$$f[y(n+1)|y(1),\ldots,y(n)] = \int_\Omega f[y(n+1)|y(1),y(2),\ldots,y(n),\theta] \times \xi(\theta|y(1),\ldots,y(n))\, d\theta,$$

where the first factor of the integrand is the conditional density of $y(n + 1)$ given θ and the previous observations, and the second factor is the posterior density of θ. The predictive density is the average of the conditional predictive density of $y(n + l)$, given θ, with respect to the posterior distribution of θ, that is if one knew θ, one would predict $y(n + 1)$ with $f[y(n + 1) |y(1), y(2),\ldots, y(n),\theta]$ but since θ is unknown this density is averaged over the posterior distribution of θ and a Bayesian forecast consists of an average of conditional predictions.

We have seen how this approach is used with the general linear model of Chapter 1. In what is to follow, the Bayesian predictive density is found for the time series models of Chapter 5 and the linear dynamic model of Chapter 6. Often with linear dynamic systems it is important to control the states (parameters) of the system at predetermined levels and to do this one must be able to forecast the future observations at each stage in the evolution of the system. The Bayesian predictive density plays an important role in the control and monitoring of physical and economic systems and its use will be demonstrated in Chapters 5 and 6.

Press (1982) and Zellner (1971) consider a control problem which is not discussed in this book. Suppose

$$Y = X\theta + Z\phi + e$$

is a linear model, where Y is the vector of observations, X is a matrix consisting of the observations on independent variables which cannot be controlled, θ is an unknown parameter vector, Z a matrix consisting of the values of control variables, ϕ another unknown parameter vector, and e an error vector. Under the assumption of normally distributed errors and a vague prior for the parameters, Zellner shows how to choose the control variable Z so that Y is "close" to some predetermined value. He adopts three control strategies to do this but in all three the Bayesian predictive density plays a crucial role. These problems, although important, will not be examined because the author is unable to add anything new and Zellner's account is excellent.

The type of control problem which occurs in Chapter 6 is similar to the following problem. Suppose $\{y(t): t = 1,2,\ldots, n,\ldots\}$ is a time series which follows the law

$$y(t) = \theta(t)y(t-1) + \varepsilon(t)$$

and

$$\theta(t) = G\theta(t-1) + H(t)x(t-1) + e(t)$$

where $t = 1,2,\ldots$ This is a dynamic AR(1) process where the time dependent autoregressive coefficients follow an AR(1) law. The observation errors $\varepsilon(t)$ and the system errors $e(t)$ are independent normally distributed random variables, $y(0)$ is known, G is known, and $H(t)$ is known, and at the initial stage, assume $\theta(0)$ is known. Among these and other assumptions, the problem is to control $\theta(1)$ at some level, say $T(l)$, but to do this one must have an idea of the value of $\theta(1)$ (as a function of $H(0)$), thus one must predict $\theta(1)$ (one-step-ahead). Since the posterior mean of $\theta(1)$ depends on $y(l)$ to predict $\theta(1)$, one averages this posterior mean over the Bayesian predictive distribution of $y(l)$, because $y(l)$ has yet to be observed. This procedure is applied at each stage in the evolution of the system.

Linear dynamic systems are very interesting models to study from a Bayesian perspective because in order to analyze them one must use the complete arsenal of Bayesian methodology.

Once the Bayesian predictive density is found, one predicts the future value $y(n + 1)$ by examining this distribution as one would do in making inferences for the parameters from their posterior distribution. If one wants one value to forecast $y(n + 1)$, perhaps the mean, median, or mode will do, and if one wants an interval forecast, one could construct an HPD (highest predictive density?) for $y(n + 1)$, analogous to the way one would find a region of highest posterior density for the parameters of the model.

The Bayesian predictive distribution for $y(n + 1)$ turns out to be a univariate, multivariate, or matrix t (depending on the dimension of $y(n + 1)$) for most of the linear models, with the exception of the linear models exhibiting structural change, where the predictive distribution is a mixture of t distributions. Other exceptions are the moving average process, the ARMA model and linear dynamic models.

SUMMARY

This chapter gives the reader a chance to see what this book is about. First, a description of each class of models is given with some applications provided illustrating their use. For example, we have seen that linear dynamic models are sophisticated ways for modeling the monitoring and control of many physical and economic variables and that models for structural change give one a degree of flexibility not possessed by the others.

Secondly, we see the way a Bayesian would analyze data which were generated by one of the linear models. The components of an analysis, namely, the probability model, prior information, the posterior and predictive analysis were carefully explained. Problems with choosing prior information were discussed and it was decided that the "principle of imaginary results" was a promising approach to assessing prior information, but that there was no easy and "quick" solution to this vexing problem. Further research along the lines suggested by Winkler (1980) and Kadane et al. (1980) should enhance the way one expresses their prior information about the parameters, and if these ideas can be made operational so that they become standard in the routine problems of regression, the analysis of variance, and time series, the Bayesian methodology of this book will easily be implemented.

EXERCISES

1. If $x_1, x_2,\ldots, x_n$ is a random sample from a population with a mass function

$$f(x|\theta) = \theta^x(1-\theta)^{1-x},\ x = 0,1,\quad 0 \leqslant \theta \leqslant 1$$

and

and

$$\xi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0 \leqslant \theta \leqslant 1, \quad \alpha < 0, \beta > 0$$

is the prior density of θ, find:

(a) The posterior density of θ,

(b) The predictive distribution of x_{n+1}, and

(c) The joint predictive distribution of x_{n+1}, x_{n+2},

(d) If $n = 4$, $x_1 = 1$, $x_2 = 1$, $x_3 = 0$, $x_4 = 0$, $\alpha = 1 = \beta$, plot the posterior density of θ and predictive mass function of x_{n+1}.

2. Consider an exponential population with density

$$f(x|\theta) = \theta e^{-\theta x}, \quad x > 0, \quad \theta > 0$$

and let $x_1, x_2, \ldots, x_n$ be a random sample,

(a) Examine the likelihood function for θ (based on $x_1, x_2, \ldots, x_n$) as a function of θ, where $\theta > 0$. What type of function is it?

(b) Upon examining the likelihood function, what is a conjugate prior density for θ?

(c) Using the conjugate prior for θ, derive the posterior density of θ and the predictive density of x_{n+1}.

(d) Suppose, a priori, that previous values of x have been observed, namely $x_1^* = 1$, $x_2^* = 6$, and $x_3^* = 4$ What values would you use for the hyperparameters of the conjugate prior density?

3. Let $x_1, x_2, \ldots, x_n$ be a random sample from a Poisson population

$$f(x|\theta) = e^{-\theta}\theta^{x}/x, \quad x > 0, 1, 2, \ldots$$

and suppose the mean θ has prior density

$$\xi(\theta) \propto \theta^{\alpha-1}e^{-\theta\beta}, \quad \alpha > 0, \quad \beta > 0, \quad \theta > 0.$$

(a) Is the above a conjugate prior density?

(b) Using the above prior density, find a HPD region for θ.

(c) Find the predictive mass function of x_{n+1}

(d) Give an interval forecast of x_{n+1}.

4. 4. Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population with density

$$f(x|\theta_1,\theta_2)=(\theta_2-\theta_1)^{-1}, \quad \theta_1 < x < \theta_2$$

where θ_1 and θ_2 are unknown real parameters. Give a conjugate prior density for θ_1 and θ_2 and use it to find the predictive density of x_{n+1} Refer to Chapter 9 of DeGroot (1970).

5. If x and y have a normal-gamma distribution, that is the conditional distribution of x given y is normal and the marginal distribution of y is gamma, find the conditional distribution of y given x. Also, find the upper, 01 point of the marginal distribution of x. Refer to the Appendix.

6. If $x_1, x_2, \ldots, x_n$ is a random sample of size n from a bivariate normal population with mean vector μ and prediction matrix P (2×2), and if the prior distribution of μ and P is a normal-Wishart, find the posterior distribution of the correlation coefficient.

7. The gamma population has density

$$f(x|\alpha,\beta)=\frac{\Gamma^{-1}(\alpha)}{\beta^{-\alpha}}x^{\alpha-1}e^{-x\beta}, \quad x > 0, \quad \beta > 0$$

where Γ is the gamma function.

(a) Find a conjugate prior density for α and β.

(b) If a is known, what is the conjugate class of prior densities for β?

(c) If α is known, and $x_1^*, x_2^*, \ldots, x_m^*$ are previous values of x (the population random variable), how would you choose values for the parameters of the conjugate prior density of β?

8. Is the sampling distribution of an estimator of θ irrelevant in making inferences about θ? Is the posterior distribution of θ relevant in making inferences about θ? Which approach is more relevant? Explain your answers carefully.

9. What are the controversial aspects of the Bayesian approach to statistical analysis?

10. Why isn't the Bayesian approach to statistics more popular? Is there an inherent defect in the Bayesian analysis of a statistical problem?

REFERENCES

Aykac, Ahmet and Carlo Brumat (1977). *New Developments in the Applications of Bayesian Methods*. North-Holland Publishing Company, New York.

Barnett, Vic (1982). *Comparative Statistical Inference*, second edition. John Wiley and Sons, New York.

Bayes, T. (1963). "An essay toward solving a problem in the doctrine of chances," *Philosophical Transactions Royal Society*, Vol. 53, pp. 320–418.

Berger, James O. (1980). *Statistical Decision Theory, Foundations, Concepts, and Methods*. Springer-Verlag, New York.

Bernardo, J. M. (1980). "A Bayesian analysis of classical hypothesis testing," in Bayesian Statistics: Proceedings of the First International Meeting, Valencia, 1979, edited by Bernardo, DeGroot, Lindley and Smith.

Bernoulli, J. (1713). *Ars Conjectandi*. Basel.

Box, G. E. P. (1980). "Sampling and Bayes inference in scientific modeling and robustness" (with discussion), *Journal of the Royal Statistical Society, Series A*, Vol. 143, part 4, pp. 383–430.

Box, G. E. P. and Gwilym M. Jenkins (1970). *Time Series Analysis, Forecasting, and Control*, Holden-Day, San Francisco.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Ma.

Box, G. E. P. and G. C. Tiao (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Ma.
Box, G. E. P. and G. C. Tiao (1981). "Modeling multiple time series," *Journal of the American Statistical Association*, Vol. 76, No. 376, pp. 802–816.
Broemeling, Lyle (1982). "Introduction," *Journal of Econometrics*, Vol. 19, pp. 1–5.
Davies, O. L. (1967). *Statistical Methods in Research and Production*, third edition, Oliver and Boyd, London.
Dawid, A. P., Stone, M. and J. W. Zedek (1973). "Marginalization paradoxes in Bayesian structural inference" (with discussion), *Journal of the Royal Statistical Society, B*, Vol. 35, pp. 189–233.
deFinetti, B. (1930). "Sulla proprieti conglomerativa delle probabilita subordinate," *Rend R. Inst. Lombardo (Milano)*, Vol. 63, pp. 414–418.
deFinetti, B. (1937). "Foresight: its logical laws, its subjective sources," *Annals de l'Institute Henri Poincare. Reprinted (in translation) in Kyburg and Smokler (1964).*
DeGroot, M. H. (1970). *Optimal Statistical Decisions*, McGraw-Hill, New York.
DeMorgan, August (1847). *Formal Logic*, Taylor and Walton, London.
Fienberg, S. E. and A. Zellner (1975). *Studies in Bayesian Econometrics and Statistics*. North-Holland Publishing Co., New York.
Fisher, R. A. (1922). "On the mathematical foundations of theoretical statistics," *Phil. Transactions Roy. Soc. London, A*, Vol. 222, pp. 309–362.