pp. 70–73). When the precisions are unequal, various tests such as that of Scheffe (1943) have been proposed.

Suppose $S_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ is a random sample of size n_i from a $n(\theta_i, \tau^{-1}{}_i)$ population, where i = 1, 2, then how does one provide posterior inferences for the parameters? The likelihood function for $\theta = (\theta_1, \theta_2)$ and $\tau = (\tau_1, \tau_2)$ is

$$L(\theta,\tau|s) \propto \prod_1^2 \tau_i^{n_i/2} e^{-\tau_i/2 \sum_{j=1}^{n_i} (x_{ij}-\theta_i)^2},$$

$$\theta \in R^2, \quad \tau_1 > 0, \quad \tau_2 > 0 \tag{3.25}$$

where $s = (S_1, S_2)$, which implies

$$\xi(\theta,\tau) \propto \prod_1^2 \tau_i^{\alpha_i - 1}\ e^{-\tau_i\beta_i}\ \tau_i^{1/2}\ e^{-(\tau_i/2)P_i(\theta_i - \mu_i)^2}, \tag{3.26}$$
$$\theta \in R^2, \quad \tau_1 > 0, \quad \tau_2 > 0$$

is the conjugate density, a product of two normal-gamma densities, thus, a priori, (θ_1, τ_1) and (θ_2, τ_2) are independent. It follows, a posteriori, that (θ_1, τ_1) and (θ_2, τ_2) are independent with density,

$$\xi(\theta,\tau) \propto \prod_{i=1}^2 \tau_i^{(n_i+2\alpha_i+1/2)-1} \exp\frac{\tau_i}{2}\{2\beta_i + \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 + P_i(\theta_i - \mu_i)^2 + n_i(\theta_i - \bar{x}_i)^2\}, \tag{3.27}$$

where $\tau_i > 0$, $\theta \in R^2$, hence (θ_i,τ_i) has a normal-gamma distribution with parameters α_i, β_i, μ_i, and P_i, i = 1, 2. The joint marginal p.d.f. of and is

$$\xi(\theta|s) \propto \prod_{i=1}^2 \{2\beta_i + \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 + \frac{n_i P_i(\mu_i - \bar{x}_i)^2}{(n_i + P_i)} + [\theta_i - (n_i + P_i)^{-1}(P_i\mu_i + n_i\bar{x}_i)]^2(n_i + P_i)\}^{-(n_i+2\alpha_i+1)/2} \tag{3.28}$$

and θ_i has a univariate t distribution with $n_i + 2\alpha_i$ degrees of freedom, location

$$E(\theta_i|s) = (n_i + P_i)^{-1}(P_i\mu_i + n_i\bar{x}_i)$$

and precision

$$P(\theta_i|s) = \frac{(n_i+2\alpha_i)(n_i+P_i)}{2\beta_i \sum_1^{n_i}(x_{ij}-\bar{x}_i)^2 + n_i P_i(\mu_i-\bar{x}_i)^2(n_i+P_i)^{-1}}, \quad i = 1, 2.$$

Furthermore, τ_1 and τ_2 are independent and τ_i is gamma with parameters $\alpha' = (n_i + 2\alpha_i)/2$ and β' where

$$2\beta' = 2\beta_i + \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 + n_i P_i(\mu_i - \bar{x}_i)^2(n_i + P_i)^{-1}, \quad i = 1, 2.$$

In this type of problem, often $\gamma = \theta_1 - \theta_2$ is the parameter of interest and and are nuisance parameters. How would one find the posterior density of γ? See Box and Tiao (1973, pages 104–109) where they explain an approximation of Patil (1964) to the posterior density of γ and illustrate the approximation with a “textile” example. When reading Box and Tiao the reader should remember, they are using an improper prior density

$$\xi(\theta,\tau) \propto \tau_1^{-1}\tau_2^{-1}, \quad \theta \in R^2, \quad \tau_1 > 0, \tau_2 > 0$$

for the parameters and not the conjugate prior density (3.26); however, their analysis can be duplicated by letting $\alpha_. \to -1/2$, $\beta_i \to 0$, and $P_. \to 0$, i = 1, 2, in the posterior density (3.27). When letting the hyperparameters approach these limits in the posterior density, the normalizing constant of the density should be ignored.

Now returning to the distribution of γ, let $\xi_i(\theta_i|s)$ be the posterior density of θ_i, then we have seen ξ_i is a t density with $n_i + 2\alpha_i$ degrees of freedom, location $E(\theta_i|s)$ and precision $P(\theta_i|s)$, i = 1, 2, hence the posterior density of γ is

$$g(\gamma|s) = \int_R \xi_1(\gamma + \Delta|s)\xi_2(\Delta|s)d\,\Delta, \quad \gamma \in R \tag{3.29}$$

and the integral cannot be expressed in terms of simple functions and must be computed by numerical integration. Patil's approximation to $g(\gamma|s)$ avoids the numerical integration of (3.29) and expresses $g(\gamma|s)$ as a t distribution. Box and Tiao compare the Patil approximation to the exact distribution of γ|s the numerical integration of (3.29) and expresses $g(\gamma | s)$ as a t distribution. Box and Tiao compare the Patil approximation to the exact distribution of $\gamma | s$ and show the approximation works well.

How would one develop a normal approximation to the distribution of $\gamma | s$?

Concluding Remarks

This chapter begins with the Bayesian analysis of two normal populations, where either the means are distinct or the precisions are distinct, or both.

In the case the means are distinct, but the precisions are equal, the marginal posterior distribution of the means is a bivariate t distribution with density (3.13) and the marginal posterior distribution of the common precision parameter is a gamma. Also, the marginal posterior distributions of θ_1, θ_2, and $\gamma = \theta_1 - \theta_2$ are univariate t densities and Bayesian inferences for these parameters are easily obtained.

If the precision parameters are distinct and the means are equal, the marginal posterior density of the common mean is a univariate 2/0 poly-t density (3.20), and the joint posterior density of the precisions is given by (3.21), which is a bivariate-type gamma density. A numerical

example gave three estimates of the common mean for various combinations of the parameters of the prior distribution and the estimates are listed in Table 3.1. One of these estimates was calculated from the θ component of the mode of the posterior distribution of all the parameters. An iterative procedure of Lindley and Smith (1972) produced this modal estimate of θ while the other estimates, the marginal mode and mean, were calculated from the marginal posterior density of θ.

This section was concluded with a Bayesian analysis of the Behrens-Fisher problem. Using a conjugate prior density for the parameters, it was shown that the joint posterior density (3.27) is such that (θ_1, τ_1) and (θ_2, τ_2) are independent and each has a normal-gamma distribution, however the distribution of $\gamma = \theta_1 - \theta_2$ must be evaluated numerically. The marginal posterior distributions for and τ_2 are each gamma, which follows from (3.27) but inferences for these parameters were not made. How would one compare τ_1 and τ_2? Perhaps the ratio of these two parameters could be used.

LINEAR REGRESSION MODELS

One of the standard models the statistician must know how to use is the linear regression model. With such models one attempts to build a connection between a dependent variable Y and a set of m ($m \geqslant 1$) independent variables. Often, but not always, the average value of Y is a linear function of a set of p unknown parameters θ, thus if there are n observations $(Y_i, x_{i1}, x_{i2}, \ldots, x_{im})$ on these $m + 1$ variables, the model relating Y to m independent variables is

$$Y = x\theta + e, \tag{3.30}$$

where Y is the $n \times 1$ observation vector, x is the $n \times p$, $(p = m + 1)$, matrix of values of the independent variables, θ a $p \times 1$ parameter vector, and e is a $n \times 1$ vector of unobservable random variables with zero means. If, in addition, the n observations on y are independent and normally distributed with a common precision τ, model (3.30) is the general linear model of Chapter 1. The prior, posterior and predictive analysis for such a model was developed in Chapter 1.

In this section, a detailed Bayesian analysis of the linear regression model will be given. That is, using the normal-gamma prior density for θ and τ:

(i) Point estimates for θ and τ will be found.
(ii) HPD interval estimates for the parameters will be constructed.
(iii) Point and interval estimates of $E(Y | x)$ for a given set X of values of the independent variables will be constructed.
(iv) Prediction intervals for future Y values for a given set of values of the independent variables are to be derived.
(v) Numerical examples will illustrate the above four techniques.

Of course, not all regression studies involve the general linear model (3.30), since sometimes the $E(Y | x)$ is a nonlinear function of x and often the observations on Y are correlated, that is the precision matrix $P(Y | x)$ is not τI_n, but is nondiagonal. Nonlinear regression models and models with correlated observations will be presented later.

Simple Linear Regression

Let

$$y_i = \theta_1 + \theta_2 + \theta_2 x_i + e_i, \quad i = 1, 2, \ldots, n$$

where y_i is the i-th observation on a dependent variable Y, x_i is the i-th observation on an independent variable x, θ_1 and are unknown intercept and slope coefficients respectively, and e_i is the i-th error term. If the e_i are n.i.d. $(0,\tau^{-1})$ variables, then the general linear model and the theory of Chapter 1 apply. In terms of this model Y is the $n \times 1$ vector of observations of the dependent variable, x is a $n \times 2$ matrix

$$x = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{pmatrix}$$

$\theta = (\theta_1, \theta_2) \in R^2$, and e is the $n \times 1$ vector of errors e_i, where $e \sim n(0, \tau^{-1} I_n)$.

If one uses a normal-gamma prior density for θ and τ, the Bayesian analysis was developed in Chapter 1.

First, how does one develop point and interval estimates for θ_1, θ_2, and τ? From Chapter 1 the marginal posterior density of θ is a t with n + 2α degrees of freedom, location vector

$$\mu* = (x'x + p)^{-1}(x'y + p\mu) \tag{3.31}$$

and precision matrix

$$P(\theta|s) = \frac{(x'x + p)(n+2\alpha)}{2\beta + y'y - (x'y + p\mu)'(x'x + p)^{-1}(x'y + p\mu)} \tag{3.32}$$

where

$$x'x = \begin{pmatrix} n & \sum_1^n x_i \\ \sum_1^n x_i & \sum_1^n x_i^2 \end{pmatrix},$$

and

$$x'x = \sum_1^n x_i y_i.$$

If one uses an improper prior density

$$\xi(\theta, \tau) \propto 1/\tau, \quad \tau > 0, \quad \theta \in R^2, \tag{3.33}$$

the marginal posterior density of θ is a t with n − 1 degrees of freedom, location vector

$$\hat{\theta} = (x'x)^{-1}x'y \tag{3.34}$$

$$= \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix},$$

where

$$\hat{\theta}_1 = \bar{y} - \bar{\theta}_2\bar{x}$$

and

$$\hat{\theta}_2 = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_1^n (x_i - \bar{x})^2},$$

which is the usual least squares estimator of θ, and the precision matrix of θ is

$$P * (\theta|s) = \frac{x'x}{s^2}, \tag{3.35}$$

where

$$s^2 = \frac{y'y - y'x(x'x)^{-1}x'y}{(n-2)}$$

$$= \sum_{i=1}^n (y_i - \hat{\theta}_1 - \hat{\theta}_2 x_i)^2/(n-2),$$

and s^2 is the usual unbiased estimator of $\sigma^2 = 1/\tau$. The dispersion matrix of the marginal posterior distribution of θ is $s^2\,(x'x)^{-1}\,(n-1).\,(n-3)^{-1}$, $n > 3$.

We see if one uses an improper prior density for the parameters, the Bayesian analysis closely resembles the least squares or maximum likelihood approach to estimating the parameters.

To estimate θ from the joint posterior distribution, the vector $\hat{\theta}$ of least squares estimators seems appropriate, if one uses an improper prior density (3.33), but if one uses a conjugate prior density, then μ*, the mean of the marginal posterior density of θ seems appropriate. By letting $\alpha \to -1$, $\beta \to 0$, $p \to 0(2 \times 2)$ in the joint posterior density of θ and τ, which corresponds to the conjugate prior density, one will obtain the joint posterior density of θ and τ corresponding to the improper prior density (3.33). In particular, $\mu \to \theta$ (the least squares estimator of θ) as $p \to 0$ and as $\alpha \to -1$ also if $p \to 0$, and β 0, $P(\theta|s) \to (x'x)/s^2$, the precision matrix of the marginal posterior distribution of θ corresponding to the improper prior density (3.33).

If one is interested in τ, how should this parameter be estimated? We know the marginal posterior density of τ is

$$\xi(\tau|s) \propto \tau^{(n+2\alpha)/2-1} e^{-(\tau/2)\ [2\beta+y'y\text{-}(x')y+p\mu'(x'x+p)^{-1}(x'y+p\mu)\}}, \tag{3.36}$$

$$\tau > 0$$

if one uses a conjugate prior density with hyperparameters μ, p, α, and β. Since τ|s has a gamma density, the mean and mode of τ|s are distinct, thus if one uses the mean to estimate τ, the estimator is

$$E\,(\tau|s) = \frac{n+2\alpha}{2\beta + y'y \text{ - } (x'y + p\mu)'(x'x + p)^{-1}(x'y + p\mu)} \tag{3.37}$$

whereas the marginal posterior mode of τ is

$$M\,(\tau|s) = \frac{n+2\alpha-2}{2\beta + y'y - (x'y + p\mu)'(x'x + p)^{-1}(x'y + p\mu)} \tag{3.38}$$

What is the marginal posterior mean of τ^{-1} (which is the variance σ^2 along the regression line) and what is E(τ|s) if an improper prior density is used to express one's prior information about the parameters?

The marginal posterior densities for θ and τ are also useful in obtaining HPD regions. For example, what is the HPD interval for θ_2?

Since $\theta_2|s$ has a univariate t density, one may use Student's t-tables to find these intervals for θ_2 as well as for θ_1. With regard to τ, HPD intervals are more difficult to construct because the gamma marginal posterior distribution is not symmetric and numerical integration is necessary in order to find an HPD interval for τ. An approximate HPD interval is easily constructed from the chi-square tables since, a posteriori, $[2\beta + y'y - (x'y + p_\mu)'(x'x + p)^{-1}(x'y + P_\mu)]\tau/s$ has a chi-square distribution with n + 2α degrees of freedom.

Often, one is interested in estimating the average value of the dependent variable Y when the independent variable x = x*, where x* is known. Now

$$E(Y|s^*) = (1, x^*) \begin{pmatrix} \theta_1 \\ \theta_1 \end{pmatrix} = \gamma$$

is the average value of Y when x = x*, and since θ has a bivariate t distribution with n + 2α degrees of freedom, location vector μ* and precision matrix P(θ|s), γ has a t distribution with n + 2α degrees of freedom, location (1, x*)μ* and precision

$$P(\gamma|s) = [(1, x*)P^{-1}(\theta|s)(1, x*)']^{-1} \tag{3.39}$$