3

THE TRADITIONAL UNEAR MODELS

INTRODUCTION

A first one-semester course in linear models introduces the student to simple and multiple regression models and fixed models of designed experiments, and if time permits to random and mixed models of designed experiments. These are the so-called standard or traditional linear models that the statistician is expected to know and be able to use.

The Bayesian analysis of the general linear model was introduced in Chapter 1 and will be used to analyze the regression models and fixed models for designed experiments, and these are to be studied in this chapter. The random and fixed models will be examined in the following chapter.

First, elementary models for one and two normal populations are given and the Bayesian approach is illustrated with the prior, posterior, and predictive analysis.

Secondly, simple and multiple linear regression models are analyzed, then nonlinear regression models, and finally models for designed experiments including models containing concomitant variables are introduced and Bayesian inferences are provided and illustrated with numerical examples.

Prior information about the parameters of these models is specified by either a vague improper Jeffreys' density or a member of a class which is conjugate to the model. We have seen the conjugate class is the normal-gamma family of distributions and that the four parameters of this distribution must be specified from one's prior knowledge of the future experiment. This chapter will use Winkler's (1977) method of the prior density of a future observation to set the values of the parameters of the normal-gamma prior density.

Once having determined the prior distribution, the joint posterior distribution of all the parameters and the marginal distribution of certain subsets of the parameters will be found. Of course, from Chapter 1 the relevant distribution theory has been derived and all that is needed is to find the parameters of the various posterior distributions. For example, in a simple linear regression model, the intercept and slope parameters will have a bivariate t distribution, while the marginal distribution of these parameters will be univariate t distributions. Having the sample information and the parameters of the prior distribution will allow one to know the precision matrix, degrees of freedom, and mean vector of the joint posterior distribution of the slope and intercept coefficients of the model. Thus, this will allow one to test hypotheses about the parameters, using HPD regions, and construct confidence intervals for these parameters as well as others such as the average response of the dependent variable at selected values of the regressor variables.

To predict a future value of the dependent variable, the Bayesian predictive density will give point and interval forecasts.

This type of analysis will be repeated for multiple linear regression models and those for some designed experiments and accompanied by numerical examples which will provide tables and graphs of posterior and predictive densities.

The reader should consult Box and Tiao (1973), Lindley (1965) and Zellner (1971) for a similar presentation along the lines developed in this chapter.

PRIOR INFORMATION

Two types of prior information will be used in this chapter. First, Jeffreys' improper density

$$\xi(\theta,\tau) \propto 1/\tau, \quad \theta \in R^p, \quad \tau > 0 \tag{1.13}$$

is to be used when very little prior information is available for the parameters θ and τ of the linear model, and when one is more informed, a priori, recall

$$\xi(\theta,\tau) = \xi_1(\theta|\tau)\xi(\tau), \quad \tau > 0, \quad \theta \in R^p \tag{1.5}$$

where

$$\xi_1(\theta|\tau) \propto \tau^{p/2}\exp\ -\tfrac{\tau}{2}(\theta-\mu)'p(\theta-\mu), \quad \theta \in R^p \tag{1.6}$$

and

$$\xi_2(\tau) \propto \tau^{\alpha-1}\ e^{-\tau\beta}, \quad \tau > 0 \tag{1.7}$$

is the normal-gamma density with known parameters μ e R^p, p a symmetric positive definite matrix, α > 0, and β > 0. Of course the difficulty with using this conjugate class is that the parameters need to be specified, thus how does one know what values to assign to the hyperparameters?

Consider one way advocated by Winkler (1977) but what is to follow is my interpretation. Thus consider the model

$$Y = X\theta + e \tag{1.1}$$

which describes the experiment which will be run in the future, where Y is n × 1, X is n × p, θ is p × 1, and e is $n(0,\tau I_n)$. Also consider the model

$$Z = X*\theta + \varepsilon, \tag{3.1}$$

where Z is m × 1, X* is m ×p, θ is p × 1, and ε ~ $(0, \tau^{-}I_m)$, and where θ and τ are the parameters of both models. Note θ and τ have the same meaning in both models. The latter model (3.1) thus describes a related experiment (related by the parameters θ and τ) which perhaps is a past experiment with past data Z when the design matrix was X*. On the other hand (3.1) is perhaps a hypothetical experiment with hypothetical observations Z when the design matrix is X*. The hypothetical data Z are supplied by the experimenter and in this situation, if Z is the experimenter's guess or prediction of Y, then the

matrix is X*. The hypothetical data Z are supplied by the experimenter and in this situation, if Z is the experimenter's guess or prediction of Y, then the design matrix is X(= X*).

What is the predictive prior density of Z? It is the marginal density of Z, which will depend on the hyperparameters μ, p, α, and β. The conditional density of Z given θ and τ is from (3.1).

$$g(z|\theta,\tau) \propto \tau^{m/2}\exp -\tfrac{\tau}{2}(z-x*\theta)'(z-x*\theta), \quad z\in R^m \tag{3.2}$$

thus the marginal density of Z is

$$h(z|\mu,p,\alpha,\beta) \propto \frac{1}{|(z-A^{-1}B)'A(z-A^{-1}B)+C-B'A^{-1}B|^{(m+2\alpha)/2}}, \quad z\in R^m, \tag{3.3}$$

where

$$A = I - X*(X*'X* + p)^{-1}X*', \tag{3.4}$$

$$B = X*(X*'X* + p)^{-1}p\mu, \tag{3.5}$$

and

$$C = \mu'p\mu + 2\beta - \mu'p(X*'X* + p)^{-1}p\mu. \tag{3.6}$$

Since the prior predictive density (3.3) of Z is a m-variate multivariate t with 2α degrees of freedom, location

$$E(Z) = A^{-1}B$$

and precision matrix

$$P(Z) = \frac{(2\alpha)A}{C-B'A^{-1}B}$$

and since these moments depend on the hyperparameters, one may choose them in such a way that the equations

$$A^{-1}B = X*(X*'X*)^{-1}X*'Z \tag{3.7}$$

and

$$\frac{(2\alpha)A}{C-B'A^{-1}B} = \frac{Z'[I-X*(X*'X*)^{-1}X*']ZI_m}{(m-p)} \tag{3.8}$$

are satisfied; however, the solution is not unique. The right-hand side of these two equations were obtained from (3.1) as follows. From (3.1) $E(Z|\theta) = X*\theta$ and this is estimated by the right-hand side of (3.7), since the usual estimate of θ is $(X*'X*)^{-1}X*'Z$. Also from (3.1), the dispersion matrix of Z is τI_m, thus its precision matrix is $\tau^{-1}I_m$ and τ^{-}, the variance is usually estimated by the residual sum of squares of (3.8).

Still a much easier way to set the values of the hyperparameters is to notice that from (3.1), the usual estimators of θ and τ^{-1} are

$$\theta* = (X*'X*)^{-1}X*'Z \tag{3.9}$$

and

$$(\tau^{-1})* = \frac{Z'Z-Z'X*(X*'X*)^{-1}X*'Z}{(m-p)}, \tag{3.10}$$

and since the prior mean of θ is μ, choose $\mu = \theta*$. In the same way, since the prior mean of τ^{-1} is $\beta(\alpha-1)$, $(\alpha>1)$, choose α and β such that $\beta(\alpha-1) = (\tau^{-1})*$. Of course the choice of α and β is not unique.

The prior dispersion of θ is $3p(\alpha-1)^{-1}$, thus choose p so that $(\tau*)^{-1}p = (\tau*)^{-1}(X*'X*)^{-1}$ or let $p = (X*'X*)^{-1}$. This determines values of the hyperparameters from the past or hypothetical experiment (3.1), although the choice of α and β is not unique. Is there a way to choose α and β uniquely?

Of course it is important that Z be known, and this is not a problem if Z is from a past experiment, however if Z must be chosen hypothetically as the future value of Y, more subjectivity is involved because if Z is the future value of Y, Z must be supplied by the experimenter in the hypothetical future experiment

$$Z = X\theta + e,$$

which is the experiment to be conducted where Y (in place of Z) will be observed.

As Winkler emphasizes, the advantage of this method is that the values of the hyperparameters are determined indirectly in terms of future (or past) values of Y and not directly in terms of the hyperparameters. It is much easier to "guess" future values of Y or to observe past values of Y than it is to "guess" the values of the hyperparameters.

"guess" the values of the hyperparameters.

For additional information on choosing the hyperparameters of the conjugate prior density (1.5), Zellner (1980) uses a g-prior distribution, while Kadane et al. (1980) choose the hyperparameters from the quantiles of the prior predictive distribution.

To recapitulate one may either use an improper prior density or a conjugate prior density with which to express one's prior information about the parameters, and in the latter case, one may use the prior predictive density to set the values of the hyperparameters.

Consider a $n(\theta, \tau)$ population and suppose a normal-gamma conjugate prior $\xi(\theta, \tau)$ is appropriate, then how does one choose the hyperparameters when m past observations $Z_1, Z_2, \ldots Z_m$ are available? From the past data, θ and τ^- would be estimated by $\theta^*=\bar{z}$ and $(\tau^{-1})^*=\sum_1^m (z_i-\bar{z})^2/(m-1)$ then the hyperparameters could be chosen as follows: μ by $\bar{z}$, α and β to satisfy $\beta(\alpha-1)^{-1}=(\tau^{-1})^*$ and p (scalar) by m^{-1} where $\alpha > 1$.

NORMAL POPULATIONS

Bayesian inferences for one normal population were discussed in Chapter 1 so now several normal populations will be examined.

Suppose one has two normal populations, $n(\theta_1,\tau_1^{-1})$ and $n(\theta_2,\tau_2^{-1})$; then four cases arise, namely:

(1) $\theta_1=\theta_2, \tau_1=\tau_2$

(2) $\theta_1 \neq \theta_2, \tau_1=\tau_2$

(3) $\theta_1=\theta_2\ \tau_1 \neq \tau_2$

(4) $\theta_1 \neq \theta_2\ \tau_1 \neq \tau_2$.

In the first case the populations are identical and there is actually only one population, therefore only the three remaining cases are of interest.

Two Normal Populations, Distinct Means

Let $s_1 = (x_{11}, x_{12}, \ldots, x_{1n_1})$ be a random sample from a $n(\theta_1, \tau^{-1})$ population and suppose $s_2 = (x_{21}, x_{22}, \ldots, x_{2n_2})$ is a random sample from the second $n(\theta_2, \tau^{-1})$ population, then the likelihood function of θ_1, θ_2 and τ is

$$L(\theta,\tau|s) \propto \tau^{(n_1+n_2)/2}\exp\ -\frac{\tau}{2}\left\{\sum_{i=1}^{n_1}(x_{1i}-\theta_1)^2+\sum_{i=1}^{n_2}(x_{2i}-\theta_2)^2\right\}, \qquad (3.11)$$
$$\theta \in R^2,\quad \tau > 0$$

where $s = (S_1, S_2)$ and $\theta = (\theta_1, \theta_2)$ one uses the normal (bivariate)-gamma prior density of (θ, τ) suggested by (3.11), with parameters $\mu(2 \times 1)$, $p(2 \times 2)$, α and β, the joint posterior density of (θ, τ) is

$$\xi(\theta,\tau|s) \propto \tau^{(n+2\alpha+2)/2-1}\exp\ -\frac{\tau}{2}[(\theta-A^{-1}B)'A(\theta-A^{-1}B) \qquad (3.12)$$
$$+\ C-B'A^{-1}B],$$

where $n = n_1 + n_2$,

$$A=\begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}+P,$$
$$B=\begin{pmatrix} \Sigma X_{1i} \\ \Sigma X_{2i} \end{pmatrix}+P\mu,$$

and

$$C = 2\beta+\Sigma x_{1i}^2+\Sigma x_{2i}^2+\mu' P\mu.$$

Furthermore, the marginal posterior density of θ is a bivariate t density

$$\xi(\theta|s) \propto [\theta\text{-}A^{-1}B)'\,A(\theta-A^{-1}B)+C-B'A^{-1}B]^{-(n+2\alpha+2)/2},\theta \in R^2, \qquad (3.13)$$

thus θ has $n + 2\alpha$ degrees of freedom, location vector $A^{-1}B$ and precision matrix

$$P(\theta|s)=\frac{(n+2\alpha)A}{C-B'A^{-1}B}. \qquad (3.14)$$

In addition the marginal posterior distribution of τ is a gamma with parameters

$$\alpha' = (n+2\alpha)/2$$

and

$$\beta' = (C - B'A^{-1}B)/2.$$

How does one make posterior inferences about θ and τ? First consider θ, then the mean $A^{-1}B$ jointly estimates θ_1 and θ_2, but suppose one wants to estimate θ_1, ignoring θ_2. If $\theta|s$ has a bivariate t distribution then $\theta_1|s$ has a univariate t distribution (see DeGroot, Chapter 5, and the Appendix of this book) with $n + 2\alpha$ degrees of freedom, location parameter which is the first component of $A^{-1}B$, and precision
$p_1(\theta_1|s) = p_{11}(\theta|s) - p_{12}(\theta|s)p_{22}^{-1}(\theta|s)p_{21}(\theta|s)$ where p_{ij} is the ij-th-component of $p(\theta|s)$ given by (3.14).

Often $\gamma = \theta_1 - \theta_2$ is the parameter of interest, because γ measures the difference in the means of the two normal populations and if $\gamma = 0$, the two populations are identical. What is the posterior distribution oí γ? Since $\theta|s$ is a bivariate t, $\gamma|s$ has a univariate t distribution with $n + 2\alpha$ degrees of freedom, location $(1, -1)A^{-1}$