Printed by: dimitar.gueorguiev@nike.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

2.13) but the systems equation is amended to give

$$\theta(t)=G\theta(t-1)+Hx(t-1)+V(t), \tag{2.14}$$

where t = 1,2, …, H is a p × s known matrix, x(t – 1) a s × 1 control vector, G a p × p matrix, and V(t) ap × 1 error vector.

With navigation problems, one must accurately predict the position θ(t + 1) of the vehicle one time unit ahead, then make the necessary adjustments so that the actual position of the vehicle is at a predetermined location. This problem of navigation is called the control problem in the literature.

At time t, one has t observations y(l), y(2), …, y(t) and one must control the vehicle so that its position θ(t + 1) one time unit in the future is to be say T(t + l), thus, how should one choose x(t) so that θ(t + 1) is "close" to T(t + 1)? A plan to do this is called a control strategy and is well-developed in the literature.

Chapter 6 introduces the dynamic linear model and how it is applied to navigation and tracking problems. The presentation is from the perspective of a statistician and should be readable to people with a statistical training. The chapter reviews the Kalman filter, the control problem, nonlinear filtering, adaptive estimation (when some of the parameters of the system and observation equations are unknown), smoothing, and prediction.

The dynamic linear model is the most general of those considered in this book and indeed all of the others are special cases.

Other Linear Models

The chapter on other models, Chapter 8, includes many of the multivariate models which are familiar to theoretical and applied statisticians. For example, multivariate regression and design models as well as vector autoregressive and moving average processes are to be considered. For each univariate model considered in this book, there is a corresponding multivariate version.

The linear dynamic model is an example of a multivariate linear model and is to be examined in Chapter 6; however it is so important, so believes the author, that a separate chapter will be devoted to its analysis.

Let us consider the multivariate version of the p-th order autoregressive process, namely

$$y'(t)=\sum_{i=1}^{p} y'(t-i)\theta_i+e'(t) \tag{2.15}$$

where t = 1,2, …,n, y(t) is a m × 1 observation vector, y(0), y(−l), …, y(l − p) are known m × 1 vectors, the θ_i (i = 1,2, …, p) are unknown m × m matrices of real numbers where θ is nonzero, and the e(t) are independent $N(0,P^{-1})$ random vectors, where P is a m × m unknown positive definite symmetric precision matrix. This model will accommodate the simultaneous analysis of m time series observed at equally spaced points, and as with the univariate AR(p) process, the main objective of our analysis is to develop the joint posterior distribution of the parameters $\theta_1, \theta_2, \ldots, \theta_p$, and P and to predict future observations y(n+l), y(n + 2), …

It will be shown that the Bayesian analysis of the multivariate AR model is a straightforward generalization of the univariate AR model if one uses the natural conjugate prior density for the parameters.

In the univariate version of the model, the conjugate family is the normal-gamma class and the normal-Wishart for the vector version of the model, however, a constraint on the posterior precision matrix of $\theta_1, \theta_2, \ldots, \theta_p$ is imposed in the vector case (m ⩾ 2). See Zellner (1971) for a discussion of this situation.

Since the multivariate autoregression model is also a multivariate linear regression model, the same constraint occurs with that model, but except for this particular nuisance, no difficulties are encountered in the posterior and predictive analysis of the multivariate linear model.

The autoregressive process is, as mentioned earlier, the most often used to model time series data and the vector version is being used mere and more along with the vector ARMA model. Jenkins (1979, page 75) gives an interesting example of a bivariate time series

$$y(t)=\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}, \quad t=1,2,\ldots \tag{2.16}$$

where y_i.(t) is the sales of product i, i = 1,2. The products are competitive, and the observations are quarterly and consist of eleven years of data. Since the sales of one product affect those of the other (and conversely), a multivariate ARMA model

$$y'(t)-\sum_{i=1}^{p} y'(t-i)\theta_i = e\prime(t)-\sum_{j=1}^{q} e\prime(t-j)\phi_j \tag{2.17}$$

seems an appropriate tentative guess to explain the data where y(t) is the vector of observations at time t (the t-th quarter) consisting of the sales of the two products at that time, θ_i is a 2 × 2 unknown matrix which explains the dependence of each series on the other at time lag i, the e(t) are bivariate normal vectors with mean zero and unknown precision matrix P, and the ϕ_j's are unknown 2×2 moving average matrix coefficients. Also θ_p and ϕ_q are nonzero matrices. Jenkins explains the Box-Jenkins methodology of identification, estimation, diagnostic checking and forecasting in building a model for the sales data.

Box and Tiao (1981) have recently expanded the univariate analysis of Box and Jenkins (1970) to the vector ARMA case where an iterative procedure of identification, estimation, and diagnostic checking is proposed. Unfortunately, the Bayesian analysis of vector ARMA processes (which is also the situation with univariate models) has not progressed as far as the Box-Tiao methodology. With regard to the AR process, the Bayesian analysis is fairly easy to derive, but it is another story with the moving average process because its likelihood function is difficult to work with.

Other models considered in Chapter 8 are the regression model with autocorrelated errors and changing multivariate regression model. The regression model is

$$y_i' = \beta' x_i' + e_i', \quad i=1,2,\ldots,n \tag{2.18}$$

Printed by: dimitar.gueorguiev@nike.com. Printing is for personal, private use only. No part of this book may be reproduced or transmitted without publisher's prior permission. Violators will be prosecuted.

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

where y_i. is a m × 1 observation vector, β is a m × 1 unknown parameter vector, x_i. a m × m known matrix, and e_i. a m × 1 random vector which follows a vector AR(1) model, namely

$$e'_i = e'_{t-1}\theta + \varepsilon'_i,$$

where e_0 is known, the ε_i are independent $N(0,P^{-1})$ random m × 1 vectors, θ is a m × m unknown coefficient matrix, and P is an unknown precision matrix.

This model is often used to model time series which have a dependence on a set of m explanatory variables and which have the same correlation structure as an autoregressive time series of order one.

The linear logistic model is in a class by itself because it is quite different from the others since it is used to analyze counting or discrete data. Curiously, the multinormal logit model has not been examined, until recently by Zellner and Rossi (1982) from a Bayesian viewpoint, however, from the classical perspective a lot of work has been accomplished. For example in econometrics, Hausman and McFadden (1981) report that the multinomial logit model is the most used model specification.

The multinomial logit model is the usual multinomial model but with a logit parametrization of the class probabilities. Suppose there are n independent trials and on each exactly one of three events can occur with probabilities p_1, p_2, and p_3 where

$$p_i = e^{\beta Z_i} / \sum_{i=1}^{3} e^{\beta Z_i}, \quad i=1,2,3,$$

Z_1 = l, $Z_2 = 0 = Z_3$, and β is a scalar parameter, then the joint distribution of n_1, n_2, and n_3, where n_1. is the frequency of the i-th event in n trials ($n=n_1 + n_2 + n_3$)is multinomial. By reparametrizing the model in terms of $\theta = e^{\beta}/2$, Zellner (1982) and Zellner and Rossi (1982) present a convenient exact small-sample posterior analysis. Their parametrization by θ allows them to identify the inverted Beta as the conjugate class to the model, hence the posterior moments are easily computed. This is an elegant way to solve the problem, and their solution is quite significant since recently very little work on non-normal models has appeared.

The foregoing should give the reader a good idea of what models are going to appear and in what situations they are useful. The linear models appearing in this work are useful in the engineering, physical, social, mathematical, economic, and biological sciences and the following chapters will give the reader a good idea of how to do a Bayesian statistical analysis when one of these linear models is appropriate for the analysis of the data. It is assumed one knows that a particular model is indeed appropriate for a particular experimental situation, and the question of choosing the appropriate model, although very important, is not at issue. What is important is how one implements a Bayesian statistical analysis.

BAYESIAN STATISTICAL INFERENCE

The Bayesian analysis of a statistical problem, using the general linear model, was illustrated in Chapter 1, where the posterior and predictive analysis was done on the basis of either a proper prior density or the proper conjugate prior density. In this section we will discuss just what constitutes our Bayesian analysis when one adopts a linear model for the probability model of the observations, and what Bayesian inference is, as well as what it is not. The results of this section will apply to all the linear models which were introduced in the first part of this chapter.

In what is to follow, we will look at the history of Bayesian inference, the subjective interpretation of probability, the main ingredients of a Bayesian analysis, the various types of prior information, the implementation of prior information, and the advantages and disadvantages of Bayesian inference.

Some History of Bayes Procedures

Our subject's beginning originated with the Rev. Thomas Bayes who published only two papers on the subject. His "An Essay Toward Solving a Problem in the Doctrine of Chances" was published in 1764 and was concerned with the problem of making inferences about the parameter θ of n Bernoulli trials. Bayes assumed θ was uniformly distributed (by construction of a "billiard" table) and presented a way to compute P(a<θ

Some have interpreted the principle of insufficient reason to mean that if nothing is known about the values of θ, then one value of θ is as likely as another, and hence a uniform prior density is appropriate for θ. According to Stigler (1982), this is not what Bayes really meant, but instead knowing nothing about θ really means the marginal distribution of x is uniform.

In any case, such controversy over Bayes' article is not unusual. Beginning with Laplace (1951), many people have commented on the essay of Bayes including Pearson (1920), Fisher (1922), and Jeffreys (1939).

In the essay, Bayes used a uniform prior density for θ but it was constructed to be so, because θ was the horizontal component of the position of a ball tossed at random on a flat table; hence the prior distribution of θ can be given a frequency interpretation and there is very little controversy. The interesting question is that if one cannot use a frequency interpretation for the distribution of θ, a priori, how does one choose a prior for θ and is it appropriate to use a uniform distribution for θ, if one knows "nothing" about the possible values of θ? For example, I have a coin and its probability of landing "heads" on a toss is θ, can I use a uniform prior density for θ if I know "nothing" about θ? There are those that say "no," because if I know nothing about θ, then I know nothing about any function of θ, say $g(\theta) = \theta^2$

DO NOT COPY