BAYESIAN INFERENCES WITH THE POLY-t DISTRIBUTION

By

MAT YUSOFF ABDULLAH

Bachelor of Science
University of Malaya
Malaysia
1971

Master of Science
University of the Philippines at Los Baños
Philippines
1974

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
July, 1982

BAYESIAN INFERENCES WITH THE

POLY-t DISTRIBUTION

Thesis Approved:

Lyle Broemeling

Thesis Adviser

Don Holbert

William H. Stewart

Dayll E. Ray

Norman N Durham

Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my gratitude to my adviser, Dr. Lyle D. Broemeling, for his encouragement, guidance and assistance. Also my appreciation to Dr. David L. Weeks, Dr. Donald Holbert, and Dr. Daryll E. Ray for being in my advisory committee. I am especially grateful to Dr. William H. Stewart for replacing Dr. Weeks in his absence during my oral examination.

I gratefully acknowledge indebtedness for the financial support from the Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia, through the years of my study at the Oklahoma State University.

Special thanks to Miss Barbara Newport for her careful and beautiful typing of my thesis.

Lastly, my love to my wife and two sons for their patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The probability density function of a p-dimensional multivariate t distribution with ν degrees of freedom, location vector μ and precision matrix T is given by

$$p(\theta|\nu,\mu,T) = \frac{\Gamma(\frac{\nu+p}{2})|T|^{1/2}}{\Gamma(\frac{\nu}{2})(\nu\pi)^{p/2}} [1 + \frac{1}{\nu}(\theta-\mu)'T(\theta-\mu)]^{-(\frac{\nu+p}{2})} , \qquad (1.1)$$

where $\theta \in R^p$, $\nu > 0$, $\mu \in R^p$ and T is p×p positive definite symmetric (PDS) matrix.

Following Dreze (1977), a Student kernel is defined as

$$t(\theta|\nu,\mu,M) = [1 + (\theta-\mu)'M(\theta-\mu)]^{-\frac{\nu}{2}} , \; \theta \in R^p ,$$

where $\nu > p > 0$, and M is a p×p PDS matrix. So (1.1) can be written as

$$p(\theta|\nu,\mu,T) = \frac{\Gamma(\frac{\nu+p}{2})|T|^{1/2}}{\Gamma(\frac{\nu}{2})(\nu\pi)^{p/2}} t(\theta|\nu+p,\mu,\nu^{-1}T) , \; \theta \in R^p .$$

The product form of p-dimensional k/0 poly-t density for θ is defined through its kernel by

$$p(\theta|\{\nu_j,\mu_j,M_j\},k) \propto \prod_{j=1}^{k} t(\theta|\nu_j,\mu_j,M_j), \qquad \theta \in R^p$$

$$\propto \prod_{j=1}^{k} [1 + (\theta-\mu_j)'M_j(\theta-\mu_j)]^{-\frac{\nu_j}{2}} \qquad (1.2)$$

which is the product of k p-dimensional Student kernels, where $\theta \in R^p$, $\nu_j > 0$, $\sum_{j=1}^{k} \nu_j > p$, M_j is p×p positive semi-definite symmetric (PSDS) matrix, and $\sum_{j=1}^{k} M_j$ PDS for every j = 1,2,...,k.

The ratio form of p-dimensional k/ℓ poly-t density for θ is defined through its kernel by

$$p(\theta|\{\nu_j,\mu_j,M_j\},k,\ell) \propto \frac{\prod_{j=1}^{k} t(\theta|\nu_j,\mu_j,M_j)}{\prod_{j=k+1}^{k+\ell} t(\theta|\nu_j,\mu_j,M_j)}$$

$$\propto \prod_{j=1}^{k} [1 + (\theta-\mu_j)'M_j(\theta-\mu_j)]^{-\frac{\nu_j}{2}} \prod_{j=k+1}^{k+\ell} [1 + (\theta-\mu_j)'M_j(\theta-\mu_j)]^{\frac{\nu_j}{2}}, \quad \theta \in R^p \qquad (1.3)$$

which is the ratio of k/0 and $\ell/0$ product form kernels, where $\theta \in R^p$, $\nu_j > 0$, $\sum_{j=1}^{k} \nu_j - \sum_{j=k+1}^{k+\ell} \nu_j > p$, M_j PSDS and $\sum_{j=1}^{k} M_j$ PDS for all j = 1,2, ...,k.

Since some of the properties of the Student distribution carry over to the poly-t distribution, the following is the review of the properties of Student distribution.

Since $(\theta-\mu)'T(\theta-\mu)$ is PD, (1.1) has a single mode at $\theta = \mu$, and also μ is the mean of the distribution which exists for $\nu > 0$. As (1.1) is symmetric about μ, this implies that all odd-order moments about μ, when they exist, will be zero. The matrix of the second order moments about the mean exists for $\nu > 2$ and is given by $\frac{\nu}{\nu-2} T^{-1}$. Also $(\theta-\mu)'T(\theta-\mu)/p$ has an F distribution with p and ν degrees of freedom. See DeGroot (1970).

If θ, μ and T are partitioned as

$$\theta = \begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \end{pmatrix} \begin{matrix} q\times 1 \\ (p-q)\times 1 \end{matrix}, \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \begin{matrix} q\times 1 \\ (p-q)\times 1 \end{matrix} \quad \text{and}$$

$$T = \begin{matrix} & q & p-q & \\ & \end{matrix} \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{matrix} q \\ p-q \end{matrix},$$

then $\theta^{(1)}$ has a q dimensional t distribution with ν degrees of freedom, location vector $\mu^{(1)}$ and precision matrix $T_{11} - T_{12}T_{22}^{-1}T_{21}$. Also if $Y = A\theta$ where A is a constant $m\times p$ matrix, then Y has a m-dimensional multivariate t distribution with ν degrees of freedom, location parameter $A\mu$ and precision matrix $(AT^{-1}A')^{-1}$. For matricvariate generalization of the multivariate t distribution, see Dickey (1967).

Now suppose θ has a poly-t distribution, and θ is partitioned as $\theta' = (\theta^{(1)'} \ \theta^{(2)'})$. Then $\theta^{(1)}|\theta^{(2)}$ and $\theta^{(2)}|\theta^{(1)}$ also have poly-t distributions, which are properties that carry over from Student distribution.

Since M_j in (1.2) and (1.3) is not necessarily PDS and ν_j is not necessarily greater than p, then this implies that individually, each t in (1.2) and (1.3) need not be the kernel of a Student density.

Poly-t densities are typically assymmetric and multimodal. Box and Tiao (1973) give an example of a 2/0 poly-t, although the individual Student kernals describe symmetric and unimodal functions, their product may produce functions which are multimodal and assymmetric when the two sample means are far enough apart.

Poly-t distributions are difficult to analyze theoretically since the normalizing constants and hence the moments are unknown analytically. So they must be evaluated numerically. The normalizing constant in the p-dimensional poly-t distribution is a p-dimensional integral which is difficult to evaluate, particularly when p is large. Similar difficulties will also be encountered if one is interested in making inference about a subset of the elements of parameters, since in this case, it does not appear possible to express the corresponding marginal posterior distribution of the subset of interest in terms of simple functions. Box and Tiao (1973) gave an alternative form in evaluating the posterior distribution of θ, the common parameters for two independent responses, in which, if we are interested in the marginal distribution of the subset of θ, $\theta_\ell(\ell\times 1)$ say, $\ell > k$, then instead of using a $(k-\ell)$ dimensional integral, only one-dimensional numerical integrations are required.

The following are two examples of poly-t distribtions: First, consider a 2/0 poly-t distribution with mean -.0057 and variance .0920 and density

$$f(\theta|s) \propto [1 + \frac{10(\theta + .2248)^2}{71.5235}]^{-\frac{10}{2}} [1 + \frac{5(\theta - .0238)^2}{.8433}]^{-\frac{5}{2}} ,$$

$\theta \in R$ and S the sample, and next suppose we have a 1/1 poly-t distribution with mean -1.2815 and variance 11.4868 and density

$$f(\theta|s) \propto [1 + \frac{10(\theta + .2623)^2}{97.3515}]^{-\frac{10}{2}} [1 + \frac{5(\theta - .0357)^2}{1.8975}]^{\frac{5}{2}} , \ \theta \in R.$$

This paper is concerned with the poly-t distribution and its application to Bayesian statistical inference. In what is to follow, we have a review of the literature, an example of how the poly-t distribution arises in the analysis of several normal populations, a proposed solution to adaptive estimation (filtering) in a linear dynamic system, and a proposed approximation to the marginal distribution of any subset of poly-t distributed random variables.

CHAPTER II

REVIEW OF THE LITERATURE

Since 1960, Bayesian ideas have gradually developed in the literature. The work of Zellner (1971) covers virtually the entire field of econometrics, and frequently gives in parallel the sampling theory and Bayesian treatments of the same problems. A sampling of subsequent work in Bayesian econometrics is given in a volume edited by Fienberg and Zellner (1975). The development of Bayesian analysis, with applications of the concepts and principles, can be found in Box and Tiao (1973). Lindley (1971) is concerned with estimation problems, devoted to the formulation and solution of technical problems in statistics, by means of Bayesian methods.

In a Bayesian analysis, ideally, we would like to have a prior probability density function represent our prior information as accurately as possible and yet be relatively simple so that the mathematical operations can be performed conveniently. Thus the conjugate prior, which is closed under sampling, is relatively simple and mathematically tractable, hence is very useful in representing prior information. The concept of a conjugate family of distributions was formalized by Raiffa and Schlaifer (1961), who also studied many families of distributions. Examples of natural conjugate densities can also be found in De Groot (1970). Zellner (1971) discussed the role and nature of prior information in analysis of data and the use of probability density function to represent prior information and gave applications encountered in economics and econometrics.

Raiffa and Schlaifer (1961) developed the distribution theory necessary to carry out the Bayesian analysis of the multivariate normal model. Ando and Kaufman (1965) generalized the above treatment by considering the process where neither the mean vector nor the variance covariance matrix is known.

Tiao and Zellner (1964) used the Bayesian approach to analyze sets of regression equations which have correlated error terms. Press (1972) discussed both multivariate regression and generalized multivariate regression models from the Bayesian view-point employing both informative and non-informative prior densities on the parameters. Zellner (1976) used a broader assumption on the distribution of the error terms in regression models. He assumed the error terms have a joint multivariate t distribution with zero location vector and scalar precision matrix, and the multivariate Cauchy and normal distributions are special cases. In the last part of the paper (Zellner, 1976), the natural conjugate prior for the 'multivariate Student-t' regression model was represented.

Poly-t densities, which are defined in Chapter I, arise with surprising insistence in the Bayesian analysis of many statistical models, particularly normal regression and simultaneous equations models. An overview of the role played by the poly-t density in statistics has been done by Sedory (1981). Dreze (1977) discussed how poly-t densities arise as Bayesian posterior densities for regression coefficients under a variety of specifications for the prior density and the data generating process and he summarized the results obtained for a number of models. As mentioned in Chapter I, no analytical expression exists for the normalizing constant and moments of poly-t densities, thus Tiao and Zellner (1964 a) expanded poly-t densities, which they call multiple t densities, in a Taylor series in order to compute approximate moments from the leading terms of the series. They described methods for taking higher order terms into account.

Since calculation of the mean requires the constant of proportionality in Bayes' formula, thus involving yet another integration which is technically difficult to execute, Lindley and Smith (1972) suggested an approximation using the mode of the posterior distribution in place of the mean, by solving the modal equations via an iterative technique. Dickey (1968) has shown that k/0 poly-t densities of θ can be analyzed by means of k-1 dimensional numerical integration, independent of the number of components in θ. Dayal and Dickey (1977) studied integrals whose integrands are product of t densities. Then Richard and Tompa (1980) expand Dickey's (1968) analysis to obtain formulae which are more convenient for computation. They developed methods and computer algorithms to evaluate integrating constants and other characteristics of poly-t densities.

CHAPTER III

BAYESIAN ANALYSIS OF NORMAL POPULATIONS WITH A COMMON MEAN θ

This chapter develops the Bayesian analysis of several normal populations each having an unknown mean θ, but having distinct precision parameters. First several normal scalar populations are considered, then several multivariate populations, and finally several regression models with common regression coefficients are studied.

In all cases, using the conjugate family of prior distributions, the marginal posterior distribution of θ has a k/0 product form poly-t distribution and the marginal posterior distribution of the precision parameters, although related to a gamma distribution, is of an unknown form.

The posterior analysis reveals that the conditional distribution of θ given the precision parameters is always a normal distribution and the conditional distribution of the precision parameters given the common mean, is that of independent gamma (or Wishart) random variables (or matrices). The conditional posterior modes are thus easy to find, hence the joint posterior mode of all the parameters, perhaps, can be calculated by iterative algorithms.

In the scalar case, the poly-t distribution can be evaluated easily with numerical plotting and integration techniques, but in the multivariate situation, some sort of approximation to the marginal distribution of subsets of θ, would be valuable.

Prior Analysis

Consider several normal populations $N(\theta,\tau_i)$ for $i = 1,2,\ldots,k$ where $\theta \in R$ is the common mean and $\tau_i > 0$ is the precision of population i. Suppose we want to estimate these k+1 parameters by means of Bayesian techniques. Let x_{ij}, $j = 1,2,\ldots,n_i$ be a random sample of size n_i from $N(\theta,\tau_i)$, then the likelihood function is

$$L(\theta,\rho|s) \propto \prod_{i=1}^{k} \tau_i^{\frac{n_i}{2}} e^{-\frac{\tau_i}{2}[\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + n_i(\theta - \bar{x}_i)^2]} \tag{3.1}$$

where $\rho = (\tau_1, \tau_2, \ldots, \tau_k)$, s is the sample, $\theta \in R$ and $\tau_i > 0$ for all $i = 1,2,\ldots,k$. We see that the form of the likelihood function suggests

$$p(\theta,\rho) \propto \prod_{i=1}^{k} \tau_i^{\alpha_i - \frac{1}{2}} e^{-\frac{\tau_i}{2}[2\beta_i + \xi_i(\theta-\mu_i)^2]} , \tag{3.2}$$

$\tau_i > 0$, $\xi_i > 0$, $\alpha_i > 0$, $\beta_i > 0$, $\mu_i \in R$ for every $i = 1,2,\ldots,k$ and $\theta \in R$ as the conjugate prior density, which is the product of k normal-gamma densities, one for each population. By completing the square on θ in (3.2), the equation can be written as

$$p(\theta,\rho) \propto \left\{ \left(\sum_{i=1}^{k} \tau_i\xi_i\right)^{\frac{1}{2}} e^{-\frac{1}{2}[\theta - (\sum_{i=1}^{k} \tau_i\xi_i)^{-1}(\sum_{i=1}^{k} \tau_i\xi_i\mu_i)]^2(\sum_{i=1}^{k} \tau_i\xi_i)} \right\}$$

$$\left\{ \prod_{i=1}^{k} \tau_i^{\alpha_i - \frac{1}{2}} e^{-\tau_i\beta_i} \frac{e^{-\frac{1}{2}[\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \tau_i\tau_j\xi_i\xi_j(\mu_i - \mu_j)^2]}}{\left(\sum_{i=1}^{k} \tau_i\xi_i\right)^{\frac{1}{2}}} \right\}$$

$$\propto p_1(\theta|\rho)p_2(\rho) .$$

The function inside the first set of brackets if regarded as a function of θ, must be proportional to the conditional density function of θ given ρ, since θ does not appear in the second set of brackets. This function is proportional to a normal distribution with mean $(\sum_{i=1}^{k} \tau_i \xi_i)^{-1} (\sum_{i=1}^{k} \tau_i \xi_i \mu_i)$ and precision $\sum_{i=1}^{k} \tau_i \xi_i$. It follows that the function inside the second set of brackets is proportional to the marginal prior density of ρ which is not that of independent gammas.

Of course, if one lets $\xi_i \longrightarrow 0$ for $i = 1,2,\ldots,k$ in (3.2), one is assuming θ has a constant density over R, then the conjugate density is the product of k gamma densities, namely

$$p(\theta,\rho) \propto \prod_{i=1}^{k} \tau_i^{\alpha_i - \frac{1}{2}} e^{-\tau_i \beta_i}, \quad \theta \in R, \quad \tau_i > 0,$$

which is improper. Also from (3.2), the marginal prior density of θ is a k/0 poly-t distribution.

$$p(\theta) \propto \prod_{i=1}^{k} [2\beta_i + \xi_i(\theta - \mu_i)^2]^{-\frac{(2\alpha_i+1)}{2}}, \quad \theta \in R \tag{3.3}$$

and in the special case when $\beta_i = \beta$, $\xi_i = \xi$ and $\mu_i = \mu$ for all $i = 1,2,\ldots,k$,

$$p(\theta) \propto [2\beta + \xi(\theta - \mu)^2]^{-\sum_{i=1}^{k} (2\alpha_i+1)/2}, \quad \theta \in R$$

which is a t density with $\sum_{i=1}^{k} 2\alpha_i + k - 1$ degrees of freedom, location parameter μ and the precision $\frac{\xi}{2\beta}(\sum_{i=1}^{k} 2\alpha_i + k - 1)$.

Posterior Analysis

By combining the likelihood function (3.1) and the conjugate prior density (3.2), we have the posterior density of the parameters.

$$p(\theta,\rho|s) \propto \prod_{i=1}^{k} [\tau_i^{\frac{n_i+2\alpha_i+1}{2}-1} e^{-\frac{\tau_i}{2}\{2\beta_i + \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + n_i(\theta - \bar{x}_i)^2 + \xi_i(\theta - \mu_i)^2\}}], \quad \theta \in R,\ \tau_i > 0. \tag{3.4}$$

Completing the square on θ and integrating with respect to $\tau_1,\tau_2,\ldots,\tau_k$, we have the marginal posterior density of θ

$$p(\theta|s) \propto \prod_{i=1}^{k} [2\beta_i + \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \frac{n_i\xi_i(\bar{x}_i - \mu_i)^2}{(n_i + \xi_i)} + \{\theta - (n_i + \xi_i)^{-1} (n_i\bar{x}_i + \xi_i\mu_i)\}^2(n_i + \xi_i)]^{-\frac{(n_i+2\alpha_i+1)}{2}}, \quad \theta \in R \tag{3.5}$$

which is a k/0 poly-t distribution.

Of course from a Bayesian standpoint, all we have to do is plot $p(\theta|s)$ vs. θ, compute the mean, mode, median, etc., and this must be done numerically. Note that there are no formulas for the integrating constants and moments of $p(\theta|s)$ (except in special cases). Also note that if

$n_i + \xi_i = a$ for all $i = 1,2,\ldots,k$ (choose ξ_i's)

$n_i\bar{x}_i + \xi_i\mu_i = b$ for all $i = 1,2,\ldots,k$ (choose μ_i's)

and $2\beta_i + \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 + \frac{n_i\xi_i(x_i - \mu_i)^2}{(n_i + \xi_i)} = c$ for all $i = 1,2,\ldots k$

(choose β_i's), then

$$p(\theta|s) \propto [c + (\theta - a^{-1}b)a]^{-\sum_{i=1}^{k} \frac{(n_i+2\alpha_i+1)}{2}}, \quad \theta \in R$$

which is a t density.

To find the marginal distribution of $\rho = (\tau_1, \tau_2, \ldots, \tau_k)$ we have to complete the square on θ from the exponent in (3.4), then integrate with respect to θ. The result is

$$p(\rho|s) \propto \{\prod_{i=1}^{k} \tau_i^{\frac{(n_i+2\alpha_i+1)}{2}-1} e^{-\frac{\tau_i}{2}[2\beta_i + \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}\}$$

$$\{\frac{e^{-\frac{A}{2}}}{[\sum_{i=1}^{k} \tau_i(n_i+\xi_i)]^{1/2}}\} \qquad (3.6)$$

where $\tau_i > 0$ for all $i = 1,2,\ldots,k$ and

$$A = [\sum_{i=1}^{k} \tau_i^2 n_i \xi_i (\bar{x}_i-\mu_i)^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \tau_i \tau_j \{n_i n_j (\bar{x}_i - \bar{x}_j)^2$$

$$+ n_i \xi_j (\bar{x}_i-\mu_j)^2 + n_j \xi_i (\bar{x}_j-\mu_i)^2 + \xi_i \xi_j (\mu_i-\mu_j)^2\}]$$

$$[\sum_{i=1}^{k} \tau_i(n_i+\xi_i)]^{-1},$$

which is a very complex distribution. The first part of (3.6) is just the density of k independent gammas, and the second part is a factor involving the squares and cross products of τ_i and τ_j, $i,j = 1,2,\ldots,k$. Thus the marginal density of ρ is not that of independent gammas.

By (3.4) we see that given θ, the conditional distribution of ρ is that of k independent gammas, that is

$$(\tau_i|\theta) \sim G\left[\frac{n_i+2\alpha_i+1}{2}, \frac{2\beta_i + \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 + n_i(\theta-\bar{x}_i)^2 + \xi_i(\theta-\mu_i)^2}{2}\right]$$

where G denotes a gamma distribution for all $i = 1, 2,\ldots, k$, and given ρ, it can be shown that θ is distributed as normal with

$$E(\theta|\rho,s) = \left[\sum_{i=1}^{k} \tau_i(n_i+\xi_i)\right]^{-1}\left[\sum_{i=1}^{k} \tau_i(n_i\bar{x}_i+\xi_i\mu_i)\right] \quad \text{and}$$

$$V(\theta|\rho,s) = \left[\sum_{i=1}^{k} \tau_i(n_i+\xi_i)\right]^{-1} .$$

The modes of the conditional distributions are

$$M(\tau_i|\tau_j\text{'s},\theta)_{j\neq i} = \frac{n_i + 2\alpha_i - 1}{2\beta_i + \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 + n_i(\theta-\bar{x}_i)^2 + \xi_i(\theta-\mu_i)^2}$$

for every $i = 1,2,\ldots,k$ and

$$M(\theta|\rho) = \frac{\sum_{i=1}^{k} \tau_i(n_i\bar{x}_i+\xi_i\mu_i)}{\sum_{i=1}^{k} \tau_i(n_i+\xi_i)} .$$

The latter is a convex combination of $\bar{x}_i$'s and μ_i's. For k=2, the modal equations are

$$M(\tau_1|\tau_2,\theta) = \frac{n_1 + 2\alpha_1 - 1}{2\beta_1 + \sum_{j=1}^{n_1} (x_{1j}-\bar{x}_1)^2 + n_1(\theta-\bar{x}_1)^2 + \xi_1(\theta-\mu_1)^2}$$

$$M(\tau_2|\tau_1,\theta) = \frac{n_2 + 2\alpha_2 - 1}{2\beta_2 + \sum_{j=1}^{n_2} (x_{2j}-\bar{x}_2)^2 + n_2(\theta-\bar{x}_2)^2 + \xi_2(\theta-\mu_2)^2}$$

and $$M(\theta|\tau_1,\tau_2) = \frac{n_1\tau_1\bar{x}_1 + n_2\tau_2\bar{x}_2 + \tau_1\xi_1\mu_1 + \tau_2\xi_2\mu_2}{n_1\tau_1 + n_2\tau_2 + \tau_1\xi_1 + \tau_2\xi_2} . \tag{3.7}$$

Now consider a Bayesian analysis using the improper prior density (Jeffreys, 1966)

$$p(\theta,\rho) \propto \prod_{i=1}^{k} \frac{1}{\tau_i} , \quad \tau_i > 0 , \quad i = 1,2,\ldots,k , \quad \theta \in R .$$

Then if this is combined with the likelihood function (3.1) we have

$$p(\theta,\rho|s) \propto \prod_{i=1}^{k} \tau_i^{\frac{n_i}{2}-1} e^{-\frac{\tau_i}{2}[\sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 + n_i(\theta-\bar{x}_i)^2]} , \quad \theta \in R, \ \tau_i > 0. \tag{3.8}$$

as the joint posterior distribution of the parameters. The marginal posterior distribution of θ is k/0 poly-t distribution, that is

$$p(\theta|s) \propto \prod_{i=1}^{k} \left[\sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 + n_i(\theta-\bar{x}_i)^2 \right]^{-\frac{n_i}{2}} , \quad \theta \in R . \tag{3.9}$$

Now, if the posterior distribution of ρ is derived, first by completing the square on θ in the exponent of (3.8) and then integrating with respect to θ,

$$p(\rho|s) \propto \frac{\prod_{i=1}^{k} \tau_i^{\frac{n_i}{2}-1} e^{-\frac{1}{2}[\tau_i \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \tau_i\tau_j n_i n_j (\bar{x}_i-\bar{x}_j)^2 / \sum_{i=1}^{k} \tau_i n_i]}}{[\sum_{i=1}^{k} \tau_i n_i]^{1/2}} , \tag{3.10}$$

$\tau_i > 0$, is the marginal distribution of ρ. Also, the conditional posterior distributions are

$$(\theta|\rho) \sim N[(\sum_{i=1}^{k} n_i\tau_i)^{-1}(\sum_{i=1}^{k} \tau_i n_i \bar{x}_i) , \sum_{i=1}^{k} \tau_i n_i] \quad \text{and}$$

$$(\tau_i|\tau_j\text{'s},\theta)_{j\neq i} \sim G\left[\frac{n_i}{2} , \frac{\sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)^2 + n_i(\theta-\bar{x}_i)^2}{2}\right]$$

for $i = 1,2,\ldots k$.

Note, that if we put $\alpha_i = -\frac{1}{2}$, $\beta_i = 0$, $\xi_i = 0$, $i = 1,2,\ldots,k$ in the posterior density (3.4), we will get the same results as the one with an improper prior.

Inferences for the Parameters

We have seen that the posterior distributions for the parameters give complex distributions; however, the marginal distribution of θ, (3.5) or (3.9) is a k/0 poly-t distribution. For $k = 2$, it is a 2/0 poly-t (one-dimensional), hence all we have to do is plot this density and calculate numerically the posterior mean, variance and other moments. On the other hand, the marginal distribution of (τ_1,τ_2) is quite complex since τ_1 and τ_2 are not independent. Hence one must numerically plot the contours of the joint density (3.6) or (3.10) for $k = 2$, over the region where $\tau_1 > 0$ and $\tau_2 > 0$.

Another approach is to find the mode of the joint density of (θ,τ_1,τ_2). Using Lindley and Smith's (1972) idea, we can consider the three conditional modes (3.7) and attempt to solve these equations using an iterative algorithm. So starting with some estimate of τ_1 and τ_2, say the sample precisions, estimate θ; then solve for τ_1 and τ_2 and continue the procedure until the estimate stabilizes. The solution, perhaps, is the mode of the joint distribution, however, the joint density may be multi-modal and the solution may not converge to the joint mode, but to a local maximum or minimum.

Prior Analysis for θ a Vector

The theoretical results can be extended from a scalar θ to a vector $\theta(p\times 1)$ where θ is the common mean vector of several normal populations $N_p(\theta,P_i)$, $i = 1,2,\ldots,k$, and P_i is an unknown precision matrix of the ith population. Let $x_1, x_2,\ldots, x_{n_i}$ be the n_i independent p-variate observations from $N_p(\theta,P_i)$, $i = 1,2,\ldots,k$. Then the likelihood function is

$$L(\theta,P|s) \propto \prod_{i=1}^{k} |P_i|^{\frac{n_i}{2}} e^{-\frac{1}{2}[\sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i) + n_i(\theta-\bar{x}_i)'P_i(\theta-\bar{x}_i)]} \tag{3.11}$$

where $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ the sample mean vector of the ith population,

$P = (P_1, P_2,\ldots,P_k)$, P_i positive definite (PD) for $i = 1,2,\ldots,k$, $\theta \in R^p$ and S is the sample.

The form of the likelihood function suggests

$$p(\theta,P) \propto \prod_{i=1}^{k} |P_i|^{\frac{\alpha_i-p}{2}} e^{-\frac{1}{2}[Tr(\tau_i P_i) + \xi_i(\theta-\mu_i)'P_i(\theta-\mu_i)} \tag{3.12}$$

$\theta \in R^p$, $\xi_i > 0$, τ_i PDS matrix, $i = 1,2,\ldots,k$, as the conjugate density which is the product of k normal Wishart densities, one for each population. Now, (3.12) can be written as

$$p(\theta,P) \propto \{\prod_{i=1}^{k} e^{-\frac{s_1}{2}}\}\{\prod_{i=1}^{k} |P_i|^{\frac{\alpha_i-p}{2}} e^{-\frac{1}{2}[s_2+Tr(\tau_i P_i)]}\} = p_1(\theta|P)p_2(P)$$

where $s_1 = \{\theta - (\sum_{i=1}^{k} \xi_i P_i)^{-1}(\sum_{i=1}^{k} \xi_i P_i \mu_i)\}'(\sum_{i=1}^{k} \xi_i P_i)\{\theta - (\sum_{i=1}^{k} \xi_i P_i)^{-1} \cdot$

$(\sum_{i=1}^{k} \xi_i P_i \mu_i)\}$

and $s_2 = \sum_{i=1}^{k} \xi_i \mu_i' P_i \mu_i - (\sum_{i=1}^{k} \xi_i P_i \mu_i)'(\sum_{i=1}^{k} \xi_i P_i)^{-1}(\sum_{i=1}^{k} \xi_i P_i \mu_i)$.

Thus the conditional prior distribution for θ, given P, is multivariate normal with mean vector $(\sum_{i=1}^{k} \xi_i P_i)^{-1}(\sum_{i=1}^{k} \xi_i P_i \mu_i)$ and precision matrix $\sum_{i=1}^{k} \xi_i P_i$. It now follows that the marginal prior density of P is proportional to the function inside the second set of braces, but is not a well-known function.

If we let $\xi_i \longrightarrow 0$ for $i = 1,2,\ldots,k$, that is assume θ has a constant density over R^p, then the conjugate density is just the product of k Wishart densities, which is

$$p(\theta,P) \propto \prod_{i=1}^{k} |P_i|^{\frac{\alpha_i - p}{2}} e^{-\frac{1}{2}Tr(\tau_i P_i)}, \quad \theta \in R^p, \ |P_i| > 0,$$

which is an improper density.

If we write (3.12) as

$$p(\theta,P) \propto \prod_{i=1}^{k} |P_i|^{\frac{\alpha_i - p}{2}} e^{-\frac{1}{2}Tr\,[\{\tau_i + \xi_i(\theta-\mu_i)(\theta-\mu_i)'\}P_i]},$$

we could see that the marginal prior density of θ is a p-dimensional k/0 poly-t distribution, that is

$$p(\theta) \propto \prod_{i=1}^{k} |\tau_i + \xi_i(\theta-\mu_i)(\theta-\mu_i)'|^{-\frac{\alpha_i+1}{2}}$$

$$\propto \prod_{i=1}^{k} [1 + \xi_i(\theta-\mu_i)'\tau_i^{-1}(\theta-\mu_i)]^{-\frac{\alpha_i+1}{2}}. \tag{3.13}$$

A special case occurs when $\mu_i = \mu$, $\xi_i = \xi$ and $\tau_i = \tau$ for all $i = 1,2,\ldots,k$, then the density of θ

$$p(\theta) \propto \{1 + \xi(\theta-\mu)'\tau^{-1}(\theta-\mu)\}^{-\frac{1}{2}\sum_{i=1}^{k}(\alpha_i+1)}$$

which is a p-variate t density.

Posterior Analysis for θ a Vector

Utilizing the prior distribution (3.12) in conjunction with the likelihood function (3,11) the posterior distribution of the parameters can be expressed as

$$p(\theta,P|s) \propto \prod_{i=1}^{k} |P_i|^{\frac{n_i+\alpha_i-p}{2}} e^{-\frac{1}{2}\mathrm{Tr}[\{\tau_i + \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'} \tag{3.14}$$

$$+ n_i(\theta-\bar{x}_i)(\theta-\bar{x}_i)' + \xi_i(\theta-\mu_i)(\theta-\mu_i)'\}P_i] \quad , \quad \theta \in R^p, \ |P_i| > 0$$

Integrating with respect to P, one has the marginal posterior density of θ, namely

$$p(\theta|s) \propto \prod_{i=1}^{k} |\tau_i + \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)' + n_i(\theta-\bar{x}_i)(\theta-\bar{x}_i)'$$

$$+ \xi_i(\theta-\mu_i)(\theta-\mu_i)'|^{-\frac{(n_i+\alpha_i+1)}{2}}$$

$$\propto \prod_{i=1}^{k} [1 + (n_i+\xi_i)(\theta - \frac{n_i\bar{x}_i + \xi_i\mu_i}{n_i + \xi_i})'\{\tau_i + \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'$$

$$+ \frac{n_i\xi_i}{n_i+\xi_i}(\bar{x}_i-\mu_i)(\bar{x}_i-\mu_i)'\}^{-1}(\theta - \frac{n_i\bar{x}_i + \xi_i\mu_i}{n_i + \xi_i})]^{-\frac{(n_i+\alpha_i+1)}{2}} \tag{3.15}$$

$$\theta \in R^p$$

which is a p-dimensional k/0 poly-t distribution.

To find the marginal posterior distribution of P, we write the joint posterior distribution (3.14) as

$$p(\theta,P|s) \propto \prod_{i=1}^{k} \{e^{-\frac{1}{2}[n_i(\theta-\bar{x}_i)'P_i(\theta-\bar{x}_i) + \xi_i(\theta-\mu_i)'P_i(\theta-\mu_i)]}.$$

$$|P_i|^{\frac{n_i+\alpha_i-p}{2}} \quad e^{-\frac{1}{2}[\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i) + Tr(\tau_i P_i)]}\},$$

and completing the square in the exponent

$$n_i(\theta-\bar{x}_i)'P_i(\theta-\bar{x}_i) + \xi_i(\theta-\mu_i)'P_i(\theta-\mu_i) = (\theta-\mu_i^*)'P_i^*(\theta-\mu_i^*) + S_i^* ,$$

where $\mu_i^* = \dfrac{n_i\bar{x}_i + \xi_i\mu_i}{n_i + \xi_i}$, $P_i^* = (n_i + \xi_i)P_i$, and

$$S_i^* = \frac{n_i\xi_i}{n_i + \xi_i}(\bar{x}_i-\mu_i)'P_i(\bar{x}_i-\mu_i) .$$

Now, the posterior distribution of the parameters can be written as

$$p(\theta,P|s) \propto \{e^{-\frac{1}{2}\sum_{i=1}^{k}(\theta-\mu_i^*)'P_i^*(\theta-\mu_i^*)}\}$$

$$\{\prod_{i=1}^{k}|P_i|^{\frac{n_i+\alpha_i-p}{2}} \quad e^{-\frac{1}{2}[S_i^*+\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i)+Tr(\tau_i P_i)]}\}$$

$$\propto \{e^{-\frac{1}{2}(\theta-\mu^{**})'P^{**}(\theta-\mu^{**})} \quad |P^{**}|^{\frac{1}{2}}\}$$

$$\{\prod_{i=1}^{k}|P_i|^{\frac{n_i+\alpha_i-p}{2}} \quad e^{-\frac{1}{2}[S_i^*+\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i)+Tr(\tau_i P_i)]}\}$$

$$\{|P^{**}|^{-\frac{1}{2}} \; e^{-\frac{1}{2}[\sum_{i=1}^{k}\mu_i^{*\prime}P_i^*\mu_i^* + (\sum_{i=1}^{k}P_i^*\mu_i^*)'(P^{**})^{-1}(\sum_{i=1}^{k}P_i^*\mu_i^*)]}\}$$

where $\mu^{**} = (\sum_{i=1}^{k}P_i^*)^{-1}(\sum_{i=1}^{k}P_i^*\mu_i^*)$, $P^{**} = \sum_{i=1}^{k}P_i^*$, $\theta \in R^p$ and P_i PD for all $i = 1,2,\ldots,k$.

Integrating $p(\theta,P|s)$ with respect to θ, one has the marginal posterior distribution of P.

$$p(P|s) \propto \{ \prod_{i=1}^{k} |P_i|^{\frac{n_i+\alpha_i-p}{2}} e^{-\frac{1}{2}[S_i^* + \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i) + Tr(\tau_i P_i)]} \}$$

$$\{ |P^{**}|^{-\frac{1}{2}} e^{-\frac{1}{2}[\sum_{i=1}^{k} \mu_i^{*\prime} P_i^* \mu_i^* + (\sum_{i=1}^{k} P_i^*\mu_i^*)'(P^{**})^{-1}(\sum_{i=1}^{k} P_i^*\mu_i^*)]} \} \quad (3.16)$$

which is not a known distribution. Note that in the first set of brackets, we have the product of k independent Wisharts, and in the second set of brackets, a factor involving squares and cross products of the P_i's.

From (3.14) it can be shown that given θ, the conditional distribution of P is that of k independent Wisharts, namely

$$p(P_i|\theta, P_j\text{'s}, s)_{j\neq i} \propto |P_i|^{\frac{n_i+\alpha_i-p}{2}} e^{-\frac{1}{2}Tr[\{\tau_i + \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)' + n_i(\theta-\bar{x}_i)(\theta-\bar{x}_i)' + \xi_i(\theta-\mu_i)(\theta-\mu_i)'\}P_i]}$$

for i = 1,2,...,k. Given P, θ is distributed as a multivariate normal with mean vector μ^{**} and precision matrix P^{**}. The conditional modes are given by

$$M(P_i|P_j\text{'s},\theta)_{j\neq i} = [n_i + \alpha_i - p][\tau_i + \sum_{j=1}^{n_i} (x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)' + n_i(\theta-\bar{x}_i)(\theta-\bar{x}_i)' + \xi_i(\theta-\mu_i)(\theta-\mu_i)']^{-1} \quad \text{for } i = 1,2,\ldots,k, \text{ and}$$

$$M(\theta|P) = \mu^{**}.$$

Now consider the Bayesian analysis using the improper prior density (Jeffreys, 1966)

$$p(\theta,P) \propto \prod_{i=1}^{k} |P_i|^{-\frac{(p+1)}{2}}, \quad \theta \in R^p,\ P_i \text{ PD for all } i = 1,2,\ldots,k.$$

Combine this with (3.11), and we have the joint posterior distribution

$$p(\theta,P|s) \propto \prod_{i=1}^{k} |P_i|^{\frac{n_i-p-1}{2}} e^{-\frac{1}{2}[\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i) + n_i(\theta-\bar{x}_i)'P_i(\theta-\bar{x}_i)]}. \tag{3.17}$$

Integrating with respect to P, we have as the marginal posterior distribution of θ, $\theta \in R^p$ as

$$p(\theta|s) \propto \prod_{i=1}^{k} \left|\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)' + n_i(\theta-\bar{x}_i)(\theta-\bar{x}_i)'\right|^{-\frac{n_i}{2}}$$

$$\propto \prod_{i=1}^{k} [1 + n_i(\theta-\bar{x}_i)'\{\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)'\}^{-1}(\theta-\bar{x}_i)]^{-\frac{n_i}{2}}, \tag{3.18}$$

which is a p-dimensional k/0 poly-t distribution.

The joint posterior distribution can also be written as

$$p(\theta,P|s) \propto \{\prod_{i=1}^{k} |P_i|^{\frac{n_i-p-1}{2}} e^{-\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i)}$$

$$\left|\sum_{i=1}^{k} n_iP_i\right|^{-\frac{1}{2}} e^{-\frac{1}{2}[\sum_{i=1}^{k} n_i\bar{x}_i'P_i\bar{x}_i - (\sum_{i=1}^{k} n_iP_i\bar{x}_i)'(\sum_{i=1}^{k} n_iP_i)(\sum_{i=1}^{k} n_iP_i\bar{x}_i)}\}$$

$$\{\left|\sum_{i=1}^{k} n_iP_i\right|^{\frac{1}{2}} e^{-\frac{1}{2}[\theta - (\sum_{i=1}^{k} n_iP_i)^{-1}(\sum_{i=1}^{k} n_iP_i\bar{x}_i)]'(\sum_{i=1}^{k} n_iP_i)}$$

$$[\theta - (\Sigma n_iP_i)^{-1}(\Sigma n_iP_i\bar{x}_i)]\}, \qquad \theta \in R^p,\ |P_i| > 0.$$

Integrating with respect to θ we have

$$p(P|s) \propto \{ \prod_{i=1}^{k} |P_i|^{\frac{n_i-p-1}{2}} e^{-\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)'P_i(x_{ij}-\bar{x}_i)} \}\{ |\sum_{i=1}^{k} n_iP_i|^{-\frac{1}{2}}$$

$$e^{-\frac{1}{2}[\sum_{i=1}^{k} n_i\bar{x}_i'P_i\bar{x}_i - (\sum_{i=1}^{k} n_iP_i\bar{x}_i)'(\sum_{i=1}^{k} n_iP_i)(\sum_{i=1}^{k} n_iP_i\bar{x}_i)]} \} \quad (3.19)$$

as the marginal posterior distribution of P, $|P_i| > 0$.

From (3.17) it can be shown that given P, θ is distributed as multivariate normal with mean vector $(\sum_{i=1}^{k} n_iP_i)^{-1}(\sum_{i=1}^{k} n_iP_i\bar{x}_i)$ and precision matrix $\sum_{i=1}^{k} n_iP_i$, and given θ and P_j's, $j \neq i$, P_i is distributed as Wishart with parameter matrix $[\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)(x_{ij}-\bar{x}_i)' + n_i(\theta-\bar{x}_i)(\theta-\bar{x}_i)']^{-1}$ and degrees of freedom n_i, for all $i = 1,2,\ldots,k$ and $p \leq n_i$.

Regression Model

Now we extend the results to several regression models, which have a common set of regression coefficients. Let

$y_i = X_i\theta + \varepsilon_i$, $i = 1,2,\ldots,k$, where

$y_i : n_i \times 1$ vector of observations,

$X_i : n_i \times p$ design matrix of full rank,

$\theta : p \times 1$ common vector of regression parameters,

$\varepsilon_i : n_i \times 1$ vector of errors distributed as normal with mean vector zero and precision matrix $\tau_i I_{n_i}$. The likelihood function can be written as

$$L(\theta,T|s) \propto \prod_{i=1}^{k} \tau_i^{\frac{n_i}{2}} e^{-\frac{\tau_i}{2}[(y_i-X_i\hat{\theta}_i)'(y_i-X_i\hat{\theta}_i) + (\theta-\hat{\theta}_i)'X_i'X_i(\theta-\hat{\theta}_i)]}, \qquad (3.20)$$

where $T = (\tau_1, \tau_2, \ldots, \tau_k)$, $\hat{\theta}_i = (X_i'X_i)^{-1}X_i'y_i$ and s is the sample. Let the prior distribution be

$$p(\theta,T) \propto \prod_{i=1}^{k} \tau_i^{\alpha_i - \frac{1}{2}} e^{-\frac{\tau_i}{2}\{2\beta_i + (\theta-\mu_i)'\eta_i(\theta-\mu_i)\}}, \theta \in R^p, \ \tau_i > 0, \qquad (3.21)$$

then the posterior density is

$$p(\theta,T|s) \propto \prod_{i=1}^{k} \tau_i^{\frac{2\alpha_i+n_i+1}{2}} e^{-\frac{\tau_i}{2}[2\beta_i + (y_i-X_i\hat{\theta}_i)'(y_i-X_i\hat{\theta}_i)}$$

$$+ (\theta-\hat{\theta}_i)'X_i'X_i(\theta-\hat{\theta}_i) + (\theta-\mu_i)'\eta_i(\theta-\mu_i)] \ . \qquad (3.22)$$

Comparing (3.22) to equation (3.14), we see they are formally identical. It follows that the entire analysis of the regression problem is formally identical to the analysis we have discussed, thus so far. The results are:

$$p(\theta|s) \propto \prod_{i=1}^{k} [2\beta_i + (y_i-X_i\hat{\theta}_i)'(y_i-X_i\hat{\theta}_i) + \hat{\theta}_i'X_i'X_i\hat{\theta}_i + \mu_i'\eta_i\mu_i$$

$$- (X_i'X_i\hat{\theta}_i+\eta_i\mu_i)'(X_i'X_i+\eta_i)^{-1}(X_i'X_i\hat{\theta}_i+\eta_i\mu_i)$$

$$+ \{\theta-(X_i'X_i+\eta_i)^{-1}(X_i'X_i\hat{\theta}_i+\eta_i\mu_i)\}'(X_i'X_i+\eta_i)\{\theta-(X_i'X_i+\eta_i)^{-1}$$

$$(X_i'X_i\hat{\theta}_i+\eta_i\mu_i)\}]^{-\frac{(2\alpha_i+\eta_i+1)}{2}}, \qquad \theta \in R^p, \qquad (3.23)$$

which is a p-dimensional k/0 poly-t distribution. The marginal density of T is

$$p(T|s) \propto \prod_{i=1}^{k} [\tau_i^{\frac{2\alpha_i+n_i+1}{2}} e^{-\frac{\tau_i}{2}\{2\beta_i + (y_i-X_i\hat{\theta}_i)'(y_i-X_i\hat{\theta}_i) + R_i\}}]$$

$$(\sum_{i=1}^{k} \tau_i^*)^{-\frac{1}{2}} e^{-\frac{1}{2}\{\sum_{i=1}^{k} \mu_i^*\tau_i^*\mu_i^* + (\sum_{i=1}^{k} \tau_i^*\mu_i^*)'(\sum_{i=1}^{k} \tau_i^*)^{-1}(\sum_{i=1}^{k} \tau_i^*\mu_i^*)\}}, \tau_i>0, \quad (3.24)$$

where $\mu_i^* = (X_i'X_i+\eta_i)^{-1}(X_i'X_i\hat{\theta}_i+\eta_i\mu_i)$,

$\tau_i^* = \tau_i(X_i'X_i+\eta_i)$

and $R_i = \tau_i[(\hat{\theta}_i'X_i'X_i\hat{\theta}_i+\mu_i'\eta_i\mu_i) - (X_i'X_i\hat{\theta}_i+\eta_i\mu_i)'(X_i'X_i+\eta_i)^{-1}$

$(X_i'X_i\hat{\theta}_i+\eta_i\mu_i)]$.

The conditional marginal density of τ_i, given the other parameters is

$$p(\tau_i|\theta,\tau_j\text{'s},s) \propto \tau_i^{\frac{2\alpha_i+n_i+1}{2}} e^{-\frac{\tau_i}{2}\{2\beta_i + (y_i-X_i\hat{\theta}_i)'(y_i-X_i\hat{\theta}_i) + (\theta-\hat{\theta}_i)'X_i'X_i(\theta-\hat{\theta}_i) + (\theta-\mu_i)'\eta_i(\theta-\mu_i)\}}, \quad j\neq i, \ \tau_i>0 \quad (3.25)$$

for $i = 1,2,\ldots,k$. Also, given T, θ is distributed as multivariate normal with mean vector $(\sum_{i=1}^{k} \tau_i^*)^{-1}(\sum_{i=1}^{k} \tau_i^*\mu_i^*)$ and precision matrix $\sum_{i=1}^{k} \tau_i^*$, and the modal equations are

$$M(\tau_i|\theta) = (2\alpha_i+n_i-1)/[2\beta_i + (y_i-X_i\hat{\theta}_i)'(y_i-X_i\hat{\theta}_i) + (\theta-\hat{\theta}_i)'X_i'X_i(\theta-\hat{\theta}_i)$$

$$+ (\theta-\mu_i)'\eta_i(\theta-\mu_i)] \quad \text{for } i = 1,2,\ldots,k \text{ and}$$

$$M(\theta|T) = (\sum_{i=1}^{k} \tau_i^*)^{-1} (\sum_{i=1}^{k} \tau_i^* \mu_i^*) .$$

Note that if we put $\beta_i = 0$, $\eta_i = 0$ and $\alpha_i = -p/2$ in the posterior density (3.22) for all $i = 1,2,\ldots,k$, we have the same results as if an improper prior density had been used. Zellner (1971) in section 4.2 gave an example for $k = 2$, using improper prior density.

CHAPTER IV

METHODS OF APPROXIMATION

The main goal of this research is to develop approximation techniques to the marginal pdf associated with poly-t. The marginal distribution of any subset of poly-t distributed random variables involves integration of the poly-t (excluding the normalizing constant). In the first method of approximation, the integrand in the marginal density is approximated by a normal density with the same mean and precision of each factor in the density of the integrand. In the second method, the integrand in the marginal density is approximated by a Bernstein polynomila (Ralston, 1965). Lastly, we use the generalization of the Weierstrass theorem (Lorentz, 1953) to directly approximate the poly-t density. From here one may obtain either marginal distribution as a mixture of Beta densities. One may also obtain the equations for the means and variances.

Approximation by Normal

Let θ be a $p\times 1$ random vector where $p = p_1+p_2$ and partition $\theta = (\theta_1', \theta_2')'$ where $\theta_1 \in R^{p_1}$ and $\theta_2 \in R^{p_2}$ and suppose θ has a k-factor poly-t distribution with density

$$h(\theta) = c \prod_{i=1}^{k} f_i(\theta) \;, \quad \theta \in R^p,$$

where c (>0) is the normalizing constant, thus

$$h(\theta) = \frac{c}{\prod_{i=1}^{k} c_i} \prod_{i=1}^{k} c_i f_i(\theta) , \tag{4.1}$$

where $f_i(\theta) = [1 + \frac{(\theta-\mu^{(i)})' T_i (\theta-\mu^{(i)})}{n_i}]^{-\frac{(n_i+p)}{2}}$, $\theta \in R^p$

and $\int_{R^p} c_i f_i(\theta) d\theta = 1$ for all $i = 1,2,\ldots,k$.

Partition μ_i and T_i for all $i = 1,2,\ldots,k$ as follows:

$$\mu^{(i)} = \begin{pmatrix} \mu_1^{(i)} \\ \mu_2^{(i)} \end{pmatrix} \begin{matrix} p_1 \times 1 \\ p_2 \times 1 \end{matrix} \quad \text{and} \quad T_i = \begin{matrix} & p_1 & p_2 & \\ & \end{matrix} \begin{pmatrix} T_{11}^{(i)} & T_{12}^{(i)} \\ T_{21}^{(i)} & T_{22}^{(i)} \end{pmatrix} \begin{matrix} p_1 \\ p_2 \end{matrix} , \quad |T_i| > 0,$$

and suppose $T_{11}^{(i)}$ is non singular for all $i = 1,2,\ldots k$.

Now $c_i f_i(\theta) = c_{i1}(\theta_2) f_{i1}(\theta_1|\theta_2) c_{i2} f_{i2}(\theta_2)$ where

$$c_{i1}(\theta_2) = \frac{|T_{11}^{(i)}|^{\frac{1}{2}} \Gamma(\frac{n_i+p}{2})}{\Gamma(n_i+p_2)[(n_i+p_2)\pi]^{\frac{p_1}{2}}} \cdot$$

$$\{\frac{n_i + p_2}{n_i+(\theta_2-\mu_2^{(i)})'(T_{22}^{(i)}-T_{21}^{(i)}T_{11}^{(i)^{-1}}T_{12}^{(i)})(\theta_2-\mu_2^{(i)})}\}^{\frac{p_1}{2}} ,$$

the normalizing constant of $f_{i1}(\theta_1|\theta_2)$,

c_{i2} = constant (the normalizing constant of $f_{i2}(\theta_2)$),

$$f_{i1}(\theta_1|\theta_2) = \{1 + \{\theta_1 - [\mu_1^{(i)} - T_{11}^{(i)^{-1}} T_{12}^{(i)}(\theta_2 - \mu_2^{(i)})]\}' T_{11}^{(i)} \{\theta_1 - [\mu_1^{(i)} - T_{11}^{(i)^{-1}} T_{12}^{(i)}(\theta_2 - \mu_2^{(i)})]\}/[n_i + (\theta_2 - \mu_2^{(i)})'(T_{22}^{(i)} - T_{21}^{(i)} T_{11}^{(i)^{-1}} T_{12}^{(i)}(\theta_2 - \mu_2^{(i)})]\}^{-\frac{(n_i+p)}{2}} \tag{4.2}$$

for all $i = 1,2,\ldots,k$, $\theta_1 \in R^{p_1}$, $\theta_2 \in R^{p_2}$. Thus $f_{i1}(\theta_1|\theta_2)\, c_{i1}$ is a t density with $n_i + p_2$ degrees of freedom, location vector $\mu_1^{(i)}(\theta_2) = \mu_1^{(i)} - T_{11}^{(i)^{-1}} T_{12}^{(i)}(\theta_2 - \mu_2^{(i)})$ and the precision matrix

$$T^{(i)*}(\theta_2) = \frac{(n_i+p_2)T_{11}^{(i)}}{n_i + (\theta_2-\mu_2^{(i)})'(T_{22}^{(i)} - T_{21}^{(i)} T_{11}^{(i)^{-1}} T_{12}^{(i)})(\theta_2-\mu_2^{(i)})} .$$

Now consider (4.1) which may be written as

$$h(\theta) = \frac{c}{\prod_{i=1}^{k} c_i} \prod_{i=1}^{k} c_{i1}(\theta_2) f_{i1}(\theta_1|\theta_2) c_{i2} f_{i2}(\theta_2), \tag{4.3}$$

$\theta \in R^p$, $\theta_1 \in R^{p_1}$, $\theta_2 \in R^{p_2}$. Thus the conditional density of $\theta_1|\theta_2$ is

$$h_1(\theta_1|\theta_2) \propto \prod_{i=1}^{k} f_{i1}(\theta_1|\theta_2) \;,\; \theta_1 \in R^{p_1},\; \theta_2 \in R^{p_2} \tag{4.4}$$

and the marginal of θ_2 is

$$h_2(\theta_2) \propto \prod_{i=1}^{k} c_{i1}(\theta_2) f_{i2}(\theta_2) \int_{R^{p_1}} \prod_{i=1}^{k} f_{i1}(\theta_1|\theta_2) d\theta_1, \tag{4.5}$$

$\theta_2 \in R^{p_2}$. Thus one must perform the integration, where the integrand is $\prod_{i=1}^{k} f_{i1}(\theta_1|\theta_2)$. Now $f_{i1}(\theta_1|\theta_2)$ is a t density (excluding the normalizing constants $c_{i1}(\theta_2)$) for i = 1,2,...,k. The integral is to be approximated by approximating $f_{i1}(\theta_1|\theta_2)$ with a normal density with the same mean and dispersion for all i = 1,2,...,k.

Now consider $f_{i1}(\theta_1|\theta_2)$, which has mean $\mu_1^{(i)}(\theta_2) = \mu_1^{(i)} - T_{11}^{(i)^{-1}} T_{12}^{(i)} (\theta_2 - \mu_2^{(i)})$ and precision $T^{(i)*}(\theta_2)$. The dispersion matrix is $\frac{n_i+p_2}{n_i+p_2-2}[T^{(i)*}(\theta_2)]^{-1}$. Thus the precision matrix to be used in the normal approximation is

$$\tilde{T}_i(\theta_2) = \frac{n_i+p_2-2}{n_i+p_2} T^{(i)*}(\theta_2).$$

Hence $f_{i1}(\theta_1|\theta_2)$ is approximated by

$$f_{i1}^*(\theta_1|\theta_2) = e^{-\frac{1}{2}[\theta_1-\mu_1^{(i)}(\theta_2)]'\tilde{T}_i(\theta_2)[\theta_1-\mu_1^{(i)}(\theta_2)]}, \qquad \theta_2 \in R^{p_2}.$$

Therefore

$$\prod_{i=1}^{k} f_{i1}^*(\theta_1|\theta_2) = e^{-\frac{1}{2}\sum_{i=1}^{k}[\theta_1-\mu_1^{(i)}(\theta_2)]'\tilde{T}_i(\theta_2)[\theta_1-\mu_1^{(i)}(\theta_2)]}, \quad \theta_2 \in R^{p_2}$$

$$= e^{-\frac{1}{2}[\{\theta_1-m(\theta_2)\}'H(\theta_2)\{\theta_1-m(\theta_2)\}}$$

$$+ \sum_{i=1}^{k} \mu_1^{(i)'}(\theta_2)\tilde{T}_i(\theta_2)\mu_1^{(i)}(\theta_2)$$

$$- (\sum_{i=1}^{k} \tilde{T}_i(\theta_2)\mu_1^{(i)}(\theta_2))'(H(\theta_2))^{-1}(\sum_{i=1}^{k} \tilde{T}_i(\theta_2)\mu_1^{(i)}(\theta_2))]$$

(4.6)

where $m(\theta_2) = [\sum_{i=1}^{k} \tilde{T}_i(\theta_2)]^{-1}[\sum_{i=1}^{k} \tilde{T}_i(\theta_2)\mu_1^{(i)}(\theta_2)]$ and $H(\theta_2) = \sum_{i=1}^{k} \tilde{T}_i(\theta_2)$.

So (4.6) is a normal density with mean $m(\theta_2)$ and precision $H(\theta_2)$. Now the approximation reduces to

$$\int_{R^{p_1}} \prod_{i=1}^{k} f_{i1}(\theta_1|\theta_2)d\theta_1 \doteq \int_{R^{p_1}} \prod_{i=1}^{k} f_{i1}^*(\theta_1|\theta_2)d\theta_1$$

$$= \frac{1}{|H(\theta_2)|^{1/2}} e^{-\frac{1}{2}\{\sum_{i=1}^{k} \mu_1^{(i)'}(\theta_2)\tilde{T}_i(\theta_2)\mu_1^{(i)}(\theta_2) -(\sum_{i=1}^{k} \tilde{T}_i(\theta_2)\mu_1^{(i)}(\theta_2))'H^{-1}(\theta_2)(\sum_{i=1}^{k} \tilde{T}_i(\theta_2)\mu_1^{(i)}(\theta_2))\}}$$

$$= c^*(\theta_2) \ , \qquad \theta_2 \in R^{p_2} \ .$$

Therefore the approximate marginal density of θ_2 is from (4.5)

$$h_2(\theta_2) \propto \prod_{i=1}^{k} c_{i1}(\theta_2)f_{i2}(\theta_2)c^*(\theta_2) \ , \qquad \theta_2 \in R^{p_2} \ .$$

Now suppose x (p×1) has a two-factor poly-t distribution with density $h(x) = cf(x)g(x)$, $x \in R^p$ where

$$f(x) = [1 + \frac{(x-\mu)'T(x-\mu)}{n_1}]^{-\frac{(n_1+p)}{2}} \ , \qquad |T| > 0, \ x \in R^p,$$

$$g(x) = [1 + \frac{(x-v)'S(x-v)}{n_2}]^{-\frac{(n_2+p)}{2}} , \qquad |S| > 0 , \; x \in R^p \qquad (4.7)$$

Partition X, μ, v, T and S as follows:

$$\underset{p\times 1}{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{matrix} p_1\times 1 \\ p_2\times 1 \end{matrix} , \quad \underset{p\times 1}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \begin{matrix} p_1\times 1 \\ p_2\times 1 \end{matrix} , \quad \underset{p\times 1}{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \begin{matrix} p_1\times 1 \\ p_2\times 1 \end{matrix}$$

$$T = \overset{\begin{matrix} p_1 & p_2 \end{matrix}}{\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}} \begin{matrix} p_1 \\ p_2 \end{matrix} , \quad S = \overset{\begin{matrix} p_1 & p_2 \end{matrix}}{\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}} \begin{matrix} p_1 \\ p_2 \end{matrix} .$$

Then the marginal density of x_2 can be written as

$$h_2(x_2) \propto c_{11}(x_2) f_2(x_2) c_{21}(x_2) g_2(x_2) c_3(x_2) , \quad x_2 \in R^{p_2} \text{ where}$$

$$c_{11}(x_2) = \frac{|T_{11}|^{\frac{1}{2}} \Gamma(\frac{n_1+p}{2})}{\Gamma(n_1+p_2)[(n_1+p_2)\pi]^{\frac{p_1}{2}}} \left\{\frac{n_1 + p_2}{n_1+(x_2-\mu_2)'(T_{22}-T_{21}T_{11}^{-1}T_{12})(x_2-\mu_2)}\right\}^{\frac{p_1}{2}}$$

$$c_{21}(x_2) = \frac{|S_{11}|^{\frac{1}{2}} \Gamma(\frac{n_2+p}{2})}{\Gamma(n_2+p_2)[(n_2+p_2)\pi]^{\frac{p_1}{2}}} \left\{\frac{n_2 + p_2}{n_2+(x_2-v_2)'(S_{22}-S_{21}S_{11}^{-1}S_{12})(x_2-v_2)}\right\}^{\frac{p_1}{2}}$$

$$f_2(x_2) = [n_1 + (x_2-\mu_2)'(T_{22}-T_{21}T_{11}^{-1}T_{12})(x_2-\mu_2)]^{-\frac{(n_1+p_2)}{2}}$$

$$g_2(x_2) = [n_2 + (x_2-v_2)'(S_{22}-S_{21}S_{11}^{-1}S_{12})(x_2-v_2)]^{-\frac{(n_2+p_2)}{2}}$$

$$c_3(x_2) = |H(x_2)|^{-\frac{1}{2}} e^{-\frac{1}{2}\{\mu_1'(x_2)\tilde{T}(x_2)\mu_1(x_2) + v_1'(x_2)\tilde{S}(x_2)v_1(x_2) - [\tilde{T}(x_2)\mu_1(x_2)}$$

$$+ \tilde{S}(x_2)v_1(x_2)]' H^{-1}(x_2)[\tilde{T}(x_2)\mu_1(x_2) + \tilde{S}(x_2)v_1(x_2)]\} \tag{4.8}$$

$$\tilde{T}(x_2) = \frac{(n_1+p_2-2)T_{11}}{n_1 + (x_2-\mu_2)'(T_{22}-T_{21}T_{11}^{-1}T_{12})(x_2-\mu_2)}$$

$$\tilde{S}(x_2) = \frac{(n_2+p_2-2)S_{11}}{n_2 + (x_2-v_2)'(S_{22}-S_{21}S_{11}^{-1}S_{12})(x_2-v_2)}$$

$$\mu_1(x_2) = \mu_1 - T_{11}^{-1}T_{12}(x_2-\mu_2)$$

$$v_1(x_2) = v_1 - S_{11}^{-1}S_{12}(x_2-v_2) \quad \text{and}$$

$$H(x_2) = \tilde{T}(x_2) + \tilde{S}(x_2) .$$

If x_2 is scalar, we can plot the marginal density of each component of x, which allows one to compute the mean vector of x. If x_2 is 2×1, one can compute the covariance between the components of x_2, which allows us to calculate the variance covariance matrix of x. Of course one can check the approximation with the exact marginal density of x_2 using the numerical integration technique.

Approximation by Bernstein Polynomial

For θ_1 a scalar, the integrand $f(\theta_1) = \prod_{i=1}^{k} f_{i1}(\theta_1|\theta_2)$ in (4.5) can be approximated by a Bernstein polynomial of degree m (Weierstrass Approximation theorem, Lorentz, 1953) which is defined by

$$B_m(f;\theta_1) = \sum_{k=0}^{m} f[(b-a)\frac{k}{m}+a]\binom{m}{k}\frac{(\theta_1 - a)^k(b - \theta_1)^{m-k}}{(b-a)^m}, \qquad \theta_1 \in [a,b],$$

where f is the function that is to be approximated, and taking a and b three standard deviations from its mean, i.e.

$$a = \min\{\mu_1^{(i)}(\theta_2) - 3[\frac{n_i + p_2}{n_i + p_2 - 2}\{T^{(i)*}(\theta_2)\}^{-1}]^{\frac{1}{2}}, \quad i = 1, 2,\ldots, k\},$$

$$b = \max\{\mu_1^{(i)}(\theta_2) + 3[\frac{n_i + p_2}{n_i + p_2 - 2}\{T^{(i)*}(\theta_2)\}^{-1}]^{\frac{1}{2}}, \quad i = 1, 2,\ldots, k\}.$$

Hence

$$\int_R \prod_{i=1}^{k} f_{i1}(\theta_1|\theta_2)d\theta_1 \doteq \sum_{k=0}^{m} f[(b-a)\frac{k}{m} + a]\binom{m}{k} \int_a^b \frac{(\theta_1-a)^k(b-\theta_1)^{m-k}}{(b-a)^m} d\theta_1$$

$$= (\frac{b-a}{m+1}) \sum_{k=0}^{m} f[(b-a)\frac{k}{m} + a] \qquad (4.9)$$

Now suppose $\theta_1 \in R^{p_1}$, $p_1 \geq 2$. By the generalization of the Weierstrass approximation theorem (Lorentz, 1953) we have

$$B(f;\theta_{11},\theta_{12},\ldots,\theta_{1p_1}) = \sum_{k_1=0}^{m_1} \cdots \sum_{k_{p_1}=0}^{m_{p_1}} f[(b_1-a_1)\frac{k_1}{m_1} + a_1,\ldots,(b_{p_1}-a_{p_1})\frac{k_{p_1}}{m_{p_1}}$$

$$+ a_{p_1}]\binom{m_1}{k_1}\frac{(\theta_{11}-a_1)^{k_1}(b_1-\theta_{11})^{m_1-k_1}}{(b_1-a_1)^{m_1}} \cdots \binom{m_{p_1}}{k_{p_1}}\frac{(\theta_{1p_1}-a_{p_1})^{k_{p_1}}(b_{p_1}-\theta_{1p_1})^{m_{p_1}-k_{p_1}}}{(b_{p_1}-a_{p_1})^{m_{p_1}}},$$

where f is the function that is to be approximated, $\theta_{ij} \in [a_j,b_j]$, $j = 1, 2,\ldots, p_1$, and taking

$$a_j = \min_j\{\mu^{(i)}(\theta_2) - 3[\frac{n_i+p_2}{n_i+p_2-2}\{T^{(i)*}(\theta_2)\}^{-1}]^{\frac{1}{2}}, \quad i = 1, 2,\ldots, k\}$$

where $\min_j$ is the minimum value of the jth element,

$$b_j = \max_j \{\mu^{(i)}(\theta_2) + 3[\frac{n_i+p_2}{n_i+p_2-2}\{T^{(i)*}(\theta_2)\}^{-1}]^{\frac{1}{2}}, \quad i = 1, 2, \ldots, k\}$$

where $\max_j$ is the maximum value of the jth element, $j = 1, 2, \ldots, p_1$.

Hence

$$\int_{R^{p_1}} \prod_{i=1}^{k} f_{i1}(\theta_1|\theta_2)d\theta_1 \doteq \{\prod_{j=1}^{p_1} (\frac{b_j-a_j}{m_j+1})\} \sum_{k_1=0}^{m_1} \cdots \sum_{k_{p_1}=0}^{m_{p_1}} f[(b_1-a_1)\frac{k_1}{m_1} + a_1, \ldots, (b_{p_1}-a_{p_1})\frac{k_{p_1}}{m_{p_1}} + a_{p_1}] \quad (4.10)$$

Now we use the generalization of the Weierstrass approximation theorem to directly approximate the poly-t density (4.1).

$$B(f;\theta_1,\theta_2,\ldots,\theta_p) = \sum_{k_1=0}^{m_1} \cdots \sum_{k_p=0}^{m_p} f[(b_1-a_1)\frac{k_1}{m_1}+a_1,\ldots(b_p-a_p)\frac{k_p}{m_p}+a_p]$$

$$\binom{m_1}{k_1}\frac{(\theta_1-a_1)^{k_1}(b_1-\theta_1)^{m_1-k_1}}{(b_1-a_1)^{m_1}} \cdots \binom{m_p}{k_p}\frac{(\theta_p-a_p)^{k_p}(b_p-\theta_p)^{m_p-k_p}}{(b_p-a_p)^{m_p}},$$

where f is the function that is to be approximated, $\theta \in R^p$, $\theta_j \in [a_j,b_j]$, $j = 1, 2, \ldots, p$, and taking

$$a_j = \min\{\mu_j^{(i)} - 3\{[\frac{n_i}{n_i-2}(T_{jj}^{(i)} - T_{jk}^{(i)}T_{kk}^{(i)-1}T_{kj}^{(i)})]^{-1}\}^{\frac{1}{2}}, \quad i = 1, 2, \ldots, k\}$$

$$b_j = \max\{\mu_j^{(i)} + 3\{[\frac{n_i}{n_i-2}(T_{jj}^{(i)} - T_{jk}^{(i)}T_{kk}^{(i)-1}T_{kj}^{(i)})]^{-1}\}^{\frac{1}{2}}, \quad i = 1, 2, \ldots, k\}.$$

So the marginal density of θ_i is given by

$$f(\theta_i) \doteq c\{\prod_{\substack{j=1 \\ j\neq i}}^{p} (\frac{b_j-a_j}{m_j+1})\}(b_i-a_i)^{-m_i} \sum_{k_1=0}^{m_1} \cdots \sum_{k_p=0}^{m_p} f[(b_1-a_1)\frac{k_1}{m_1}+a_1, \ldots,$$

$$(b_p-a_p)\frac{k_p}{m_p}+a_p] \binom{m_i}{k_i}(\theta_i-a_i)^{k_i}(b_i-\theta_i)^{m_i-k_i} \qquad (4.11)$$

$$\text{where } c = [\{\prod_{j=1}^{p} (\frac{b_j-a_j}{m_j+1})\} \sum_{k_1=0}^{m_1} \cdots \sum_{k_p=0}^{m_p} f[(b_1-a_1)\frac{k_1}{m_1}+a_1, \ldots, (b_p-a_p)\frac{k_p}{m_p}+a_p]]^{-1}$$

$$(4.12)$$

Therefore,

$$E(\theta_i) = a_i + c\{\prod_{\substack{j=1 \\ j\neq i}}^{p} (\frac{b_j-a_j}{m_j+1})\}\frac{(b_i-a_i)^2}{(m_i+1)(m_i+2)} \sum_{k_1=0}^{m_1} \cdots \sum_{k_p=0}^{m_p} f[(b_1-a_1)\frac{k_1}{m_1}+a_1, \ldots$$

$$(b_p-a_p)\frac{k_p}{m_p}+a_p](k_i+1) \qquad (4.13)$$

and

$$V(\theta_i) = c\{\prod_{\substack{j=1 \\ j\neq i}}^{p} (\frac{b_j-a_j}{m_j+1})\} \frac{(b_i-a_i)^3}{(m_i+1)(m_i+2)(m_i+3)} \sum_{k_1=0}^{m_1} \cdots \sum_{k_p=0}^{m_p} f[(b_1-a_1)\frac{k_1}{m_1}+a_1,$$

$$\ldots, (b_p-a_p)\frac{k_p}{m_p}+a_p](k_i+1)(k_i+2) - \{E(\theta_i) - a_i\}^2 \qquad (4.14)$$

CHAPTER V

LINEAR DYNAMIC SYSTEMS

Linear dynamic systems are linear models which are used to solve communication and control problems in electrical engineering. They consist of two equations, namely an observation and systems equation.

From a statistical viewpoint, much has been accomplished in adaptive estimation, non-linear filtering, and control, however, very little has appeared from a Bayesian viewpoint.

The purpose in introducing the linear dynamic model is to demonstrate that, as with other linear models which are familiar to statisticians, this one is no different, in that the Bayesian analysis produces well-known posterior distributions and that consequently, the Bayesian approach, perhaps, will provide convenient solutions to some engineering problems. As will be shown, the current state of the system, $\theta(t)$, is a scalar subset of the vector of state vectors, which has a poly-t distribution, thus the filtering problem is solved once the marginal distribution of $\theta(t)$ is obtained.

Consider a sequence of observations $Y(i)$, $i = 1,2,\ldots$ where

$$Y(i) = F(i)\theta(i) + v(i) \tag{5.1}$$

and $$\theta(i) = G\theta(i-1) + w(i). \tag{5.2}$$

The observations are $n\times 1$ vectors, $F(i)$ is a $n\times p$ known matrix, and the

$v(i)$ are observational $n \times 1$ error vectors. The equation (5.1) is called the observational equation and it expresses the observations as linear functions of the state vector $\theta(i)$ and an additive error term. The systems equation (5.2) is a first order autoregressive process which explains the evolution of the system through time, where the states $\{\theta(i) : i = 1,2,...\}$ are unknown $p \times 1$ real parameter vectors, G is a $p \times p$ known systems design matrix, and $w(i)$ a $p \times 1$ systems error vector.

Suppose the $v(i)$ have mean zero and precision $P_{v(i)}$, a $n \times n$ positive definite matrix and let zero be the mean of $w(i)$ and suppose $P_{w(i)}$ is a $p \times p$ precision matrix of $w(i)$. Furthermore, suppose $\theta(0)$, the initial state is either known or is random with a $N(\mu_0, P_0)$ distribution, where μ_0 and P_0 are known.

The main goal is to estimate past, present, and future states, given t observations, $Y(i)$, $i = 1,2,...,t$. In the engineering literature, estimation consists of three things, namely:

1) Interpolation, that is estimate $\theta(i)$, $i = 1,2,...,t-1$, the past states of the system,
2) Filtering, that is estimate $\theta(t)$, the present state, and
3) Prediction, that is to estimate the future states $\theta(t+1)$, $\theta(t+2),....$

The foundation of estimation is the Kalman filter, which was presented in 1963. This is a sequential or recursive algorithm which estimates the present state $\theta(t)$, and it recursively updates the estimate of $\theta(t)$ from the past estimate of $\theta(t-1)$. See Harrison and Stevens (1976) for details of the filter.

Briefly, Kalman assumed the following. First, all the precision matrices $P_{v(i)}$, $P_{w(i)}$, $i = 1,2,...,t$ are known and the error vectors are independent normal vectors and independent of the initial state vector $\theta(0) \sim N(\mu_0, P_0)$. Kalman estimates $\theta(t)$ by the mean of the posterior distribution of $\theta(t)$, which is a normal distribution.

Since engineers are capable of designing systems in such a way to control the noise characteristics $v(i)$ and $w(i)$, the assumptions of known error precisions (dispersions) is not an unrealistic assumption. Note, statisticians assume the noise precision is usually unknown when they use regression models, because they analyze data where one is unable to control the noise. Nevertheless, there are situations where the noise precisions are unknown and when this is encountered, the term adaptive estimation is used by control engineers.

For example, suppose the observations and state vectors are scalar (n=1=p) and that the precisions are constant but unknown, that is

$$P_{v(i)} = \tau_v (>0)$$

$$P_{w(i)} = \tau_w (>0) \qquad \text{for } i = 1,2,\ldots.$$

Also assume that given $\theta(1), \theta(2),\ldots, \theta(t)$ and τ_v, that the observations $Y(1), Y(2),\ldots, Y(t)$ are independent normal variables and that given τ_w and $\theta(0)$, that $\theta(1), \theta(2),\ldots, \theta(t)$ follow a first order autoregressive process. Let the prior density of (τ_v, τ_w) be

$$p_0(\tau_v, \tau_w) \propto \tau_v^{\alpha_v - 1} e^{-\tau_v \beta_v} \tau_w^{\alpha_w - 1} e^{-\tau_w \beta_w}$$

and suppose (τ_v, τ_w) is independent, a priori, of $\theta(0)$. Then the posterior density of $\theta(1), \theta(2),\ldots, \theta(t), \tau_v, \tau_w$ is

$$p(\theta^t,\tau_v,\tau_w) \propto \tau_v^{\frac{t}{2}+\alpha_v-1} e^{-\frac{\tau_v}{2}\{\sum_{i=1}^{t}[Y(i)-F(i)\theta(i)]^2+2\beta_v\}}$$

$$\tau_w^{\frac{t}{2}+\alpha_w-1} e^{-\frac{\tau_w}{2}\{\sum_{i=1}^{t}[\theta(i)-G\theta(i-1)]^2+2\beta_w\}} \qquad (5.3)$$

for $\tau_v > 0$, $\tau_w > 0$, and $\theta^t = [\theta(1), \theta(2),\ldots, \theta(t)] \in R^t$.

We see from (5.3) that the marginal density of θ^t is

$$p(\theta^t) \propto \{2\beta_v + \sum_{i=1}^{t}[Y(i)-F(i)\theta(i)]^2\}^{-\frac{(2\alpha_v+t)}{2}} \{2\beta_w +$$

$$\sum_{i=1}^{t}[\theta(i)-G\theta(i-1)]^2\}^{-\frac{(2\alpha_w+t)}{2}} \qquad (5.4)$$

$$\propto \{1+\frac{(\theta-\hat{\theta})'F'F(\theta-\hat{\theta})}{2\beta_v+(Y-F\hat{\theta})'(Y-F\hat{\theta})}\}^{-\frac{(2\alpha_v+t)}{2}}$$

$$\{1+\frac{(\theta-A^{-1}B)'A(\theta-A^{-1}B)}{2\beta_w+C-B'A^{-1}B}\}^{-\frac{(2\alpha_w+t)}{2}}, \qquad (5.5)$$

where $\theta' = (\theta(1), \theta(2),\ldots, \theta(t))$, $\hat{\theta} = (F'F)^{-1}F'Y$

$$\underset{t\times t}{F} = \begin{pmatrix} F(1) & & \phi \\ & F(2) & \\ \phi & & \ddots \; F(t) \end{pmatrix}, \quad Y = \begin{pmatrix} Y(1) \\ Y(2) \\ \vdots \\ Y(t) \end{pmatrix},$$

F(1), F(2),..., F(t) must be nonsingular,

$$\underset{t\times t}{A} = \begin{pmatrix} 1+G^2 & -G & 0 & \cdots & \cdots & 0 \\ -G & 1+G^2 & -G & \cdots & \cdots & \vdots \\ 0 & -G & 1+G^2 & -G & \cdots & \vdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & \cdots & -G & 1+G^2 & -G \\ 0 & \cdots & \cdots & 0 & -G & 1 \end{pmatrix}$$

$$\underset{1\times t}{B'} = (G\theta(0) \quad 0 \quad 0 \; . \; . \; . \; . \; 0)$$

$$C = G^2\theta(0),$$

for $\theta^t \in R^t$, and the posterior distribution of θ^t is a 2/0 t-dimensional poly-t distribution. Each factor is a proper t density, the first with $2\alpha_v$ degrees of freedom, and the second with $2\alpha_w$.

Thus, the filtering problem is to find the marginal posterior distribution of $\theta(t)$ for all $t = 1, 2, \ldots,$ and it would be necessary to use the approximation. Note, the degrees of freedom of each factor is independent of the number of observations, and the degrees of freedom depend only on the prior information for τ_v and τ_w.

Adaptive Estimation of Linear Dynamic Models Using a Normal Approximation to the Conditional of $\theta(k)$ Given $\theta(1), \theta(2), \ldots, \theta(k-1)$

Consider (5.1) and (5.2) where the observations $Y(i)$, $i = 1,2,\ldots,$

are n×1 vectors, F(i) is a n×p known matrix. Assume that θ(0) is known, V(i) are normally independently distributed with mean zero and precision $P_{v(i)}$, a n×n positive definite matrix and W(i) are normally independently distributed with mean zero and precision $P_{w(i)}$, a p×p positive definite matrix, F(i) and G are known, v(1), v(2),..., w(1), w(2),..., are independent sequences, $P_{v(i)}$, $P_{w(i)}$ are unknown and $P_{v(i)} = P_v$, $P_{w(i)} = P_w$, Cov[v(i),θ(i-1)] = 0, i = 1,2,....

Our interest is to build a recursive formula by which one may estimate θ(k), given k observations Y(1), Y(2),... Y(k). One way of using the Bayesian approach is to find the conditional posterior distribution of θ(k), given θ(1), θ(2),..., θ(k-1), for k = 1, 2,....

Let k = 1. Assume that P_v and P_w are apriori independent and the prior density of P_v is

$$|P_v|^{\frac{\alpha_v-n-1}{2}} e^{-\frac{1}{2}\mathrm{Tr}\,(T_vP_v)}, \qquad |P_v| > 0$$

and

$$|P_w|^{\frac{\alpha_w-p-1}{2}} e^{-\frac{1}{2}\mathrm{Tr}\,(T_wP_w)}, \qquad |P_w| > 0$$

for the prior density of P_w. The joint density of Y(1), θ(1), P_v and P_w is proportional to

$$|P_v|^{\frac{\alpha_v-n}{2}} e^{-\frac{1}{2}\mathrm{Tr}\,\{[Y(1)-F(1)\theta(1)][Y(1)-F(1)\theta(1)]' + T_v\}P_v}$$

$$|P_w|^{\frac{\alpha_w - p}{2}} e^{-\frac{1}{2}\mathrm{Tr}\{[\theta(1)-G\theta(0)][\theta(1)-G\theta(0)]' + T_w\}P_w},$$

where P_v and P_w are PDS, $Y(1) \in R^n$ and $\theta(1) \in R^p$. Eliminate P_v and P_w by properties of Wishart distribution, we have

$$g[Y(1),\theta(1)] \propto |T_v + [Y(1) - F(1)\theta(1)][Y(1) - F(1)\theta(1)]'|^{-\frac{(\alpha_v+1)}{2}}$$

$$|T_w + [\theta(1) - G\theta(0)][\theta(1) - G\theta(0)]'|^{-\frac{(\alpha_w+1)}{2}}.$$

The posterior density of $\theta(1)$ is a 2/0 p-variate poly-t

$$g[\theta(1)|Y(1)] \propto \{1 + [\theta(1) - \mu_{11}]'P_{11}[\theta(1) - \mu_{11}]\}^{-\frac{(\alpha_v+1)}{2}}$$

$$\{1 + [\theta(1) - \mu_{12}]'P_{12}[\theta(1) - \mu_{12}]\}^{-\frac{(\alpha_w+1)}{2}} \quad \theta(1) \in R^p \tag{5.6}$$

where $\mu_{11} = [F'(1)T_v^{-1}F(1)]'F'(1)T_v^{-1}Y(1)$,

$$P_{11} = F'(1)T_v^{-1}F'(1)/\{1 + Y'(1)T_v^{-1}Y(1) - \mu_{11}'F'(1)T_v^{-1}Y(1)\},$$

$$\mu_{12} = G\theta(0)$$

and $P_{12} = T_w^{-1}$.

So $\theta(1)$ has a 2/0 p-variate poly-t distribution. Therefore $\theta(1)$ has an approximate normal distribution with mean

$$m(1) = (P^*_{11} + P^*_{12})^{-1}(P^*_{11}\mu_{11} + P^*_{12}\mu_{12}) \tag{5.7}$$

and precision $P(1) = (P^*_{11} + P^*_{12})$ (5.8)

where $P^*_{11} = (\alpha_v - p - 1)P_{11}$,

$$P^*_{12} = (\alpha_w - p - 1)P_{12}.$$

Now let $k \geq 2$. The joint density of $Y(1), Y(2), \ldots, Y(k), \theta(1), \theta(2), \ldots, \theta(k), P_v$ and P_w is

$$g[Y^{(k)}, \theta^{(k)}, P_v, P_w] \propto$$

$$|P_v|^{\frac{\alpha_v+k-n-1}{2}} e^{-\frac{1}{2}Tr\{\sum_{i=1}^{k}[Y(i)-F(i)\theta(i)][Y(i)-F(i)\theta(i)]' + T_v\}P_v}$$

$$|P_w|^{\frac{\alpha_w+k-p-1}{2}} e^{-\frac{1}{2}Tr\{\sum_{i=1}^{k}[\theta(i)-G\theta(i-1)][\theta(i)-G\theta(i-1)]' + T_w\}P_w}$$

where $Y^{(k)} = \{Y(1), Y(2), \ldots, Y(k)\}$, $\theta^{(k)} = \{\theta(1), \theta(2), \ldots, \theta(k)\}$.

Eliminating P_v and P_w we have,

$$h[Y^{(k)},\theta^{(k)}] \propto \left|T_v + \sum_{i=1}^{k} [Y(i) - F(i)\theta(i)][Y(i) - F(i)\theta(i)]'\right|^{-\frac{(\alpha_v+k)}{2}}$$

$$\left|T_w + \sum_{i=1}^{k} [\theta(i) - G\theta(i-1)][\theta(i) - G\theta(i-1)]'\right|^{-\frac{(\alpha_w+k)}{2}}.$$

Therefore,

$$h[\theta(k)|\underset{i\neq k}{\theta(i)'s},Y^{(k)}] \propto |Q_1|^{-\frac{(\alpha_v+k)}{2}} \{1+[\theta(k)-\mu_{k1}]'P_{k1}[\theta(k)-\mu_{k1}]\}^{-\frac{(\alpha_v+k)}{2}} \tag{5.9}$$

$$|Q_2|^{-\frac{(\alpha_w+k)}{2}} \{1+[\theta(k)-\mu_{k2}]'P_{k2}[\theta(k)-\mu_{k2}]\}^{-\frac{(\alpha_w+k)}{2}}, \quad \theta \in R^p$$

where $\mu_{k1} = [F'(k)Q_1^{-1}F(k)]^{-1}F'(k)Q_1^{-1}Y(k)$.

$$Q_1 = T_v + \sum_{i=1}^{k-1} [Y(i) - F(i)\theta(i)][Y(i) - F(i)\theta(i)]',$$

$$P_{k1} = F'(k)Q_1^{-1}F(k)/\{1 + Y'(k)Q_1^{-1}Y(k) - \mu_{k1}'F'(k)Q_1^{-1}Y(k)\},$$

$$\mu_{k2} = G\theta(k-1),$$

$$Q_2 = T_w + \sum_{i=1}^{k-1} [\theta(i) - G\theta(i-1)][\theta(i) - G\theta(i-1)]',$$

$$P_{k2} = Q_2^{-1}.$$

So the conditional posterior density of $\theta(k)$ given $\theta(1)$, $\theta(2)$,..., $\theta(k-1)$ for k = 2, 3,..., is a 2/0 p-variate poly-t. This conditional distribution is approximately normal with mean

$$m(k) = (P^*_{k1} + P^*_{k2})^{-1}(P^*_{k1}\mu_{k1} + P^*_{k2}\mu_{k2}) \tag{5.10}$$

and precision $P(k) = (P^*_{k1} + P^*_{k2})$ (5.11)

where $P^*_{k1} = (\alpha_v + k - p - 2)P_{k1}$,

and $P^*_{k2} = (\alpha_w + k - p - 2)P_{k2}$.

So for each k, one may compute m(k) which estimates $E\{\theta(k)|\hat{\theta}(1),\hat{\theta}(2),\ldots,\hat{\theta}(k-1)\}$ where $\hat{\theta}(i)$ is the conditioning value of $\theta(i)$. For example, one might use

$$\begin{aligned}
\hat{\theta}(1) &= E\{\theta(1)|\theta(0)\} \\
\hat{\theta}(2) &= E\{\theta(2)|\theta(0),\hat{\theta}(1)\} \\
&\vdots \\
\hat{\theta}(k) &= E\{\theta(k)|\theta(0),\hat{\theta}(1),\ldots,\hat{\theta}(k-1)\},
\end{aligned} \tag{5.12}$$

then the conditioning values are the conditional means.

CHAPTER VI

NUMERICAL STUDY

Whenever a bivariate 2/0 poly-t is used in this chapter, we refer to the following distribution:

$$f(\theta) \propto [1 + \frac{1}{n_1}(\theta-u)'T(\theta-u)]^{-\frac{(n_1+2)}{2}} [1 + \frac{1}{n_2}(\theta-v)'S(\theta-v)]^{-\frac{(n_2+2)}{2}}, \quad \theta \in R^p$$

where $\theta = (\theta_1 \;\; \theta_2)'$, $u = (u_1 \;\; u_2)'$, $v = (v_1 \;\; v_2)'$, $T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$ and $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$; u and v are called the "means" of the first and second factor, respectively. Similarly, T and S are called the "precisions" and n_1 and n_2 "degrees of freedom" of each factor.

The first part of the numerical study is based on the iterative method (Lindley and Smith, 1972) applied to univariate two factor poly-t, then to the two normal linear regression equations (Zellner, 1971, p. 102). Then we look at the effect of changing the "precision" for various values of "means" of the two factors of the bivariate 2/0 poly-t in several regression models. From here we proceed to the approximation by normal and Bernstein polynomials, and the last part of the chapter is the study of the adaptive estimation of Linear Dynamic Models using normal approximation to the conditional distribution of the present state.

Table I presents the comparison of the mode of the marginal density of θ (equation (3.5) for k=2) by numerical integration, the mode by iteration from the modal equation (3.7), and the marginal mean of θ by numerical integration. The mode obtained through the iterative method is the mode of the joint density of all the parameters in the model and in general, it is not the mode of the marginal density (O'Hagan, A. 1976). Sample values used are $n_1 = n_2 = 10$,

$\sum_{j=1}^{n_1} (X_{1j}-\bar{X}_1)^2 = 31.7882$, $\sum_{j=1}^{n_2} (X_{2j}-\bar{X}_2)^2 = 9.999\overline{9}$, $\bar{X}_1 = -.1499$ and $\bar{X}_2 =$ .3611

From Table 1 we observe that, if β's and ξ's are equal to zero, changing the value of μ's would not change the modes and mean of θ and this is obvious from equations (3.4) and (3.5). The joint mode obtained through iteration is quite close to the mode of marginal by numerical integration whatever the values of α, β, ξ and μ's. These modes can be used to estimate means because the distribution is unimodal and n_1 and n_2 are fairly large.

Direct application of the iterative method to the bivariate problem was done using the two normal linear regression equations from Zellner (1971, p. 102) with improper prior. The modal equations are:

$$M(\tau_i|\theta) = (n_i+1)/[\nu_i s_i^2 + (\theta - \hat{\theta}_i)'X_i'X_i(\theta - \hat{\theta}_i)]$$

and $M(\theta|\tau_i) = (\sum_{i=1}^{2} \tau_i X_i'X_i)^{-1}(\sum_{i=1}^{2} X_i'X_i\hat{\theta}_i)$, where $n_i = 19$, $\nu_i = 17$, $\hat{\theta}_i =$

$(X_i'X_i)^{-1}X_i'Y_i$ for $i = 1, 2$, $s_1^2 = 777.4463$, $s_2^2 = 104.3079$, $X_1'X_1 = s_1^2 S$,

$$X_2'X_2 = s_2^2 T, \quad S = \begin{pmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 1535.0640 \end{pmatrix}, \text{ and } \quad T = \begin{pmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 706.3320 \end{pmatrix}$$

The mode of the joint distribution is found to be (.0371, .1448) which is very close to the mode of the marginal distribution of θ given in Zellner (1971), i. e. (.0373, .1446).

With slight changes in the values of n_1, n_2, S and T from the two normal linear regression problem, the iterative method was done for different values of $\hat{\theta}_i$, i = 1, 2. Using the same modal equations with $n_1 = n_2 = 5$, $\nu_1 = \nu_2 = 3$, the results of this study are given in Table 2 (only the mode corresponding to θ_2 is presented) for different values of S and T and different values of $\hat{\theta}_1 = (u_1 \;\; u_2)'$ and $\hat{\theta}_2 = (v_1 \;\; v_2)'$. When $\hat{\theta}_1$ and $\hat{\theta}_2$ are far enough apart, the distribution becomes bimodal, so the mode by iteration depends on the starting value. From Table 2 we notice that the mode of θ_2 by iteration is close to the mode of marginal of θ_2 by numerical integration.

Approximation by the normal distribution was developed in Chapter IV and was applied to two normal linear regression problems (Zellner 1971, p. 102) and the results are:

	Approximation (Normal)	Asymptotic Expansion (Zellner)
Mode:	.0373, .1447	.0373, .1447
Mean:	.0374	.0373
Variance:	10.1964×10^{-5}	9.6158×10^{-5}

The results are very close to the one obtained by Zellner based on the leading normal term in the asymptotic expansion.

Then approximation by normal was applied to the marginal density of θ_2 which is a bimodal, for various values of u and v, S and T. Comparisons of the mode of the exact with the approximation by normal are given in Tables 3, 4, and 5.

Since approximation by the normal distribution is not that good for the bimodal marginal density of a subset of θ (Tables 3 - 5 and Figures 6 - 10), we use a Bernstein polynomial to approximate the integration in (4.5). We have to determine the degree of polynomial in order to achieve a good approximation. A bivariate 2/0 poly-t was considered where $u = (-.5 \quad -.5)'$, $v = (.5 \quad .5)'$,

$$T = \begin{pmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.00 \end{pmatrix}, \qquad S = \begin{pmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 1500.00 \end{pmatrix}, \text{ and } n_1 = n_2 = 3.$$

The exact marginal of θ_2(4.5) is compared with the approximate marginal using a Bernstein polynomial (4.9) for different degrees of the polynomial m. From Figures 1 - 4, the value of m that gives a good approximation (in this particular example) must be greater than 120.

Approximation by Bernstein polynomial of the integration in the marginal density of θ_1 and θ_2 was also applied to the example on the two normal linear regression problem (Zellner 1971, p. 102) with m = 250. The results are: mode of θ_1 = .0370, mode of θ_2 = .1450, mean of θ_1 = .0373, and variance of $\theta_1 = 10.240 \times 10^{-5}$.

Comparison of the exact, and the approximations by the normal distribution and the Bernstein polynomial (m = 250) of the marginal density of θ_2 (4.5) were done, and the sample quantities are given together with the results in Tables 3 - 5. The graphs of these comparisons, for a few cases, are shown in Figures 5 - 10. The conclusion is that the approximation by Bernstein polynomial is more accurate than that of normal.

If a generalized Bernstein polynomial is applied directly to a bivariate 2/0 poly-t, it involves double summation. Let the first summation be from 0 to m_1 and the second one from 0 to m_2. This approximation is unlikely to be satisfied unless the values of m_1 and m_2 are fairly large, about 200 say. Therefore, it will be a very expensive method. In any case, the approximation by generalized Bernstein polynomial was done on the two normal linear regression equations (Zellner, 1971, p. 102) using $m_1 = m_2 = 25$. The mode of the marginal density of θ_1 is found to be .0369 which is a good approximation, but the value of the density at this point is 21.6607 which is about half of the true value (40.647). Similarly, the mode of the marginal distribution of θ_2 is .1439 which is a good approximation, but the value of the density at this point is 7.7434 while the true value is 17.068. The value of the approximating polynomial function at the mode increases with the values of m_1 and m_2.

With regard to adaptive estimation of a Linear Dynamic Model using a normal approximation to the conditional distribution of $\theta(k)$ given $\theta(1)$, $\theta(2)$,..., $\theta(k-1)$, the study was done using the following procedure:

To generate Linear Dynamic Model variables, two sets of bivariate normal random vectors were generated, the first set V, was generated from a normal population with mean vector zero and observational precision P_v and the second set W, was generated from a normal population with mean vector zero and system precision P_w. Then using the system equation we have $\theta(1) = G\theta(0) + W(1)$, and by observational equation we have $Y(1) = F(1)\theta(1) + V(1)$. Using equations (5.7) and (5.8), the mean of $\theta(1)$, called $m(1)$ and precision $P(1)$ were obtained.

Then $\theta(2)$, $Y(2)$, $m(2)$ and $P(2)$ were calculated using $\theta(k) = Gm(k-1) + W(k)$, $Y(k) = F(k)\theta(k) + V(k)$, equations (5.10) and (5.11), respectively, for $k = 2$. This recursive process was repeated for all k till $k = 30$, m(k) of (5.10) estimates θ(k), of (5.2).

Sample quantities used: $F = \begin{pmatrix} 3.5 & 3 \\ 3 & 3 \end{pmatrix}$, $G' = \begin{pmatrix} .5 & .5 \\ .5 & .5 \end{pmatrix}$, $\alpha_v = \alpha_w = 4$, $T_v = \alpha_v P_v^{-1}$, $T_w = \alpha_w P_w^{-1}$, and P_v, P_w and θ(0) are given together with the results in Tables 6 - 10. Table presents the results of the recursive process when both observational precision P_v and system precision P_w are small. Although the estimate is not good, we notice that it follows the right trend, i.e., most of the time positive m(k) corresponds to positive θ(k). Table 7 presents the results of this recursive process when P_v is large and P_w is small. We notice that there is a lot of variation in θ(k) and m(k), because P_w is small. In Table 8, there is little variation in θ(k) and m(k), because P_w is large and the estimates are good. Table 9 presents the results of this recursive process when P_v and P_w are fairly large and Table 10 presents the results of this recursive process when P_v and P_w are quite large. The conclusion is that the larger the system precision, the better estimate we have. Precisions of θ(k) are not given in the table, but it is found that the larger the system precision, the larger the precision of θ(k), and vice versa.

CHAPTER VII

CONCLUSIONS

From the numerical study in Chapter VI, we can conclude that the iterative method is a good approximation for estimating the mode of the joint distribution and can be used to estimate the mean of a unimodal distribution which is nearly symmetric. In the bimodal case, the modal value will converge to a local mode, depending on the starting value.

Approximation by normal distribution of the integral in the marginal density of a subset of poly-t random variable is good for the unimodal case when one has large "degrees of freedom" and high "precision" of each factor. When the "means" of each factor are far enough apart, the marginal becomes bimodal and the approximation by the normal distribution does not give good results, so one way of improving it is by using Bernstein polynomial to approximate the integral. The results using Bernstein polynomial approximation are satisfactory. The higher the degree of polynomial used in Bernstein approximation, the more accurate results we have, but the more expensive the method becomes, especially when it involves double summation, for example, in the direct application of generalized Bernstein polynomial to the bivariate 2/0 poly-t.

With regard to adaptive estimation of Linear Dynamic Systems, the joint posterior distribution of $\theta(1)$, $\theta(2),\ldots,$ $\theta(k)$ is a product of matricvariate t distributions, and we look at the conditional distribution of $\theta(k)$ given $\theta(1)$, $\theta(2),\ldots,$ $\theta(k-1)$ which is distributed as a multivariate product form poly-t. From the numerical study using a normal approximation to the conditional distribution of the present state, we notice that the system precision plays an important role in deciding whether the estimate is good or not. The higher the system precision the better estimate we have and vice versa.

REFERENCES CITED

1. Ando, A., and G. M. Kaufman (1965). Bayesian Analysis of the Independent Multinormal Process--Neither Mean nor Precision Known. Journal of the American Statistical Association, 60, 347-358.

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

3. Dayal, H. H., and J. M. Dickey (1977). Numerical Evaluation of Integrals Involving Products of t-densities. Journal of Statistical Computation and Simulation, G, 19-28.

4. DeGroot, M. H. (1970). Optimal Statistical Decisions. New York: McGraw-Hill Book Co.

5. Dickey, J. M. (1967). Matricvariate Generalization of the Multivariate t Distribution and the Inverted Multivariate t Distribution. Annals of Mathematical Statistics, 38, 511-518.

6. Dickey, J. M. (1968). Three Multidimensional Integral Identities with Bayesian Application. Annals of Mathematical Statistics, 39, 1615-1627.

7. Dreze, J. H. (1977). Bayesian Regression Analysis Using Poly-t Densities. Journal of Econometrics, 6, 329-354.

8. Fienberg, S. E., and A. Zellner (1975). Studies in Bayesian Econometrics and Statistics. Amsterdam: North-Holland Pub. Co.

9. Harrison, P. J., and C. F. Stevens (1976). Bayesian Forecasting. Journal of the Royal Statistical Society B, 38, 205-228.

10. Jeffreys, H. (1966). Theory of Probability. (3rd ed.) Oxford: Clarendon.

11. Kalman, R. E. (1963). New Methods in Wiener Filtering Theory, Proceedings of the Symposium on Engineering Applications of Random Function Theory and Probability (J. L. Bogdanoff and F. Kozin, eds.). New York: John Wiley and Sons, Inc.

12. Lindley, D. V. (1971). The Estimation of Many Parameters. In Foundation of Statistical Inference (V. P. Goodambe and D. A. Sprott, eds.), pp. 435-455. Toronto: Holt, Rinehart and Winston.

13. Lindley, D. V., and A. F. M. Smith (1972). Bayes Estimates for the Linear Model. Journal of the Royal Statistical Society B, 34, 1-18.

14. Lorentz, G. G. (1953). Bernstein Polynomials. Toronto: University of Toronto Press.

15. O'Hagan, A. (1976). On Posterior Joint and Marginal Modes. Biometrika, 63, 2, 329-333.

16. Press, S. I. (1972). Applied Multivariate Analysis. New York: Holt, Rinehart and Winston.

17. Raiffa, H., and R. Schlaifer (1961). Applied Statistical Decision Theory. Boston, Massachusetts: Harvard University Press.

18. Ralston, A. (1965). A First Course in Numerical Analysis. New York: McGraw-Hill Book Company.

19. Richard, J. F., and H. Tompa (1980). On the Evaluation of Poly-t Density Functions. Journal of Econometrics, 12, 335-351.

20. Sedory, S. A. (1981). Pursuing the Poly-t. Unpublished Masters Report, Department of Statistics, Oklahoma State University.

21. Tiao, G. C., and A. Zellner (1964a). Bayes Theorem and the Use of Prior Knowledge in Regression Analysis. Biometrika, 51, 219-230.

22. Tiao, G. C., and A. Zellner (1964b). On the Bayes Estimation of Multivariate Regression. Journal of the Royal Statistical Society B, 26, 277-285.

23. Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics. New York: John Wiley & Sons, Inc.

24. Zellner, A. (1976). Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms. Journal of the American Statistical Association, 71, 400-405.

APPENDIXES

APPENDIX A

TABLES

TABLE I

COMPARISON OF THE MODE (FROM MARGINAL DISTRIBUTION), MODE (BY ITERATION) AND THE MEAN

α_1	α_2	β_1	β_2	ξ_1	ξ_2	μ_1	μ_2	Mode (Marginal)	Mode (Iterative)	Mean
0	0	0	0	0	0	0	0	.2458	.2418	.2261
0	0	0	0	0	0	2	2	.2458	.2418	.2261
0	0	0	0	0	0	4	4	.2458	.2418	.2261
0	0	0	0	0	0	6	6	.2458	.2418	.2261
0	0	0	0	0	0	8	8	.2458	.2418	.2261
0	0	0	0	0	0	10	10	.2458	.2418	.2261
0	0	0	0	2	2	0	0	.2060	.1995	.1867
0	0	0	0	4	4	0	0	.1661	.1698	.1589
0	0	0	0	6	6	0	0	.1528	.1478	.1384
0	0	0	0	8	8	0	0	.1262	.1308	.1225
0	0	0	0	10	10	0	0	.1130	.1173	.1099
0	0	0	0	12	12	0	0	.1130	.1064	.0997
0	0	0	0	14	14	0	0	.0997	.0973	.0911
0	0	0	0	16	16	0	0	.0864	.0896	.0840
0	0	0	0	18	18	0	0	.0864	.0831	.0779
0	0	1	3	0	0	0	0	.1894	.1989	.1867
0	0	2	6	0	0	0	0	.1528	.1679	.1592
0	0	3	9	0	0	0	0	.1528	.1443	.1389
0	0	4	12	0	0	0	0	.1162	.1260	.1231
0	0	5	15	0	0	0	0	.1162	.1113	.1105
0	0	6	18	0	0	0	0	.1162	.0992	.1001
0	0	7	21	0	0	0	0	.0797	.0891	.0915
0	0	8	24	0	0	0	0	.0797	.0806	.0841
0	0	9	27	0	0	0	0	.0797	.0733	.0778
2	2	0	0	0	0	0	0	.2458	.2418	.2305
4	4	0	0	0	0	0	0	.2458	.2418	.2330
6	6	0	0	0	0	0	0	.2458	.2418	.2346
8	8	0	0	0	0	0	0	.2458	.2418	.2357
10	10	0	0	0	0	0	0	.2458	.2418	.2365
12	12	0	0	0	0	0	0	.2458	.2418	.2371
14	14	0	0	0	0	0	0	.2458	.2418	.2376
16	16	0	0	0	0	0	0	.2458	.2418	.2380
2	2	1	1	3	3	-8	-8	-1.7741	-1.7634	-1.7636
2	2	1	1	3	3	-6	-6	-1.2957	-1.2967	-1.2974
2	2	1	1	3	3	-4	-4	-.8173	-.8225	-.8245
2	2	1	1	3	3	-2	-2	-.3389	-.3299	-.3349
2	2	1	1	3	3	0	0	.1794	.1745	.1664
2	2	1	1	3	3	2	2	.6179	.6262	.6187
2	2	1	1	3	3	4	4	1.0565	1.0572	1.0520
2	2	1	1	3	3	6	6	1.4950	1.4999	1.4966
2	2	1	1	3	3	8	8	1.9336	1.9512	1.9489
2	2	1	1	0	0	1	1	.2359	.2299	.2191
2	2	1	1	3	3	1	1	.4186	.4078	.3996

TABLE I (Continued)

α_1	α_2	β_1	β_2	ξ_1	ξ_2	μ_1	μ_2	Mode (Marginal)	Mode (Iterative)	Mean
2	2	1	1	6	6	1	1	.5382	.5189	.5124
2	2	1	1	9	9	1	1	.5781	.5949	.5895
2	2	1	1	12	12	1	1	.6578	.6502	.6455
2	2	1	1	15	15	1	1	.6977	.6922	.6881
2	2	1	1	18	18	1	1	.7375	.7252	.7215
2	2	1	3	2	2	2	2	.4983	.4947	.4877
4	4	2	6	4	4	4	4	1.2425	1.2501	1.2477
6	6	3	9	6	6	6	6	2.3322	2.3318	2.3301
8	8	4	12	8	8	8	8	3.6346	3.6236	3.6229

TABLE II

MODES OF θ_2 BY ITERATIVE METHOD

				$S = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.00 \end{bmatrix}$, $T = \begin{bmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 1500.00 \end{bmatrix}$				$S = \begin{bmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 1500.00 \end{bmatrix}$, $T = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.00 \end{bmatrix}$				$S = T = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.00 \end{bmatrix}$			
u_1	u_2	v_1	v_2	Exact	App.	Exact	App.	Exact	App.	Exact	App.	Exact	App.	Exact	App.
.00	.00	.00	.00	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
.05	.05	.05	.05	-.0400	-.0374	-.0400	.0506	.0400	-.0506	.0400	.0374	.0000	-.0439	.0000	.0439
-.10	-.10	.10	.10	-.1199	-.0942	.0800	.1003	-.0800	-.1003	.1199	.0942	-.0800	-.0971	.0800	.0971
-.15	-.15	.15	.15	-.1600	-.1462	.1600	.1502	-.1600	-.1502	.1600	.1461	-.1600	-.1481	.1600	.1481
-.20	-.20	.20	.20	-.2000	-.1971	.2000	.2002	-.2000	-.2002	.2000	.1971	-.2000	-.1986	.2000	.1986
-.25	-.25	.25	.25	-.2399	-.2477	.2399	.2501	-.2399	-.2501	.2399	.2477	-.2399	-.2489	.2399	.2489
-.30	-.30	.30	.30	-.3200	-.2981	.2800	.3001	-.2800	-.3001	.3200	.2981	-.2800	-.2991	.2800	.2991
-.35	-.35	.35	.35	-.3600	-.3484	.3600	.3501	-.3600	-.3501	.3600	.3484	-.3600	-.3492	.3600	.3492
-.40	-.40	.40	.40	-.4000	-.3986	.4000	.4001	-.4000	-.4001	.4000	.3986	-.4000	-.3993	.4000	.3993
-.45	-.45	.45	.45	-.4400	-.4487	.4400	.4501	-.4400	-.4501	.4400	.4487	-.4400	-.4494	.4400	.4494
-.50	-.50	.50	.50	-.5200	-.4988	.4800	.5001	-.4800	-.5001	.5200	.4989	-.4800	-.4994	.4800	.4994
-.75	-.75	.75	.75	-.7600	-.7492	.7600	.7500	-.7600	-.7500	.7600	.7492	-.7600	-.7496	.7600	.7496
-1.00	-1.00	1.00	1.00	-.9999	-.9994	1.0000	1.0000	-.9999	-1.0000	1.0000	.9994	-.9999	-.9997	1.0000	.9997
-1.25	-1.25	1.25	1.25	-1.2399	-1.2495	1.2399	1.2500	-1.2399	-1.2500	1.2399	1.2495	-1.2399	-1.2498	1.2399	1.2498

TABLE III

MODES, MEAN AND VARIANCE OF THE EXACT, APPROXIMATE (NORMAL) AND APPROXIMATE (BERNSTEIN) MARGINAL POSTERIOR DISTRIBUTION OF θ_2

$S = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.0000 \end{bmatrix}$, $T = \begin{bmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 1500.0000 \end{bmatrix}$. $n_1 = n_2 = 3$

				Modes			Mean			Variance		
u_1	u_2	v_1	v_2	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Nor.)	App. (Ber.)
.00	.00	.00	.00	.0000	.0000	.0000	.0000	.0000	.0000	.0008	.0008	.0008
-.05	-.05	.05	.05	-.0400	-.0800 .0800	-.0400	-.0348	-.0622	-.0348	.0028	.0043	.0028
-.10	-.10	.10	.10	-.1199 .0800	-.1199 .1600	-.1199 .0800	-.0716	-.1344	-.0716	.0076	.0122	.0076
-.15	-.15	.15	.15	-.1600 .1600	-.2000 .2000	-.1600 .1600	-.1011	-.1927	-.1012	.0145	.0237	.0145
-.20	-.20	.20	.20	-.2000 .2000	-.2399 .2800	-.2000 .2000	-.1402	-.2391	-.1402	.0226	.0350	.0227
-.25	-.25	.25	.25	-.2399 .2399	-.2800 .3200	-.2399 .2399	-.1708	-.2859	-.1708	.0367	.0452	.0367
-.30	-.30	.30	.30	-.3200 .2800	-.3200 .3600	-.3200 .2800	-.2035	-.3290	-.2035	.0499	.0577	.0499
-.35	-.35	.35	.35	-.3600 .3600	-.3600 .4000	-.3600 .3600	-.2446	-.3696	-.2446	.0662	.0630	.0663
-.40	-.40	.40	.40	-.4000 .4000	-.4000 .4400	-.4000 .4000	-.2696	-.4139	-.2696	.0901	.0758	.0901
-.45	-.45	.45	.45	-.4400 .4400	-.4800 .4800	-.4400 .4400	-.3107	-.4528	-.3107	.1069	.0828	.1069
-.50	-.50	.50	.50	-.5200 .4800	-.5200 .5200	-.5200 .4800	-.3440	-.4977	-.3440	.1364	.0890	.1364
-.75	-.75	.75	.75	-.7600 .7600	-.7600 .7600	-.7600 .7600	-.5217	-.7121	-.5218	.2951	.1322	.2951

TABLE III (Continued)

$$S = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.0000 \end{bmatrix},\quad T = \begin{bmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 1500.0000 \end{bmatrix}.\quad n_1 = n_2 = 3$$

				Modes			Mean			Variance		
u_1	u_2	v_1	v_2	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Nor.)	App. (Ber.)
-1.0	-1.0	1.0	1.0	$-.999\overline{9}$ $1.000\underline{0}$	$-.999\overline{9}$ $1.000\underline{0}$	$-.999\overline{9}$ $1.000\underline{0}$	-.6981	-.9318	-.7011	.5148	.1936	.5107
-1.25	-1.25	1.25	1.25	$-1.239\overline{9}$ $.239\overline{9}$	$-1.239\overline{9}$ $1.279\overline{9}$	$-1.239\overline{9}$ $1.239\overline{9}$	-.8651	-1.1526	-.8565	.8120	.2840	.8283

TABLE IV

MODES, MEAN AND VARIANCE OF THE EXACT, APPROXIMATE (NORMAL) AND APPROXIMATE (BERNSTEIN) MARGINAL POSTERIOR DISTRIBUTION OF θ_2

$$S = \begin{bmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 1500.0000 \end{bmatrix}, \quad T = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.0000 \end{bmatrix}. \quad n_1 = n_2 = 3$$

u_1	u_2	v_1	v_2	Modes: Exact	Modes: App. (Normal)	Modes: App. (Berns.)	Mean: Exact	Mean: App. (Normal)	Mean: App. (Berns.)	Variance: Exact	Variance: App. (Nor.)	Variance: App. (Ber.)
.00	.00	.00	.00	.0000	.0000	.0000	.0000	.0000	.0000	.0008	.0008	.0008
-.05	-.05	.05	.05	.0400	-.0800 .0800	.0400	.0348	.0622	.0348	.0028	.0043	.0028
-.10	-.10	.10	.10	-.0800 .1199	-.1600 .1199	-.0800 .1199	.0716	.1344	.0716	.0076	.0122	.0076
-.15	-.15	.15	.15	-.1600 .1600	-.2000 .2000	-.1600 .1600	.1011	.1927	.1012	.0145	.0237	.0145
-.20	-.20	.20	.20	-.2000 .2000	-.2300 .2399	-.2000 .2000	.1402	.2391	.1402	.0226	.0350	.0227
-.25	-.25	.25	.25	-.2399 .2399	-.3200 .2800	-.2399 .2399	.1708	.2859	.1708	.0367	.0452	.0367
-.30	-.30	.30	.30	-.2800 .3200	-.3600 .3200	-.2800 .3200	.2035	.3290	.2035	.0499	.0577	.0499
-.35	-.35	.35	.35	-.3600 .3600	-.4000 .3600	-.3600 .3600	.2446	.3696	.2446	.0662	.0630	.0663
-.40	-.40	.40	.40	-.4000 .4000	-.4400 .4000	-.4000 .4000	.2696	.4139	.2096	.0901	.0758	.0901
-.45	-.45	.45	.45	-.4400 .4400	-.4800 .4800	-.4400 .4400	.3107	.4527	.3107	.1069	.0828	.1069
-.50	-.50	.50	.50	-.4800 .5200	-.5200 .5200	-.4800 .5200	.3440	.4976	.3440	.1364	.0890	.1364
-.75	-.75	.75	.75	-.7600 .7600	-.7600 .7600	-.7600 .7600	.5217	.7121	.5218	.2951	.1322	.2951

TABLE IV (Continued)

$S = \begin{bmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 1500.0000 \end{bmatrix}$, $T = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.0000 \end{bmatrix}$. $n_1 = n_2 = 3$

				Modes			Mean			Variance		
u_1	u_2	v_1	v_2	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Nor.)	App. (Ber.)
-1.0	-1.0	1.0	1.0	-.9999 1.0000	-.9999 1.0000	-.9999 1.0000	.6981	.9318	.7011	.5148	.1938	.5107
-1.25	-1.25	1.25	1.25	-1.2399 1.2399	-1.2799 1.2399	-1.2399 1.2399	.8651	1.1526	.8565	.8120	.2840	.8283

TABLE V

MODES, MEAN AND VARIANCE OF THE EXACT, APPROXIMATE (NORMAL) AND APPROXIMATE (BERNSTEIN) MARGINAL POSTERIOR DISTRIBUTION OF θ_2

$$S = T = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.0000 \end{bmatrix}, \quad n_1 = n_2 = 3$$

u_1	u_2	v_1	v_2	Modes: Exact	Modes: App. (Normal)	Modes: App. (Berns.)	Mean: Exact	Mean: App. (Normal)	Mean: App. (Berns.)	Variance: Exact	Variance: App. (Nor.)	Variance: App. (Ber.)
.0	.0	.0	.0	.0000	.0000	.0000	.0000	.0000	.0000	.0012	.0012	.0012
-.05	-.05	.05	.05	.0000	-.0800 .0800	.0000	.0000	.0000	.0000	.0044	.0073	.0044
-.1	-.1	.1	.1	-.0800 .0800	-.1600 .1600	-.0800 .0800	.0000	.0000	.0000	.0129	.0343	.0129
-.15	-.15	.15	.15	-.1600 .1600	-.2000 .2000	-.1600 .1600	.0000	.0000	.0000	.0255	.0690	.0255
-.2	-.2	.2	.2	-.2000 .2000	-.2399 .2399	-.2000 .2000	.0000	.0000	.0000	.0431	.1073	.0432
-.25	-.25	.25	.25	-.2399 .2399	-.2800 .2800	-.2399 .2399	.0000	.0000	.0000	.0660	.1479	.0661
-.3	-.3	.3	.3	-.2800 .2800	-.3200 .3200	-.2800 .2800	.0000	.0000	.0000	.0929	.1918	.0929
-.35	-.35	.35	.35	-.3600 .3600	-.4000 .4000	-.3600 .3600	.0000	.0000	.0000	.1261	.2354	.1261
-.4	-.4	.4	.4	-.4000 .4000	-.4400 .4400	-.4000 .4000	.0000	.0000	.0000	.1634	.2835	.1634
-.45	-.45	.45	.45	-.4400 .4400	-.4800 .4800	-.4400 .4400	.0000	.0000	.0000	.2053	.3320	.2054
-.5	-.5	.5	.5	-.4800 .4800	-.5200 .5200	-.4800 .4800	.0000	-.0001	.0000	.2539	.3834	.2540
-.75	-.75	.75	.75	-.7600 .7600	-.7600 .7600	-.7600 .7600	.0000	-.0004	.0000	.5663	.6938	.5663

TABLE V (Continued)

$S = T = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 500.0000 \end{bmatrix}$, $n_1 = n_2 = 3$											
				Modes			Mean			Variance	
u_1	u_2	v_1	v_2	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Normal)	App. (Berns.)	Exact	App. (Nor.)	App. (Ber.)
-1.0	-1.0	1.0	1.0	-.9999 1.0000	-.9999 1.0000	-.9999 1.0000	.0000	-.0007	.0000	1.0029	1.1138	1.0029
-1.25	-1.25	1.25	1.25	-1.2399 1.2399	-1.2799 1.2799	-1.2399 1.2399	.0000	-.0011	.0000	1.5637	1.6594	1.5640

TABLE VI

ADAPTIVE ESTIMATION OF LINEAR DYNAMIC SYSTEM

k	Y(2×1)		θ(k)(2×1)		m(k)(2×1)	
1	9.1904	8.3826	.5528	2.3827	1.3359	1.5842
2	15.8108	12.5371	2.1137	1.9474	2.4765	1.9208
3	13.9354	10.9933	2.8114	.6664	2.2235	1.7833
4	-8.9547	-8.6536	-1.2862	-1.2015	-.3826	-.2517
5	-2.1338	-2.6767	-.2337	-1.0013	-.3079	-.4820
6	.7983	-.5539	.2608	-.9347	.0292	-.3257
7	-16.4680	-15.2119	-3.3120	-1.8764	-1.9323	-1.7768
8	-12.8372	-10.4074	-2.0927	-1.2436	-1.9984	-1.6240
9	-2.6355	-1.2617	-.3335	-.2536	-.9214	-.6967
10	-23.8043	-23.2740	-4.6590	-3.9810	-2.6536	-2.8531
11	-.5123	-1.1460	-1.4735	.5755	-1.1091	-1.3906
12	-13.9216	-12.4296	-2.9443	-1.4696	-1.8208	-1.6687
13	-9.1700	-8.6425	-1.2594	-2.9697	-1.5369	-1.5991
14	-8.9245	-6.9377	-1.7036	-1.1055	-1.5826	-1.2302
15	1.3559	2.0875	-.1908	.3053	-.6039	-.4265
16	6.0636	4.0691	2.4328	.0096	.4000	-.0830
17	16.7220	17.6292	3.6291	2.1593	1.2849	1.8355
18	-14.6151	-13.4243	-1.4901	-3.6807	-.5013	-.3787
19	-10.7431	-9.2827	-1.6751	-1.3145	-1.1855	-.9551
20	-22.2969	-21.4714	-2.2303	-4.7058	-2.2927	-2.5014
21	-5.7949	-4.1827	-.4491	-.8034	-1.7375	-1.4061
22	-12.2283	-9.6707	-1.4381	-.8028	-1.9015	-1.3767
23	-13.8386	-13.6016	-3.4687	-.7416	-1.8072	-2.0717
24	-46.4360	-41.4877	-7.4726	-6.0131	-4.6229	-4.0666
25	-14.1466	-13.1173	-2.0633	-1.8412	-3.3804	-3.4502
26	-23.0247	-22.7963	-2.4840	-4.7020	-3.2498	-3.7753
27	-25.1401	-22.3110	-4.0806	-3.5812	-3.8182	-3.4927
28	-11.2475	-9.5568	-1.0990	-1.6422	-3.1463	-2.8848
29	-9.4424	-9.5759	-.6657	-1.9627	-2.3466	-2.6830
30	-31.4292	-28.6890	-5.2937	-4.4210	-3.3200	-3.1634
	$P_v = \begin{bmatrix} .8148 & -.7407 \\ -.7407 & 1.0370 \end{bmatrix}$		$P_w = \begin{bmatrix} .8108 & -.7568 \\ -.7568 & .9730 \end{bmatrix}$		$\theta(0) = (3.5 \quad 5)$	

TABLE VII

ADAPTIVE ESTIMATION OF LINEAR DYNAMIC SYSTEM

k	Y(2×1)		θ(k)(2×1)		m(k)(2×1)	
1	-6.2470	-5.3009	-2.0187	.2715	-1.7886	.0073
2	-2.8369	-2.8967	-.3569	-.5617	0.1711	-1.0364
3	-8.7679	-8.8775	-.1034	-2.8284	-.4053	-2.7329
4	-27.4728	-25.3970	-4.2550	-4.1827	-3.8115	-3.5974
5	-26.5015	-24.7432	-3.6363	-4.6128	-3.4781	-4.2120
6	-25.9566	024.4178	-3.3096	-4.8249	-3.1248	-4.3479
7	-35.6669	-32.5466	-6.3196	-4.5256	-5.0648	-3.8705
8	-27.1346	-24.6571	-4.6620	-3.5834	-5.0297	-4.0316
9	-21.2804	-19.4803	-3.3242	-3.2064	-4.5411	-4.1322
10	-45.9876	-42.4126	-7.4801	-6.6584	-4.8146	-4.7252
11	-16.2322	-14.5011	-3.7249	-1.1018	-5.1964	-3.9506
12	-33.1120	-30.1215	-5.9570	-4.0975	-5.2469	-4.1224
13	-34.2356	-32.0541	-4.2883	-6.4610	-4.1151	-5.3017
14	-29.0301	-26.4786	-4.8191	-4.0589	-5.1326	-4.3284
15	-22.3201	-20.3607	-3.7379	-3.0931	-4.9735	-4.3302
16	-23.8466	-22.8484	-2.2449	-5.2986	-3.4472	-5.5370
17	-16.3748	-15.2448	-1.6584	-3.4925	-4.1791	-4.4819
18	-52.7424	-49.2581	-6.8210	-9.6436	-3.7749	-5.1845
19	-34.3711	-31.5627	-5.4881	-5.0424	-4.7626	-4.3596
20	-45.3125	-42.5980	-5.5082	-8.6732	-3.6413	-5.4908
21	-20.7000	-19.0655	-2.9756	-3.4056	-4.5341	-4.4563
22	-26.2591	-23.9003	-4.3861	-3.5757	-4.9362	-4.1329
23	-28.8056	-25.9668	-6.0284	-2.5754	-5.3793	-3.8961
24	-54.1995	-49.5016	-9.1555	-7.3568	-5.5199	-4.2653
25	-18.8243	-17.3265	-3.0298	-2.7223	-4.8775	-4.7454
26	-34.7836	-32.8837	-4.0510	-6.8655	-3.7928	-5.6437
27	-31.7982	-29.1128	-5.1820	-4.5522	-5.0457	-4.4837
28	-19.4884	-18.0410	-2.6774	-3.3414	-4.6268	-4.7481
29	-22.9223	-21.6092	-2.7689	-4.3950	-4.0422	-5.0999
30	-41.2032	-37.7675	-6.8401	-5.7589	-4.9974	-4.3651
	$P_v = \begin{bmatrix} 211.2676 & -28.1690 \\ -28.1690 & 70.4225 \end{bmatrix}$		$P_w = \begin{bmatrix} .60 & -.40 \\ -.40 & .60 \end{bmatrix}$		$\theta(0) = (1,1)$	

TABLE VIII

ADAPTIVE ESTIMATION OF LINEAR DYNAMIC SYSTEM

k	Y(2×1)		θ(k)(2×1)		m(k)(2×1)	
1	5.9989	4.8692	.8768	.9436	.9921	.9920
2	8.9545	5.7053	1.0139	1.0086	.9956	.9950
3	8.2783	5.8912	1.0157	.9276	.9969	.9964
4	4.9645	4.0146	.8870	.8819	.9952	.9951
5	7.9533	6.7002	.9979	.9660	.9960	.9959
6	8.9607	7.0257	1.0178	.9695	.9971	.9968
7	6.6476	5.7266	.8915	.9433	.9970	.9969
8	4.8836	6.0885	.9890	1.0240	.9964	.9968
9	6.1650	7.1815	1.0459	1.0531	.9965	.9969
10	9.8147	7.3130	.8684	.8866	.9977	.9973
11	9.7303	7.6041	1.0402	1.1293	.9984	.9981
12	7.0266	6.6348	.9418	.9990	.9984	.9984
13	10.2231	9.8594	1.0146	.9443	.9992	.9993
14	6.8682	7.8868	.9948	1.0194	.9993	.9997
15	7.8576	8.2482	1.0400	1.0640	.9997	1.0000
16	4.5608	2.8225	1.0981	1.0045	.9996	.9992
17	4.8155	7.6715	1.1151	1.0627	.9989	.9997
18	7.0563	7.3036	.8977	.7979	.9994	.9996
19	5.3973	5.6742	.9583	.9701	.9993	.9994
20	5.5787	4.6925	.9607	.8540	.9992	.9991
21	5.2094	6.3802	1.0641	1.0544	.9989	.9993
22	2.1724	3.7670	1.0036	1.0305	.9984	.9988
23	6.9102	4.5081	.9376	1.0470	.9988	.9983
24	3.3553	4.3913	.8141	.8607	.9981	.9984
25	5.7313	5.0971	1.0743	1.0896	.9982	.9981
26	6.2108	4.3487	1.0292	.9388	.9982	.9979
27	6.3166	6.8564	.9791	.9985	.9980	.9982
28	4.7026	5.4953	1.0833	1.0673	.9978	.9981
29	5.7181	4.4848	1.0763	1.0282	.9980	.9978
30	6.3066	6.0010	.9052	.9344	.9978	.9979
	$P_v = \begin{bmatrix} .60 & -.40 \\ -.40 & .60 \end{bmatrix}$		$P_w = \begin{bmatrix} 555.555 & -444.44\overline{4} \\ -444.444 & 555.555 \end{bmatrix}$		θ(0) = (1,1)	

TABLE IX

ADAPTIVE ESTIMATION OF LINEAR DYNAMIC SYSTEM

k	Y(2×1)		θ(k)(2×1)		m(k)(2×1)	
1	3.1057	2.6788	.3250	.6395	.5678	.4154
2	5.1415	3.6361	.6110	.6044	.6130	.5712
3	3.6664	2.4684	.7040	.0774	.5249	.4352
4	-1.7813	-1.8493	-.1206	-.3222	.1275	-.0042
5	.5801	.1424	.0769	-.1568	.0502	.0167
6	1.2581	.4621	.1531	-.1743	.0668	.0251
7	-2.4216	-2.3854	-.5317	-.3032	-.1663	-.2489
8	-1.7118	-.8293	-.2511	-.0027	-.1873	-.1232
9	.7992	1.3637	.1146	.2419	.0180	.1250
10	-2.2126	-2.9784	-.6314	-.6868	-.1888	-.3523
11	3.2949	2.5109	-.0369	.6910	.0722	.1571
12	.1289	.1998	-.1947	.1479	.0721	.0633
13	1.4361	1.1188	.1563	-.3435	.1313	.1378
14	1.4094	1.8926	.1098	.2870	.2148	.2898
15	4.3125	4.3355	.4743	.7133	.4768	.5855
16	4.1485	3.3469	1.0694	.5179	.5510	.5200
17	5.8083	6.6651	1.1692	.9513	.7613	.9470
18	.0082	.0023	.2972	-.5975	.5563	.4417
19	1.4247	1.6349	.2735	.3004	.4217	.4181
20	-1.3779	-1.6064	.2081	-.6445	.1912	.0765
21	2.4161	2.9318	.4895	.5140	.2578	.3631
22	.5829	1.5318	.3348	.5423	.2979	.3750
23	2.4311	1.6745	.0024	.7263	.3147	.2585
24	-5.5265	-4.4704	-.7236	-.6519	-.0393	-.1151
25	2.2390	2.1322	.3393	.5667	.0680	.1291
26	-.2402	-.8592	.2686	-.3587	.0124	-.0742
27	-.5787	-.1209	-.1346	-.0182	-.0240	.0135
28	1.8773	2.3168	.4615	.4692	.1314	.2338
29	2.6839	2.2062	.6117	.3700	.2431	.2437
30	-1.3096	-1.1485	-.2640	-.1844	.0983	.0466
	$P_v = \begin{bmatrix} 2.8985 & -2.6087 \\ -2.6087 & 4.3478 \end{bmatrix}$		$P_w = \begin{bmatrix} 16.6666 & -10.0 \\ -10.0 & 10.0 \end{bmatrix}$		$\theta(0) = (1,1)$	

TABLE X

ADAPTIVE ESTIMATION OF LINEAR DYNAMIC SYSTEM

k	Y(2×1)		θ(k)(2×1)		m(k)(2×1)	
1	26.862	24.804	4.0365	4.2424	4.1480	4.1226
2	27.249	25.010	4.1730	4.1576	4.1554	4.0614
3	26.234	24.058	4.1438	3.8637	4.1313	4.0193
4	25.129	23.154	3.8853	3.8623	4.0818	4.0481
5	26.261	24.187	4.0698	3.9681	4.0736	4.0309
6	26.306	24.186	4.0901	3.9373	4.0680	3.9911
7	25.531	23.582	3.8469	4.0056	4.0265	4.0411
8	26.344	24.428	4.0200	4.1309	4.0109	4.1223
9	27.019	25.031	4.1519	4.1780	4.0484	4.1361
10	25.602	23.559	3.8699	3.9203	4.1066	4.0378
11	28.005	25.847	4.1461	4.4331	4.0730	4.0697
12	26.418	24.439	3.9735	4.1528	4.0618	4.1072
13	26.337	24.302	4.1125	3.8892	4.0861	4.0785
14	26.743	24.796	4.0745	4.1529	4.0626	4.1563
15	27.478	25.445	4.1796	4.2586	4.0946	4.1651
16	26.907	24.679	4.3000	4.0072	4.1601	4.0172
17	27.235	25.283	4.2890	4.1289	4.0600	4.1957
18	24.823	22.893	3.9518	3.6274	4.1327	4.1095
19	26.367	24.384	4.0498	4.0851	4.1105	4.1607
20	25.393	23.334	4.0686	3.7259	4.1621	4.0368
21	27.176	25.164	4.2119	4.1849	4.0778	4.1804
22	26.829	25.869	4.1368	4.2230	4.0995	4.2392
23	27.485	25.341	4.0637	4.4089	4.1776	4.1388
24	25.214	23.372	3.8387	3.9759	4.1308	4.2596
25	28.192	26.019	4.3269	4.3804	4.1966	4.1900
26	26.732	24.528	4.2471	3.9606	4.2368	4.0319
27	26.835	24.844	4.1016	4.1623	4.1156	4.2042
28	27.696	25.613	4.3075	4.2616	4.1472	4.2070
29	27.505	25.305	4.3128	4.1641	4.2005	4.0901
30	26.188	24.206	3.9849	4.0724	4.1378	4.1732
	$P_v = \begin{bmatrix} 169.4915 & -135.5932 \\ -135.5932 & 175.1412 \end{bmatrix}$		$P_w = \begin{bmatrix} 90.9091 & -36.3636 \\ -36.3636 & 54.5455 \end{bmatrix}$		$\theta(0) = (3.5, 5)$	

APPENDIX B

FIGURES

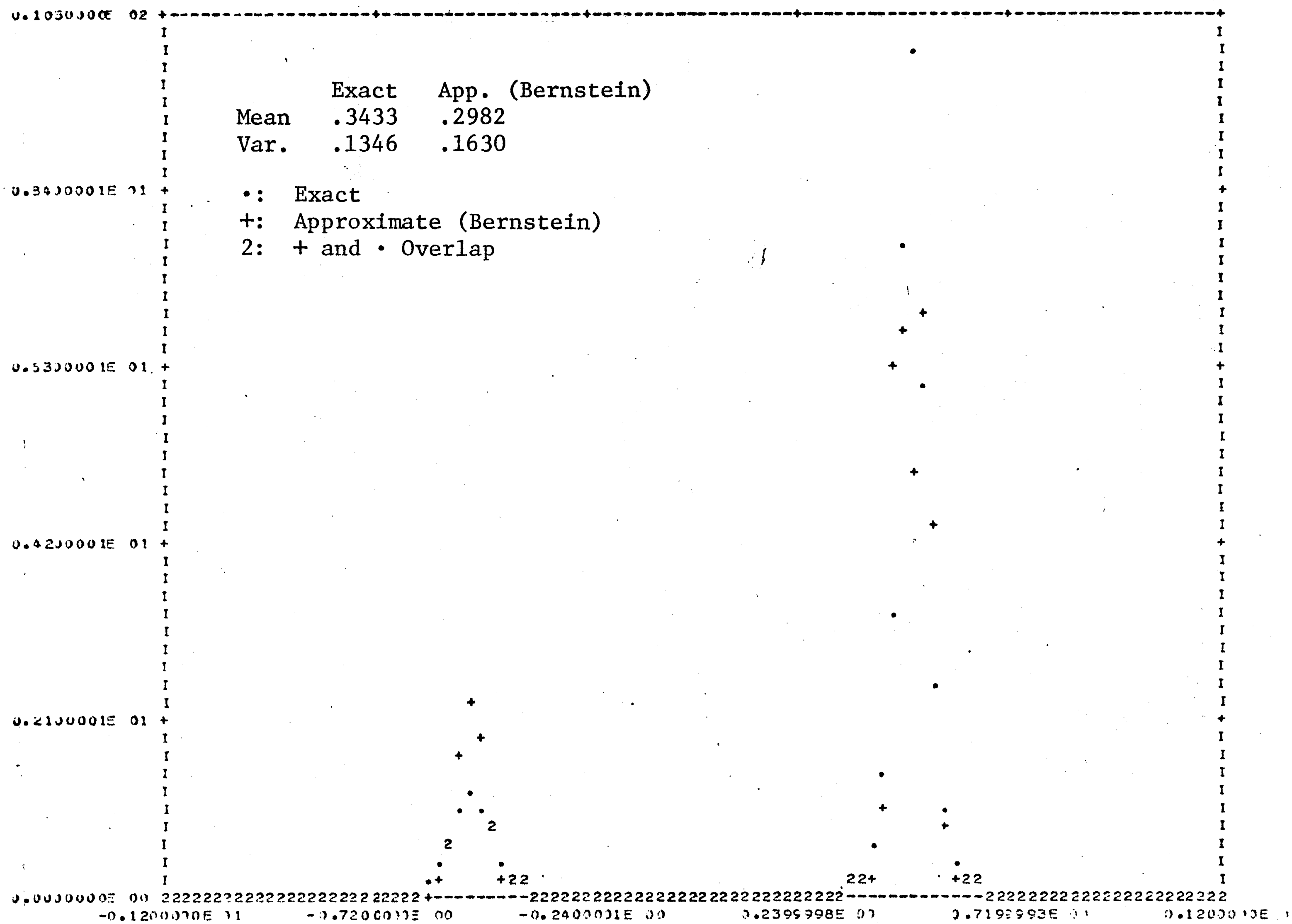

Figure 1. Comparison of the Exact and Approximate (Bernstein) Marginal Posterior Density of θ_2 When M = 30

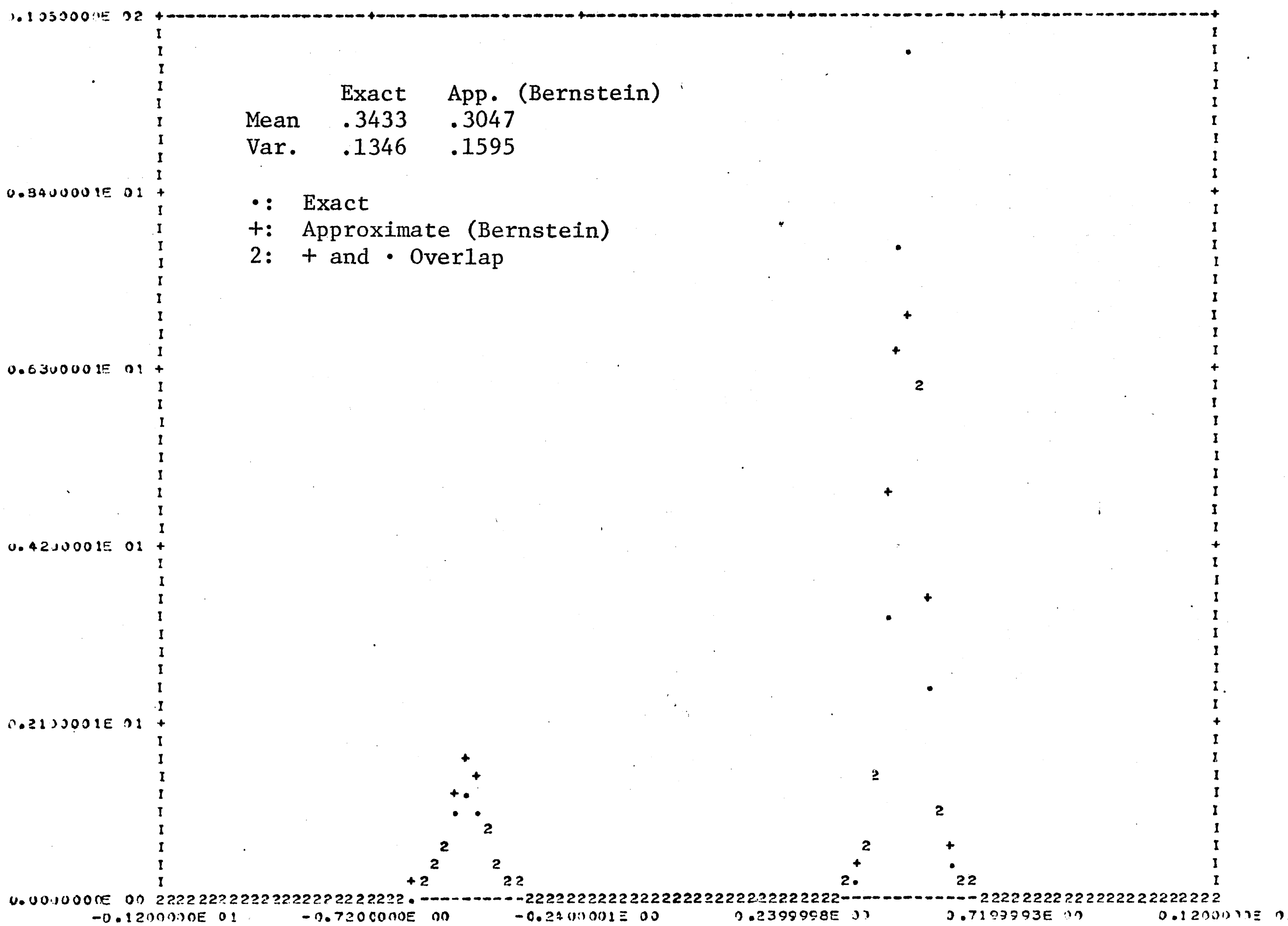

Figure 2. Comparison of the Exact and Approximate (Bernstein) Marginal Posterior Density of θ_2 When M = 60

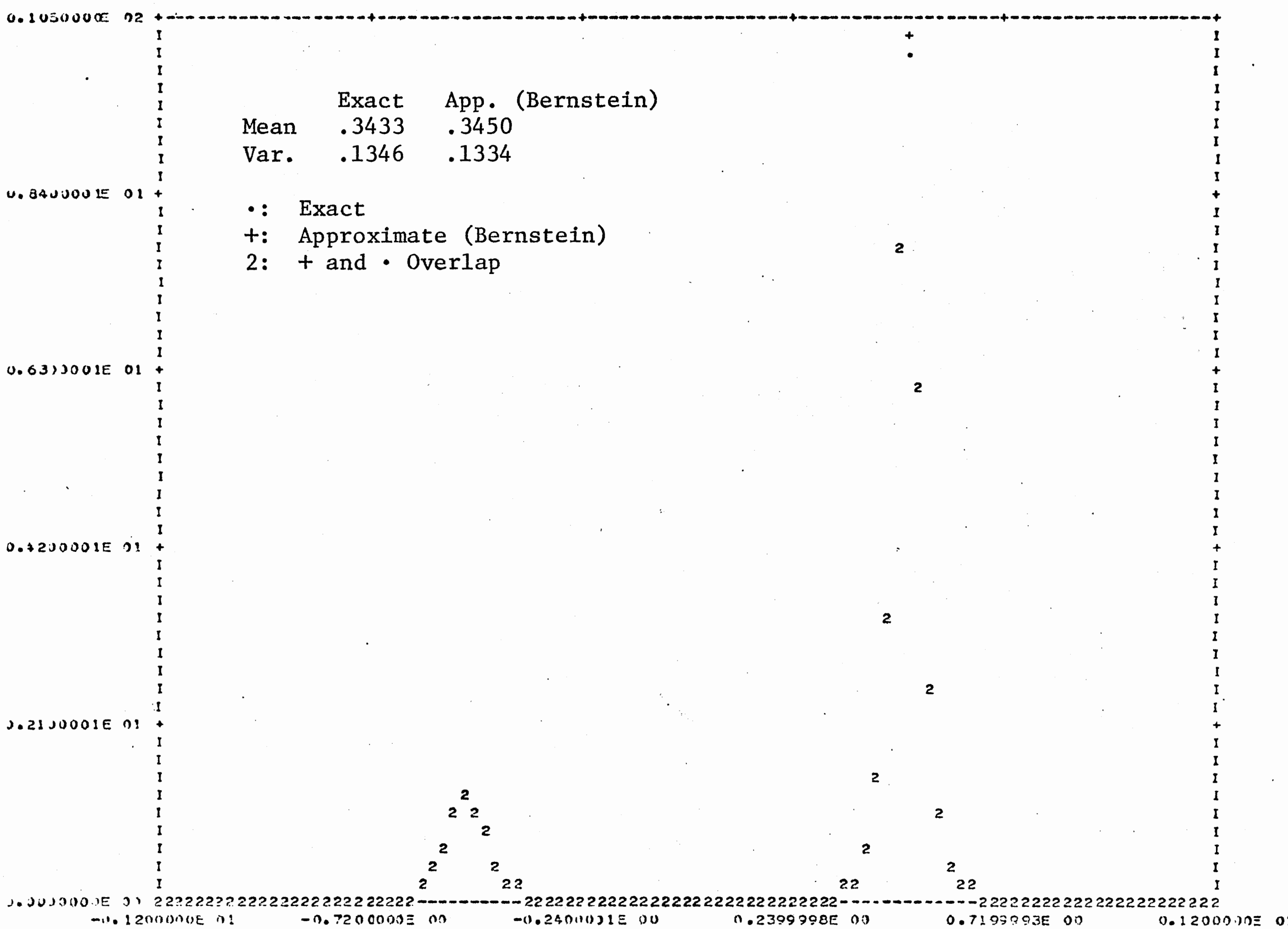

Figure 3. Comparison of the Exact and Approximate (Bernstein) Marginal Posterior Density of θ_2 When M = 120

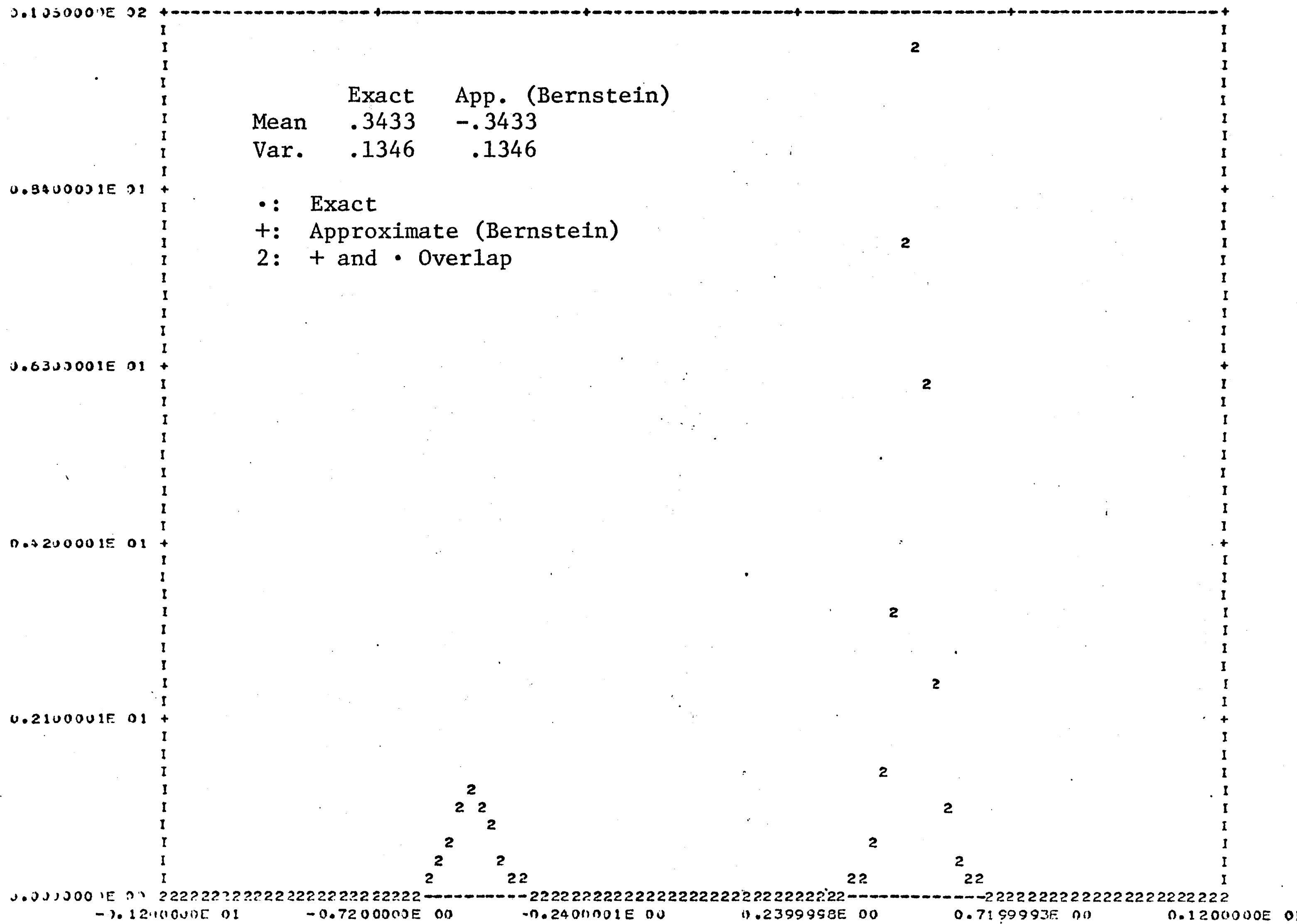

Figure 4. Comparison of the Exact and Approximate (Bernstein) Marginal Posterior Density of θ_2 When M = 240

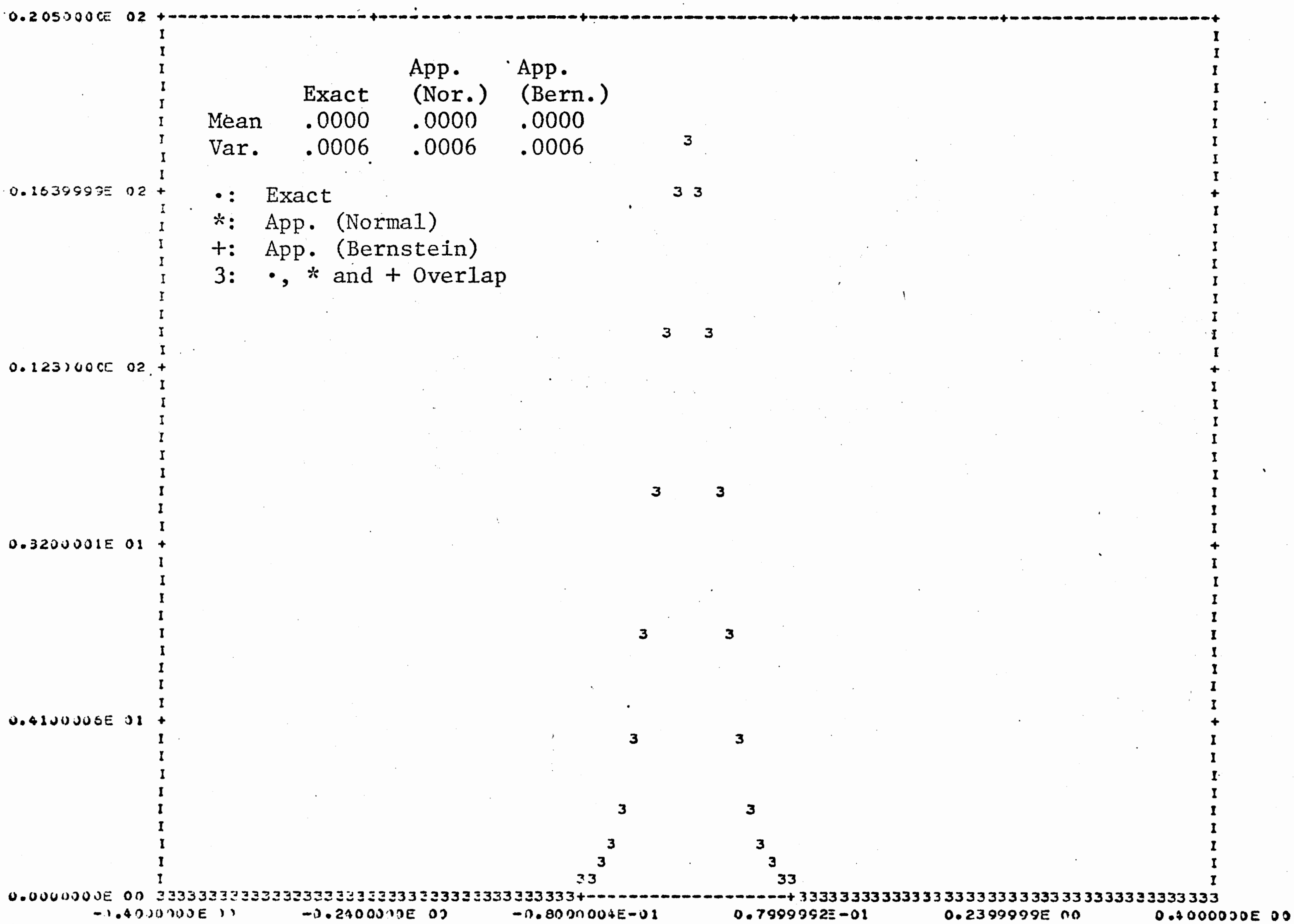

Figure 5. Comparison of the Exact, Approximate (Normal) and Approximate (Bernstein) Marginal Posterior Density of θ_2 When $u_1 = u_2 = 0.0$, $v_1 = v_2 = 0.0$

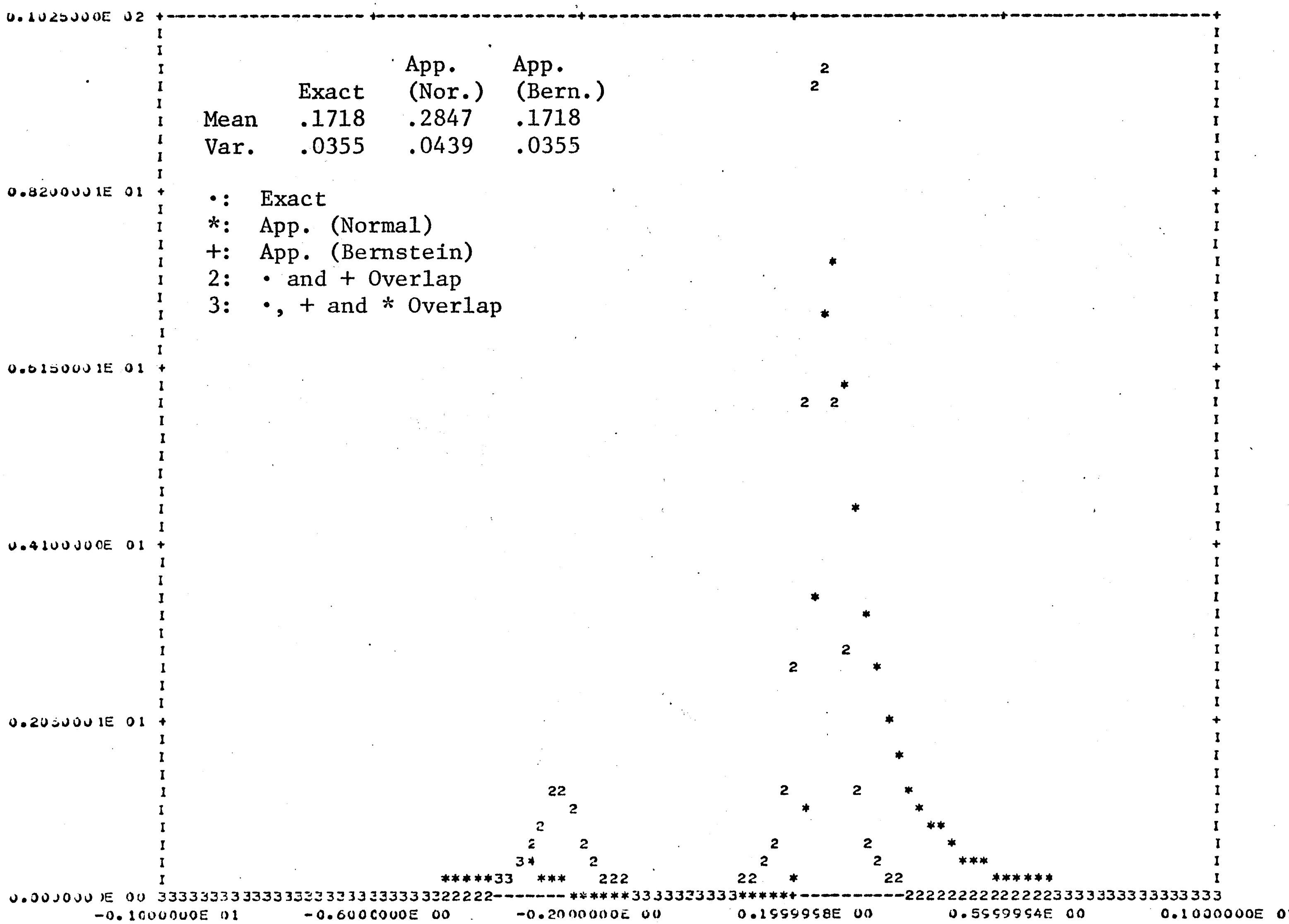

Figure 6. Comparison of the Exact, Approximate (Normal) and Approximate (Bernstein) Marginal Posterior Density of θ_2 When $u_1 = u_2 = -.25$, $v_1 = v_2 = .25$

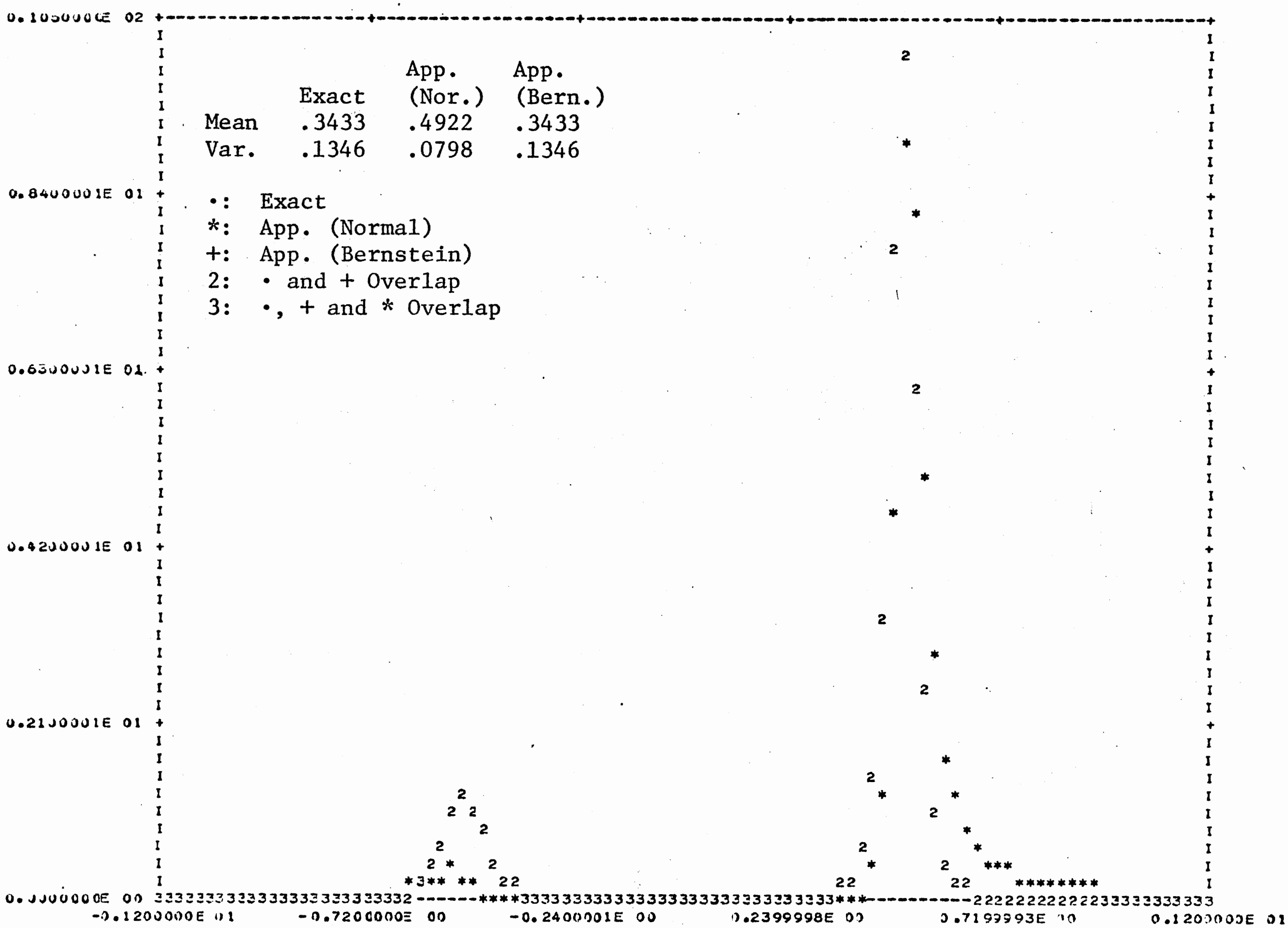

Figure 7. Comparison of the Exact, Approximate (Normal) and Approximate (Bernstein) Marginal Posterior Density of θ_2 When $u_1 = u_2 = -.5$, $v_1 = v_2 = .5$

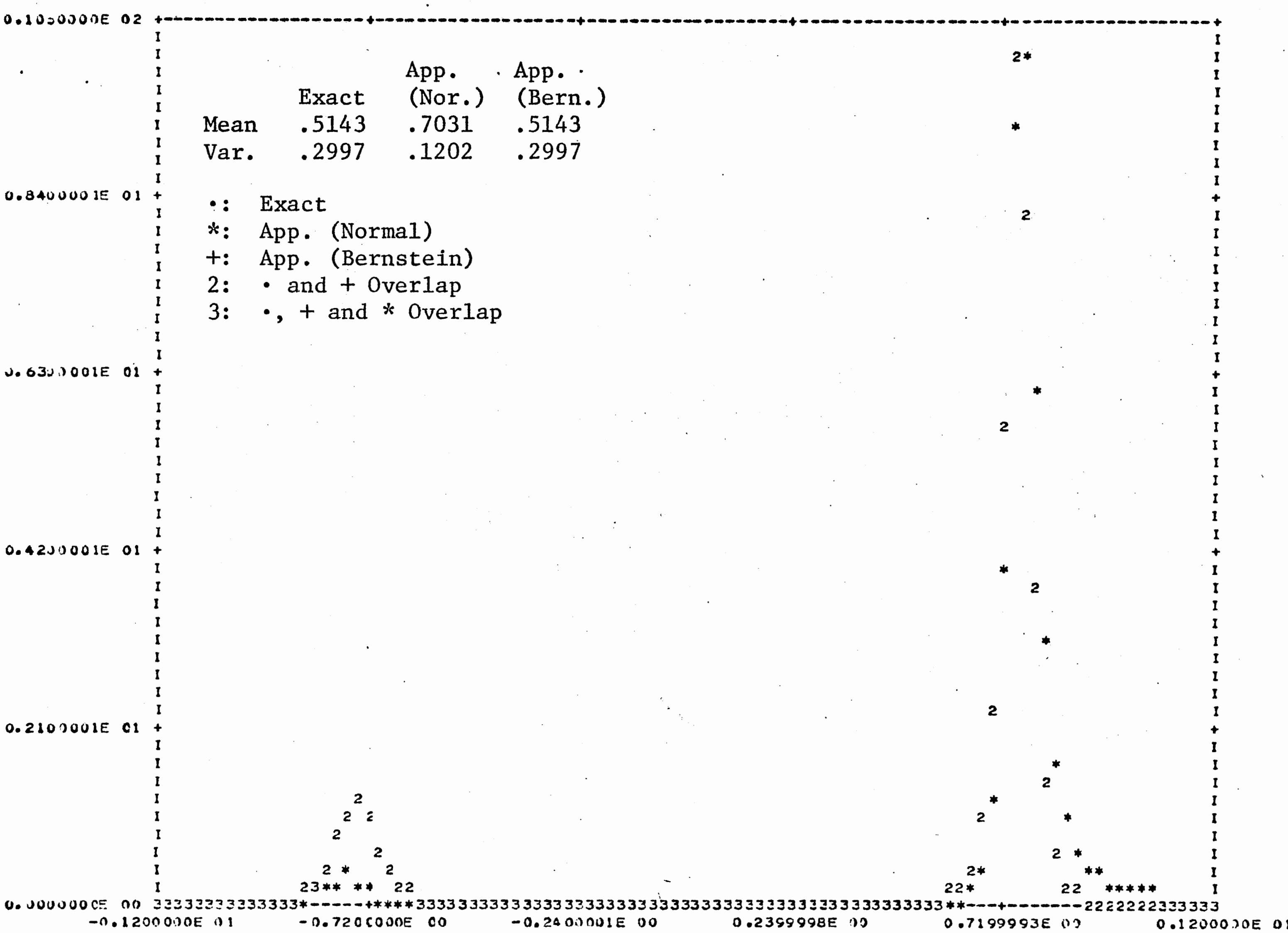

Figure 8. Comparison of the Exact, Approximate (Normal) and Approximate (Bernstein) Marginal Posterior Density of θ_2 When $u_1 = u_2 = -.75$, $v_1 = v_2 = .75$

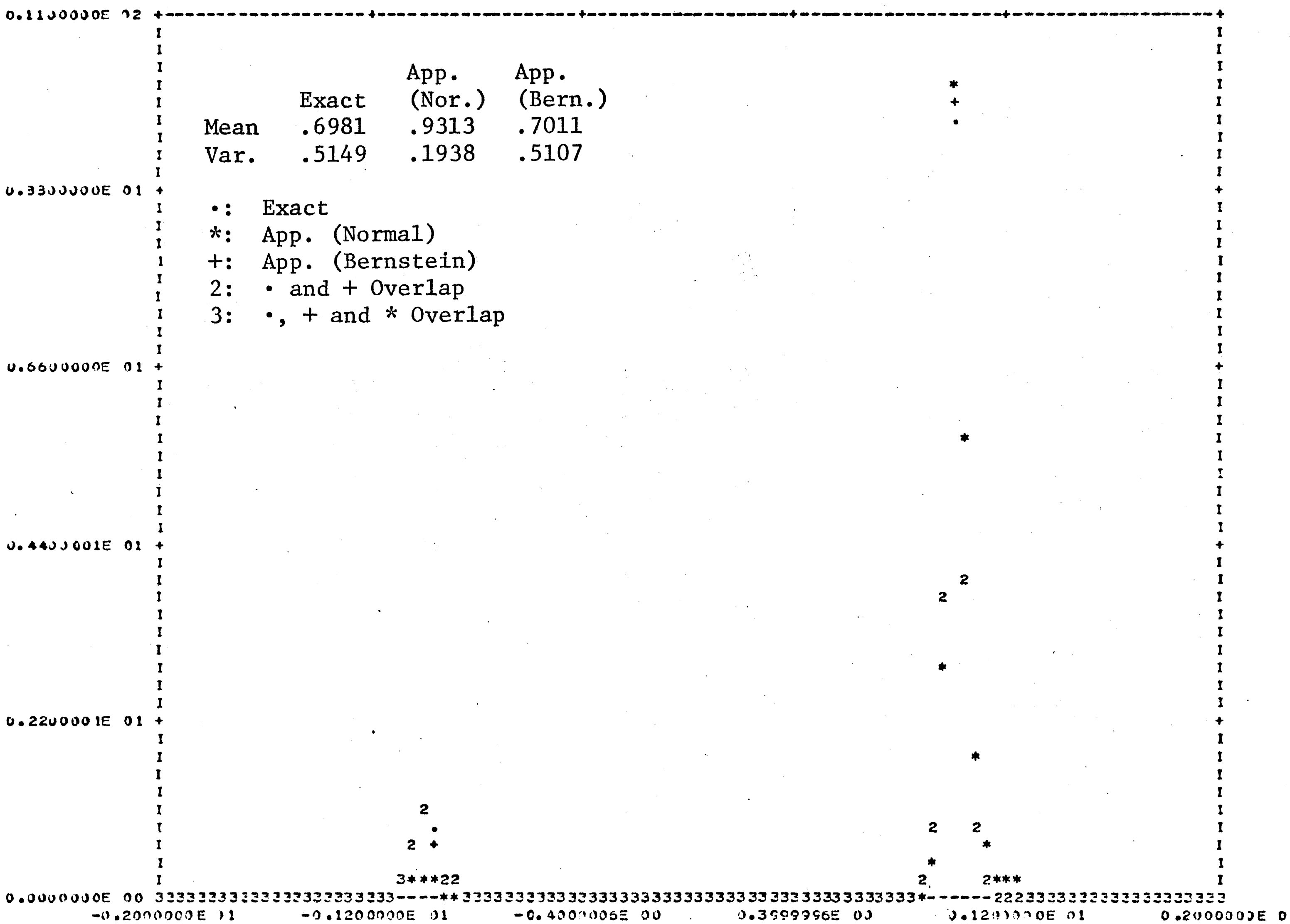

Figure 9. Comparison of the Exact, Approximate (Normal) and Approximate (Bernstein) Marginal Posterior Density of θ_2 When $u_1 = u_2 = -1.0$, $v_1 = v_2 = 1.0$

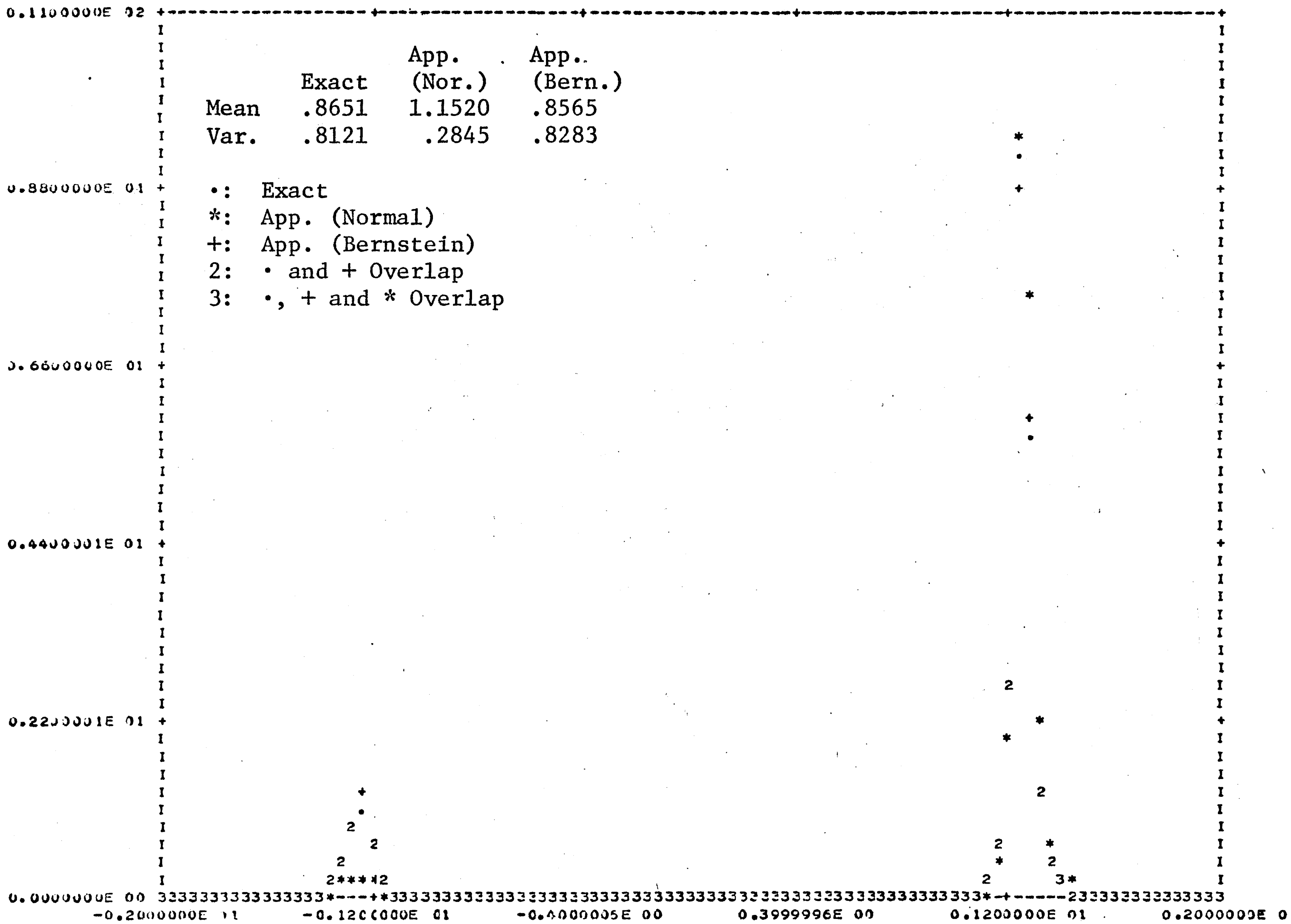

Figure 10. Comparison of the Exact, Approximate (Normal) and Approximate (Bernstein) Marginal Posterior Density of θ_2 When $u_1 = u_2 = -1.25$, $v_1 = v_2 = 1.25$

VITA

Mat Yusoff Abdullah

Candidate for the Degree of

Doctor of Philosophy

Thesis: BAYESIAN INFERENCES WITH THE POLY-t DISTRIBUTION

Major Field: Statistics

Biographical:

Personal Data: Born in K. Bharu, Kelantan, Malaysia, December 23, 1946.

Education: Received Bachelor of Science degree in Mathematics from the University of Malaya, Malaysia, in 1971; Master of Science degree in Experimental Statistics from the University of the Philippines at Los Baños in 1974; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in July, 1982.

Professional Experience: Lecturer at the Kolej Pertanian Malaysia, 1971-1972; Lecturer at the Universiti Pertanian Malaysia, 1974-1978 and held position as the Acting Head of the Department of Mathematics at the Universiti Pertanian Malaysia, 1976-1978.

Professional Organization: American Statistical Association.