2

LINEAR STATISTICAL MODELS AND BAYESIAN INFERENCE

INTRODUCTION

This chapter will introduce the various linear models which are to be examined in the remaining chapters of this book. One class of models was introduced in Chapter 1, namely the so-called general linear model, which includes, as special cases, the fixed models, which are used for regression analysis and the analysis of designed experiments.

The regression and design models are the traditional ones in that they are the most frequently used in statistical practice. Many of the software packages such as SAS and BMD allow one to routinely analyze data with one of the traditional models.

A first cousin to the design model is the mixed model, which is often employed when some of the experimental factors are random, that is the levels of the factors are selected at random from a population of levels and one is interested in making inferences about the factor populations.

The time series models to be analyzed in this book are special Gaussian linear models, namely the regression model with autocorrelated errors, the ARMA class (autoregressive moving average models), and the distributed lag models of econometrics. During the past decade the ARMA models have been thoroughly examined from the classical analysis of Box and Jenkins (1970), while the other time series models, in addition to being analyzed from a classical perspective, have been approached, principally by Zellner (1971), from the Bayesian viewpoint.

Linear dynamic models are now being studied by statisticians, but in the past have received the most attention of engineers who are interested in communication theory, navigation systems, and tracking of satellites. This class of models should prove to be very useful in the statistical analysis of time series.

Linear models which have unstable parameters are studied in Chapter 7 under the title structural change, which is a term from the field of economics. If one is confident, a priori, that a change in model (population) parameters will occur, special techniques are necessary.

We will see that the Bayesian analysis of such models has been a major contribution to the analysis of data in this area.

Most of this book deals with univariate linear models and it is only in Chapter 8 that multivariate models are first encountered. Multivariate regression, design, time series models, and some econometric models are examined in detail.

The linear logistic model is also mentioned in this chapter but will not be studied.

In what is to follow, each of the above models will be defined. An example of each is given, and a preview of the Bayesian analysis of the particular model is described.

LINEAR STATISTICAL MODELS

Regression Models

The regression model is a special case of the general linear model

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

of Chapter 1, where Y is a $n \times 1$ vector of observations, X is a $n \times p$ known matrix, θ a $p \times 1$ unknown real parameter vector, and e a $n \times 1$ vector of observation errors. Regression models are employed to examine the relationship between a dependent variable y and q independent variables $x_1, x_2, \ldots, x_q$, thus one observes $(Y, x_1, x_2, \ldots, x_q)$ and the n observations are denoted by $(y_{i\cdot}, x_{1i}, x_{2i}, \ldots, x_q)$, i= 1,2,…,n, where $y_{i\cdot}$ is the value of Y when one observes $x_{1i}, x_{2i}, \ldots, x_{qi}$. Presumably Y is observed with an error, but the independent variables are observed without error.

In terms of the general model (1.1)

$$y_i = \beta_0 + \sum_{j=1}^{q} x_{ji}\beta_j + e_i, \quad i = 1, 2, \ldots, n, \tag{2.1}$$

and the i-th component of Y is $y_{i\cdot}$, of e, $e_{i\cdot}$, the first column of X is j, and the second column is $(x_{11}, x_{12}, \ldots, x_{1n})'$, etc., where $P = q + 1$.

Thus, one is assuming the average value of the dependent variable is a linear function of q independent variables if the n errors each have a zero mean. The regression model is examined in detail in Chapter 3, where a complete Bayesian analysis is performed. It is shown how to estimate and test hypotheses about the parameters of the model and how to forecast future observations, using either a vague density or a conjugate density to express prior information. Chapter 1 explained how one is to do a Bayesian analysis and these ideas are applied to the regression model and design models of Chapter 3.

The regression and design models are quite similar and each is expressed in terms of the general linear model, however, with the design model the design matrix X of (1.1) is used to indicate the presence or absence of the levels of several factors of the experiment, and the components of X are either zero or one.

The Design Models

If the experiment has only one factor with, say, k levels, the model is

$$y_{ij} = \theta_i + e_{ij}, \tag{2.2}$$

where $y_{ij\cdot}$ is the j-th observation on the i-th level, where i = 1,2, …,k, and j=1,2,…,n., and $n = \Sigma_{i=1}^{k} n_i$ is the number of observations and $e_{ij\cdot}$ is the error associated with the ij-th observation. Of course, this example can be expressed in terms of the general linear model, where $y = (y_{11}, y_{12}, \ldots, y_{1n_1}; \ldots; y_{k1}, y_{k2}, \ldots, y_{nk_k})$,

$$X = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 1 & 0 & \ldots & 0 \\ \vdots & & & \\ 1 & 0 & \ldots & 0 \\ & & \vdots & \\ 0 & 0 & \ldots & 1 \\ 0 & 0 & \ldots & 1 \\ & & \vdots & \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

is a n × k matrix, where the first n_1 components of the first column are ones and the remaining elements are zero. The last n_k components of the k-th column are ones, thus the elements of X are either zeros or ones with a one in each row. The elements indicate the presence or absence of the k levels of the factor for each of the n observations, and the theory of Chapter 1 is applied in Chapter 3 in order to analyze one- and two-factor experiments, such as the completely randomized design and completely randomized block design.

If the k levels of the experiment are the only ones of interest to the experimenter, that is inferences are confined to the k means θ., the model is called fixed, otherwise if the k levels are selected at random from a population of levels, the model is said to be mixed or random.

In the fixed version of the one-way model (2.2), the $e_{ij\cdot}$'s are independent normal variables with zero mean and unknown precision τ, and the k parameters θ_i represent the effects of the k levels of the factor. In the classical approach to this problem, the θ_i's are regarded as fixed unknown constants, if the model is fixed, otherwise they are thought of as unobservable random normal variables with zero mean and constant variance (the variance component) if the k levels are selected at random from a population of levels, hence if the variance component is zero the population is degenerate and there is no treatment effect.

From the Bayesian viewpoint in the fixed case, the θ_i's are treated as random variables with some known prior distribution, but in the random case, a prior density is put on the variance component of the model. Thus if one adopts a Bayesian approach the distinction between fixed and random is not as meaningful as it is when one adopts the classical approach.

Consider an example by Davies (1967, page 105), which Box and Tiao (1973, page 216) analyze from a Bayesian approach.

The experiment consists of learning what effect the batch to batch variation of a raw material has on the yield of the product produced. Five samples from each of six batches were examined for product yield and the experiment is thought of as random, because the six batches were selected at random from a population of batches. Clearly, one's inferences should not be confined only to the six batches of the experiment, but to the population of batches, which is characterized by a variance component which explains the batch-to-batch variation.

The Bayesian analysis of random and mixed models is given in Chapter 4, but the approach differs from the method of Box and Tiao (1973) who use numerical integration to isolate the marginal posterior distribution of each of the variance components. The results of Chapter 4 are based on the dissertation of Rajagopalan (1980) who found a way to isolate the posterior density of each variance component via analytical approximation.

The mixed model is given by

$$y = x\theta + \sum_{i=1}^{c} u_i b_i + e \tag{2.3}$$

where y is a n × 1 vector of observations, x is a n × p known matrix, θ a p × 1 unknown parameter vector, $u_{i\cdot}$ is a n × m. known matrix, the $b_{i\cdot}$'s are independent normal unobservable vectors each with a zero mean vector and b_i has precision matrix $\tau_i J_{m_i}$, and e is independent of the $b_{i\cdot}$'s and is a normal random vector with zero mean and precision matrix τI_n. Also, $\theta \in R^p$, $\tau_{i\cdot} > 0$, $\tau > 0$, and the c variance components are $\sigma_1^2 = \tau_i^{-1}$, while $\sigma^2 = \tau^{-1}$ is called the error variance.

The mixed model accommodates both fixed and random factors of a designed experiment and the θ vector has the levels of the fixed factors, while the levels of the c random factors are represented by $b_1, b_2, \ldots, b_c$. Box and Tiao (1973, page 341) introduce the additive mixed model

$$y_{ij} = \theta_i + c_j + e_{ij} \tag{2.4}$$

to analyze a car-driver experiment with eight drivers and six cars. The main response was the mileage per gallon of gasoline and each driver drove each of the six cars, where i = 1,2, …, 8 and j = 1,2, …, 6. Thus y_{ij} is the gasoline mileage when driver i drives car j, the θ_i's are unknown constants, the c_j's are normal independent random variables with zero mean and variance component σ_c^2, and the e_{ij}'s are independent normal variables with zero means and error variance σ^2. Thus model (2.4) is seen to be a special case of the mixed model (2.3), where the cars were selected from a population of cars with variance σ_c^2, but inferences about the drivers are confined to the eight drivers of the experiment.

Chapter 4 is concerned with such experiments and the posterior analysis consists of determining the posterior distribution of each variance component and the vector of fixed effects.

Time Series Models

One of the most useful models to analyze time series data is the p-th order autoregressive model

$$y(t)=\sum_{i=1}^{p}\theta_i y(t\text{-}i)+e_t, \tag{2.5}$$

where $y(t)$ is the observation at time t, θ_i is an unknown parameter vector, the e_t's, $t = 1,2, \ldots, n$, are independent $n(0, \tau^{-1})$ random variables, and $y(0)$, $y(-l), \ldots, y(l-p)$ are known constants.

Time series data are often correlated through time and models such as (2.5) allow one to introduce correlation into the model. For example, if $p = 1$, the correlation between observations s units apart is

$$\rho[y(t),y(t+s)]=\theta_1^s, |\theta_1| < 1, \tag{2.6}$$

and if $p \geqslant 2$, more complex correlations can be studied.

Still another useful model to analyze time series data is the q-th order moving average class

$$y_t = e_t - \sum_{i=1}^{q}\phi_i e_{t-1}, \quad t=1,2,\ldots,n, \tag{2.7}$$

where y_t is the observation at time t, the ϕ_i.'s are unknown real parameters (the moving average coefficients), and the e_t 's are independent $n(0, \tau^{-1})$ random variables (white noise). If $q = 1$, the correlation, see Box and Jenkins (1970), is given by

$$\rho(y_t,y_{t+s})=\begin{cases} -\frac{\phi_1}{1+\phi_1^2}, & s=1 \\ 0, & s\geqslant 2 \end{cases} \tag{2.8}$$

and the moving average class introduces another form of correlation for the observations.

The autoregressive model has been used for many years for analysis of time series data and has received attention from both the Bayesians, see Zellner (1971), and from classical procedures, see for example Box and Jenkins (1970). On the other hand the moving average model has not been explored from a Bayesian viewpoint but has received a lot of attention from other perspectives.

In Chapter 5, the Bayesian analysis of Zellner will be expanded for the autoregressive model, and the first and second order moving average models will be studied, where the posterior and predictive analysis will be derived.

The autoregressive and moving average models are combined into the ARMA class, thus an ARMA(1,1) model is given by