CHAPTER 9

ESTIMATION OF COMMON REGRESSION COEFFICIENTS

9.1 INTRODUCTION: PRACTICAL IMPORTANCE OF COMBINING INFORMATION FROM DIFFERENT SOURCES

Problems occur, notably in engineering, economics, and business, where it is difficult to obtain sufficiently reliable estimates of critical parameters. One way to improve estimates is to seek more data of the same kind. But in engineering this may mean running further costly experiments while in economics and business further data from the same source may not be available. Another way in which estimates can sometimes be improved is by combining information from *different kinds* of data. Thus, in Section 8.2.6 we considered a chemical system producing three different products where the yield of each product depended on one or both of two chemical rate parameters. It was demonstrated that, by appropriately utilizing observations on all three products, much more precise inferences could be drawn than would have been possible with only one.

As a further illustration, in a study of the effect of interest rates on the demand for durables, an economist might wish to appropriately combine demand data from m different communities. Let $\mathbf{y}_i = (y_{1i}, \ldots, y_{ni})'$ be an $n \times 1$ vector, representing demand over n specific periods, for the ith community. Suppose that

$$\begin{aligned} \mathbf{y}_i &= E(\mathbf{y}_i) + \boldsymbol{\varepsilon}_i, \\ E(\mathbf{y}_i) &= \boldsymbol{\eta}_i(\boldsymbol{\xi}_i, \theta_i), \qquad i = 1, \ldots, m, \end{aligned} \tag{9.1.1}$$

where θ_i is a parameter for the ith community, measuring the effect on demand of a change in the interest rate, $\boldsymbol{\xi}_i$ is the corresponding $n \times 1$ vector of interest rates, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \ldots, \varepsilon_{ni})'$ is the associated $n \times 1$ vector of errors.

To complete the model we must make specific assumptions about (i) the relationship among $(\theta_1, \ldots, \theta_m)$, (ii) the nature of the expectation functions $\boldsymbol{\eta}_i(\boldsymbol{\xi}_i, \theta_i)$ and (iii) the form of the probability distribution of $\boldsymbol{\varepsilon}_i$. We discuss these assumptions below:

(i) If there is reason to believe that the parameters $(\theta_1, \ldots, \theta_m)$ for the m different communities resemble one another, then, irrespective of whether or not there are correlations among the m vectors of errors, $(\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_m)$, information about θ_i from the ith response vector $\mathbf{y}_i$ ought to contribute to knowledge about θ_j. A formulation which takes account of one kind of resemblence, supposes that $\theta_1, \ldots, \theta_m$

are m independent observations from some population distributed as $f(\theta)$. In particular, the random effect model discussed in Section 7.2 was of this kind. If the variance of $f(\theta)$ was expected to be small, the m vectors of expectation functions η_i might then be regarded as containing a *common* parameter $\theta_1 = \ldots = \theta_m = \theta_c$ and data from all m communities could be used to make inferences about θ_c. We have already considered a multivariate setup of this kind in Section 8.2.6. It was there possible to show that, even in the general situation where the expectation functions were nonlinear in the parameters, exact results could be obtained yielding the posterior distribution of the common parameters in a form which was mathematically simple.

ii) Mathematical simplicity does not invariably lead to tractible computation however, and we now consider the problem with the added simplification that the expectation functions are *linear* in the parameter θ_c. As usual, consideration of the more manageable linear case has more than one motivation. On the one hand, practical problems do occur where the appropriate expectation functions are indeed linear in the parameters. On the other hand, for nonlinear problems, we rarely need to concern ourselves with nonlinearity over the whole permissible region of variation of θ_c, but only in some smaller sub-region of the parameter space—for example, over a $(1 - \alpha)$ H.P.D. region for some small α. A function which is "globally" highly nonlinear in θ can be, in this sense, "locally" nearly linear. In such a case, as is indicated in Section 8.3.1, linear distribution theory can provide useful approximations. In the Bayesian formulation, we are concerned only with approximate linearity in relation to the one sample of observations which has in fact occurred and not with some hypothetical "repeated samples from the same population" which have not. It follows that we can, *for any given data sample*, check the adequacy of the assumption of local linearity by direct numerical computation as part of the statistical analysis.

iii) We shall suppose that the n vectors $\boldsymbol{\varepsilon}_{(u)} = (\varepsilon_{u1}, \ldots, \varepsilon_{um})$, $u = 1, \ldots, n$, are a sample of independent observations from the m-dimensional Normal population $N_m(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$.

In general, the relationships connecting demand with interest rate will contain, not one, but k common parameters $\boldsymbol{\theta}_c$. Thus, with the above assumptions we are led to a linear, Normal, common parameter model, which can be written

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta}_c + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, m \tag{9.1.2}$$

with $\mathbf{X}_i$ an $n \times k$ matrix appropriate to the ith community and $\boldsymbol{\theta}_c$ a $k \times 1$ vector of parameters common to all communities.

This model (9.1.2) will be recognized as a special case of the general linear model set out in (8.3.2)

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, m \tag{9.1.3}$$

appropriate to the circumstance that the regression coefficients $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ associated with the m vectors of responses are *common*, so that

$$\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_m = \boldsymbol{\theta}_c. \tag{9.1.4}$$

In what follows we shall first discuss the situation, where the m response vectors $\mathbf{y}_1, \ldots, \mathbf{y}_m$ are supposed independent of one another so that Σ is diagonal. The posterior distribution of $\boldsymbol{\theta}_c$ is then a product of m multivariate t distributions. A useful and simple approximation procedure is developed for making inferences about subsets of $\boldsymbol{\theta}_c$ for the case of two responses. The more general situation is next considered where the m responses are correlated. Except when the derivative matrices $\mathbf{X}_1, \ldots, \mathbf{X}_m$ are identical, the resulting posterior distribution of $\boldsymbol{\theta}_c$ is rather complicated analytically. A useful approximation procedure will, however, be developed for $m = 2$.

As we have pointed out in Section 8.2.6, care must be exercised in the pooling of information to estimate common parameters. A natural question that arises in the investigator's mind will be whether the evidence from the data are *compatible* with the supposition that $\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_m = \boldsymbol{\theta}_c$. In the following section, this question is considered for the important special case of estimating the assumed common mean of two Normal populations which have possibly unequal variances—a problem which has an interesting history.

9.2 THE CASE OF m INDEPENDENT RESPONSES

Suppose there are m independent responses related to the same parameters $\boldsymbol{\theta}_c = (\theta_1, \ldots, \theta_k)'$ by possibly different linear models. Suppose further that the number of observations for each response is not necessarily the same. Then the likelihood function is

$$l(\boldsymbol{\theta}_c, \sigma_{11}, \ldots, \sigma_{mm} \mid \mathbf{y}) \propto \left\{\prod_{i=1}^{m} \sigma_{ii}^{-\frac{1}{2}n_i}\right\} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m} \sigma_{ii}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_c)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_c)\right\}, \tag{9.2.1}$$

where $\mathbf{y}_i$ is an $n_i \times 1$ vector of observations and $\mathbf{X}_i$ a $n_i \times k$ matrix of fixed elements. Relative to the noninformative reference prior

$$p(\boldsymbol{\theta}_c, \sigma_{11}, \ldots, \sigma_{mm}) \propto \prod_{i=1}^{m} \sigma_{ii}^{-1},$$

the posterior distribution is

$$p(\boldsymbol{\theta}_c, \sigma_{11}, \ldots, \sigma_{mm} \mid \mathbf{y}) \propto \left\{\prod_{i=1}^{m} \sigma_{ii}^{-(\frac{1}{2}n_i+1)}\right\}$$
$$\times \exp\left[-\frac{1}{2}\sum_{i=1}^{m} \frac{a_{ii} + (\boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_i)'\mathbf{X}_i'\mathbf{X}_i(\boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_i)}{\sigma_{ii}}\right], \tag{9.2.2}$$

where $$(\mathbf{X}_i'\mathbf{X}_i)\hat{\boldsymbol{\theta}}_i = \mathbf{X}_i'\mathbf{y}_i \qquad \text{and} \qquad a_{ii} = (\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\theta}}_i)'(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\theta}}_i).$$

Integrating out $\sigma_{11}, \ldots, \sigma_{mm}$, the posterior distribution of the common parameters $\boldsymbol{\theta}_c$ is then proportional to the product of m multivariate t distributions

$$p(\boldsymbol{\theta}_c \mid \mathbf{y}) \propto \prod_{i=1}^{m} [a_{ii} + (\boldsymbol{\theta}_c - \hat{\boldsymbol{\theta}}_i)'\mathbf{X}_i'\mathbf{X}_i(\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)]^{-\frac{1}{2}n_i}, \qquad -\infty < \boldsymbol{\theta}_c < \infty. \tag{9.2.3}$$

Properties of this distribution have been considered by Dickey (1967a, 1968) and others.

9.2.1 The Weighted Mean Problem

The simplest problem of this kind has come to be called the problem of the *weighted mean*, Yates (1939). Suppose that $m = 2$, $k = 1$ and $(\mathbf{X}_1 = \mathbf{1}_{n_1}, \mathbf{X}_2 = \mathbf{1}_{n_2})$ are columns of ones. Typically we have two sets of Normally distributed observations $\mathbf{y}_1 = (y_{11}, \ldots, y_{n_1 1})'$ and $\mathbf{y}_2 = (y_{12}, \ldots, y_{n_2 2})'$ independent of one another and having different variances $\sigma_1^2 = \sigma_{11}$ and $\sigma_2^2 = \sigma_{22}$ but both assumed to have the same mean θ. The posterior distribution of θ is then proportional to the product of two t distributions

$$p(\theta \mid \mathbf{y}) \propto p(\theta \mid \mathbf{y}_1)\, p(\theta \mid \mathbf{y}_2) \propto \left[1 + \frac{n_1(\theta - \bar{y}_1)^2}{\nu_1 s_1^2}\right]^{-\frac{1}{2}(\nu_1+1)}$$

$$\times \left[1 + \frac{n_2(\theta - \bar{y}_2)^2}{\nu_2 s_2^2}\right]^{-\frac{1}{2}(\nu_2+1)}, \qquad -\infty < \theta < \infty, \tag{9.2.4}$$

where

$$\nu_i = n_i - 1, \qquad \nu_i s_i^2 = \Sigma(y_{ui} - \bar{y}_i)^2, \qquad i = 1, 2,$$

a distribution given by Jeffreys (1961). Simultaneously Fisher (1961a, b) obtained an identical distribution using a fiducial argument. Except for the normalizing constant, the density (9.2.4) can be easily calculated from a table of the density function of Student's t—e.g. Bracken and Schleifer (1964).

An Example

The following results were obtained in estimating a physical constant by two independent methods

	Method 1	*Method* 2
$\bar{y}$	107.9	109.5
s^2	12.1	39.4
n	11	13

Fig. 9.2.1 Individual posterior distributions of θ: estimation of a common physical constant.

Fig. 9.2.2 Combined posterior distribution of θ: estimation of a common physical constant.

If θ_1 and θ_2 are the means for the two methods, then crucial to what follows is the assumption that $\theta_1 = \theta_2 = \theta$. We shall entertain this assumption for the moment and later discuss it further in Section 9.2.3. With this assumption, then, the two t distributions shown in Fig. 9.2.1 centered at 107.9 and 109.5 with scale factors $s_1/\sqrt{n_1} = 1.03$ and $s_2/\sqrt{n_2} = 1.74$ are the individual posterior distributions $p(\theta \mid \mathbf{y}_1)$ and $p(\theta \mid \mathbf{y}_2)$ obtained respectively from the results of methods A and B. The *unnormalized* posterior distribution of θ which combines information from both samples, is obtained by multiplying the t densities and is shown by the "t-like" distribution in Fig. 9.2.2. As might be expected, the combined distribution is "located" between the individual t distributions. For this example, $p(\theta \mid \mathbf{y}_1)$ has a narrower spread than $p(\theta \mid \mathbf{y}_2)$ and consequently exerts a stronger influence on the final posterior distribution.

9.2.2 Some Properties of $p(\theta \mid \mathbf{y})$

Although the distribution $p(\theta \mid \mathbf{y})$ in (9.2.4) is the product of two distributions which are symmetric, the distribution itself will not be symmetric unless either $\bar{y}_1 = \bar{y}_2$ or $n_1 = n_2$ and $s_1^2 = s_2^2$. We see for example in Fig. 9.2.2 that the distribution is slightly skewed to the right.

Upon differentiating the logarithm of $p(\theta \mid \mathbf{y})$ with respect to θ, it is easy to see that for $\bar{y}_1 < \bar{y}_2$

$$\frac{\partial}{\partial\theta} \log p(\theta \mid \mathbf{y}) \gtrless 0 \qquad \text{for} \qquad \begin{matrix} \theta < \bar{y}_1 \\ \theta > \bar{y}_2 \end{matrix}, \tag{9.2.5}$$

where

$$\frac{\partial}{\partial\theta} \log p(\theta \mid \mathbf{y}) = -\frac{n_1^2(\theta - \bar{y}_1)}{\nu_1 s_1^2 + n_1(\theta - \bar{y}_1)^2} - \frac{n_2^2(\theta - \bar{y}_2)}{\nu_2 s_2^2 + n_2(\theta - \bar{y}_2)^2}, \tag{9.2.6}$$

so that the distribution is increasing in $(-\infty, \bar{y}_1)$ and decreasing in $(\bar{y}_2, \infty)$ and consequently the mode(s) of the distribution will lie between $(\bar{y}_1, \bar{y}_2)$.

Bimodality

When $\bar{y}_1$ and $\bar{y}_2$ are wide enough apart, the distribution becomes bimodal. By setting $(\partial/\partial\theta) \log p(\theta \mid \mathbf{y}) = 0$, it can be verified that when $n_1 = n_2 = n$ (so that $\nu_1 = \nu_2 = \nu$) a sufficient and necessary condition for bimodality is

$$\left[d - 2\nu\left(\frac{s_1^2}{n} + \frac{s_2^2}{n}\right)\right]^3 - 27d\left(\frac{\nu s_1^2}{n} - \frac{\nu s_2^2}{n}\right)^2 > 0, \tag{9.2.7}$$

where

$$d = (\bar{y}_1 - \bar{y}_2)^2.$$

To consider the implications of this condition, suppose for the moment that the sample

variances are nearly equal, so that $s_1^2 \doteq s_2^2 = s^2$ and the second term on the left can be ignored, then (9.2.7) implies that the distribution will be bimodal only if

$$|\bar{y}_2 - \bar{y}_1| > 2\sqrt{v}\,(s/\sqrt{n}), \tag{9.2.8}$$

i.e., if

$$|\bar{y}_2 - \bar{y}_1| > 2\sqrt{v-2}\,(\text{s.d.}),$$

where

$$\text{s.d.} = \frac{s}{\sqrt{n}}\sqrt{\frac{v}{v-2}}$$

is the standard deviation of the individual posterior distributions $p(\theta \mid \mathbf{y}_1)$ and $p(\theta \mid \mathbf{y}_2)$. For example, when $v = 11$ this means that $\bar{y}_1$ and $\bar{y}_2$ must be at least 6 standard deviations apart before the distribution becomes bimodal.

The practical implication of this analysis is now considered. A question mentioned already and to be discussed in more detail in Section 9.2.3 is the appropriateness of the assumption of *compatibility* of the means. That is, the assumption that, to a sufficient approximation, $\theta_1 = \theta_2 = \theta$. One might at first think that bimodality in the posterior distribution of θ or its absence could provide a guide as to whether the compatibility assumption was reasonable. Expression (9.2.8) makes it clear that while bimodality would normally imply incompatibility, lack of bimodality certainly need not imply the converse.

To see why this is so, we have only to remember that for sufficiently large v the component t distributions will approach the Normal form. Thus, for large sample sizes the right-hand side of (9.2.4) approaches the product of two Normal densities. But the product of two Normal densities, whatever their separation, is proportional to another Normal density, which is of course unimodal. It is not surprising then that, relative to their spread, the separation of the two component posterior density functions necessary to induce bimodality increases without limit as v is increased.

Normal Approximation to $p(\theta \mid \mathbf{y})$

When the two sample means $\bar{y}_1$ and $\bar{y}_2$ are not widely discrepant, the combined posterior distribution (9.2.4) will be unimodal and located between the two individual t distributions. Although the unnormalized density function of θ can be easily determined, it does not seem possible to express the posterior probability integrals in a simple form. Thus, for example, exact evaluation of the probability content of a specific H.P.D. interval would require the use of numerical methods. However, we see that when both v_1 and v_2 are large the distribution in (9.2.4) is approximately

$$p(\theta \mid \mathbf{y}) \dot{\propto} \exp\left[-\tfrac{1}{2}n'(\theta - \bar{\theta})^2\right], \tag{9.2.9}$$

where

$$n' = \frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}, \qquad \bar{\theta} = \frac{1}{n'}\left(\frac{n_1}{s_1^2}\bar{y}_1 + \frac{n_2}{s_2^2}\bar{y}_2\right).$$

As shown by Fisher (1961b), the above expression can be made the leading term of an asymptotic expansion of the distribution in (9.2.4) in powers of ν_1^{-1} and ν_2^{-1}. The "correction" terms in the asymptotic series are polynomials in θ. In Appendix A9.1 the asymptotic procedure is given for the more general situation when $m = 2$ but $\boldsymbol{\theta}_c$ consists of more than one element. An alternative and simpler approximation to the distribution of $\boldsymbol{\theta}_c$ based upon multivariate t distribution will be developed in Section 9.3.

Returning to the distribution shown in Fig. 9.2.2 for the physical constant example, we see that the distribution is only slightly asymmetrical. Using the Normal approximation in (9.2.9), we have

$$\bar{\theta} = \left(\frac{11}{12.1} + \frac{13}{39.4}\right)^{-1} \left[\frac{11}{12.1}\,107.9 + \frac{13}{39.4}\,109.5\right] = 108.34,$$

which nearly coincides with the mode of the distribution in the figure. This itself, of course, does not imply that the distribution can be closely approximated by a Normal distribution without "correction" terms. A convenient device advocated by Barnard† for checking Normality is to graph the first derivative of the logarithm of the posterior density function against the parameter value. A unique characteristic of the Normal distribution is that this plot produces a straight line. A plot of the derivative in (9.2.6) in the range where the density is appreciable is shown by the broken line in Fig. 9.2.2. It is nearly linear, so that the posterior distribution $p(\theta \mid \mathbf{y})$ may be treated approximately as $N(108.34, 0.807)$.

9.2.3 Compatibility of the Means

Before information from m sets of measurements obtained by m different methods are combined, one should first consider whether the methods are measuring the same thing. In particular, it is helpful to compare individual posterior distributions, as was done for the numerical example in Fig. 9.2.1. More formally, questions of compatibility may be studied by considering the distribution of contrasts among the means derived from the joint "unconstrained" distribution of $\theta_1, \ldots, \theta_m$, when these parameters are *not* regarded as common. We illustrate with the case $m = 2$ and, as before, use an asterisk $*$ to denote unconstrained distributions, the constraint being that $\theta_1 = \theta_2$. As we have seen in Section 2.5, relative to the noninformative reference prior distribution

$$p^*(\theta_1, \theta_2, \sigma_1, \sigma_2) \propto (\sigma_1\sigma_2)^{-1}, \tag{9.2.10}$$

the joint posterior distribution of θ_1 and θ_2 is

$$p^*(\theta_1, \theta_2 \mid \mathbf{y}) \propto \left[1 + \frac{n_1(\theta_1 - \bar{y}_1)^2}{\nu_1 s_1^2}\right]^{-\frac{1}{2}(\nu_1+1)} \left[1 + \frac{n_2(\theta_2 - \bar{y}_2)^2}{\nu_2 s_2^2}\right]^{-\frac{1}{2}(\nu_2+1)},$$
$$-\infty < \theta_1 < \infty, \quad -\infty < \theta_2 < \infty. \tag{9.2.11}$$

† Personal communication.

Figure (9.2.3) shows the contours of (θ_1, θ_2) as well as various other aspects of the distribution.

Fig. 9.2.3 Various aspects of the unconstrained posterior distribution of (θ_1, θ_2).

To obtain the distribution of the difference in means we write

$$\delta = \theta_2 - \theta_1 \qquad \text{and} \qquad \theta_w = w_1\theta_1 + w_2\theta_2, \tag{9.2.12}$$

where

$$0 \leqslant w_1 \leqslant 1 \qquad \text{and} \qquad w_1 + w_2 = 1.$$

Then the joint distribution of δ and θ_w is

$$p^*(\delta, \theta_w \mid \mathbf{y}) \propto \left[1 + \frac{n_1(\theta_w - w_2\delta - \bar{y}_1)^2}{\nu_1 s_1^2}\right]^{-\frac{1}{2}(\nu_1+1)} \left[1 + \frac{n_2(\theta_w + w_1\delta - \bar{y}_2)^2}{\nu_2 s_2^2}\right]^{-\frac{1}{2}(\nu_2+1)}, \qquad -\infty < \delta < \infty, \quad -\infty < \theta_w < \infty. \tag{9.2.13}$$

The distribution can be written as the product

$$p^*(\delta, \theta_w \mid \mathbf{y}) = p^*(\delta \mid \mathbf{y})\, p^*(\theta_w \mid \delta, \mathbf{y}). \tag{9.2.14}$$

Whatever the choice of the weights w_1 and w_2, the density $p^*(\delta \mid \mathbf{y})$ will be proportional to the Behrens–Fisher distribution in (2.5.12) appropriate for making

inferences about the difference $\delta = \theta_2 - \theta_1$. The distribution $p(\theta \mid \mathbf{y})$ of the common mean (9.2.4) is $p^*(\theta_w \mid \delta = 0, \mathbf{y})$, the conditional distribution of θ_w given *that it is known* that $\delta = \theta_1 - \theta_2 = 0$.

To shed light on the question of whether a supposition that δ was close to zero was, or was not, tenable in the light of the data one could inspect the distribution $p^*(\delta \mid \mathbf{y})$, the appropriately scaled Behrens–Fisher distribution for the difference in means. In a case like that illustrated in Fig. 9.2.4 the relevance of the distribution $p(\theta \mid \mathbf{y})$ in (9.2.4) would be thrown in doubt. Such evidence might suggest that, instead of attempting to pool incompatible information, the investigator ought to direct his efforts to explain possible bias in the methods.

Fig. 9.2.4 Posterior distribution of the difference δ in relation to the value $\delta = 0$.

9.2.4 The Fiducial Distributions of θ Derived by Yates and by Fisher

Yates (1939), in his discussion of the problem of the weighted mean, considered the fiducial distribution of

$$\theta_w = w_1\theta_1 + w_2\theta_2$$

in the special case where

$$w_1 = \frac{n_1}{s_1^2} \bigg/ \left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}\right), \qquad w_2 = \frac{n_2}{s_2^2} \bigg/ \left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}\right) \tag{9.2.15}$$

and found that

$$\theta_w = \left(\frac{n_1}{s_1^2} + \frac{n_2}{s_2^2}\right)^{-1} \left(\frac{n_1}{s_1^2}\theta_1 + \frac{n_2}{s_2^2}\theta_2\right)$$

had a Behrens–Fisher distribution.

However, Yates' fiducial distribution is *not* the same as the fiducial distribution subsequently derived by Fisher. This, at first sight perplexing, situation is clarified by the Bayes analysis.

We have already discussed Fisher's result corresponding to (9.2.4). We now consider Yates' result from this same viewpoint. The marginal posterior distribution of

$\theta_w = w_1\theta_1 + w_2\theta_2$ is obtained in general by integrating out δ in (9.2.14),

$$p^*(\theta_w \mid \mathbf{y}) = \int_{-\infty}^{\infty} p^*(\theta_w \mid \delta, \mathbf{y})\, p^*(\delta \mid \mathbf{y})\, d\delta. \tag{9.2.16}$$

To see that this yields a density function numerically identical to Yates' fiducial distribution, in (9.2.11) write

$$t_1 = \frac{\theta_1 - \bar{y}_1}{s_1/\sqrt{n_1}}, \quad t_2 = \frac{\theta_2 - \bar{y}_2}{s_2/\sqrt{n_2}} \tag{9.2.17}$$

and

$$\tau = (\theta_w - \bar{y}_w) \Big/ \left(\frac{s_1^2}{n_1} w_1^2 + \frac{s_2^2}{n_2} w_2^2\right)^{1/2} = t_2 \cos\phi - t_1 \sin\phi \tag{9.2.18}$$

$$z = t_2 \sin\phi + t_1 \cos\phi$$

where

$$\bar{y}_w = w_1\bar{y}_1 + w_2\bar{y}_2 \quad \text{and} \quad \tan\phi = -\frac{(s_1/\sqrt{n_1})\, w_1}{(s_2/\sqrt{n_2})\, w_2},$$

Then, whatever weights w_1 and $w_2 = 1 - w_1$ are employed, the distribution of τ is

$$p(\tau \mid \nu_1, \nu_2, \phi)$$
$$\propto \int_{-\infty}^{\infty} \left[1 + \frac{(\tau \sin\phi - z\cos\phi)^2}{\nu_1}\right]^{-\frac{1}{2}(\nu_1+1)} \left[1 + \frac{(z\sin\phi + \tau\cos\phi)^2}{\nu_2}\right]^{-\frac{1}{2}(\nu_2+1)} dz,$$
$$-\infty < \tau < \infty, \tag{9.2.19}$$

which is a Behrens–Fisher distribution. Yates result is obtained by using the particular weighting of (9.2.15).

The basic difference between the two results $p^*(\theta_w \mid \delta = 0, \mathbf{y})$ and $p^*(\theta_w \mid \mathbf{y})$ (see Fig. 9.2.3) is now readily seen. Fisher's result for the distribution of the common mean, which corresponds with Jeffreys' Bayesian solution, is the *conditional* distribution of θ_w (or θ_1 or θ_2), *given that* $\theta_1 = \theta_2$. That is to say, given that $\delta = \theta_1 - \theta_2 = 0$. On the other hand, Yates' result corresponds not to any conditional distribution but to a *marginal* distribution, that of $\theta_w = w_1\theta_1 + w_2\theta_2$.

The relevance of Yates' result, is clearly open to question. For, *if* it can be assumed that $\theta_1 = \theta_2$ and hence that $\delta = \theta_1 - \theta_2 = 0$, one should employ the conditional distribution (9.2.4). Conversely, if θ_1 and θ_2 cannot be assumed equal (implying that one or both contain unknown *systematic* error), there seems to be no basis for the weighting

$$\frac{w_1}{w_2} = \frac{n_1/s_1^2}{n_2/s_2^2}$$

which is based only on *sampling* variation.

There will be cases where it cannot be plausibly asserted that $\theta_1 = \theta_2$ exactly, but nevertheless we may feel fairly sure that the difference $\delta = \theta_1 - \theta_2$ is small. In this case

the Fisher–Jeffreys solution may be used as an approximation. Specifically, if an *informative* prior distribution $p(\delta)$ is introduced which might for example be a Normal distribution centered at $\delta = 0$ with standard deviation σ_δ, then, on the same assumptions as before, the posterior distribution of θ_w will be

$$p^*(\theta_w \mid \mathbf{y}) \propto \int_{-\infty}^{\infty} p^*(\theta_w \mid \delta, \mathbf{y})\, p^*(\delta \mid \mathbf{y})\, p(\delta)\, d\delta. \tag{9.2.20}$$

Now (i) if we make σ_δ very large corresponding to the assertion that little is known about δ, we obtain the marginal distribution in (9.2.16), which however will depend on the *particular* weights w_1 and w_2 adopted; (ii) on the other hand, if we make σ_δ small so that $p(\delta)$ approaches a delta function at $\delta = 0$, then the distribution of θ_w tends to the Fisher–Jeffreys solution (9.2.4) *whatever the values of the weights* w_1 *and* w_2.

9.3 SOME PROPERTIES OF THE DISTRIBUTION OF THE COMMON PARAMETERS FOR TWO INDEPENDENT RESPONSES

Until further notice, we shall use the symbol $\boldsymbol{\theta}$ to denote the $k \times 1$ vector of regression coefficients common to all m responses, so that

$$\boldsymbol{\theta}' = \boldsymbol{\theta}_c' = (\theta_1, \ldots, \theta_k).$$

When there are two independednt responses, from (9.2.3) the posterior distribution of $\boldsymbol{\theta}$ is proportional to the product of two multivariate t distributions with different degrees of freedom,

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \prod_{i=1}^{2} \left[1 + \frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)' \mathbf{X}_i' \mathbf{X}_i (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)}{\nu_i s_i^2}\right]^{-\frac{1}{2}(\nu_i + k)}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{9.3.1}$$

where

$$\nu_i = n_i - k, \qquad s_i^2 = \frac{1}{\nu_i} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\theta}}_i)'(\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\theta}}_i),$$

and $\hat{\boldsymbol{\theta}}_i$ satisfies the normal equations $(\mathbf{X}_i'\mathbf{X}_i)\hat{\boldsymbol{\theta}}_i = \mathbf{X}_i'\mathbf{y}_i$. To obtain the marginal distribution of a subset of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_l' = (\theta_1, \ldots, \theta_l)$, directly it would be necessary to evaluate, for each value of $\boldsymbol{\theta}_l$, a $k - l$ dimensional integral. The difficulty may be avoided by writing the distribution in a different form which, in turn, leads to a simple approximation.

9.3.1 An Alternative Form for $p(\boldsymbol{\theta} \mid \mathbf{y})$

Writing $\sigma_1^2 = \sigma_{11}$ and $\sigma_2^2 = \sigma_{22}$ in (9.2.2) with $m = 2$ the posterior distribution of $(\sigma_1^2, \sigma_2^2, \boldsymbol{\theta})$ is

$$p(\sigma_1^2, \sigma_2^2, \boldsymbol{\theta} \mid \mathbf{y}) \propto (\sigma_1^2)^{-(\frac{1}{2}n_1 + 1)} (\sigma_2^2)^{-(\frac{1}{2}n_2 + 1)} \exp\left\{-\frac{1}{2}\left[\frac{\nu_1 s_1^2 + R_1(\boldsymbol{\theta})}{\sigma_1^2} + \frac{\nu_2 s_2^2 + R_2(\boldsymbol{\theta})}{\sigma_2^2}\right]\right\},$$
$$\sigma_1^2 > 0, \quad \sigma_2^2 > 0, \quad -\infty < \boldsymbol{\theta} < \infty, \tag{9.3.2}$$

where $$R_i(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)' \mathbf{X}_i'\mathbf{X}_i(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i), \qquad i = 1, 2.$$

If we integrate out (σ_1^2, σ_2^2) directly from (9.3.2), we obtain the form (9.3.1). Alternatively, we may make the transformation from $(\sigma_1^2, \sigma_2^2, \boldsymbol{\theta})$ to $(\sigma_1^2, w, \boldsymbol{\theta})$ where

$$w = \sigma_1^2/\sigma_2^2 \tag{9.3.3}$$

and integrate out σ_1^2 to obtain

$$p(\boldsymbol{\theta}, w \mid \mathbf{y}) \propto w^{\frac{1}{2}n_2 - 1} \{v_1 s_1^2 + R_1(\boldsymbol{\theta}) + w[v_2 s_2^2 + R_2(\boldsymbol{\theta})]\}^{-\frac{1}{2}(n_1+n_2)},$$
$$0 < w < \infty, \quad -\infty < \boldsymbol{\theta} < \infty. \tag{9.3.4}$$

This distribution can be written as the product

$$p(\boldsymbol{\theta}, w \mid \mathbf{y}) = p(\boldsymbol{\theta} \mid w, \mathbf{y})\, p(w \mid \mathbf{y}). \tag{9.3.5}$$

Assuming that at least one of the matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ is of full rank, the identity (A7.1.1) in Appendix A7.1 may be used to combine the two quadratic forms $R_1(\boldsymbol{\theta})$ and $wR_2(\boldsymbol{\theta})$ in (9.3.4) into

$$R_1(\boldsymbol{\theta}) + wR_2(\boldsymbol{\theta}) = R(\boldsymbol{\theta} \mid w) + S_c(w) \tag{9.3.6}$$

where

$$R(\boldsymbol{\theta} \mid w) = [\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)]'(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)[\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)],$$
$$S_c(w) = w(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)'\mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_2(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2),$$

and

$$\bar{\boldsymbol{\theta}}(w) = (\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1}(\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\theta}}_1 + w\mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\theta}}_2).$$

It follows that the conditional distribution of $\boldsymbol{\theta}$ given w is

$$p(\boldsymbol{\theta} \mid w, \mathbf{y}) \propto [(n_1 + n_2 - k)s^2(w) + R(\boldsymbol{\theta} \mid w)]^{-\frac{1}{2}(n_1+n_2)}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{9.3.7}$$

where

$$s^2(w) = \frac{1}{n_1 + n_2 - k}[v_1 s_1^2 + wv_2 s_2^2 + S_c(w)],$$

which is the k-dimensional multivariate t distribution

$$t_k[\bar{\boldsymbol{\theta}}(w),\ s^2(w)(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1},\ n_1 + n_2 - k].$$

To obtain the marginal distribution $p(w \mid \mathbf{y})$, we substitute (9.3.6) into (9.3.4) and integrate out $\boldsymbol{\theta}$ yielding

$$p(w \mid \mathbf{y}) = d\,|\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2|^{-1/2}\, w^{\frac{1}{2}n_2 - 1}\,[s^2(w)]^{-\frac{1}{2}(n_1+n_2-k)}, \qquad 0 < w < \infty, \tag{9.3.8}$$

where d is the appropriate normalizing constant such that

$$d^{-1} = \int_0^\infty |\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2|^{-1/2}\, w^{\frac{1}{2}n_2 - 1}\,[s^2(w)]^{-\frac{1}{2}(n_1+n_2-k)}\, dw.$$

An alternative form for the posterior distribution of $\boldsymbol{\theta}$ is therefore

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \int_0^\infty p(\boldsymbol{\theta} \mid w, \mathbf{y})\, p(w \mid \mathbf{y})\, \mathrm{d}w, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{9.3.9}$$

where $p(\boldsymbol{\theta} \mid w, \mathbf{y})$ and $p(w \mid \mathbf{y})$ are given in (9.3.7) and (9.3.8), respectively.

By comparing (9.3.9) with (9.3.1), it can be verified that the normalizing constant for the expression (9.3.1) is

$$\frac{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)(n_1 + n_2 - k)^{\frac{1}{2}(n_1+n_2-k)}}{[\Gamma(\frac{1}{2})]^k\,\Gamma\left(\dfrac{n_1 + n_2 - k}{2}\right)(v_1 s_1^2)^{\frac{1}{2}n_1}(v_2 s_2^2)^{\frac{1}{2}n_2}}\, d \tag{9.3.10}$$

where d^{-1} is the integral given in (9.3.8).

Using properties of the multivariate t distribution, and partitioning

$$\mathbf{C}(w) = (\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1} = \begin{array}{c} \begin{array}{cc} \quad l & \quad\quad p \end{array} \\ \left[\begin{array}{c:c} \mathbf{C}_{ll}(w) & \mathbf{C}_{lp}(w) \\ \hdashline \mathbf{C}_{pl}(w) & \mathbf{C}_{pp}(w) \end{array}\right] \end{array}\begin{array}{c} l \\ p \end{array} \qquad \bar{\boldsymbol{\theta}}(w) = \left[\begin{array}{c} \bar{\boldsymbol{\theta}}_l(w) \\ \hdashline \bar{\boldsymbol{\theta}}_p(w) \end{array}\right]\begin{array}{c} l \\ p \end{array},$$

$$p = k - l, \tag{9.3.11}$$

the marginal distribution of the subset $\boldsymbol{\theta}_l' = (\theta_1, \ldots, \theta_l)$, $(l \leqslant k)$, is

$$p(\boldsymbol{\theta}_l \mid \mathbf{y}) = \int_0^\infty p(\boldsymbol{\theta}_l \mid w, \mathbf{y})\, p(w \mid \mathbf{y})\, dw, \qquad -\infty < \boldsymbol{\theta}_l < \infty, \tag{9.3.12}$$

where $p(\boldsymbol{\theta}_l \mid w, \mathbf{y})$ is the $\mathbf{t}_l[\bar{\boldsymbol{\theta}}_l(w), s^2(w)\mathbf{C}_{ll}(w), n_1 + n_2 - k]$ distribution. A similar expression can be obtained for the remaining elements $\boldsymbol{\theta}_p' = (\theta_{l+1}, \ldots, \theta_k)$. Although it does not seem possible to express (9.3.12) in closed form, its evaluation only requires numerical integration of one-dimensional integrals instead of the $(k - l)$-dimensional integration implied by (9.3.1)—a very significant simplification.

9.3.2 Further Simplifications for the Distribution of $\boldsymbol{\theta}$

In evaluating the posterior distribution of $\boldsymbol{\theta}$ and its associated marginals for subsets of $\boldsymbol{\theta}$, if we were to employ directly the expressions in (9.3.9) and (9.3.12), it would be necessary to calculate the determinant and the inverse of the matrix $(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)$ for every w. We now show how this problem can be made more manageable.

The special case where $\mathbf{X}_1'\mathbf{X}_1$ *and* $\mathbf{X}_2'\mathbf{X}_2$ *are proportional*

Consider first the special case where

$$\mathbf{X}_2'\mathbf{X}_2 = a\mathbf{X}_1'\mathbf{X}_1. \tag{9.3.13}$$

The expressions in (9.3.6) then become

$$\begin{aligned} R(\boldsymbol{\theta} \mid w) &= (1 + aw)[\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)]' \mathbf{X}_1'\mathbf{X}_1 [\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)] \\ S_c(w) &= \frac{aw}{1 + aw} (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)' \mathbf{X}_1'\mathbf{X}_1 (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2) \\ \bar{\boldsymbol{\theta}}(w) &= \frac{1}{1 + aw} (\hat{\boldsymbol{\theta}}_1 + aw\hat{\boldsymbol{\theta}}_2), \end{aligned} \tag{9.3.14}$$

so that (a) the conditional distribution of $\boldsymbol{\theta}$ given w, $p(\boldsymbol{\theta} \mid w, \mathbf{y})$, is

$$p(\boldsymbol{\theta} \mid w, \mathbf{y}) \propto \{(n_1 + n_2 - k)s^2(w) + (1 + aw)[\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)]' \mathbf{X}_1'\mathbf{X}_1[\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)]\}^{-\frac{1}{2}(n_1+n_2)}, \quad -\infty < \boldsymbol{\theta} < \infty, \tag{9.3.15}$$

where

$$(n_1 + n_2 - k)s^2(w) = \nu_1 s_1^2 + w\nu_2 s_2^2 + \frac{aw}{1 + aw} (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)' \mathbf{X}_1'\mathbf{X}_1 (\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2),$$

(b) the conditional distribution for the subset $\boldsymbol{\theta}_l$, $p(\boldsymbol{\theta}_l \mid w, \mathbf{y})$, is

$$p(\boldsymbol{\theta}_l \mid w, \mathbf{y}) \propto \{(n_1 + n_2 - k)s^2(w) + (1 + aw)[\boldsymbol{\theta}_l - \bar{\boldsymbol{\theta}}_l(w)]' \mathbf{D}_{ll}^{-1} [\boldsymbol{\theta}_l - \bar{\boldsymbol{\theta}}_l(w)]\}^{-\frac{1}{2}(n_1+n_2-k+l)}, \quad -\infty < \boldsymbol{\theta}_l < \infty, \tag{9.3.16}$$

where

$$(\mathbf{X}_1'\mathbf{X}_1)^{-1} = \begin{array}{c} \begin{array}{cc} \quad l & \quad p \end{array} \\ \left[\begin{array}{c:c} \mathbf{D}_{ll} & \mathbf{D}_{lp} \\ \hdashline \mathbf{D}_{pl} & \mathbf{D}_{pp} \end{array}\right] \end{array} \begin{array}{c} p \\ l \end{array},$$

and (c) the marginal distribution of w, $p(w \mid \mathbf{y})$, reduces to

$$p(w \mid \mathbf{y}) = d|\mathbf{X}_1'\mathbf{X}_1|^{-1/2} (1 + aw)^{-k/2} w^{\frac{1}{2}n_2 - 1} [s^2(w)]^{-\frac{1}{2}(n_1+n_2-k)}, \quad 0 < w < \infty. \tag{9.3.17}$$

Employing (9.3.15)–(9.3.17), the posterior distribution of $\boldsymbol{\theta}$ in (9.3.9) and its associated marginal distributions for subsets of $\boldsymbol{\theta}$ in (9.3.12) can now be conveniently evaluated on a computer with little numerical complication.

The posterior distribution of $\boldsymbol{\theta}$ thus obtained is of very much the same form as the posterior distribution of $\boldsymbol{\phi}$ in (7.4.33) and (7.4.41) for the BIBD model. In both cases we are combining information about regression coefficients from two sources where the variance ratio is not assumed known. The main distinction lies in the fact that for the BIBD model we have the added constraint $w = \sigma_{be}^2/\sigma_e^2 > 1$ on the variance ratio while no such restriction exists here.

The General Case when $\mathbf{X}_1'\mathbf{X}_1$ *and* $\mathbf{X}_2'\mathbf{X}_2$ *are not Proportional*

In obtaining (9.3.6), we have assumed that at least one of the two matrices, $\mathbf{X}_1'\mathbf{X}_1$ and $\mathbf{X}_2'\mathbf{X}_2$, is positive definite. To be specific, suppose that $\mathbf{X}_1'\mathbf{X}_1$ is positive definite. Then there exists a non-singular $k \times k$ matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{X}_1'\mathbf{X}_1\mathbf{P} = \mathbf{I} \quad \text{and} \quad \mathbf{P}'\mathbf{X}_2'\mathbf{X}_2\mathbf{P} = \boldsymbol{\Lambda} \tag{9.3.18}$$

where $\boldsymbol{\Lambda} = \{\lambda_{jj}\}$ is a $k \times k$ diagonal matrix with non-negative diagonal elements. It can then be shown that, in the integral of (9.3.9) and (9.3.12),

a) the conditional distribution $p(\boldsymbol{\theta} \mid w, \mathbf{y})$ becomes

$$p(\boldsymbol{\theta} \mid w, \mathbf{y}) \propto \{(n_1 + n_2 - k)s^2(w) + [\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)]'(\mathbf{P}')^{-1}(\mathbf{I} + w\boldsymbol{\Lambda})\mathbf{P}^{-1}[\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(w)]\}^{-\frac{1}{2}(n_1+n_2)}, \quad -\infty < \boldsymbol{\theta} < \infty, \tag{9.3.19}$$

with

$$\bar{\boldsymbol{\theta}}(w) = \mathbf{P}(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}(\hat{\boldsymbol{\phi}}_1 + w\boldsymbol{\Lambda}\hat{\boldsymbol{\phi}}_2), \qquad \hat{\boldsymbol{\phi}}_i = \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}_i, \quad i = 1, 2$$

and

$$(n_1 + n_2 - k)s^2(w) = v_1 s_1^2 + w v_2 s_2^2 + \sum_{j=1}^{k} \frac{w\lambda_{jj}}{1 + w\lambda_{jj}} (\hat{\phi}_{1j} - \hat{\phi}_{2j})^2,$$

b) the conditional distribution $p(\boldsymbol{\theta}_l \mid w, \mathbf{y})$ of the subset $\boldsymbol{\theta}_l$ is

$$p(\boldsymbol{\theta}_l \mid w, \mathbf{y}) \propto \{(n_1 + n_2 - k)s^2(w) + [\boldsymbol{\theta}_l - \bar{\boldsymbol{\theta}}_l(w)]'\mathbf{C}_{ll}^{-1}(w)[\boldsymbol{\theta}_l - \bar{\boldsymbol{\theta}}_l(w)]\}^{-\frac{1}{2}(n_1+n_2-k+l)}, \quad -\infty < \boldsymbol{\theta}_l < \infty, \tag{9.3.20}$$

where

$$\mathbf{C}_{ll}(w) = \mathbf{P}_1(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}\mathbf{P}_1'$$

and

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \hdashline \mathbf{P}_2 \end{bmatrix}\begin{matrix} l \\ p \end{matrix},$$

and (c) the marginal distribution $p(w \mid \mathbf{y})$ reduces to

$$p(w \mid \mathbf{y}) = d\,|\mathbf{X}_1'\mathbf{X}_1|^{-1/2} \left[\prod_{j=1}^{k} (1 + w\lambda_{jj})\right]^{-1/2} w^{\frac{1}{2}n_2 - 1}\,[s^2(w)]^{-\frac{1}{2}(n_1+n_2-k)}, \quad 0 < w < \infty. \tag{9.3.21}$$

To obtain these results, we may write

$$\mathbf{X}_1'\mathbf{X}_1 = (\mathbf{P}')^{-1}\mathbf{P}^{-1} \quad \text{and} \quad \mathbf{X}_2'\mathbf{X}_2 = (\mathbf{P}')^{-1}\boldsymbol{\Lambda}\mathbf{P}^{-1} \tag{9.3.22}$$

so that

$$(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1} = \mathbf{P}(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}\mathbf{P}'. \tag{9.3.23}$$

It follows that for the expressions of $\bar{\boldsymbol{\theta}}(w)$ and $S_c(w)$ in (9.3.6) we now have

$$\begin{aligned}\bar{\boldsymbol{\theta}}(w) &= \mathbf{P}(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}(\mathbf{P}^{-1}\hat{\boldsymbol{\theta}}_1 + w\boldsymbol{\Lambda}\mathbf{P}^{-1}\hat{\boldsymbol{\theta}}_2)\\ &= \mathbf{P}(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}(\hat{\boldsymbol{\phi}}_1 + w\boldsymbol{\Lambda}\hat{\boldsymbol{\phi}}_2)\end{aligned} \tag{9.3.24}$$

with

$$\hat{\boldsymbol{\phi}}_i = \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}_i, \qquad i = 1, 2$$

and

$$\begin{aligned}S_c(w) &= w(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)'(\mathbf{P}')^{-1}\mathbf{P}^{-1}\mathbf{P}(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}\mathbf{P}'(\mathbf{P}')^{-1}\boldsymbol{\Lambda}\mathbf{P}^{-1}(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)\\ &= w(\hat{\boldsymbol{\phi}}_1 - \hat{\boldsymbol{\phi}}_2)'(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}\boldsymbol{\Lambda}(\hat{\boldsymbol{\phi}}_1 - \hat{\boldsymbol{\phi}}_2)\\ &= \sum_{j=1}^{k}\left(\frac{w\lambda_{jj}}{1 + w\lambda_{jj}}\right)(\hat{\phi}_{1j} - \hat{\phi}_{2j})^2\end{aligned} \tag{9.3.25}$$

Further, the determinant $|\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2|$ is

$$\begin{aligned}|\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2| &= |(\mathbf{P}')^{-1}\mathbf{P}^{-1} + w(\mathbf{P}')^{-1}\boldsymbol{\Lambda}\mathbf{P}^{-1}| = |\mathbf{P}'|^{-1}|\mathbf{P}^{-1}|\,|\mathbf{I} + w\boldsymbol{\Lambda}|\\ &= |\mathbf{X}_1'\mathbf{X}_1|\prod_{j=1}^{k}(1 + w\lambda_{jj})\end{aligned} \tag{9.3.26}$$

Substituting (9.3.24)—(9.3.26) into (9.3.9) and (9.3.12), we obtain the desired results.

In computing the posterior distributions in (9.3.9) and (9.3.12), standard numerical methods can be used to obtain first the matrices $\mathbf{P}$ and $\boldsymbol{\Lambda}$. The distribution of $\boldsymbol{\theta}$ and its marginals of subsets of the elements of $\boldsymbol{\theta}$ can then be conveniently calculated using (9.3.19)–(9.3.21) with the appropriate normalizing constants inserted. All that is required is numerical evaluation of one-dimensional integrals involving simple functions. With the availability of a computer, this presents no problem.

9.3.3 Approximations to the Distribution of $\boldsymbol{\theta}$

We have shown in (7.4.50) that the distribution of $\boldsymbol{\phi}$ for the BIBD model can be closely approximated by setting

$$p(\boldsymbol{\phi} \mid \mathbf{y}) \doteq p(\boldsymbol{\phi} \mid \hat{w}, \mathbf{y})$$

where $\hat{w}$ is the mode of the posterior distribution of $\log w$. A similar argument can be applied to the distribution of $\boldsymbol{\theta}$ in (9.3.9). We can write

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \underset{\log w}{E}\, p(\boldsymbol{\theta} \mid w, \mathbf{y}) \doteq p(\boldsymbol{\theta} \mid \hat{w}, \mathbf{y}) \tag{9.3.27}$$

where $p(\boldsymbol{\theta} \mid \hat{w}, \mathbf{y})$ is obtained by inserting $\hat{w}$ for w in (9.3.19) and $\hat{w}$ is the mode of the density

$$p(\log w \mid \mathbf{y}) \propto w^{n_2/2}\left[\prod_{j=1}^{k}(1 + w\lambda_{jj})\right]^{-1/2}[s^2(w)]^{-\frac{1}{2}(n_1+n_2-k)},$$

$$-\infty < \log w < \infty, \tag{9.3.28}$$

which is derived from (9.3.21) by inserting the Jacobian $[(d/dw)\log w]^{-1} = w$. Upon taking logarithms of (9.3.28) and differentiating, it can be verified that $\hat{w}$ is the appropriate root of the equation

$$h(w) = 0, \tag{9.3.29}$$

$$h(w) = n_2 w^{-1} - \sum_{j=1}^{k} \lambda_{jj}(1 + w\lambda_{jj})^{-1} - [s^2(w)]^{-1}\left[v_2 s_2^2 + \sum_{j=1}^{k} \lambda_{jj}(1 + \lambda_{jj} w)^{-2}(\hat{\phi}_{1j} - \hat{\phi}_{2j})^2\right],$$

which maximizes (9.3.28).

To obtain the root $\hat{w}$, it seems most convenient to employ the standard Newton–Raphson iteration procedure. That is, with an initial guessed value w_0, expand

$$h(w) \doteq h(w_0) + h'(w_0)(w_1 - w_0) = 0 \tag{9.3.30}$$

where

$$h'(w) = -n_2 w^{-2} + \sum_{j=1}^{k}\left[\frac{\lambda_{jj}}{1 + w\lambda_{jj}}\right]^2 + 2[s^2(w)]^{-1}\left[\sum_{j=1}^{k} \lambda_{jj}^2(1 + w\lambda_{jj})^{-3}(\hat{\phi}_{1j} - \hat{\phi}_{2j})^2\right] + (n_1 + n_2 - k)[s^2(w)]^{-2}\left[v_2 s_2^2 + \sum_{j=1}^{k} \lambda_{jj}(1 + w\lambda_{jj})^{-2}(\hat{\phi}_{1j} - \hat{\phi}_{2j})^2\right]^2.$$

Thus, we find a new guessed value w_1 such that

$$w_1 = w_0 - \frac{h(w_0)}{h'(w_0)} \tag{9.3.31}$$

which can now be used as the next guessed value, and the process repeated until convergence occurs. A convenient choice for the initial value is $w_0 = s_1^2/s_2^2$.

To this degree of approximation, then, $\boldsymbol{\theta}$ is distributed as the k dimensional multivariate t distribution $t_k[\bar{\boldsymbol{\theta}}(\hat{w}), s^2(\hat{w})(\mathbf{X}_1'\mathbf{X}_1 + \hat{w}\mathbf{X}_2'\mathbf{X}_2)^{-1}, n_1 + n_2 - k]$ from which corresponding approximations to marginal distributions of subsets of $\boldsymbol{\theta}$ can also be obtained. The accuracy of this approximation will be illustrated by an example in the next section.

An alternative approximation employing an asymptotic expansion of the distribution in (9.3.1) in powers of (v_1^{-1}, v_2^{-1}) is given in Appendix A.9.1.

9.3.4 An Econometric Example

For illustration, we analyse a simple econometric investment model with annual time series data, 1935–54, taken from Boot and De Witt (1960), relating to two

large corportaions, General Electric and Westinghouse. In this model, price deflated gross investment is assumed to be a linear function of expected profitability and beginning of year real capital stock. Following Grunfeld (1958), the value of outstanding shares at the beginning of the year is taken as a measure of a firm's expected profitability. The two investment relations are

$$\begin{aligned} y_{u1} &= \theta_{01} + \theta_1 x_{u11} + \theta_2 x_{u21} + \varepsilon_{u1}, \\ y_{u2} &= \theta_{02} + \theta_1 x_{u12} + \theta_2 x_{u22} + \varepsilon_{u2}, \end{aligned} \tag{9.3.32}$$

where u denotes the value of a variable in year u, $u = 1, 2, \ldots, 20$, and

Variable	General Electric	Westinghouse
Annual real gross investment	y_{u1}	y_{u2}
Value of shares at beginning of year	x_{u11}	x_{u12}
Real capital stock at beginning of year	x_{u21}	x_{u22}
Error term	ε_{u1}	ε_{u2}

The parameters θ_1 and θ_2 in (9.3.32) are taken to be the same for the two firms; however, θ_{01} and θ_{02} are assumed to be different to allow for certain possible differences in the investment behavior of the two firms. Further, ε_{u1} and ε_{u2} are assumed to be independently and Normally distributed for all u as $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$, respectively.

We may write the model in (9.3.32) in the form

$$\mathbf{y}_i = \mathbf{T}_i \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i \tag{9.3.33}$$

where

$$\boldsymbol{\gamma}_i = \begin{bmatrix} \theta_{0i} \\ \hdashline \boldsymbol{\theta} \end{bmatrix}, \qquad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix},$$

$$\mathbf{T}_i = \begin{bmatrix} 1 & x_{11i} & x_{12i} \\ \vdots & \vdots & \vdots \\ 1 & x_{u1i} & x_{u2i} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1i} & x_{n2i} \end{bmatrix}, \qquad \begin{aligned} \mathbf{y}_i' &= (y_{1i}, \ldots, y_{ui}, \ldots, y_{ni}) \\ \boldsymbol{\varepsilon}_i' &= (\varepsilon_{1i}, \ldots, \varepsilon_{ui}, \ldots, \varepsilon_{ni}), \\ i &= 1, 2 \quad \text{and} \quad n = 20. \end{aligned}$$

Thus, the two vectors of responses $(\mathbf{y}_1, \mathbf{y}_2)$ are independently Normally distributed, and contain *some* (but not all) common parameters $\boldsymbol{\theta}$. On the assumption that $(\theta_{01}, \theta_{02}, \boldsymbol{\theta})$ and $\log \sigma_1$ and $\log \sigma_2$ are locally uniform and independent *a priori*, the posterior distribution of the parameters is

$$p(\theta_{01}, \theta_{02}, \boldsymbol{\theta}, \sigma_1^2, \sigma_2^2 \mid \mathbf{y}) \propto \prod_{i=1}^{2} (\sigma_i^2)^{-(\frac{1}{2}n+1)} \exp\left\{ -\frac{1}{2\sigma_i^2} [\nu s_i^2 + (\boldsymbol{\gamma}_i - \hat{\boldsymbol{\gamma}}_i)' \mathbf{T}_i' \mathbf{T}_i (\boldsymbol{\gamma}_i - \hat{\boldsymbol{\gamma}}_i)] \right\},$$
$$\sigma_i^2 > 0, \quad -\infty < \theta_{0i} < \infty, \quad -\infty < \boldsymbol{\theta} < \infty, \tag{9.3.34}$$

where
$$\hat{\gamma}_i = \begin{bmatrix} \hat{\theta}_{0i} \\ \hat{\boldsymbol{\theta}}_i \end{bmatrix} = (\mathbf{T}_i'\mathbf{T}_i)^{-1}\mathbf{T}_i'\mathbf{y}_i, \qquad \nu = n - 3$$

and
$$\nu s_i^2 = (\mathbf{y}_i - \mathbf{T}_i\hat{\gamma}_i)'(\mathbf{y}_i - \mathbf{T}_i\hat{\gamma}_i)$$

Integrating out (θ_{01} and θ_{02}), we obtain

$$p(\boldsymbol{\theta}, \sigma_1^2, \sigma_2^2 \mid \mathbf{y}) \propto \prod_{i=1}^{2} (\sigma_i^2)^{-[\frac{1}{2}(n-1)+1]} \exp\left\{-\frac{1}{2\sigma_i^2}[\nu s_i^2 + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)'\mathbf{X}_i'\mathbf{X}_i(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)]\right\},$$
$$\sigma_i^2 > 0, \quad -\infty < \boldsymbol{\theta} < \infty, \qquad (9.3.35)$$

where
$$\mathbf{X}_i = \begin{bmatrix} x_{11i} - \bar{x}_{1i} & x_{12i} - \bar{x}_{2i} \\ \vdots & \vdots \\ x_{u1i} - \bar{x}_{1i} & x_{u2i} - \bar{x}_{2i} \\ \vdots & \vdots \\ x_{n1i} - \bar{x}_{1i} & x_{n2i} - \bar{x}_{2i} \end{bmatrix}, \qquad \bar{x}_{ji} = \frac{1}{n}\sum_{u=1}^{n} x_{uji}, \qquad j = 1, 2.$$

This distribution is of exactly the same form as that in (9.3.2) with $n_i = n - 1$, $\nu_i = \nu$ and $k = 2$. Eliminating σ_1^2 and σ_2^2 by integration, the distribution of $\boldsymbol{\theta}$ then takes the form of the product of two bivariate t distributions as shown in (9.3.1) or alternatively can be expressed in the integral form (9.3.9). Numerical values for sample quantities needed in our subsequent analysis are given below.

General Electric	*Westinghouse*
$\mathbf{X}_1'\mathbf{X}_1 = 10^6\begin{bmatrix} 3.254 & 0.233 \\ 0.233 & 1.193 \end{bmatrix}$	$\mathbf{X}_2'\mathbf{X}_2 = 10^6\begin{bmatrix} 0.940 & 0.195 \\ 0.195 & 0.074 \end{bmatrix}$
$\hat{\boldsymbol{\theta}}_1 = \begin{pmatrix} \hat{\theta}_{11} \\ \hat{\theta}_{21} \end{pmatrix} = \begin{pmatrix} 0.02655 \\ 0.15170 \end{pmatrix}$	$\hat{\boldsymbol{\theta}}_2 = \begin{pmatrix} \hat{\theta}_{12} \\ \hat{\theta}_{22} \end{pmatrix} = \begin{pmatrix} 0.05289 \\ 0.09241 \end{pmatrix}$
$s_1^2 = 0.777 \times 10^3$	$s_2^2 = 0.104 \times 10^3$
$\nu_1 = 17$	$\nu_2 = 17$
$n_1 = 19$	$n_2 = 19$

A plot of the contours of the distribution of $\boldsymbol{\theta}$ is shown in Fig. 9.3.1. The three contours shown by the solid curves are drawn such that

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = c\,p(\hat{\boldsymbol{\theta}} \mid \mathbf{y}) \qquad \text{for} \qquad c = 0.50, 0.25, 0.05$$

where $\hat{\boldsymbol{\theta}} \doteq (0.037, 0.145)$ is the mode of the distribution. In other words, they are boundaries of the H.P.D. regions containing approximately 50, 75 and 95 per cent of the probability mass, respectively. Also shown in the same figure are the 95

per cent contours (labelled by G.E. and W.H.) of the two factors of the distribution of $\boldsymbol{\theta}$ together with the corresponding centers $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$. As expected, the distribution of $\boldsymbol{\theta}$ is located "between" the two individual bivariate t distributions, and the spread of the combined distribution is smaller than either of its components. Further, the influence of the first factor—G.E.—is seen to be greater in determining the overall distribution because

$$s_1^4\,|\mathbf{X}_1'\mathbf{X}_1|^{-1} < s_2^4|\mathbf{X}_2'\mathbf{X}_2|^{-1}.$$

The solid contours in Fig. 9.3.1 are nearly elliptical suggesting that the distribution of $\boldsymbol{\theta}$ might well be approximated by the bivariate t distribution in (9.3.27). Since comparison of exact and approximate contours is rather difficult to make, the exact and approximate marginal distributions of θ_1 and θ_2 will be compared.

Fig. 9.3.1 Contours of the posterior distribution of $\boldsymbol{\theta}$ and its component factors: the investment data.

The solid curves in Fig. 9.3.2a and b show, respectively, the distributions of θ_1 and θ_2 calculated from (9.3.12). The two factors in the integrand were

Fig. 9.3.2 Comparison of the exact and the approximate posterior distributions of θ_1 and θ_2: the investment data.

determined from the expressions in (9.3.20) and (9.3.21). The matrices $\mathbf{P}$ and $\mathbf{\Lambda}$ in (9.3.18) and the vectors $\hat{\boldsymbol{\phi}}_i$ in (9.3.19) were found to be

$$\mathbf{P} = 10^{-3}\begin{bmatrix} -0.19814 & 0.52196 \\ 0.89469 & 0.22306 \end{bmatrix} \qquad \mathbf{\Lambda} = \begin{bmatrix} 0.02703 & 0 \\ 0 & 0.30513 \end{bmatrix}$$

$$\hat{\boldsymbol{\phi}}_1 = \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}_1 = 10^3\begin{bmatrix} 0.14331 \\ 0.10527 \end{bmatrix}, \qquad \hat{\boldsymbol{\phi}}_2 = \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}_2 = 10^3\begin{bmatrix} 0.07128 \\ 0.12839 \end{bmatrix}.$$

The approximating distributions shown by the broken curves in Figs. 9.3.2 and 9.3.3 were determined using the result in (9.3.27). The root $\hat{w}$ of the equation $h(w) = 0$ in (9.3.29) was calculated by employing the iterative process in (9.3.30) and (9.3.31). Starting with the initial value $w_0 = s_1^2/s_2^2 \doteq 7.47$, we found after two iterations, $\hat{w} \doteq 7.349$. Thus, $\boldsymbol{\theta}$ is approximately distributed as $t_2[\bar{\boldsymbol{\theta}}(\hat{w}), s^2(\hat{w})\mathbf{P}(\mathbf{I} + \hat{w}\mathbf{\Lambda})^{-1}\mathbf{P}', 36]$ where

$$\bar{\boldsymbol{\theta}}(\hat{w}) = \begin{bmatrix} 0.03726 \\ 0.14460 \end{bmatrix} \quad \text{and} \quad s^2(\hat{w})\mathbf{P}(\mathbf{I} + \hat{w}\mathbf{\Lambda})^{-1}\mathbf{P}' = 10^{-4}\begin{bmatrix} 0.8898 & -0.8535 \\ -0.8535 & 5.2060 \end{bmatrix}.$$

It follows that to this degree of approximation, the quantities

$$\frac{\theta_1 - 0.03726}{0.00943} \quad \text{and} \quad \frac{\theta_2 - 0.1446}{0.0228}$$

are individually distributed as $t(0, 1, 36)$ from which the broken curves were drawn. For this example, the agreement between the exact distribution obtained from numerical integration, and the approximating distributions are sufficiently close for most practical purposes. This, together with the near elliptical shape of the contours in Fig. 9.3.1, suggests that the approximation in (9.3.27) will be useful for higher dimensional distributions.

9.4 INFERENCES ABOUT COMMON PARAMETERS FOR THE GENERAL LINEAR MODEL WITH A COMMON DERIVATIVE MATRIX

Returning to the linear multivariate model in (9.1.2), we now consider the case where not only the parameters but also the derivative matrices are assumed common, i.e., $\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_m = \boldsymbol{\theta}_c$ and $\mathbf{X}_1 = \ldots = \mathbf{X}_m = \mathbf{X}$. The elements of the multivariate observation vectors $\boldsymbol{\varepsilon}_{(u)} = (\varepsilon_{u1}, \ldots, \varepsilon_{um})'$, $u = 1, \ldots, n$ are, however, not assumed independent.

For example, suppose observations were made by $m = 3$ different observers of a temperature which was linearly increasing with time. An appropriate model for the observations y_{u1}, y_{u2}, y_{u3} made at time t_u might be

$$\begin{aligned} y_{u1} &= \theta_1 + \theta_2 t_u + \varepsilon_{u1} \\ y_{u2} &= \theta_1 + \theta_2 t_u + \varepsilon_{u2} \\ y_{u3} &= \theta_1 + \theta_2 t_u + \varepsilon_{u3} \end{aligned} \tag{9.4.1}$$

If observations were made at n distinct times, we should have a model of the common parameter, common derivative form with

$$\boldsymbol{\theta}_1' = \boldsymbol{\theta}_2' = \boldsymbol{\theta}_3' = \boldsymbol{\theta}_c' = (\theta_1, \theta_2) \qquad \text{and} \qquad \mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3 = \mathbf{X},$$

where $\mathbf{X}$ is the $n \times 2$ matrix whose uth row is $(1, t_u)$, $u = 1, \ldots, n$.

In general, the model may be written in the form of equation (8.4.1) with the $k \times m$ matrix $\boldsymbol{\theta}$ of (8.4.2) having common parameters within every row so that

$$\boldsymbol{\theta} = \boldsymbol{\theta}_c \mathbf{1}_m', \qquad \boldsymbol{\theta}_c' = (\theta_1, \theta_2, \ldots, \theta_k). \tag{9.4.2}$$

From (8.4.8), it is readily shown that the posterior distribution of $\boldsymbol{\theta}_c$ is

$$p(\boldsymbol{\theta}_c \mid \mathbf{y}) \propto [1 + d(\boldsymbol{\theta}_c - \bar{\boldsymbol{\theta}}_c)'\mathbf{V}^{-1}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}_c)]^{-\frac{1}{2}n}, \qquad -\infty < \boldsymbol{\theta}_c < \infty, \tag{9.4.3}$$

where

$$d = \mathbf{1}_m'\mathbf{A}^{-1}\mathbf{1}_m, \qquad \bar{\boldsymbol{\theta}}_c = \frac{1}{d}\hat{\boldsymbol{\theta}}\mathbf{A}^{-1}\mathbf{1}_m,$$

and

$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1} + \hat{\boldsymbol{\theta}}\left[\mathbf{A}^{-1} - \frac{1}{d}\mathbf{A}^{-1}\mathbf{1}_m\mathbf{1}_m'\mathbf{A}^{-1}\right]\hat{\boldsymbol{\theta}}',$$

which is the k-dimensional multivariate t distribution

$$t_k\left[\bar{\boldsymbol{\theta}}_c, \frac{1}{(n-k)d}\mathbf{V}, n-k\right].$$

To show this, from (8.4.8) the posterior distribution of $\boldsymbol{\theta}_c$ is

$$p(\boldsymbol{\theta}_c \mid \mathbf{y}) \propto |\mathbf{A} + (\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})'\,\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})|^{-\frac{1}{2}n}, \qquad -\infty < \boldsymbol{\theta}_c < \infty, \tag{9.4.4}$$

where it is to be remembered that $\hat{\boldsymbol{\theta}}$ is a $k \times m$ matrix. Now, using (8.4.11)

$$|\mathbf{A} + (\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})| = |\mathbf{A}|\,|\mathbf{X}'\mathbf{X}|\,|(\mathbf{X}'\mathbf{X})^{-1} + (\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})\mathbf{A}^{-1}(\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})'|. \tag{9.4.5}$$

We can write

$$(\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})\mathbf{A}^{-1}(\boldsymbol{\theta}_c\mathbf{1}_m' - \hat{\boldsymbol{\theta}})' = d(\boldsymbol{\theta}_c - \bar{\boldsymbol{\theta}}_c)(\boldsymbol{\theta}_c - \bar{\boldsymbol{\theta}}_c)' + \hat{\boldsymbol{\theta}}\mathbf{B}\hat{\boldsymbol{\theta}}', \tag{9.4.6}$$

where

$$\mathbf{B} = \mathbf{A}^{-1} - \frac{1}{d}\mathbf{A}^{-1}\mathbf{1}_m\mathbf{1}_m'\mathbf{A}^{-1}.$$

Making use of (9.4.6) and (8.4.11), we see that the third factor on the right-hand side of (9.4.5) is proportional to

$$[1 + d(\boldsymbol{\theta}_c - \bar{\boldsymbol{\theta}}_c)'\,\mathbf{V}^{-1}(\boldsymbol{\theta}_c - \bar{\boldsymbol{\theta}}_c)], \tag{9.4.7}$$

where

$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1} + \hat{\boldsymbol{\theta}}\,\mathbf{B}\hat{\boldsymbol{\theta}}',$$

and the desired result follows.

The special case for which $k = 1$ and $\mathbf{X}$ a $n \times 1$ column of ones was discussed by Geisser (1965b). The result for the situation when we have common parameters within rows of a block submatrix of $\boldsymbol{\theta}$ is similar to (9.4.3) and is left to the reader.

9.5 GENERAL LINEAR MODEL WITH COMMON PARAMETERS

We now discuss certain aspects of the problem of making inferences about the common parameters $\boldsymbol{\theta}_c$ in the general situation when the derivative matrices $\mathbf{X}_i$ are not common and the responses are correlated. For simplicity in notation, we shall again suppress the subscript c and denote $\boldsymbol{\theta} = \boldsymbol{\theta}_c$. By setting $\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_m = \boldsymbol{\theta}$ in (8.2.23), the posterior distribution of the common parameters $\boldsymbol{\theta}$ is then

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto |\mathbf{S}(\boldsymbol{\theta})|^{-\frac{1}{2}n}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{9.5.1}$$

where

$$\left.\begin{aligned} \mathbf{S}(\boldsymbol{\theta}) &= \{S_{ij}(\boldsymbol{\theta})\} \\ S_{ij}(\boldsymbol{\theta}) &= (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta})'(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\theta}) \\ \boldsymbol{\theta}' &= (\theta_1, \ldots, \theta_k) \end{aligned}\right\} \qquad i = 1, \ldots, m, \quad j = 1, \ldots, m.$$

Since $\mathbf{X}_i \neq \mathbf{X}_j$, we cannot in general express $S_{ij}(\boldsymbol{\theta})$ in (9.5.1) in the form given in (8.4.4), and thus it is no longer possible to reduce the distribution of $\boldsymbol{\theta}$ to the much simpler form in (9.4.3). When the vector $\boldsymbol{\theta}$ consists of only one or two elements, whether or not the model is linear in the parameters the distribution can be plotted and the contributions from the individual vector of responses $\mathbf{y}_i$ assessed, as was done for example in Section 8.2.6.

As always, however, as the number of parameters increases, the situation becomes more and more complicated. Further, to obtain the marginal distribution of a subset, say $\boldsymbol{\theta}_l = (\theta_1, \ldots, \theta_l)'$, of $\boldsymbol{\theta}$, it is necessary to evaluate a $k - l$ dimensional integral for every value of $\boldsymbol{\theta}_l$. In what follows we shall discuss the important special case $m = 2$ for the linear model and develop an approximation procedure for the distribution of $\boldsymbol{\theta}$.

9.5.1 The Case $m = 2$

When $m = 2$, the distribution in (9.5.1) can be written

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto [S_{11}(\boldsymbol{\theta})]^{-\frac{1}{2}n} \left[S_{22}(\boldsymbol{\theta}) - \frac{S_{12}^2(\boldsymbol{\theta})}{S_{11}(\boldsymbol{\theta})} \right]^{-\frac{1}{2}n}, \qquad -\infty < \boldsymbol{\theta} < \infty. \tag{9.5.2}$$

The first factor $[S_{11}(\boldsymbol{\theta})]^{-\frac{1}{2}n}$ is clearly in the form of the multivariate t distribution $t[\hat{\boldsymbol{\theta}}_1, s_1^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}, v]$ where, assuming $\mathbf{X}_1$ is of full rank,

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1, \\ s_1^2 &= (\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1)'(\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1)/v \end{aligned} \tag{9.5.3}$$

and $v = n - k$. This factor can be thought of as representing the contribution from the first response vector $\mathbf{y}_1$. The nature of the second factor, $[S_{22}(\boldsymbol{\theta}) - S_{12}^2(\boldsymbol{\theta})/S_{11}(\boldsymbol{\theta})]^{-\frac{1}{2}n}$, which provides the "extra information" from $\mathbf{y}_2$, is not easily recognizable. We now proceed to develop an alternative representation of the distribution of $\boldsymbol{\theta}$.

It will be recalled that if (y_1, y_2) are jointly Normal $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the distribution can be written

$$p(y_1, y_2 \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = p(y_1 \mid \mu_1, \sigma_{11})\, p(y_2 \mid y_1, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{9.5.4}$$

where

$$p(y_1 \mid \mu_1, \sigma_{11}) = \frac{1}{\sqrt{2\pi\sigma_{11}}} \exp\left[-\frac{(y_1 - \mu_1)^2}{2\sigma_{11}} \right], \qquad -\infty < y_1 < \infty,$$

$$p(y_2 \mid y_1, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{2\pi\sigma_{22\cdot 1}}} \exp\left[-\frac{(y_2 - \mu_{2\cdot 1})^2}{2\sigma_{22\cdot 1}} \right], \qquad -\infty < y_2 < \infty,$$

$$\sigma_{22\cdot 1} = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}, \qquad \mu_{2\cdot 1} = \mu_2 + \beta(y_1 - \mu_1), \qquad \beta = \frac{\sigma_{12}}{\sigma_{11}}.$$

For $m = 2$ and with common parameters $\boldsymbol{\theta}$, the likelihood function of $\boldsymbol{\theta}$ and $(\sigma_{11}, \sigma_{22\cdot1}, \beta)$ is

$$l(\boldsymbol{\theta}, \sigma_{11}, \sigma_{22\cdot1}, \beta \mid \mathbf{y}) \propto \sigma_{11}^{-\frac{1}{2}n} \sigma_{22\cdot1}^{-\frac{1}{2}n} \exp\left\{-\frac{1}{2\sigma_{11}}(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})\right.$$

$$\left. - \frac{1}{2\sigma_{22\cdot1}}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta} - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})]'[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta} - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})]\right\}. \tag{9.5.5}$$

From the noninformative reference prior distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ in (8.2.8) and (8.2.14), we make the transformation from $(\sigma_{11}, \sigma_{22}, \sigma_{12})$ to $(\sigma_{11}, \sigma_{22\cdot1}, \beta)$ to obtain the prior

$$p(\boldsymbol{\theta}, \sigma_{11}, \sigma_{22\cdot1}, \beta) \propto \sigma_{11}^{-1/2} \sigma_{22\cdot1}^{-3/2}. \tag{9.5.6}$$

Consequently, the posterior distribution of $(\boldsymbol{\theta}, \sigma_{11}, \sigma_{22\cdot1}, \beta)$ is

$$p(\boldsymbol{\theta}, \sigma_{11}, \sigma_{22\cdot1}, \beta \mid \mathbf{y}) \propto \sigma_{11}^{-[\frac{1}{2}(n-1)+1]} \sigma_{22\cdot1}^{-[\frac{1}{2}(n+1)+1]} \exp\left[-\frac{S_{11}(\boldsymbol{\theta})}{2\sigma_{11}} - \frac{S_{22\cdot1}(\boldsymbol{\theta}, \beta)}{2\sigma_{22\cdot1}}\right],$$

$$\sigma_{11} > 0,\ \sigma_{22\cdot1} > 0, \quad -\infty < \beta < \infty, \quad -\infty < \boldsymbol{\theta} < \infty, \tag{9.5.7}$$

where

$$S_{11}(\boldsymbol{\theta}) = (\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})$$

$$S_{22\cdot1}(\boldsymbol{\theta}, \beta) = [\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta} - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})]'[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta} - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})].$$

If we were to integrate out $(\sigma_{11}, \sigma_{22\cdot1}, \beta)$ from (9.5.7), we would of course obtain the form given in (9.5.2). Alternatively, we may make the transformation from $(\sigma_{11}, \sigma_{22\cdot1})$ to (σ_{11}, w) where

$$w = \sigma_{11}/\sigma_{22\cdot1} \tag{9.5.8}$$

to obtain

$$p(\boldsymbol{\theta}, w, \beta \mid \mathbf{y}) \propto w^{\frac{1}{2}(n+1)-1}[S_{11}(\boldsymbol{\theta}) + wS_{22\cdot1}(\boldsymbol{\theta}, \beta)]^{-n},$$

$$0 < w < \infty, \quad -\infty < \beta < \infty, \quad -\infty < \boldsymbol{\theta} < \infty. \tag{9.5.9}$$

This distribution can be written as

$$p(\boldsymbol{\theta}, w, \beta \mid \mathbf{y}) = p(\boldsymbol{\theta} \mid w, \beta, \mathbf{y})p(w, \beta \mid \mathbf{y}). \tag{9.5.10}$$

Now, we may write

$$S_{11}(\boldsymbol{\theta}) + wS_{22\cdot1}(\boldsymbol{\theta}, \beta) = (2n - k)s^2(w, \beta) + R(\boldsymbol{\theta} \mid w, \beta), \tag{9.5.11}$$

where

$$R(\boldsymbol{\theta} \mid w, \beta) = [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(w, \beta)]'[\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot1}'(\beta)\mathbf{X}_{2\cdot1}(\beta)][\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(w, \beta)],$$

$$(2n - k)s^2(w, \beta) = [\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}(w, \beta)]'[\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}(w, \beta)]$$

$$+ w[\mathbf{y}_{2\cdot1}(\beta) - \mathbf{X}_{2\cdot1}(\beta)\hat{\boldsymbol{\theta}}(w, \beta)]'[\mathbf{y}_{2\cdot1}(\beta) - \mathbf{X}_{2\cdot1}(\beta)\hat{\boldsymbol{\theta}}(w, \beta)],$$

$$\mathbf{X}_{2\cdot1}(\beta) = \mathbf{X}_2 - \beta\mathbf{X}_1, \qquad \mathbf{y}_{2\cdot1}(\beta) = \mathbf{y}_2 - \beta\mathbf{y}_1$$

and

$$\hat{\boldsymbol{\theta}}(w, \beta) = [\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)]^{-1}[\mathbf{X}_1'\mathbf{y}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{y}_{2\cdot 1}(\beta)]$$

Making use of (9.5.11), we see that

a) the conditional distribution of $\boldsymbol{\theta}$, given (w, β), is

$$p(\boldsymbol{\theta} \mid w, \beta, \mathbf{y}) \propto [(2n-k)s^2(w, \beta) + R(\boldsymbol{\theta} \mid w, \beta)]^{-n}, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{9.5.12}$$

which is the

$$t_k\{\hat{\boldsymbol{\theta}}(w, \beta), s^2(w, \beta)[\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)]^{-1}, 2n-k\}$$

distribution.

b) partitioning

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_l \\ \hline \boldsymbol{\theta}_p \end{bmatrix}\begin{matrix} l \\ p \end{matrix} \quad \hat{\boldsymbol{\theta}}(w, \beta) = \begin{bmatrix} \hat{\boldsymbol{\theta}}_l(w, \beta) \\ \hline \hat{\boldsymbol{\theta}}_p(w, \beta) \end{bmatrix}\begin{matrix} l \\ p \end{matrix}$$

and (9.5.13)

$$[\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)]^{-1} = \begin{array}{c} \begin{array}{cc} \quad l \quad & \quad p \quad \end{array} \\ \left[\begin{array}{c:c} \mathbf{C}_{ll}(w, \beta) & \mathbf{C}_{lp}(w, \beta) \\ \hdashline \mathbf{C}_{lp}'(w, \beta) & \mathbf{C}_{pp}(w, \beta) \end{array}\right]\begin{matrix} l \\ p \end{matrix} \end{array},$$

the conditional distribution of the subset $\boldsymbol{\theta}_l$, given (w, β), is

$$p(\boldsymbol{\theta}_l \mid w, \beta, \mathbf{y})$$
$$\propto \{(2n-k)s^2(w, \beta) + [\boldsymbol{\theta}_l - \hat{\boldsymbol{\theta}}_l(w, \beta)]'\mathbf{C}_{ll}^{-1}(w, \beta)[\boldsymbol{\theta}_l - \hat{\boldsymbol{\theta}}_l(w, \beta)]\}^{-\frac{1}{2}(2n-k+l)},$$
$$-\infty < \boldsymbol{\theta}_l < \infty, \tag{9.5.14}$$

a $t_l[\hat{\boldsymbol{\theta}}_l(w, \beta), s^2(w, \beta)\mathbf{C}_{ll}(w, \beta), 2n-k]$ distribution, and

c) the posterior distribution of (w, β) is

$$p(w, \beta \mid \mathbf{y}) \propto |\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)|^{-1/2}\, w^{\frac{1}{2}(n+1)-1}\,[s^2(w, \beta)]^{-\frac{1}{2}(2n-k)},$$
$$0 < w < \infty, \quad -\infty < \beta < \infty. \tag{9.5.15}$$

Thus, we may write

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \int_0^\infty \int_{-\infty}^\infty p(\boldsymbol{\theta} \mid w, \beta, \mathbf{y})\, p(w, \beta \mid \mathbf{y})\, d\beta\, dw, \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{9.5.16}$$

which is an alternative representation of the distribution in (9.5.2). The usefulness of this representation lies chiefly in the associated form for a subset $\boldsymbol{\theta}_l$,

$$p(\boldsymbol{\theta}_l \mid \mathbf{y}) = \int_0^\infty \int_{-\infty}^\infty p(\boldsymbol{\theta}_l \mid w, \beta, \mathbf{y})\, p(w, \beta \mid \mathbf{y})\, d\beta\, dw, \qquad -\infty < \boldsymbol{\theta}_l < \infty. \tag{9.5.17}$$

If exact evaluation of the marginal distribution of $\boldsymbol{\theta}_l$ is desired, it will only be necessary to calculate the double integral in (9.5.17), whatever the value of k, instead

of a $k - l$ dimensional integral implied by the form in (9.5.2). This is particularly useful whenever $k - l$ is greater than two. In the situation where the responses were uncorrelated, (9.3.12) could be simplified according to (9.3.20) and (9.3.21). It does not seem possible in general here to find a matrix $\mathbf{P}$ to similarly diagonalize $[\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot1}'(\beta)\mathbf{X}_{2\cdot1}(\beta)]$ for all values of (β, w), and consequently it is necessary to evaluate its determinant and inverse as functions of β and w.

9.5.2 Approximations to the Distribution of $\boldsymbol{\theta}$ for $m = 2$

From (9.5.16), one might attempt to approximate the distribution of $\boldsymbol{\theta}$ by writing

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \underset{\mathbf{f}(w,\beta)}{E}\; p(\boldsymbol{\theta} \mid w, \beta, \mathbf{y}) \doteq p(\boldsymbol{\theta} \mid \hat{w}, \hat{\beta}, \mathbf{y}) \tag{9.5.18}$$

where $(\hat{w}, \hat{\beta})$ is the mode of the distribution of some functions $\mathbf{f}(w, \beta)$ of (w, β). However, this modal value $(\hat{w}, \hat{\beta})$ is not easy to determine since the distribution $p(w, \beta \mid \mathbf{y})$ in (9.5.15) is a rather complicated function of (w, β). We now develop an alternative approach.

The distribution in (9.5.9) may be written as

$$p(\boldsymbol{\theta}, \beta, w \mid \mathbf{y}) = p(\boldsymbol{\theta}, \beta \mid w, \mathbf{y})p(w \mid \mathbf{y}) \tag{9.5.19}$$

where the first factor is

$$p(\boldsymbol{\theta}, \beta \mid w, \mathbf{y}) \propto [S_{11}(\boldsymbol{\theta}) + wS_{22\cdot1}(\boldsymbol{\theta}, \beta)]^{-n}, \qquad -\infty < \beta < \infty, \quad -\infty < \boldsymbol{\theta} < \infty. \tag{9.5.20}$$

For a given value of w, let $(\hat{\boldsymbol{\theta}}_w, \hat{\beta}_w)$ be the mode of the conditional distribution (9.5.20). Employing Taylor's theorem, we may expand $S_{11}(\boldsymbol{\theta}) + wS_{22\cdot1}(\boldsymbol{\theta}, \beta)$ into

$$S_{11}(\boldsymbol{\theta}) + wS_{22\cdot1}(\boldsymbol{\theta}, \beta) \doteq (2n - k - 1)s^2(w) + Q(\boldsymbol{\theta}, \beta \mid w) \tag{9.5.21}$$

where

$$Q(\boldsymbol{\theta}, \beta \mid w) = (\boldsymbol{\theta}' - \hat{\boldsymbol{\theta}}_w', \beta - \hat{\beta}_w)\mathbf{H}(w)\begin{pmatrix}\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_w \\ \beta - \hat{\beta}_w\end{pmatrix}$$

$$\mathbf{H}(w) = \begin{array}{c} \begin{array}{cc} k & 1 \end{array} \\ \left[\begin{array}{c:c} \mathbf{H}_{11}(w) & \mathbf{h}_{12}(w) \\ \hdashline \mathbf{h}_{12}'(w) & h_{22}(w) \end{array}\right] \end{array}\begin{array}{c} k \\ 1 \end{array},$$

$$\mathbf{H}_{11}(w) = \mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot1}'(\hat{\beta}_w)\mathbf{X}_{2\cdot1}(\hat{\beta}_w)$$

$$h_{22}(w) = w(\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}_w)'(\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}_w)$$

$$\mathbf{h}_{12}(w) = w\{\mathbf{X}_{2\cdot1}'(\hat{\beta}_w)(\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}_w) + \mathbf{X}_1'[\mathbf{y}_{2\cdot1}(\hat{\beta}_w) - \mathbf{X}_{2\cdot1}(\hat{\beta}_w)\hat{\boldsymbol{\theta}}_w]\}$$

and

$$(2n - k - 1)s^2(w) = S_{11}(\hat{\boldsymbol{\theta}}_w) + wS_{22\cdot1}(\hat{\boldsymbol{\theta}}_w, \hat{\beta}_w).$$

To this degree of approximation, the conditional distribution of $(\boldsymbol{\theta}, \beta)$ given w is in the form of a $(k+1)$-dimensional t distribution. Integrating out β from (9.5.20) using the approximating form (9.5.21), we have that

$$p(\boldsymbol{\theta} \mid w, \mathbf{y}) \dot{\propto} [(2n-k-1)s^2(w) + Q(\boldsymbol{\theta} \mid w)]^{-\frac{1}{2}(2n-1)}, \tag{9.5.22}$$

where

$$Q(\boldsymbol{\theta} \mid w) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_w)' \, \mathbf{G}(w) \, (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_w)$$

$$\mathbf{G}(w) = \mathbf{H}_{11}(w) - h_{22}^{-1}(w)\mathbf{h}_{12}(w)\mathbf{h}'_{12}(w),$$

which is the $t_k[\hat{\boldsymbol{\theta}}_w, s^2(w)\mathbf{G}^{-1}(w), 2n-k-1]$ distribution. Further, substituting (9.5.21) into (9.5.9) and integrating out $(\boldsymbol{\theta}, \beta)$, the posterior distribution of w is, approximately,

$$p(w \mid \mathbf{y}) \dot{\propto} |\mathbf{H}(w)|^{-1/2} \, w^{\frac{1}{2}(n+1)-1} \, [s^2(w)]^{-\frac{1}{2}(2n-k-1)}, \qquad 0 < w < \infty. \tag{9.5.23}$$

We may now adopt an argument similar to that in (9.3.27) to approximate the distribution of $\boldsymbol{\theta}$ as

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \underset{\log w}{E} \; p(\boldsymbol{\theta} \mid w, \mathbf{y}) \doteq p(\boldsymbol{\theta} \mid \hat{w}, \mathbf{y}), \tag{9.5.24}$$

where $p(\boldsymbol{\theta} \mid \hat{w}, \mathbf{y})$ is obtained by inserting $\hat{w}$ in (9.5.22) and $\hat{w}$ is the mode of the density

$$p(\log w \mid \mathbf{y}) \dot{\propto} |\mathbf{H}(w)|^{-1/2} \, w^{\frac{1}{2}(n+1)} \, [s^2(w)]^{-\frac{1}{2}(2n-k-1)}, \qquad -\infty < \log w < \infty. \tag{9.5.25}$$

Calculation of $(\hat{\boldsymbol{\theta}}_w, \hat{\beta}_w, \hat{w})$.

In obtaining the approximating form (9.5.22), the basic problem is to find, for a given value of w, the conditional mode $\hat{\boldsymbol{\theta}}_w, \hat{\beta}_w$ of (9.5.20). From (9.5.7), we see that the quantity $S_{11}(\boldsymbol{\theta}) + wS_{22\cdot1}(\boldsymbol{\theta}, \beta)$ is quadratic in either β or $\boldsymbol{\theta}$ given the other. Thus, conditional on $\boldsymbol{\theta}$, $S_{11}(\boldsymbol{\theta}) + wS_{22\cdot1}(\boldsymbol{\theta}, \beta)$ is minimized when

$$\hat{\beta}(\boldsymbol{\theta}) = \frac{(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})' \, (\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta})}{(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})' \, (\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta})}, \tag{9.5.26}$$

and, conditional on β, it is minimized for

$$\hat{\boldsymbol{\theta}}(w, \beta) = [\mathbf{X}'_1\mathbf{X}_1 + w\mathbf{X}'_{2\cdot1}(\beta)\mathbf{X}_{2\cdot1}(\beta)]^{-1}[\mathbf{X}'_1\mathbf{y}_1 + w\mathbf{X}'_{2\cdot1}(\beta)\mathbf{y}_{2\cdot1}(\beta)] \tag{9.5.27}$$

as given in (9.5.11) Thus, given an initial choice of $\boldsymbol{\theta}$, say $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, expression (9.5.26) can be employed to obtain $\beta_0 = \hat{\beta}(\boldsymbol{\theta}_0)$ which in turn can be used to calculate $\hat{\boldsymbol{\theta}}(w, \beta_0)$ from (9.5.27) and so on. A convenient choice for $\boldsymbol{\theta}_0$ is the center of $S_{11}(\boldsymbol{\theta})$, namely $\boldsymbol{\theta}_0 = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y}_1$. This iterative procedure will converge under favorable circumstances.

Once the conditional mode $(\hat{\boldsymbol{\theta}}_w, \hat{\beta}_w)$ is determined for a given value of w, the corresponding density of $\log w$ in (9.5.25) for that value can be readily obtained. The

mode $\hat{w}$ may thus be found from (9.5.25) by employing standard numerical search procedures on a computer. A convenient preliminary estimate of $\hat{w}$ is given by

$$w_0 = s_1^2/s_{2\cdot 1}^2, \tag{9.5.28}$$

where s_1^2 is given in (9.5.3), and

$$s_{2\cdot 1}^2 = s_2^2 - \frac{s_{12}^2}{s_1^2} \tag{9.5.29}$$

with

$$s_2^2 = \frac{1}{\nu}(\mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2)'(\mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2), \qquad s_{12} = \frac{1}{\nu}(\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}_1)'(\mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\theta}}_2),$$

$$\hat{\boldsymbol{\theta}}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}_2 \qquad \text{and} \qquad \nu = n - k.$$

The accuracy of the approximation will be illustrated by an example in the next section.

9.5.3 An Example

To illustrate the theory developed above, we consider the following example.

An experiment was conducted to assess the effect of change in pressure from 30 psi to 50 psi on a chemical process yield, known to be subject to large errors. Twelve batches of raw materials were randomly selected and, from each batch, a pair of samples were taken by two chemists A and B for separate experimentation. Each chemist ran half of his experiments at high pressure and half at low pressure and they arranged their pressure settings such that each level of pressure run by chemist A was paired with each level of pressure run by chemist B an equal number of times. The data are given in Table 9.5.1.

Table 9.5.1

Pressure effect independently assessed by two chemists

	Chemist A		Chemist B	
Batch no.	Pressure	Yield y_{1u}	Pressure	Yield y_{2u}
1	50	79.5	30	73.2
2	50	76.4	50	85.0
3	50	77.2	30	60.3
4	50	76.0	50	72.9
5	50	78.6	30	66.5
6	50	82.9	50	80.0
7	30	61.4	30	63.6
8	30	67.4	50	78.3
9	30	63.5	30	73.3
10	30	69.1	50	81.3
11	30	61.2	30	54.2
12	30	69.7	50	76.3

It is assumed that in the region of the experimentation, the following model is appropriate

$$\begin{aligned} y_{u1} &= \theta_{01} + \theta x_{u1} + \varepsilon_{u1} \\ y_{u2} &= \theta_{02} + \theta x_{u2} + \varepsilon_{u2} \end{aligned} \tag{9.5.30}$$

where

$$x_{ui} = \frac{\text{Pressure} - 40}{10}$$

so that the x_{ui} assume only two values $(1, -1)$. The pressure effect θ is assumed common for both chemists but the parameters θ_{01} and θ_{02} are taken to be different to allow for possible systematic differences in mean yield. The error terms $(\varepsilon_{u1}, \varepsilon_{u2})$ are assumed correlated because each pair of observations (y_{u1}, y_{u2}) is made from the same batch of material. In terms of the linear model in (8.3.2) we have

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, 2, \tag{9.5.31}$$

where

$$\boldsymbol{\theta}_i = \begin{pmatrix} \theta_{0i} \\ \theta \end{pmatrix}, \qquad \mathbf{X}_i = [\mathbf{1} \,\vdots\, \mathbf{x}_i], \qquad \mathbf{x}_i' = (x_{1i}, \ldots, x_{ni}), \qquad n = 12$$

and $\mathbf{1}$ is a 12×1 vector of ones.

We are thus in a situation in which some of the parameters $\boldsymbol{\theta}_i$ and part of the derivative matrix $\mathbf{X}_i$ are common between the two responses.

Assuming that $(\varepsilon_{u1}, \varepsilon_{u2})$ has the bivariate Normal distribution $N_2(\mathbf{0}, \boldsymbol{\Sigma})$ and on the basis of the reference prior distribution in (9.5.6), the posterior distribution of $(\theta_{01}, \theta_{02}, \theta, \sigma_{11}, \sigma_{22\cdot 1}, \beta)$ is

$$\begin{aligned} &p(\theta_{01}, \theta_{02}, \theta, \sigma_{11}, \sigma_{22\cdot 1}, \beta \mid \mathbf{y}) \propto g_1(\theta_{01}, \theta, \sigma_{11})\, g_2(\theta_{02}, \theta, \sigma_{22\cdot 1}, \beta), \\ &\quad -\infty < \theta_{01} < \infty, \quad -\infty < \theta_{02} < \infty, \quad -\infty < \theta < \infty, \ \sigma_{11} > 0, \quad \sigma_{22\cdot 1} > 0, \\ &\quad -\infty < \beta < \infty, \end{aligned} \tag{9.5.32}$$

where

$$g_1(\theta_{01}, \theta, \sigma_{11}) \propto \sigma_{11}^{-[(11/2)+1]} \exp\left[-\frac{1}{2\sigma_{11}}(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1)'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1)\right]$$

$$\begin{aligned} g_2(\theta_{02}, \theta, \sigma_{22\cdot 1}, \beta) &\propto \sigma_{22}^{-[(13/2)+1]} \\ &\times \exp\left\{-\frac{1}{2\sigma_{22\cdot 1}}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta}_2 - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1)]'[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta}_2 - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1)]\right\} \end{aligned}$$

Since main interest centers on the pressure effect θ, the parameters $(\theta_{01}, \theta_{02})$ need first to be eliminated. Writing

$$\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_i = \dot{\mathbf{y}}_i - \mathbf{x}_i\theta + (\theta_{0i} - \bar{y}_i)\mathbf{1}, \tag{9.5.33}$$

where

$$\dot{\mathbf{y}}_i = \mathbf{y}_i - \bar{y}_i\mathbf{1}, \ \bar{y}_i = \frac{1}{n}\sum_{u=1}^{n} y_{ui}, \qquad i = 1, 2,$$

and noting that $\mathbf{1}'\dot{\mathbf{y}}_i = \mathbf{1}'\mathbf{x}_i = 0$, we have that

$$(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1)'(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1) = (\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta)'(\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta) + 12(\theta_{01} - \bar{y}_1)^2 \tag{9.5.34}$$

and

$$\begin{aligned}[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta}_2 - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1)]'[\mathbf{y}_2 - \mathbf{X}_2\boldsymbol{\theta}_2 - \beta(\mathbf{y}_1 - \mathbf{X}_1\boldsymbol{\theta}_1)] \\ = [\dot{\mathbf{y}}_2 - \mathbf{x}_2\theta - \beta(\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta)]'[\dot{\mathbf{y}}_2 - \mathbf{x}_2\theta - \beta(\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta)] \\ + 12[\theta_{02} - \bar{y}_2 - \beta(\theta_{01} - \bar{y}_1)]^2.\end{aligned}$$

Substituting (9.5.34) into (9.5.32), $(\theta_{01}, \theta_{02})$ can be easily integrated out, yielding

$$\begin{aligned}p(\theta, \sigma_{11}, \sigma_{22\cdot 1}, \beta \mid \mathbf{y}) \propto g_1^*(\theta, \sigma_{11})\, g_2^*(\theta, \sigma_{22\cdot 1}, \beta), \\ -\infty < \theta < \infty, \quad -\infty < \beta < \infty, \quad \sigma_{11} > 0,\ \sigma_{22\cdot 1} > 0,\end{aligned} \tag{9.5.35}$$

where

$$g_1^*(\theta, \sigma_{11}) \propto \sigma_{11}^{-[(10/2)+1]} \exp\left[-\frac{1}{2\sigma_{11}}(\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta)'(\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta)\right]$$

and

$$\begin{aligned}g_2^*(\theta, \sigma_{22\cdot 1}, \beta) \propto \sigma_{22}^{-[(12/2)+1]} \\ \times \exp\left\{-\frac{1}{2\sigma_{22\cdot 1}}[\dot{\mathbf{y}}_2 - \mathbf{x}_2\theta - \beta(\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta)]'[\dot{\mathbf{y}}_2 - \mathbf{x}_2\theta - \beta(\dot{\mathbf{y}}_1 - \mathbf{x}_1\theta)]\right\}\end{aligned}$$

which is of the form in (9.5.7) with $n = 11$, and the results in Sections 9.5.1–2 are now applicable.

Figure 9.5.1 shows the posterior distribution of θ calculated from (9.5.2). The normalizing constant was computed by numerical integration. Also shown in the same figure are the normalized curves of the two component factors

Fig. 9.5.1 Posterior distribution of θ and its component factors: the pressure data.

of the distribution. The first factor, which is a univariate t distribution centered at $\theta = 6.53$, represents information about θ coming from the results of chemist A, and the second factor, having mode at $\theta = 5.49$ and a long tail toward the right, represents the "extra information" provided by the results of chemist B. As expected, the overall distribution is located between the two component factors. It is a t-like distribution centered at $\theta \doteq 6.22$ and skewed slightly to the right.

Although for this example it will be scarcely necessary to approximate the distribution of θ, for illustrative purposes we have obtained the approximate distribution from (9.5.24). Starting with the preliminary estimate

$$w_0 = s_1^2/s_{2\cdot 1}^2 = 0.312$$

we found by trial and error that $\hat{w} \doteq 0.377$. The corresponding $\hat{\beta}_{\hat{w}}$ and $\hat{\theta}_{\hat{w}}$ are

$$\hat{\beta}_{\hat{w}} \doteq 0.6948, \qquad \hat{\theta}_{\hat{w}} \doteq 6.189$$

from which we obtained

$$p(\theta \mid \mathbf{y}) \dot{\propto} [227.9735 + 16.4171(\theta - 6.189)^2]^{-21/2}, \qquad -\infty < \theta < \infty.$$

That is, the quantity

$$\frac{\theta - 6.189}{0.833}$$

is approximately distributed as $t(0, 1, 20)$. Figure 9.5.2 compares the accuracy of the approximation for this example. The agreement between the exact and the approximate distributions is very close.

Fig. 9.5.2 Comparison of the exact and the approximate posterior distributions of θ: the pressure data.

9.6 A SUMMARY OF VARIOUS POSTERIOR DISTRIBUTIONS FOR COMMON PARAMETERS θ

In Table 9.6.1, we provide a summary of the various posterior distributions for making inferences about the common parameters $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_k)$ discussed in the preceding sections.

Table 9.6.1

A summary of various posterior distributions for the common parameters $\boldsymbol{\theta}$

Independent Responses:

1. The model is

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta} + \boldsymbol{\varepsilon}_i, \qquad i=1, \ldots, m,$$

where $\mathbf{y}_i$ is a $n_i \times 1$ vector of observations $\boldsymbol{\theta}$ is a $k \times 1$ vector of parameters, $\mathbf{X}_i$ is a $n_i \times k$ matrix of fixed elements, and $\boldsymbol{\varepsilon}_i$ is a $n_i \times 1$ vector of errors distributed as Normal $N_{n_i}(\mathbf{0}, \sigma_{ii}\mathbf{I})$. The $\boldsymbol{\varepsilon}_i$'s are assumed to be independent of one another.

2. The distribution of $\boldsymbol{\theta}$ is, from (9.2.3)

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \prod_{i=1}^{m} [a_{ii} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)'\mathbf{X}_i'\mathbf{X}_i(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)]^{-\frac{1}{2}n_i}, \qquad -\infty < \boldsymbol{\theta} < \infty,$$

where

$$a_{ii} = (\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\theta}}_i)'(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\theta}}_i) \quad \text{and} \quad (\mathbf{X}_i'\mathbf{X}_i)\hat{\boldsymbol{\theta}}_i = \mathbf{X}_i'\mathbf{y}_i.$$

3. For $m=2$, $\mathbf{X}_1'\mathbf{X}_1$ positive definite, an alternative expression for $p(\boldsymbol{\theta} \mid \mathbf{y})$ is in (9.3.12)

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \int_0^\infty p(\boldsymbol{\theta} \mid w, \mathbf{y}) p(w \mid \mathbf{y})\, dw$$

where from (9.3.8),

$$p(w \mid \mathbf{y}) \propto |\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2|^{-1/2}\, w^{\frac{1}{2}n_2 - 1}[s^2(w)]^{-\frac{1}{2}(n_1+n_2-k)}, \qquad 0 < w < \infty,$$

and, from (9.3.7), the conditional distribution $p(\boldsymbol{\theta} \mid w, \mathbf{y})$ is the multivariate t distribution

$$t_k[\bar{\boldsymbol{\theta}}(w),\ s^2(w)(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1},\ n_1 + n_2 - k]$$

with

$$\bar{\boldsymbol{\theta}}(w) = (\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1}(\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\theta}}_1 + w\mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\theta}}_2)$$

$$s^2(w) = \frac{1}{n_1 + n_2 - k}[\nu_1 s_1^2 + w\nu_2 s_2^2 + S_c(w)],$$

$$S_c(w) = w(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2)'\, \mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1}\, \mathbf{X}_2'\mathbf{X}_2(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_2),$$

$$\nu_i = n_i - k, \qquad \nu_i s_i^2 = a_{ii}, \qquad i=1, 2.$$

Table 9.6.1 *Continued*

4. In particular, partitioning

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_l \\ \hdashline \boldsymbol{\theta}_p \end{bmatrix}\begin{matrix} l \\ p \end{matrix} \qquad \bar{\boldsymbol{\theta}}(w) = \begin{bmatrix} \bar{\boldsymbol{\theta}}_l(w) \\ \hdashline \bar{\boldsymbol{\theta}}_p(w) \end{bmatrix}; \qquad (\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1} = \begin{array}{c} \begin{matrix} l & \quad p \end{matrix} \\ \left[\begin{array}{c:c} \mathbf{C}_{ll}(w) & \mathbf{C}_{lp}(w) \\ \hdashline \mathbf{C}_{pl}(w) & \mathbf{C}_{pp}(w) \end{array}\right] \end{array}\begin{matrix} l \\ p \end{matrix}$$

the marginal distribution of $\boldsymbol{\theta}_l$ is

$$p(\boldsymbol{\theta}_l \mid \mathbf{y}) = \int_0^\infty p(\boldsymbol{\theta}_l \mid w, \mathbf{y})\, p(w \mid \mathbf{y})\, dw,$$

where $p(\boldsymbol{\theta}_l \mid w, \mathbf{y})$ is the multivariate t distribution

$$t_l[\bar{\boldsymbol{\theta}}_l(w),\ s^2(w)\mathbf{C}_{ll}(w),\ n_1 + n_2 - k].$$

5. From (9.3.18) to (9.2.21) computation of $p(\boldsymbol{\theta} \mid \mathbf{y})$ and $p(\boldsymbol{\theta}_l \mid \mathbf{y})$ can be simplified by first finding a $k \times k$ non-singular matrix $\mathbf{P}$ and a $k \times k$ diagonal matrix $\boldsymbol{\Lambda} = \{\lambda_{jj}\}$, $j = 1, \ldots, k$, such that

$$\mathbf{X}_1'\mathbf{X}_1 = (\mathbf{P}')^{-1}\mathbf{P}^{-1} \qquad \text{and} \qquad \mathbf{X}_2'\mathbf{X}_2 = (\mathbf{P}')^{-1}\boldsymbol{\Lambda}\mathbf{P}^{-1}$$

so that

$$\bar{\boldsymbol{\theta}}(w) = \mathbf{P}(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}(\hat{\boldsymbol{\phi}}_1 + w\boldsymbol{\Lambda}\hat{\boldsymbol{\phi}}_2), \qquad \hat{\boldsymbol{\phi}}_i = \mathbf{P}^{-1}\hat{\boldsymbol{\theta}}_i, \qquad i = 1, 2$$

$$S_e(w) = \sum_{j=1}^{k} \left(\frac{w\lambda_{jj}}{1 + w\lambda_{jj}}\right)(\hat{\phi}_{1j} - \hat{\phi}_{2j})^2$$

$$|\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2| = |\mathbf{X}_1'\mathbf{X}_1| \prod_{j=1}^{k} (1 + w\lambda_{jj})$$

$$(\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_2'\mathbf{X}_2)^{-1} = \mathbf{P}(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}\mathbf{P}'$$

and

$$\mathbf{C}_{ll}(w) = \mathbf{P}_1(\mathbf{I} + w\boldsymbol{\Lambda})^{-1}\mathbf{P}_1' \qquad \text{where} \qquad \mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \hdashline \mathbf{P}_2 \end{bmatrix}\begin{matrix} l \\ p \end{matrix}$$

6. A simple approximation to $p(\boldsymbol{\theta} \mid \mathbf{y})$ is given by (9.3.27)

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \doteq p(\boldsymbol{\theta} \mid \hat{w}, \mathbf{y}),$$

i.e.

$$\boldsymbol{\theta} \sim t_k[\bar{\boldsymbol{\theta}}(\hat{w}),\ s^2(\hat{w})(\mathbf{X}_1'\mathbf{X}_1 + \hat{w}\mathbf{X}_2'\mathbf{X}_2)^{-1},\ n_1 + n_2 - k]$$

where from (9.3.28) and (9.3.29), $\hat{w}$ is the root of the equation

$$n_2 w^{-1} - \sum_{j=1}^{k} \lambda_{jj}(1 + w\lambda_{jj})^{-1} - [s^2(w)]^{-1}\left[v_2 s_2^2 + \sum_{j=1}^{k} \lambda_{jj}(1 + w\lambda_{jj})^{-2}(\hat{\phi}_{1j} - \hat{\phi}_{2j})^2\right] = 0$$

maximizing $wp(w \mid \mathbf{y})$. An iterative procedure for finding w is described in (9.3.30) and (9.3.31).

Table 9.6.1 *Continued*

Correlated Responses with Common Derivative Matrix **X**:

7. The model is

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\theta}_c + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, m,$$

where $\mathbf{y}_i$ is a $n \times 1$ vector of observations, $\boldsymbol{\theta}_c$ a $k \times 1$ vector of parameters, $\mathbf{X}$ a $n \times k$ matrix of fixed elements with rank k and $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \ldots, \varepsilon_{ni})'$ a $n \times 1$ vector of errors. It is assumed that $\boldsymbol{\varepsilon}_{(u)} = (\varepsilon_{u1}, \ldots, \varepsilon_{um})'$, $u = 1, \ldots, m$, are independently distributed as $N_m(\mathbf{0}, \boldsymbol{\Sigma})$.

8. Let

$$\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_m], \qquad \hat{\boldsymbol{\theta}}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i, \qquad i = 1, \ldots, m,$$

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

and $\mathbf{1}_m$ be a $m \times 1$ vector of ones. Then, from (9.4.3) the posterior distribution of $\boldsymbol{\theta}_c$ is the multivariate t distribution

$$\boldsymbol{\theta}_c \sim t_k\left[\bar{\boldsymbol{\theta}}_c, \frac{1}{(n-k)d}\mathbf{V}, n-k\right]$$

where

$$\bar{\boldsymbol{\theta}}_c = \frac{1}{d}\hat{\boldsymbol{\theta}}\mathbf{A}^{-1}\mathbf{1}_m, \qquad d = \mathbf{1}_m'\mathbf{A}^{-1}\mathbf{1}_m$$

and

$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1} + \hat{\boldsymbol{\theta}}\left[\mathbf{A}^{-1} - \frac{1}{d}\mathbf{A}^{-1}\mathbf{1}_m\mathbf{1}_m'\mathbf{A}^{-1}\right]\hat{\boldsymbol{\theta}}'$$

General Linear Model:

9. The model is

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta} + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, m,$$

where $\mathbf{y}_i$ is a $n \times 1$ vector of observations, $\mathbf{X}_i$ is a $n \times k$ matrix of fixed elements, $\boldsymbol{\theta}$ a $k \times 1$ vector of parameters and $\boldsymbol{\varepsilon}_i$ a $n \times 1$ vector of errors. It is assumed that $\boldsymbol{\varepsilon}_{(u)} = (\varepsilon_{u1}, \ldots, \varepsilon_{um})'$, $u = 1, \ldots, m$, are independently distributed as $N_m(\mathbf{0}, \boldsymbol{\Sigma})$.

10. The posterior distribution of $\boldsymbol{\theta}$ is, from (9.5.1),

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

where

$$\mathbf{S}(\boldsymbol{\theta}) = \{S_{ij}(\boldsymbol{\theta})\} \qquad i, j = 1, \ldots, m$$

$$S_{ij}(\boldsymbol{\theta}) = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta})'(\mathbf{y}_j - \mathbf{X}_j\boldsymbol{\theta}).$$

Table 9.6.1 *Continued*

11. For $m=2$, an alternative expression of the posterior distribution of $\boldsymbol{\theta}$ is given in (9.5.16)

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \int_0^\infty \int_{-\infty}^\infty p(\boldsymbol{\theta} \mid w, \beta, \mathbf{y}) p(w, \beta \mid \mathbf{y})\, d\beta\, dw, \qquad -\infty < \boldsymbol{\theta} < \infty,$$

where, from (9.5.15),

$$p(w, \beta \mid \mathbf{y}) \propto |\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)|^{-1/2}\, w^{\frac{1}{2}(n+1)-1}[s^2(w)]^{-\frac{1}{2}(2n-k)},$$
$$0 < w < \infty, \quad -\infty < \beta < \infty$$

and, from (9.5.12), the conditional posterior distribution $p(\boldsymbol{\theta} \mid w, \beta, \mathbf{y})$ is the multivariate t distribution

$$\boldsymbol{\theta} \sim t_k\{\hat{\boldsymbol{\theta}}(w, \beta), s^2(w, \beta)[\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)]^{-1}, 2n-k\}$$

with

$$\hat{\boldsymbol{\theta}}(w, \beta) = [\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)]^{-1}[\mathbf{X}_1'\mathbf{y}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{y}_{2\cdot 1}(\beta)]$$
$$\mathbf{X}_{2\cdot 1}(\beta) = \mathbf{X}_2 - \beta\mathbf{X}_1, \qquad \mathbf{y}_{2\cdot 1}(\beta) = \mathbf{y}_2 - \beta\mathbf{y}_1$$

and

$$(2n-k)s^2(w, \beta) = [\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}(w, \beta)]'[\mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\theta}}(w, \beta)]$$
$$+ w[\mathbf{y}_{2\cdot 1}(\beta) - \mathbf{X}_{2\cdot 1}(\beta)\hat{\boldsymbol{\theta}}(w, \beta)]'[\mathbf{y}_{2\cdot 1}(\beta) - \mathbf{X}_{2\cdot 1}(\beta)\hat{\boldsymbol{\theta}}(w, \beta)].$$

12. By partitioning

$$\boldsymbol{\theta} = \begin{array}{c} \left[\begin{array}{c} \boldsymbol{\theta}_l \\ \hline \boldsymbol{\theta}_p \end{array}\right] \end{array}\begin{array}{c} l \\ p \end{array} \qquad \hat{\boldsymbol{\theta}}(w, \beta) = \left[\begin{array}{c} \hat{\boldsymbol{\theta}}_l(w, \beta) \\ \hline \hat{\boldsymbol{\theta}}_p(w, \beta) \end{array}\right]\begin{array}{c} l \\ p \end{array}$$

and

$$[\mathbf{X}_1'\mathbf{X}_1 + w\mathbf{X}_{2\cdot 1}'(\beta)\mathbf{X}_{2\cdot 1}(\beta)]^{-1} = \begin{array}{c} \begin{array}{cc} \quad l & \qquad p \end{array} \\ \left[\begin{array}{c:c} \mathbf{C}_{ll}(w, \beta) & \mathbf{C}_{lp}(w, \beta) \\ \hdashline \mathbf{C}_{lp}'(w, \beta) & \mathbf{C}_{pp}(w, \beta) \end{array}\right] \end{array}\begin{array}{c} l \\ p \end{array},$$

the marginal posterior distribution of $\boldsymbol{\theta}_l$ is in (9.5.17),

$$p(\boldsymbol{\theta}_l \mid \mathbf{y}) = \int_0^\infty \int_{-\infty}^\infty p(\boldsymbol{\theta}_l \mid w, \beta, \mathbf{y}) p(w, \beta \mid \mathbf{y})\, d\beta\, dw, \qquad -\infty < \boldsymbol{\theta}_l < \infty,$$

where $p(\boldsymbol{\theta}_l \mid w, \beta, \mathbf{y})$ is the multivariate t distribution

$$\boldsymbol{\theta}_l \sim t_l\{\hat{\boldsymbol{\theta}}_l(w, \beta),\ s^2(w, \beta)\mathbf{C}_{ll}(w, \beta),\ 2n-k\}.$$

13. A procedure for approximating the distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$ is developed in Section 9.5.2.

APPENDIX A9.1

Asymptotic Expansion of the Posterior Distribution of θ for Two Independent Responses

We now develop an asymptotic expansion for the distribution of $\boldsymbol{\theta}$ in (9.3.1),

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto \prod_{i=1}^{2} \left[1 + \frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)' \mathbf{X}_i' \mathbf{X}_i (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_i)}{\nu_i s_i^2}\right]^{-\frac{1}{2}(\nu_i + k)}, \qquad -\infty < \boldsymbol{\theta} < \infty. \tag{A9.1.1}$$

The results can be used to approximate the distribution as well as the associated marginal distributions of subsets of $\boldsymbol{\theta}$. The procedure is a generalization of Fisher's result (1961b) for the case of the weighted mean discussed in Section 9.2.1.

For simplicity in notation, we write

$$Q_i = (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)' \mathbf{M}_i (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i), \qquad i = 1, 2, \tag{A9.1.2}$$

where

$$\mathbf{M}_i = \frac{1}{s_i^2} \mathbf{X}_i' \mathbf{X}_i.$$

Expression (A9.1.1) then becomes

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = c^{-1} g(Q_1, Q_2), \qquad -\infty < \boldsymbol{\theta} < \infty,$$

where

$$g(Q_1, Q_2) = \left(1 + \frac{Q_1}{\nu_1}\right)^{-\frac{1}{2}(\nu_1 + k)} \left(1 + \frac{Q_2}{\nu_2}\right)^{-\frac{1}{2}(\nu_2 + k)} \tag{A9.1.3}$$

and

$$c = \int_{-\infty < \boldsymbol{\theta} < \infty} g(Q_1, Q_2) d\boldsymbol{\theta}$$

The expression $(1 + Q_1/\nu_1)^{-\frac{1}{2}(\nu_1 + k)}$ can be written

$$\left(1 + \frac{Q_1}{\nu_1}\right)^{-\frac{1}{2}(\nu_1 + k)} = \exp\left(-\tfrac{1}{2}Q_1\right) \exp\left[\tfrac{1}{2}Q_1 - \frac{\nu_1 + k}{2} \log\left(1 + \frac{Q_1}{\nu_1}\right)\right].$$

Expanding the second factor on the right in powers of ν_1^{-1}, we obtain

$$\left(1 + \frac{Q_1}{\nu_1}\right)^{-\frac{1}{2}(\nu_1 + k)} = \exp\left(-\tfrac{1}{2}Q_1\right) \sum_{i=0}^{\infty} p_i \nu_1^{-i}, \tag{A9.1.4}$$

where

$$p_0 = 1,$$
$$p_1 = \tfrac{1}{4}(Q_1^2 - 2kQ_1),$$
$$p_2 = \tfrac{1}{96}[3Q_1^4 - 4(3k + 4)Q_1^3 + 12k(k + 2)Q_1^2], \text{ etc.}$$

Similarly, we have that

$$\left(1 + \frac{Q_2}{\nu_2}\right)^{-\frac{1}{2}(\nu_2 + k)} = \exp\left(-\tfrac{1}{2}Q_2\right) \sum_{i=0}^{\infty} q_i \nu_2^{-i}, \tag{A9.1.5}$$

where
$$q_0 = 1,$$
$$q_1 = \tfrac{1}{4}(Q_2^2 - 2kQ_2),$$
$$q_2 = \tfrac{1}{96}[3Q_2^4 - 4(3k+4)Q_2^3 + 12k(k+2)Q_2^2], \text{ etc.}$$

Substituting (A9.1.4) and (A9.1.5) into (A9.1.3) and after a little reduction, we can express the posterior distribution as

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = w^{-1}\, h(\boldsymbol{\theta}), \qquad -\infty < \boldsymbol{\theta} < \infty, \tag{A9.1.6}$$

where

$$h(\boldsymbol{\theta}) = \frac{|\mathbf{M}|^{1/2}}{(2\pi)^{\frac{1}{2}k}} \exp\left(-\tfrac{1}{2}Q\right) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_i q_j v_1^{-i} v_2^{-j},$$
$$Q = (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \mathbf{M} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}),$$
$$\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2, \qquad \tilde{\boldsymbol{\theta}} = \mathbf{M}^{-1}(\mathbf{M}_1\hat{\boldsymbol{\theta}}_1 + \mathbf{M}_2\hat{\boldsymbol{\theta}}_2) \qquad \text{and} \qquad w = \int_{-\infty < \boldsymbol{\theta} < \infty} h(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{A9.1.7}$$

The integral w in (A9.1.7) can be evaluated term by term. From (A9.1.4) and (A9.1.5), we see that each term is a bivariate polynomial in the mixed moments of the quadratic forms Q_1 and Q_2 where the variables $\boldsymbol{\theta}$ have a multivariate Normal distribution $N_k(\tilde{\boldsymbol{\theta}}, \mathbf{M}^{-1})$. For this problem, it appears much simpler to obtain the mixed moments indirectly by first finding the mixed cumulants. It is straightforward to verify that the joint cumulant generating function of Q_1 and Q_2 is

$$\begin{aligned}\kappa(t_1, t_2) &= \log \int_{-\infty < \boldsymbol{\theta} < \infty} \frac{|\mathbf{M}|^{1/2}}{(2\pi)^{\frac{1}{2}k}} \exp\left(t_1 Q_1 + t_2 Q_2 - \tfrac{1}{2}Q\right) d\boldsymbol{\theta} \\ &= -\tfrac{1}{2}\log|\mathbf{I} - 2\mathbf{M}^{-1}(t_1\mathbf{M}_1 + t_2\mathbf{M}_2)| + t_1\boldsymbol{\eta}_1'\mathbf{M}_1\boldsymbol{\eta}_1 + t_2\boldsymbol{\eta}_2'\mathbf{M}_2\boldsymbol{\eta}_2 \\ &\quad + 2(t_1\mathbf{M}_1\boldsymbol{\eta}_1 + t_2\mathbf{M}_2\boldsymbol{\eta}_2)'(\mathbf{M} - 2t_1\mathbf{M}_1 - 2t_2\mathbf{M}_2)^{-1}(t_1\mathbf{M}_1\boldsymbol{\eta}_1 + t_2\mathbf{M}_2\boldsymbol{\eta}_2),\end{aligned} \tag{A9.1.8}$$

where

$$\boldsymbol{\eta}_1 = \tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_1 \qquad \text{and} \qquad \boldsymbol{\eta}_2 = \tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_2.$$

Upon differentiating (A9.1.8) and after some algebraic reduction, we find (see Appendix A9.2)

$$\left.\begin{aligned}\kappa_{10} &= \operatorname{tr}\mathbf{M}^{-1}\mathbf{M}_1 + \boldsymbol{\eta}_1'\mathbf{M}_1\boldsymbol{\eta}_1, \\ \kappa_{01} &= \operatorname{tr}\mathbf{M}^{-1}\mathbf{M}_2 + \boldsymbol{\eta}_2'\mathbf{M}_2\boldsymbol{\eta}_2, \\ \kappa_{rs} &= 2^{r+s-1}(r+s-2)![(r+s-1)\operatorname{tr}\mathbf{M}^{-1}\mathbf{G}^{rs} \\ &\quad + (r\boldsymbol{\eta}_1 + s\boldsymbol{\eta}_2)'\mathbf{G}^{rs}(r\boldsymbol{\eta}_1 + s\boldsymbol{\eta}_2) - r\boldsymbol{\eta}_1'\mathbf{G}^{rs}\boldsymbol{\eta}_1 - s\boldsymbol{\eta}_2'\mathbf{G}^{rs}\boldsymbol{\eta}_2], \qquad r + s \geqslant 2,\end{aligned}\right\} \tag{A9.1.9}$$

where

$$\mathbf{G}^{rs} = \mathbf{M}(\mathbf{M}^{-1}\mathbf{M}_1)^r(\mathbf{M}^{-1}\mathbf{M}_2)^s.$$

A9.1

Employing the bivariate moment–cumulant inversion formulae as given by Cook (1951), the integral w in (A9.1.7) can be written as

$$w = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} b_{ij} v_1^{-i} v_2^{-j}, \tag{A9.1.10}$$

where

$$b_{00} = 1,$$

$$b_{10} = \tfrac{1}{4}(\kappa_{20} + \kappa_{10}^2 - 2k\kappa_{10}),$$

$$b_{01} = \tfrac{1}{4}(\kappa_{02} + \kappa_{01}^2 - 2k\kappa_{01}),$$

$$\begin{aligned} b_{11} = \tfrac{1}{16}[&\kappa_{22} + \kappa_{20}\kappa_{02} + 2\kappa_{11}^2 + 4\kappa_{11}\kappa_{01}\kappa_{10} + \kappa_{10}^2\kappa_{01}^2 + 2\kappa_{21}\kappa_{01} + 2\kappa_{12}\kappa_{10} \\ &+ \kappa_{20}\kappa_{01}^2 + \kappa_{02}\kappa_{10}^2 - 2k(\kappa_{12} + \kappa_{21} + \kappa_{02}\kappa_{10} + \kappa_{20}\kappa_{01} + 2\kappa_{11}\kappa_{10} \\ &+ 2\kappa_{11}\kappa_{01} + \kappa_{10}\kappa_{01}^2 + \kappa_{01}\kappa_{10}^2) + 4k^2(\kappa_{11} - \kappa_{01}\kappa_{10})], \end{aligned}$$

$$\begin{aligned} b_{20} = \tfrac{1}{96}[&3(\kappa_{40} + 3\kappa_{20}^2 + 4\kappa_{30}\kappa_{10} + 6\kappa_{20}\kappa_{10}^2 + \kappa_{10}^4) \\ &- 4(3k + 4)(\kappa_{30} + 3\kappa_{20}\kappa_{10} + \kappa_{10}^3) + 12k(k + 2)(\kappa_{20} + \kappa_{10}^2)], \end{aligned}$$

$$\begin{aligned} b_{02} = \tfrac{1}{96}[&3(\kappa_{04} + 3\kappa_{02}^2 + 4\kappa_{03}\kappa_{01} + 6\kappa_{02}\kappa_{01}^2 + \kappa_{01}^4) - 4(3k + 4) \\ &- 4(3k + 4)(\kappa_{03} + 3\kappa_{02}\kappa_{01} + \kappa_{01}^3) + 12k(k + 2)(\kappa_{02} + \kappa_{01}^2)], \text{ etc.} \end{aligned}$$

Substituting the results in (A9.1.10) into (A9.1.6), we obtain the following asymptotic expression for the posterior distribution of $\boldsymbol{\theta}$,

$$p(\boldsymbol{\theta} \mid \mathbf{y}_2) = \frac{|\mathbf{M}|^{1/2}}{(2\pi)^{\frac{1}{2}k}} \exp[-\tfrac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \mathbf{M}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} d_{ij} v_1^{-i} v_2^{-j}, \quad -\infty < \boldsymbol{\theta} < \infty, \tag{A9.1.11}$$

where

$$d_{00} = 1, \qquad d_{10} = p_1 - b_{10}, \qquad d_{01} = q_1 - b_{01},$$

$$d_{11} = (p_1 - b_{10})(q_1 - b_{01}) + b_{10}b_{01} - b_{11},$$

$$d_{20} = p_2 - b_{20} - p_1 b_{10} + b_{10}^2, \qquad d_{02} = q_2 - b_{02} - q_1 b_{01} + b_{01}^2,$$

and $\tilde{\boldsymbol{\theta}}$ and $\mathbf{M}$ are given in (A9.1.7).

Expressions for additional terms d_{12}, d_{21}, d_{22}, etc., can similarly be found if desired.

The posterior distribution is thus expressed in the form of a multivariate Normal distribution multiplied by a power series in v_1^{-1} and v_2^{-1}. When both v_1 and v_2 tend to infinity, all terms of the power series except the leading one vanish so that, in the limit, the posterior distribution is multivariate Normal $N_k(\tilde{\boldsymbol{\theta}}, \mathbf{M}^{-1})$. For finite values of v_1 and v_2, the terms in the power series can be regarded as "corrections" in a Normal approximation to the distribution in (A9.1.1). From

(A9.1.4), (A9.1.5) and (A9.1.9), we see that numerical evaluation of the coefficients in the power series involves merely matrix inversions and multiplications, operations which are easily performed on an electronic computer.

We note that when the posterior distribution is a univariate distribution as in (9.2.4), the results in (A9.1.11) are in exact agreement with those obtained by Fisher (1961b). In Fisher's derivation, each term of the integral w in (A9.1.7) was expressed in terms of the moments of a univariate Normal distribution. It can therefore be evaluated directly without making use of the mixed-cumulant formulae given in (A9.1.9) which seem more convenient for the multivariate case considered here.

For the univariate case, posterior probabilities can be calculated using the formulae given in Fisher's paper. When $k > 1$, the corresponding formulae for the evaluation of joint probabilities become exceedingly cumbersome and are not given here.

The Marginal Posterior Distribution

When interest centers on a subset of the elements of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_l' = (\theta_1, \ldots, \theta_l)$, an asymptotic expression for the corresponding marginal posterior distribution can be obtained by integrating out the remaining elements, $\boldsymbol{\theta}_r' = (\theta_{l+1}, \ldots, \theta_k)$ from the joint distribution in (A9.1.11). We have that

$$p(\boldsymbol{\theta}_l \mid \mathbf{y}) = \frac{|\mathbf{M}|^{1/2}}{(2\pi)^{\frac{1}{2}k}} \int_{-\infty < \boldsymbol{\theta}_r < \infty} \exp\left(-\tfrac{1}{2}Q\right) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} d_{ij} v_1^{-i} v_2^{-j} d\boldsymbol{\theta}_r. \tag{A9.1.12}$$

Partitioning $\tilde{\boldsymbol{\theta}}$ into $\tilde{\boldsymbol{\theta}}' = (\tilde{\boldsymbol{\theta}}_l' \,\vdots\, \tilde{\boldsymbol{\theta}}_r')$ and the matrices $\mathbf{M}$ and $\mathbf{M}^{-1}$ into

$$\mathbf{M} = \begin{array}{c} \begin{array}{cc} l & r \end{array} \\ \left[\begin{array}{c:c} \mathbf{M}_{ll} & \mathbf{M}_{lr} \\ \hdashline \mathbf{M}_{rl} & \mathbf{M}_{rr} \end{array}\right] \end{array} \begin{array}{c} l \\ r \end{array} \qquad \mathbf{M}^{-1} = \begin{array}{c} \begin{array}{cc} l & r \end{array} \\ \left[\begin{array}{c:c} \mathbf{V}_{ll} & \mathbf{V}_{lr} \\ \hdashline \mathbf{V}_{rl} & \mathbf{V}_{rr} \end{array}\right] \end{array} \begin{array}{c} l \\ r \end{array}$$

we can write the marginal posterior distribution as

$$p(\boldsymbol{\theta}_l \mid \mathbf{y}) = \frac{|\mathbf{V}_{ll}^{-1}|^{1/2}}{(2\pi)^{\frac{1}{2}l}} \exp\left[-\tfrac{1}{2}(\boldsymbol{\theta}_l - \tilde{\boldsymbol{\theta}}_l)'\mathbf{V}_{ll}^{-1}(\boldsymbol{\theta}_l - \tilde{\boldsymbol{\theta}}_l)\right] f(\boldsymbol{\theta}_l), \qquad -\infty < \boldsymbol{\theta}_l < \infty, \tag{A9.1.13}$$

where

$$f(\boldsymbol{\theta}_l) = \frac{|\mathbf{M}_{rr}|^{1/2}}{(2\pi)^{\frac{1}{2}r}} \int_{-\infty < \boldsymbol{\theta}_r < \infty} \exp\left[-\tfrac{1}{2}(\boldsymbol{\theta}_r - \bar{\boldsymbol{\theta}}_r)'\mathbf{M}_{rr}(\boldsymbol{\theta}_r - \bar{\boldsymbol{\theta}}_r)\right] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} d_{ij}\, v_1^{-i} v_2^{-j}\, d\boldsymbol{\theta}_r$$

with

$$\bar{\boldsymbol{\theta}}_r = \tilde{\boldsymbol{\theta}}_r - \mathbf{M}_{rr}^{-1}\mathbf{M}_{rl}(\boldsymbol{\theta}_l - \tilde{\boldsymbol{\theta}}_l).$$

From the expression for d_{ij} given in (A9.1.11), we see that each term in the integral $f(\boldsymbol{\theta}_l)$ is a bivariate polynomial in the quadratic form Q_1 and Q_2 where $\boldsymbol{\theta}_l$ is now considered fixed and $\boldsymbol{\theta}_r$ has a multivariate Normal distribution

$N_r(\tilde{\theta}_r, \mathbf{M}_{rr}^{-1})$. Adopting the same procedure as that described in the preceding section and by setting

$$\hat{\theta}_1' = (\hat{\theta}_{1l}' \vdots \hat{\theta}_{1r}'), \qquad \hat{\theta}_2' = (\hat{\theta}_{2l}' \vdots \hat{\theta}_{2r}') \tag{A9.1.14}$$

$$\mathbf{M}_1 = \begin{array}{c} \begin{array}{cc} l & r \end{array} \\ \left[\begin{array}{c:c} \mathbf{B}_{ll} & \mathbf{B}_{lr} \\ \hdashline \mathbf{B}_{rl} & \mathbf{B}_{rr} \end{array}\right] \end{array} \begin{array}{c} l \\ r \end{array}, \qquad \mathbf{M}_1^{-1} = \begin{array}{c} \begin{array}{cc} l & r \end{array} \\ \left[\begin{array}{c:c} \mathbf{E}_{ll} & \mathbf{E}_{lr} \\ \hdashline \mathbf{E}_{rl} & \mathbf{E}_{rr} \end{array}\right] \end{array} \begin{array}{c} l \\ r \end{array},$$

$$\mathbf{M}_2 = \begin{array}{c} \begin{array}{cc} l & r \end{array} \\ \left[\begin{array}{c:c} \mathbf{C}_{ll} & \mathbf{C}_{lr} \\ \hdashline \mathbf{C}_{rl} & \mathbf{C}_{rr} \end{array}\right] \end{array} \begin{array}{c} l \\ r \end{array}, \qquad \mathbf{M}_2^{-1} = \begin{array}{c} \begin{array}{cc} l & r \end{array} \\ \left[\begin{array}{c:c} \mathbf{F}_{ll} & \mathbf{F}_{lr} \\ \hdashline \mathbf{F}_{rl} & \mathbf{F}_{rr} \end{array}\right] \end{array} \begin{array}{c} l \\ r \end{array},$$

we obtain the mixed cumulants of Q_1 and Q_2

$$\begin{aligned}
\omega_{10} &= \operatorname{tr} \mathbf{M}_{rr}^{-1} \mathbf{B}_{rr} + \gamma_1' \mathbf{B}_{rr} \gamma_1 + (\theta_l - \hat{\theta}_{1l})' \mathbf{E}_{ll}^{-1} (\theta_l - \hat{\theta}_{1l}) \\
\omega_{01} &= \operatorname{tr} \mathbf{M}_{rr}^{-1} \mathbf{C}_{rr} + \gamma_2' \mathbf{C}_{rr} \gamma_2 + (\theta_l - \hat{\theta}_{2l})' \mathbf{F}_{ll}^{-1} (\theta_l - \hat{\theta}_{2l}) \\
\omega_{rs} &= 2^{r+s-1} (r+s-2)! \, [(r+s-1) \operatorname{tr} \mathbf{M}_{rr}^{-1} \mathbf{H}^{rs} \\
&\quad + (r\gamma_1 + s\gamma_2)' \mathbf{H}^{rs} (r\gamma_1 + s\gamma_2) - r\gamma_1' \mathbf{H}^{rs} \gamma_1 - s\gamma_2' \mathbf{H}^{rs} \gamma_2], \qquad r + s \geqslant 2,
\end{aligned} \tag{A9.1.15}$$

where

$$\begin{aligned}
\mathbf{H}^{rs} &= \mathbf{M}_{rr} (\mathbf{M}_{rr}^{-1} \mathbf{B}_{rr})^r (\mathbf{M}_{rr}^{-1} \mathbf{C}_{rr})^s \\
\gamma_1 &= (\tilde{\theta}_r - \hat{\theta}_{1r}) + (\mathbf{B}_{rr}^{-1} \mathbf{B}_{rl} - \mathbf{M}_{rr}^{-1} \mathbf{M}_{rl})(\theta_l - \hat{\theta}_{1l}) \\
\gamma_2 &= (\tilde{\theta}_r - \hat{\theta}_{2r}) + (\mathbf{C}_{rr}^{-1} \mathbf{C}_{rl} - \mathbf{M}_{rr}^{-1} \mathbf{M}_{rl})(\theta_l - \hat{\theta}_{2l}).
\end{aligned}$$

Using the results in (A9.1.15), we can express the marginal posterior distribution of θ_l as

$$p(\theta_l \mid \mathbf{y}) = \frac{|\mathbf{V}_{ll}|^{-1/2}}{(2\pi)^{\frac{1}{2}l}} \exp\left[-\tfrac{1}{2}(\theta_l - \tilde{\theta}_l)' \mathbf{V}_{ll}^{-1} (\theta_l - \tilde{\theta}_l)\right] \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \delta_{ij} \, v_1^{-i} \, v_2^{-j}, \qquad -\infty < \theta_l < \infty, \tag{A9.1.16}$$

with

$$\begin{aligned}
&\delta_{00} = 1, \qquad \delta_{10} = g_{10} - b_{10}, \qquad \delta_{01} = g_{01} - b_{01}, \\
&\delta_{11} = g_{11} - b_{11} - g_{10} b_{01} - g_{01} b_{10} + 2 b_{01} b_{10}, \\
&\delta_{20} = g_{20} - b_{20} - g_{10} b_{10} + b_{10}^2, \qquad \delta_{02} = g_{02} - b_{02} - g_{01} b_{01} + b_{01}^2, \text{ etc.,}
\end{aligned}$$

and the quantities g_{ij} are functions of the mixed cumulants ω_{ij} with the functional relationships exactly the same as those between b_{ij} and κ_{ij} shown in (A9.1.10).

It is seen that the leading term in the expansion is a multivariate Normal distribution $N_l(\tilde{\theta}_l, \mathbf{V}_{ll})$. When θ_l consists of only one variable, the quantities

δ_{ij} in (A9.1.16) are simply polynomials in that variable. Employing the well-known expression for the moments of a Normal variable, one can easily derive an asymptotic expression for the moments. In addition, probability integrals can also be approximated using methods given in Fisher's previously cited paper (1961b).

APPENDIX A9.2

Mixed Cumulants of the Two Quadratic Forms Q_1 and Q_2

From the joint cumulant generating function of the quadratic forms Q_1 and Q_2 given in (A9.1.8), we now derive the expressions for the mixed cumulants shown in (A9.1.9). In our development, we shall make use of the following lemma.

Lemma Let $\mathbf{P}_1$ be a $n \times n$ positive definite symmetric matrix and $\mathbf{P}_2$ be a $n \times n$ nonnegative definite symmetric matrix. Then, for sufficiently small i, we have

$$\log |\mathbf{I} - i\mathbf{P}_1\mathbf{P}_2| = -\sum_{r=1}^{\infty} \frac{i^r}{r} \operatorname{tr} (\mathbf{P}_1\mathbf{P}_2)^r. \tag{A9.2.1}$$

Employing the above lemma and for sufficiently small values of t_1 and t_2, we can expand the first term on the right of (A9.1.8) into

$$-\tfrac{1}{2} \log |\mathbf{I} - 2\mathbf{M}^{-1}(t_1\mathbf{M}_1 + t_2\mathbf{M}_2)| = \sum_{r=1}^{\infty} \frac{2^{r-1}}{r} \operatorname{tr} (t_1\mathbf{M}^{-1}\mathbf{M}_1 + t_2\mathbf{M}^{-1}\mathbf{M}_2)^r. \tag{A9.2.2}$$

In (A9.1.8), the quadratic form $t_1\boldsymbol{\eta}_1'\mathbf{M}_1\boldsymbol{\eta}_1$ can be written

$$\begin{aligned} t_1\boldsymbol{\eta}_1'\mathbf{M}_1\boldsymbol{\eta}_1 &= t_1\boldsymbol{\eta}_1'\mathbf{M}_1(\mathbf{M} - 2t_1\mathbf{M}_1 - 2t_2\mathbf{M}_2)^{-1}(\mathbf{M} - 2t_1\mathbf{M}_1 - 2t_2\mathbf{M}_2)\boldsymbol{\eta}_1 \\ &= t_1\boldsymbol{\eta}_1'\mathbf{M}_1(\mathbf{M} - 2t_1\mathbf{M}_1 - 2t_2\mathbf{M}_2)^{-1}\mathbf{M}\boldsymbol{\eta}_1 - 2t_1^2\boldsymbol{\eta}_1'\mathbf{M}_1(\mathbf{M} - 2t_1\mathbf{M}_1 \\ &\quad -2t_2\mathbf{M}_2)^{-1}\mathbf{M}_1\boldsymbol{\eta}_1 - 2t_1t_2\boldsymbol{\eta}_1'\mathbf{M}_1(\mathbf{M} - 2t_1\mathbf{M}_1 - 2t_2\mathbf{M}_2)^{-1}\mathbf{M}_2\boldsymbol{\eta}_1. \end{aligned} \tag{A9.2.3}$$

Similarly,

$$\begin{aligned} t_2\boldsymbol{\eta}_2'\mathbf{M}_2\boldsymbol{\eta}_2 &= t_2\boldsymbol{\eta}_2'\mathbf{M}_2(\mathbf{M} - 2t_1\mathbf{M}_1 - 2t_2\mathbf{M}_2)^{-1}\mathbf{M}\boldsymbol{\eta}_2 - 2t_2^2\boldsymbol{\eta}_2'\mathbf{M}_2(\mathbf{M} - 2t_1\mathbf{M}_1 \\ &\quad - 2t_2\mathbf{M}_2)^{-1}\mathbf{M}_2\boldsymbol{\eta}_2 - 2t_1t_2\boldsymbol{\eta}_2'\mathbf{M}_2(\mathbf{M} - 2t_1\mathbf{M}_1 - 2t_2\mathbf{M}_2)^{-1}\mathbf{M}_1\boldsymbol{\eta}_2. \end{aligned} \tag{A9.2.4}$$

Thus, the expression in (A9.1.8) becomes

$$\begin{aligned} \kappa(t_1, t_2) &= \sum_{r=1}^{\infty} \frac{2^{r-1}}{r} \operatorname{tr} (t_1\mathbf{M}^{-1}\mathbf{M}_1 + t_2\mathbf{M}^{-1}\mathbf{M}_2)^r + t_1\boldsymbol{\eta}_1'\mathbf{M}_1(\mathbf{I} - 2t_1\mathbf{M}^{-1}\mathbf{M}_1 \\ &\quad -2t_2\mathbf{M}^{-1}\mathbf{M}_2)^{-1}\boldsymbol{\eta}_1 + t_2\boldsymbol{\eta}_2'\mathbf{M}_2(\mathbf{I} - 2t_1\mathbf{M}^{-1}\mathbf{M}_1 - 2t_2\mathbf{M}^{-1}\mathbf{M}_2)^{-1}\boldsymbol{\eta}_2 \\ &\quad - 2t_1t_2(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2)'\mathbf{M}_1(\mathbf{I} - 2t_1\mathbf{M}^{-1}\mathbf{M}_1 - 2t_2\mathbf{M}^{-1}\mathbf{M}_2)^{-1}\mathbf{M}^{-1}\mathbf{M}_2(\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2). \end{aligned} \tag{A9.2.5}$$

Since $\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2$, it is easy to see that the matrix $\mathbf{M}_1\mathbf{M}^{-1}\mathbf{M}_2$ is symmetric. By virtue of this property, we have

$$(t_1\mathbf{M}^{-1}\mathbf{M}_1 + t_2\mathbf{M}^{-1}\mathbf{M}_2)^r = \sum_{i=0}^{r} \binom{r}{i} t_1^i t_2^{r-i}(\mathbf{M}^{-1}\mathbf{M}_1)^i(\mathbf{M}^{-1}\mathbf{M}_2)^{r-i} \tag{A9.2.6}$$

and, for sufficiently small values of t_1 and t_2,

$$(\mathbf{I} - 2t_1\mathbf{M}^{-1}\mathbf{M}_1 - 2t_2\mathbf{M}^{-1}\mathbf{M}_2)^{-1} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} 2^{i+j}t_1^i t_2^j \binom{i+j}{i}(\mathbf{M}^{-1}\mathbf{M}_1)^i(\mathbf{M}^{-1}\mathbf{M}_2)^j. \tag{A9.2.7}$$

Substituting (A9.2.6–7) into (A9.2.5) and after a little rearrangement, we find

$$\begin{aligned}\kappa(t_1, t_2) = 1 &+ \sum_{r=1}^{\infty} 2^{r-1}t_1^r \left[\frac{1}{r}\operatorname{tr}(\mathbf{M}^{-1}\mathbf{M}_1)^r + \boldsymbol{\eta}_1'\mathbf{M}(\mathbf{M}^{-1}\mathbf{M}_1)^r\boldsymbol{\eta}_1\right] \\ &+ \sum_{r=1}^{\infty} 2^{r-1}t_2^r \left[\frac{1}{r}\operatorname{tr}(\mathbf{M}^{-1}\mathbf{M}_2)^r + \boldsymbol{\eta}_2'\mathbf{M}(\mathbf{M}^{-1}\mathbf{M}_2)^r\boldsymbol{\eta}_2\right] \\ &+ \sum_{r=1}^{\infty}\sum_{s=1}^{\infty} 2^{r+s-1}t_1^r t_2^s \frac{(r+s-2)!}{r!s!}[(r+s-1)\operatorname{tr}\mathbf{M}^{-1}\mathbf{G}^{rs} \\ &+ (r\boldsymbol{\eta}_1 + s\boldsymbol{\eta}_2)'\mathbf{G}^{rs}(r\boldsymbol{\eta}_1 + s\boldsymbol{\eta}_2) - r\boldsymbol{\eta}_1'\mathbf{G}^{rs}\boldsymbol{\eta}_1 - s\boldsymbol{\eta}_2'\mathbf{G}^{rs}\boldsymbol{\eta}_2],\end{aligned} \tag{A9.2.8}$$

where

$$\mathbf{G}^{rs} = \mathbf{M}(\mathbf{M}^{-1}\mathbf{M}_1)^r(\mathbf{M}^{-1}\mathbf{M}_2)^s.$$

Upon differentiating (A9.2.8), we obtain

$$\kappa_{r0} = 2^{r-1}(r-1)![\operatorname{tr}(\mathbf{M}^{-1}\mathbf{M}_1)^r + r\boldsymbol{\eta}_1'\mathbf{M}(\mathbf{M}^{-1}\mathbf{M}_1)^r\boldsymbol{\eta}_1], \tag{A9.2.9}$$

$$\kappa_{0s} = 2^{s-1}(s-1)![\operatorname{tr}(\mathbf{M}^{-1}\mathbf{M}_2)^s + s\boldsymbol{\eta}_2'\mathbf{M}(\mathbf{M}^{-1}\mathbf{M}_2)^s\boldsymbol{\eta}_2], \tag{A9.2.10}$$

$$\begin{aligned}\kappa_{rs} = 2^{r+s-1}(r+s-2)![&(r+s-1)\operatorname{tr}\mathbf{M}^{-1}\mathbf{G}^{rs} + (r\boldsymbol{\eta}_1 + s\boldsymbol{\eta}_2)'\mathbf{G}^{rs}(r\boldsymbol{\eta}_1 + s\boldsymbol{\eta}_2) \\ &- r\boldsymbol{\eta}_1'\mathbf{G}^{rs}\boldsymbol{\eta}_1 - s\boldsymbol{\eta}_2'\mathbf{G}^{rs}\boldsymbol{\eta}_2], \qquad r, s \geqslant 1\end{aligned} \tag{A9.2.11}$$

which can then be combined into the expressions given in (A9.1.9).