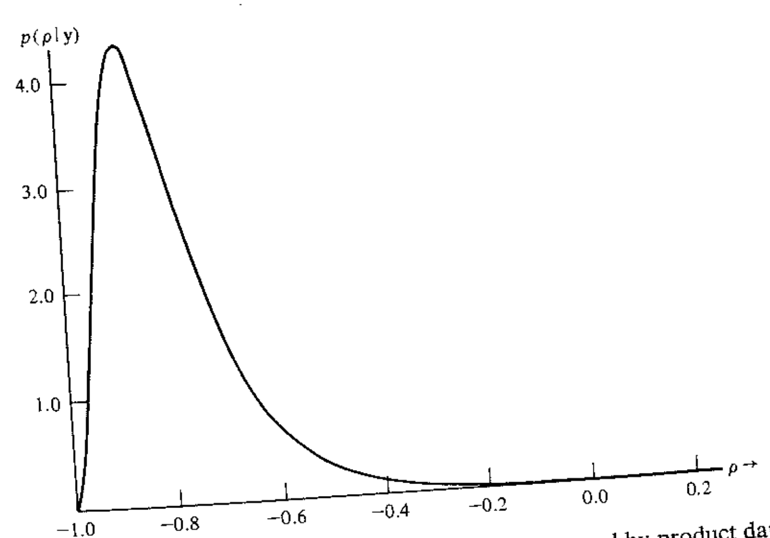

Fig. 8.5.1 Posterior distributions of ρ for the product and by-product data.

Series Expansions of $p(\rho \mid \mathbf{y})$

The distribution in (8.5.35) can be put in various different forms. In particular, it can be expressed as

$$p(\rho \mid \mathbf{y}) \propto \frac{(1-\rho^2)^{\frac{1}{2}(\nu-2)}}{(1-\rho r)^{\nu+\frac{1}{2}}} S_\nu(\rho, r), \qquad -1 < \rho < 1, \tag{8.5.43}$$

where

$$S_\nu(\rho, r) = 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \left(\frac{1+\rho r}{8}\right)^l \prod_{s=1}^{l} \frac{(2s-1)^2}{(\nu + s + \frac{1}{2})}$$

is a hypergeometric series, and the normalizing constant is

$$(1-r^2)^{\frac{1}{2}(\nu+1)} [\Gamma(\nu+1)]^2 \Big/ \left[2^{(\nu-\frac{1}{2})} \Gamma\left(\frac{\nu}{2}\right) \Gamma\left(\frac{\nu+1}{2}\right) \Gamma(\nu + \tfrac{3}{2})\right].$$

To see this, the integral in (8.5.35) can be written

$$\int_0^\infty \omega^{-1} \left[\omega + \frac{1}{\omega} - 2\rho r\right]^{-(\nu+1)} d\omega = 2 \int_1^\infty \omega^{-1} \left[\omega + \frac{1}{\omega} - 2\rho r\right]^{-(\nu+1)} d\omega. \tag{8.5.44}$$

On the right-hand side of (8.5.44), we may make the substitution, again due to Fisher,

$$u = 1 - \frac{1-\rho r}{\frac{1}{2}[\omega + (1/\omega)] - \rho r} = \frac{\frac{1}{2}[\omega + (1/\omega)] - 1}{\frac{1}{2}[\omega + (1/\omega)] - \rho r}. \tag{8.5.45}$$

Noting that

$$\frac{1}{4}\left(\omega + \frac{1}{\omega}\right)^2 - \frac{1}{4}\left(\omega - \frac{1}{\omega}\right)^2 = 1,$$

we have

$$\frac{\partial u}{\partial \omega} = \omega^{-1}(1-u)(2u)^{1/2}[1 - \tfrac{1}{2}(1+\rho r)u]^{1/2}(1-\rho r)^{-1/2} \tag{8.5.46}$$

so that

$$p(\rho \mid \mathbf{y}) \propto \frac{(1-\rho^2)^{\frac{1}{2}(\nu-2)}}{(1-\rho r)^{\nu+\frac{1}{2}}} \int_0^1 \frac{(1-u)^\nu}{(2u)^{1/2}} [1 - \tfrac{1}{2}(1+\rho r)u]^{-1/2}\, du, \qquad -1 < \rho < 1. \tag{8.5.47}$$

Expanding the last term in the integrand in powers of u and integrating term by term, each term being a complete beta function, we obtain the result in (8.5.43).

We remark here that an alternative series representation of the distribution of ρ can be obtained as follows. In the integral of (8.5.35), since

$$\omega^{-1}\left(\omega + \frac{1}{\omega} - 2\rho r\right)^{-(\nu+1)} = \omega^{\nu}(\omega^2 - 2\rho r\omega + 1)^{-(\nu+1)}, \tag{8.5.48}$$

by completing the square in the second factor on the right-hand side of (8.5.48), we can write the distribution in (8.5.35) as

$$p(\rho \mid \mathbf{y}) \propto \frac{(1-\rho^2)^{\frac{1}{2}(\nu-2)}}{(1-\rho^2 r^2)^{(\nu+1)}} \int_1^{\infty} \omega^{\nu}\left[1 + \frac{(\omega - \rho r)^2}{1-\rho^2 r^2}\right]^{-(\nu+1)} d\omega \tag{8.5.49}$$

Upon repeated integration by parts, the above integral can be expressed as a finite series involving powers of $[(1-\rho r)/(1+\rho r)]^{1/2}$ and Student's t integrals. The density function of ρ can thus be calculated from a table of t distribution. This process becomes very tedious when ν is moderately large so its practical usefulness is limited.

The series $S_\nu(\rho, r)$ in (8.5.43) has its leading term equal to one, followed by terms of order $\nu^{-l}, l = 1, 2, \ldots$. When ν is moderately large, we may simply take the first term so that approximately

$$p(\rho \mid \mathbf{y}) \doteq c\,\frac{(1-\rho^2)^{\frac{1}{2}(\nu-2)}}{(1-\rho r)^{\nu+\frac{1}{2}}}, \qquad -1 < \rho < 1, \tag{8.5.50}$$

where c is the normalizing constant

$$c^{-1} = \int_{-1}^{1} \frac{(1-\rho^2)^{\frac{1}{2}(\nu-2)}}{(1-\rho r)^{\nu+\frac{1}{2}}}\, d\rho.$$

Although evaluation of c would still require the use of numerical methods, it is much simpler to calculate the distribution of ρ using (8.5.50) than to evaluate the integral in (8.5.35) for every value of ρ. Table 8.5.1 compares the exact distribution with the approximation using the data in Table 8.4.1. In spite of the fact that ν is only 9, the agreement is very close.

It is easily seen that the density function (8.5.50) is greatest when ρ is near r. However, except when $r = 0$, the distribution is asymmetrical.

The asymmetry can be reduced by making the transformation,

$$\zeta = \tanh^{-1}\rho = \tfrac{1}{2}\log\frac{1+\rho}{1-\rho}, \tag{8.5.51}$$

due to Fisher (1921). Following his argument, it is found that ζ is approximately Normal

$$N\left[\tanh^{-1} r - \frac{5r}{2(\nu+1)}, (\nu+1)^{-1}\right].$$

Setting $m = 2$, $k = 1$ so that $\nu = n - 2$, the distribution in (8.5.43) is identical to that given by Jeffreys for the case of sampling from a bivariate Normal population. Finally, we note that while we have obtained above the distribution of the specific correlation coefficient $\rho = \rho_{12}$, it is clear that the distribution of any correlation ρ_{ij}, $i \neq j$, is given simply by setting $r = r_{ij} = a_{ij}/(a_{ii}\,a_{jj})^{1/2}$ in (8.5.35) and its associated expressions.

Table 8.5.1

Comparison of the exact and the approximate distributions of ρ for $\nu = 9$ and $r = -0.81$

	$p(\rho \mid \mathbf{y})$	
ρ	Exact	Approximate
−0.98	0.2286	0.2283
−0.96	1.2164	1.2150
−0.94	2.4867	2.4844
−0.92	3.5159	3.5134
−0.90	4.1200	4.1180
−0.88	4.3281	4.3271
−0.86	4.2448	4.2448
−0.84	3.9782	3.9790
−0.82	3.6141	3.6157
−0.80	3.2127	3.2149
−0.70	1.5182	1.5210
−0.60	0.6584	0.6603
−0.50	0.2853	0.2865
−0.40	0.1260	0.1267
−0.30	0.0569	0.0572
−0.20	0.0261	0.0263

8.6 A SUMMARY OF FORMULAE AND CALCULATIONS FOR MAKING INFERENCES ABOUT $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$

Using the product, by-product data in Table 8.4.1 for illustration, Table 8.6.1 below provides a short summary of the formulae and calculations required for making inferences about the elements of $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ for the linear model with common derivative matrix defined in (8.4.1). Specifically, the model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{8.6.1}$$

where $\mathbf{y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]$ is a $n \times m$ matrix of observations, $\mathbf{X}$ is a $n \times k$ matrix of fixed elements with rank k, $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m]$ is a $k \times m$ matrix of parameters and $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_{(1)}, \ldots, \boldsymbol{\varepsilon}_{(n)}]'$ is a $n \times m$ matrix of errors. It is assumed that $\boldsymbol{\varepsilon}_{(u)}$, $u = 1, \ldots, n$ are independently distributed as $N_m(\mathbf{0}, \boldsymbol{\Sigma})$.

Table 8.6.1

Summarized calculations for the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$

1. From (8.4.1), (8.4.4), (8.4.7) and (8.4.13), obtain

$$m=2, \quad k=2, \quad n=12, \quad \nu=n-(m+k)+1=9$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 12 & 0 \\ 0 & 0.15 \end{bmatrix}, \qquad \mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.08 & 0 \\ 0 & 6.88 \end{bmatrix}$$

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 71.14 & 18.07 \\ 54.44 & -20.09 \end{bmatrix}. \qquad \mathbf{A} = \{a_{ij}\} \quad i,j=1,\ldots,m$$

$$a_{ij} = (\mathbf{y}_i - \mathbf{X}\hat{\boldsymbol{\theta}}_i)'(\mathbf{y}_j - \mathbf{X}\hat{\boldsymbol{\theta}}_j), \qquad \text{and} \qquad \mathbf{A} = \begin{bmatrix} 61.41 & -38.48 \\ -38.48 & 36.74 \end{bmatrix}.$$

2. Inferences about a specific column or row of $\boldsymbol{\theta}$:

Writing

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m] = \begin{bmatrix} \boldsymbol{\theta}'_{(1)} \\ \vdots \\ \boldsymbol{\theta}'_{(k)} \end{bmatrix}, \qquad \hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_m] = \begin{bmatrix} \hat{\boldsymbol{\theta}}'_{(1)} \\ \vdots \\ \hat{\boldsymbol{\theta}}'_{(k)} \end{bmatrix}$$

then from (8.4.28) and (8.4.34),

$$\boldsymbol{\theta}_i \sim t_k\left(\hat{\boldsymbol{\theta}}_i, \frac{a_{ii}}{\nu}\mathbf{C}, \nu\right), \qquad i=1,\ldots,m,$$

and

$$\boldsymbol{\theta}_{(g)} \sim t_m\left(\hat{\boldsymbol{\theta}}_{(g)}, \frac{c_{gg}}{\nu}\mathbf{A}, \nu\right), \qquad g=1,\ldots,k.$$

Thus, for $i=1$ and $g=2$

$$\boldsymbol{\theta}_1 \sim t_2\left\{\begin{pmatrix} 71.14 \\ 54.44 \end{pmatrix}, 6.83 \times \begin{bmatrix} 0.08 & 0 \\ 0 & 6.88 \end{bmatrix}, 9\right\}$$

$$\boldsymbol{\theta}_{(2)} \sim t_2\left\{\begin{pmatrix} 54.44 \\ 20.09 \end{pmatrix}, 0.76 \times \begin{bmatrix} 61.41 & -38.48 \\ -38.48 & 36.74 \end{bmatrix}, 9\right\}.$$

3. H.P.D. regions of $\boldsymbol{\theta}$: To decide if a general parameter point $\boldsymbol{\theta}_0$ lies inside or outside the $(1-\alpha)$ H.P.D. region, from (8.4.63) and (8.4.66), use the approximation

$$-\phi\nu \log U \,\dot{\sim}\, \chi^2_{mk},$$

where

$$U = U(\boldsymbol{\theta}) = |\mathbf{A}|\,|\mathbf{A} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-1} \quad \text{and} \quad \phi = 1 + \frac{m+k-3}{2\nu},$$

so that $\boldsymbol{\theta}_0$ lies inside the $1-\alpha$ region if

$$-\phi\nu \log U(\boldsymbol{\theta}_0) < \chi^2(mk, \alpha).$$

Table 8.6.1 *Continued*

For the example, $\phi = 19/18$. Thus, if

$$\boldsymbol{\theta}_0 = \begin{bmatrix} 70 & 17 \\ 65 & -30 \end{bmatrix}, \qquad \text{then} \qquad U(\boldsymbol{\theta}_0) = 0.17$$

and the point lies inside the $1-\alpha$ region if

$$-9.5 \log 0.17 = 16.7 < \chi^2(4, \alpha).$$

4. H.P.D. regions for a block submatrix of $\boldsymbol{\theta}$: Let

$$\boldsymbol{\theta} = \begin{array}{cc} m_1 & m_2 \\ \left[\begin{array}{c|c} \boldsymbol{\theta}_{11} & \boldsymbol{\theta}_{12} \\ \hline \boldsymbol{\theta}_{21} & \boldsymbol{\theta}_{22} \end{array}\right] & \begin{array}{c} k_1 \\ k_2 \end{array} \end{array} \qquad \hat{\boldsymbol{\theta}} = \begin{array}{cc} m_1 & m_2 \\ \left[\begin{array}{c|c} \hat{\boldsymbol{\theta}}_{11} & \hat{\boldsymbol{\theta}}_{12} \\ \hline \hat{\boldsymbol{\theta}}_{21} & \hat{\boldsymbol{\theta}}_{22} \end{array}\right] & \begin{array}{c} k_1 \\ k_2 \end{array} \end{array}$$

$$\mathbf{A} = \begin{array}{cc} m_1 & m_2 \\ \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] & \begin{array}{c} m_1 \\ m_2 \end{array} \end{array} \qquad \mathbf{C} = \begin{array}{cc} k_1 & k_2 \\ \left[\begin{array}{c|c} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \hline \mathbf{C}_{21} & \mathbf{C}_{22} \end{array}\right] & \begin{array}{c} k_1 \\ k_2 \end{array} \end{array}$$

To decide, for example, if the parameter point $\boldsymbol{\theta}_{11,0}$ lies inside or outside the $(1-\alpha)$ H.P.D. region for $\boldsymbol{\theta}_{11}$, use the approximation in (8.4.68),

$$\phi_1 \nu \log U(\boldsymbol{\theta}_{11}) \sim \chi^2_{m_1 k_1},$$

where

$$U(\boldsymbol{\theta}_{11}) = |\mathbf{A}_{11}|\, |\mathbf{A}_{11} + (\boldsymbol{\theta}_{11} - \hat{\boldsymbol{\theta}}_{11})' \mathbf{C}_{11}^{-1} (\boldsymbol{\theta}_{11} - \hat{\boldsymbol{\theta}}_{11})|^{-1}$$

and $\phi_1 = 1 + (1/2\nu)(m_1 + k_1 - 3)$, so that $\boldsymbol{\theta}_{11,0}$ lies inside the $(1-\alpha)$ region if

$$-\phi_1 \nu \log U(\boldsymbol{\theta}_{11,0}) < \chi^2(m_1 k_1, \alpha),$$

Thus if $m_1 = k_1 = 1$, $\theta_{11,0} = 70$, then $\phi_1 = 17/18$ and

$$U(\theta_{11,0} = 70) = 61.41 \Big/ \left[61.41 + \frac{(70 - 71.14)^2}{0.08}\right] = 0.79,$$

so that $\theta_{11,0}$ lies inside the $(1-\alpha)$ region if

$$-8.5 \log 0.79 = 2.0 < \chi^2(1, \alpha).$$

Note that since $m_1 = k_1 = 1$, exact results could be obtained and the above is for illustration only.

5. Inferences about the diagonal elements of $\boldsymbol{\Sigma}$:
From (8.5.22)

$$\sigma_{ii} \sim a_{ii} \chi_\nu^{-2}, \qquad i = 1, \ldots, m.$$

Thus, for $i = 1$, $\sigma_{11} \sim 61.41 \chi_9^{-2}$

6. H.P.D. regions for the "regression matrix" $\mathbf{T}$: Let

$$\boldsymbol{\Sigma} = \begin{array}{cc} m_1 & m_2 \\ \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right] & \begin{array}{c} m_1 \\ m_2 \end{array} \end{array}, \qquad \mathbf{T} = \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}, \qquad \hat{\mathbf{T}} = \mathbf{A}_{11}^{-1} \mathbf{A}_{12},$$

$$\mathbf{A}_{22\cdot 1} = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}$$

Table 8.6.1 *Continued*

To decide if a parameter point $\mathbf{T}_0$ lies inside or outside the $(1-\alpha)$ H.P.D. region, from (8.5.23), use the approximation

$$-\phi'(\nu+m_1)\log U(\mathbf{T}) \sim \chi^2_{m_1 m_2}$$

where

$$\phi' = 1 + \frac{m_1+m_2-3}{2(\nu+m_1)}$$

and

$$U(\mathbf{T}) = |\mathbf{A}_{22\cdot 1}|\,|\mathbf{A}_{22\cdot 1}+(\mathbf{T}-\hat{\mathbf{T}})'\mathbf{A}_{11}(\mathbf{T}-\hat{\mathbf{T}})|^{-1}$$

so that $\mathbf{T}_0$ lies inside the region if

$$-\phi'(\nu+m_1)\log U(\mathbf{T}_0) < \chi^2(m_1 m_2, \alpha).$$

Thus, for $m_1=m_2=1$, $\phi'(\nu+m_1)=9.5$, $\mathbf{A}_{22\cdot 1}=a_{22\cdot 1}=12.63$ and $\hat{T}=0.63$. If $T_0=0$, then

$$U(T_0=0) = 12.63/[12.63+(0.63)^2\times 61.41]=0.34$$

and the point $T_0=0$ will lie inside the $(1-\alpha)$ region if

$$-9.5\log 0.34 = 10.3 < \chi^2(1,\alpha).$$

Note that (*i*) $U(\mathbf{0})=|\mathbf{A}|\{|\mathbf{A}_{11}|\,|\mathbf{A}_{22}|\}^{-1}$, and (ii) since $m_1=m_2=1$, exact results are, of course, available.

7. Inferences about the correlation coefficient ρ_{ij}: Use the approximating distribution in (8.5.50),

$$p(\rho\mid \mathbf{y}) \,\dot{\propto}\, \frac{(1-\rho^2)^{\frac{1}{2}(\nu-2)}}{(1-\rho r)^{\nu+\frac{1}{2}}}, \qquad -1<\rho<1,$$

where

$$\rho=\rho_{ij} \quad \text{and} \quad r=r_{ij}=a_{ij}/(a_{ii}a_{jj})^{1/2}, \qquad i,j=1,\ldots,m.$$

Thus,

$$p(\rho\mid \mathbf{y}) \,\dot{\propto}\, \frac{(1-\rho^2)^{3.5}}{(1+0.81\rho)^{9.5}}, \quad -1<\rho<1.$$

The normalizing constant of the distribution can be obtained by numerical integration when desired.

APPENDIX A8.1

The Jacobians of Some Matrix Transformations

We here give the Jacobians of some matrix transformations useful in multivariate problems. In what follows the signs of the Jacobians are ignored.

a) Let $\mathbf{X}$ be a $k\times m$ matrix of km distinct random variables. Let $\mathbf{A}$ and $\mathbf{B}$ be, respectively, a $k\times k$ and $m\times m$ matrices of fixed elements. If

$$\mathbf{Z}_1 = \mathbf{A}\mathbf{X}, \qquad \mathbf{Z}_2 = \mathbf{X}\mathbf{B}, \qquad \mathbf{Z}_3 = \mathbf{A}\mathbf{X}\mathbf{B},$$

then the Jacobians are, respectively,

$$\left|\frac{\partial \mathbf{Z}_1}{\partial \mathbf{X}}\right| = |\mathbf{A}|^m, \qquad \left|\frac{\partial \mathbf{Z}_2}{\partial \mathbf{X}}\right| = |\mathbf{B}|^k, \qquad \left|\frac{\partial \mathbf{Z}_3}{\partial \mathbf{X}}\right| = |\mathbf{A}|^m\,|\mathbf{B}|^k. \quad \text{(A8.1.}$$

b) Let $\mathbf{X}$ be a $m \times m$ symmetric matrix consisting of $\frac{1}{2}m(m+1)$ distin elements and let $\mathbf{C}$ be a $m \times m$ matrix of fixed elements. If

$$\mathbf{Y} = \mathbf{C}\,\mathbf{X}\,\mathbf{C}',$$

then

$$\left|\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}\right| = |\mathbf{C}|^{m+1}. \quad \text{(A8.1.2}$$

For proofs, see Deemer and Olkin (1951), based on results given b P. L. Hsu, also Anderson (1958, p. 162).

c) Let $\mathbf{X}$ be a $m \times m$ nonsingular matrix of random variables and $\mathbf{Y} = \mathbf{X}^{-1}$ Then,

$$\frac{\partial \mathbf{Y}}{\partial z} = -\,\mathbf{Y}\frac{\partial \mathbf{X}}{\partial z}\mathbf{Y}, \quad \text{(A8.1.3}$$

where z is any function of the elements of $\mathbf{X}$.

Proof: Since $\mathbf{X}\,\mathbf{Y} = \mathbf{I}$, it follows that

$$\frac{\partial}{\partial z}(\mathbf{X}\,\mathbf{Y}) = \left(\frac{\partial \mathbf{X}}{\partial z}\right)\mathbf{Y} + \mathbf{X}\frac{\partial \mathbf{Y}}{\partial z} = \mathbf{0}.$$

Hence

$$\frac{\partial \mathbf{Y}}{\partial z} = -\,\mathbf{Y}\left(\frac{\partial \mathbf{X}}{\partial z}\right)\mathbf{Y}.$$

The Jacobians of two special cases of $\mathbf{X}$ are of particular interest.

a) If $\mathbf{X}$ has m^2 distinct random variables, then from (A8.1.1)

$$\left|\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}\right| = |\mathbf{Y}|^{2m}. \quad \text{(A8.1.4)}$$

b) If $\mathbf{X}$ is symmetric, and consists of $\frac{1}{2}m(m+1)$ distinct random variables, then from (A8.1.2)

$$\left|\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}\right| = |\mathbf{Y}|^{m+1}. \quad \text{(A8.1.5)}$$

APPENDIX A8.2

The Determinant of the Information Matrix of Σ^{-1}

We now obtain the determinant of the information matrix of Σ^{-1} for the m-dimensional Normal distribution $N_m(\boldsymbol{\mu}, \Sigma)$. The density is

$$p(\mathbf{y} \mid \boldsymbol{\mu}, \Sigma) = (2\pi)^{-m/2} |\Sigma^{-1}|^{1/2} \exp[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'], \qquad -\infty < \mathbf{y} < \infty, \tag{A8.2.1}$$

where $\mathbf{y} = (y_1, \ldots, y_m)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)'$, $\Sigma = \{\sigma_{ij}\}$ and $\Sigma^{-1} = \{\sigma^{ij}\}$, $i, j = 1, \ldots, m$. We assume that Σ (and Σ^{-1}) consists of $\frac{1}{2}m(m+1)$ distinct elements. Taking logarithms of the density function, we have

$$\log p = -\frac{m}{2} \log(2\pi) + \tfrac{1}{2} \log |\Sigma^{-1}| - \tfrac{1}{2} \operatorname{tr} \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'. \tag{A8.2.2}$$

Differentiating $\log p$ with respect to σ^{ij}, $i, j = 1, \ldots, m$, $i \geqslant j$,

$$\frac{\partial \log p}{\partial \sigma^{ij}} = \frac{1}{2} \frac{1}{|\Sigma^{-1}|} \frac{\partial |\Sigma^{-1}|}{\partial \sigma^{ij}} - (y_i - \mu_i)(y_j - \mu_j). \tag{A8.2.3}$$

Since $\partial |\Sigma^{-1}| / \partial \sigma^{ij} = \alpha^{ij}$ where α^{ij} is the cofactor of σ^{ij}, it follows that the first term on the right-hand side of (A8.2.3) is simply $\frac{1}{2}\sigma_{ij}$. Thus, the second derivatives are

$$\frac{\partial^2 \log p}{\partial \sigma^{ij} \, \partial \sigma^{kl}} = \frac{1}{2} \frac{\partial \sigma_{ij}}{\partial \sigma^{kl}}, \qquad \begin{pmatrix} i, j = 1, \ldots, m; & i \geqslant j \\ k, l = 1, \ldots, m; & k \geqslant l \end{pmatrix}. \tag{A8.2.4}$$

Consequently, the determinant of the information matrix is proportional to

$$|\mathscr{I}(\Sigma^{-1})| = \left| -E \left\{ \frac{\partial^2 \log P}{\partial \sigma^{ij} \, \partial \sigma^{kl}} \right\} \right| \propto \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right|. \tag{A8.2.5}$$

From (A8.1.5), we have that

$$\left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right| = |\Sigma|^{m+1}. \tag{A8.2.6}$$

APPENDIX A8.3

The Normalizing Constant of the $t_{km}[\hat{\boldsymbol{\theta}}, (\mathbf{X}'\mathbf{X})^{-1}, \mathbf{A}, \nu]$ Distribution

Let $\boldsymbol{\theta}$ be a $k \times m$ matrix of variables. We now show that

$$\int_{-\infty < \boldsymbol{\theta} < \infty} |\mathbf{I}_m + \mathbf{A}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{X}'\mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})|^{-\frac{1}{2}(\nu + k + m - 1)} d\boldsymbol{\theta} = c(m, k, \nu) \, |\mathbf{X}'\mathbf{X}|^{-m/2} \, |\mathbf{A}^{-1}|^{-k/2}, \tag{A8.3.1}$$

where $\nu > 0$, $\hat{\boldsymbol{\theta}}$ is a $k \times m$ matrix, $\mathbf{X}'\mathbf{X}$ and $\mathbf{A}^{-1}$ are, respectively, a $k \times k$ and a $m \times m$ positive definite symmetric matrix,

$$c(m, k, \nu) = [\Gamma(\tfrac{1}{2})]^{mk} \frac{\Gamma_m[\frac{1}{2}(\nu + m - 1)]}{\Gamma_m[\frac{1}{2}(\nu + k + m - 1)]}$$

and $\Gamma_p(b)$ is the generalized Gamma function defined in (8.2.22).

Since $\mathbf{X}'\mathbf{X}$ and $\mathbf{A}$ are assumed definite, there exist a $k \times k$ and a $m \times m$ nonsingular matrices $\mathbf{G}$ and $\mathbf{H}$ such that

$$\mathbf{X}'\mathbf{X} = \mathbf{G}'\mathbf{G} \quad \text{and} \quad \mathbf{A}^{-1} = \mathbf{H}\mathbf{H}'. \tag{A8.3.2}$$

Let $\mathbf{T}$ be a $k \times m$ matrix such that

$$\mathbf{T} = \mathbf{G}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{H}. \tag{A8.3.3}$$

Using the identity (8.4.11) we can write

$$\begin{aligned}|\mathbf{I}_m + \mathbf{A}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})| &= |\mathbf{I}_m + \mathbf{H}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{G}'\mathbf{G}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\mathbf{H}| = |\mathbf{I}_m + \mathbf{T}'\mathbf{T}| \\ &= |\mathbf{I}_k + \mathbf{T}\mathbf{T}'|. \end{aligned} \tag{A8.3.4}$$

From (A8.1.1) the Jacobian of the transformation (A8.3.3) is

$$\left| \frac{\partial \mathbf{T}}{\partial \boldsymbol{\theta}} \right| = |\mathbf{G}|^m \, |\mathbf{H}|^k = |\mathbf{X}'\mathbf{X}|^{m/2} |\mathbf{A}^{-1}|^{k/2}. \tag{A8.3.5}$$

Thus, the integral on the left-hand side of (A8.3.1) is

$$|\mathbf{X}'\mathbf{X}|^{-m/2} |\mathbf{A}^{-1}|^{-k/2} \, Q_m, \tag{A8.3.6}$$

where

$$Q_m = \int_{-\infty < \mathbf{T} < \infty} |\mathbf{I}_k + \mathbf{T}\mathbf{T}'|^{-\frac{1}{2}(\nu + k + m - 1)} \, d\mathbf{T}. \tag{A8.3.7}$$

Let $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_m] = [\mathbf{T}_1, \mathbf{t}_m]$ where $\mathbf{t}_i$ is a $k \times 1$ vector, $i = 1, \ldots, m$. Then,

$$\begin{aligned}|\mathbf{I}_k + \mathbf{T}\mathbf{T}'| &= |\mathbf{I}_k + \mathbf{T}_1\mathbf{T}_1' + \mathbf{t}_m\mathbf{t}_m'| \\ &= |\mathbf{I}_k + \mathbf{T}_1\mathbf{T}_1'| \, [1 + \mathbf{t}_m'(\mathbf{I}_k + \mathbf{T}_1\mathbf{T}_1')^{-1}\mathbf{t}_m]. \end{aligned} \tag{A8.3.8}$$

It follows that

$$Q_m = \int_{-\infty < \mathbf{T}_1 < \infty} |\mathbf{I}_k + \mathbf{T}_1\mathbf{T}_1'|^{-\frac{1}{2}(\nu + k + m - 1)} \, q_m d\mathbf{T}_1, \tag{A8.3.9}$$

where

$$q_m = \int_{-\infty < \mathbf{t}_m < \infty} [1 + \mathbf{t}_m'(\mathbf{I}_k + \mathbf{T}_1\mathbf{T}_1')^{-1}\mathbf{t}_m]^{-\frac{1}{2}(\nu + m - 1 + k)} d\mathbf{t}_m.$$

From (A2.1.12) in Appendix A2.1,

$$q_m = [\Gamma(\tfrac{1}{2})]^k \frac{\Gamma[\frac{1}{2}(\nu + m - 1)]}{\Gamma[\frac{1}{2}(\nu + k + m - 1)]} \, |\mathbf{I}_k + \mathbf{T}_1\mathbf{T}_1'|^{1/2}.$$

Thus,
$$Q_m = [\Gamma(\tfrac{1}{2})]^k \frac{\Gamma[\frac{1}{2}(\nu + m - 1)]}{\Gamma[\frac{1}{2}(\nu + k + m - 1)]} Q_{m-1},$$

where
$$Q_{m-1} = \int_{-\infty < \mathbf{T}_1 < \infty} |\mathbf{I}_k + \mathbf{T}_1\mathbf{T}_1|^{-\frac{1}{2}(\nu+m-2+k)} \, d\mathbf{T}_1.$$

The result in (A8.3.1) follows by repeating the process $m - 1$ times.

APPENDIX A8.4

The Kronecker Product of Two Matrices

We summarize in this appendix some properties of the Kronecker product of two matrices.

Definition: If $\mathbf{A}$ is a $m \times m$ matrix and $\mathbf{B}$ is a $n \times n$ matrix, then the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ in that order is the $(mn) \times (mn)$ matrix
$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ \vdots & & \\ a_{m1}\mathbf{B} & & a_{mm}\mathbf{B} \end{bmatrix}.$$

Properties:

i) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$

ii) If $\mathbf{A}$ and $\mathbf{B}$ are symmetric, then $\mathbf{A} \otimes \mathbf{B}$ is symmetric.

iii) When $\mathbf{A}$ and $\mathbf{B}$ are non-singular, then
$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$$

iv) $\mathrm{tr}\, \mathbf{A} \otimes \mathbf{B} = \mathrm{tr}\, \mathbf{A}\, \mathrm{tr}\, \mathbf{B}$

v) $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n |\mathbf{B}|^m$.

vi) If $\mathbf{C}$ is a $m \times m$ matrix and $\mathbf{D}$ is a $n \times n$ matrix, then
$$(\mathbf{A} + \mathbf{C}) \otimes \mathbf{B} = \mathbf{A} \otimes \mathbf{B} + \mathbf{C} \otimes \mathbf{B}$$
and
$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}).$$