1990

IMBEDDED INTEGRATION RULES AND THEIR APPLICATIONS IN BAYESIAN ANALYSIS

DELLAPORTAS, PETROS

IMBEDDED INTEGRATION RULES AND THEIR APPLICATIONS IN BAYESIAN ANALYSIS

PETROS DELLAPORTAS

A thesis submitted in partial fulfilment of the requirements of the Council for National Academic Awards for the degree of Doctor of Philosophy

July 1990

Sponsoring Establishment: Polytechnic South West
Department of Mathematics and Statistics

Collaborating Establishment: University of Sheffield

Στους Γονείς μου

Imbedded Integration Rules and their Applications in Bayesian analysis

by

Petros Dellaportas

This thesis deals with the development and application of numerical integration techniques for use in Bayesian Statistics. In particular, it describes how imbedded sequences of positive interpolatory integration rules (PIIR's) obtained from Gauss-Hermite product rules can extend the applicability and efficiency of currently available numerical methods.

The numerical strategy suggested by Naylor and Smith (1982) is reviewed, criticised and applied to some examples with real and artificial data. The performance of this strategy is assessed from the viewpoint of 3 criteria: reliability, efficiency and accuracy.

The imbedded sequences of PIIR's are introduced as an alternative and an extension to the above strategy for two major reasons. Firstly, they provide a rich class of spatially ditributed rules which are particularly useful in high dimensions. Secondly, they provide a way of producing more efficient integration strategies by enabling approximations to be updated sequentially through the addition of new nodes at each step rather than through changing to a completely new set of nodes.

Finally, the improvement in the reliability and efficiency achieved by the adaption of an integration strategy based on PIIR's is demonstrated with various illustrative examples. Moreover, it is directly compared with the Gibbs sampling approach introduced recently by Gelfand and Smith (1988).

Acknowledgements

I am indebted to my supervisor, Dr David Wright, for his guidance, encouragement and friendship throughout my work on this thesis.

Thanks are also due to Dave Stephens for a careful proof-reading of a final draft of the thesis, and to Polytechnic South West (formerly Plymouth Polytechnic) which sponsored my post as a Research Assistant during the period of my study.

Last but not least, I would like to thank Effie Angelonidi for her invaluable encouragment and support.

CONTENTS

Chapter 1: Numerical methods in Bayesian Statistics

1.1. The role of Numerical integration in Bayesian Statistics

1.1.1 Introduction

A major impediment to many practical applications of Bayesian methods is the difficulty in evaluating the various integrals required. This has led to much research into numerical methods of integration for Bayesian analysis on which this thesis is focussed. This chapter is concerned with the description of the Bayesian paradigm, the technical difficulties encountered in its practical application, the need for its numerical implementation, and the numerical methods currently in use. The remainder of this section contains a brief summary of the Bayesian paradigm highlighting the need for its numerical implementation. The intention is to fix notation and introduce definitions to be used in subsequent sections rather than to give detailed account or mathematical rigour. For more details, some standard textbooks are given as references throughout this section. In section 1.2 we give a brief account of the extensively used quadrature strategy introduced by Naylor and Smith (1982). See for example Naylor and Smith (1988a,b). In section 1.3 we present other numerical integration methods currently in use. In particular spherical rules and Monte Carlo methods are outlined and their practical difficulties are emphasised. Finally, we end this chapter with a brief description of the imbedded sequences of integration rules and a discussion of the motivation for the researh work described in the rest of the thesis.

1.1.2 The Bayesian paradigm

Following Naylor and Smith (1982), we consider parametric statistical models; given sample data $\underline{x}$ generated from the model we focus on the parameter vector $\underline{\theta}' = (\theta_1, \ldots, \theta_k)$ assuming that the model gives rise to a well defined likelihood function $\ell(\underline{x}; \underline{\theta})$ through which we receive the information about $\underline{\theta}$ from the data $\underline{x}$. Choice of suitable parametric models, of likelihood construction and interpretation are issues of primary importance which have been discussed widely in the statistical literature. See for example Cox and Hinkley (1974), Barnett (1982). Adopting a Bayesian approach we place probabilities or plausibilities on various $\underline{\theta}$ in the form of a known prior density function $p(\underline{\theta})$. The construction and interpretation of this prior has been the subject of much debate in the literature. See for example Barnett (1982), Cox and Hinkley (1974), Box and Tiao (1973), Berger (1987).

Under the above assumptions Bayes' theorem can be applied to update the prior density function via information of the data to obtain the joint posterior density

$$p(\underline{\theta}/\underline{x}) = \frac{\ell(\underline{x}; \underline{\theta})\, p(\underline{\theta})}{\int_{E_k} \ell(\underline{x}; \underline{\theta})\, p(\underline{\theta})\, d\underline{\theta}} \qquad (1.1)$$

where E_k is the k-dimensional Euclidean space.

The posterior density as expressed in (1.1) is regarded as a complete description of what is known about $\underline{\theta}$ from the prior information and the data. It therefore provides the basis for inferences about $\underline{\theta}$.

If we are interested in a subset of $\underset{\sim}{\theta}$,

$$\underset{\sim}{\theta}_I = (\theta_1, \theta_2, \ldots, \theta_r) \tag{1.2}$$

where $I = (1,2,\ldots,r) \subset (1,2,\ldots,k)$,

then we can obtain the marginal posterior density of $\underset{\sim}{\theta}_I$ simply by integrating over $\underset{\sim}{\theta}_{\bar{I}}$, the complement of $\underset{\sim}{\theta}_I$ with respect to $\underset{\sim}{\theta}$. Thus, we have

$$p(\underset{\sim}{\theta}_I/\underset{\sim}{x}) = \int_{E_{k-r}} p(\underset{\sim}{\theta}/\underset{\sim}{x}) d\underset{\sim}{\theta}_{\bar{I}} . \tag{1.3}$$

Posterior expectations, such as posterior means variances or covariances as well as predictive densities, can then be derived. Many of the integrals required in Bayesian analysis can be written, perhaps after an initial parameter transformation, in the form

$$S_I(q(\underset{\sim}{\theta})) = \int_{E_{k-r}} q(\underset{\sim}{\theta}) \ell(\underset{\sim}{x};\underset{\sim}{\theta}) . p(\underset{\sim}{\theta}) d\underset{\sim}{\theta}_{\bar{I}} \tag{1.4}$$

with appropriate choice of I and $q(\underset{\sim}{\theta})$ where S_I is written as S if we have r=k in (1.2).

For example, S(1) provides the normalising constant of the joint posterior density; $S(\theta)/S(1)$ provides the posterior mean μ; $S(\underset{\sim}{\theta}-\mu)^2)/S(1)$ provides the posterior variance; $S(p(\underset{\sim}{y}/\underset{\sim}{\theta}))/S(1)$, where

$p(\underline{y}/\underline{\theta})$ is the density of the distribution of future data $\underline{y}$, provides the predictive density of $\underline{y}$. In practice the need to evaluate the integral (1.4), particularly in high dimensions, places a serious technical barrier to many applications. Attempts to avoid or overcome this technical barrier can usefully be classified as follows.

. restriction of models to a suitably tractable class.

. analytic approximations (usually based on asymptotic theory)

. numerical approximations (including Monte-Carlo methods)

In the remainder of this section we give a brief overview of each of these in turn.

1.1.3 Restricted models

If $\ell(\underline{x};\underline{\theta})$ and $p(\underline{\theta})$ belong to exponential and conjugate family respectively, the integral (1.4) can be evaluated analytically. For more details and examples, see Lindley (1972), Cox and Hinkley (1974), Box and Tiao (1973). However, the use of likelihood within the exponential family and conjugate priors often represents an unrealistically simplistic approach. Notable cases where this approach fails are when censoring or group effects or outliers are encountered in the data. Moreover, the choices of the population and prior densities are very often restricted to particular parametric families or are selected from an enriched family for reasons of analytic convenience. This is an essential disadvantage because it

eliminates important benefits of the Bayesian approach and the validity of the paradigm as described in (1.1)-(1.4).

1.1.4 Analytic approximations

Analytic approximations based on asymptotic theory have received considerable attention in the literature. In essence, the theory shows that, under certain regularity conditions and for large samples, the posterior distribution of $\underset{\sim}{\theta}$ is approximately

$$N(\hat{\underset{\sim}{\theta}}, \underset{\sim}{\Sigma})$$

where $\hat{\underset{\sim}{\theta}}$ is the maximum likelihood vector and $\underset{\sim}{\Sigma}^{-1}$ has elements

$$(\underset{\sim}{\Sigma}^{-1})_{ij} = \frac{-\partial^2 \ell n \ell(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial\theta_i \theta_j}\Bigg|_{\underset{\sim}{\theta}=\hat{\underset{\sim}{\theta}}}.$$

These results, which resemble the asymptotic results for maximum likelihood inference, lead to convenient approximate solutions to the problem for large samples. For details of typical regularity conditions see Walker (1969), Le Cam (1970), Dawid (1970) and Heyde and Johnson (1979). The major difficulty with solutions based on asymptotic properties is that of checking the assumption of approximate normality particularly when dealing with a complex, multiparameter problem. See Smith and Naylor (1987) for a case-study illustration of the problem.

At the time of writing, the ultimate analytic treatment in Bayesian analysis is the method suggested by Tierney and Kadane (1986), see also Kass *et al.* (1989) and Tierney et al. (1987). This requires the evaluation of first and second derivatives of slightly modified likelihood functions. Other asymptotic approximations of particular interest are due to Johnson (1970), Lindley (1980).

The practical potential of the analytic methods depends clearly on the context in which a particular task is to be performed. If we are dealing with a specific application where the task is to be performed repeatedly, an analytic approximation with a detailed analysis of its accuracy is probably preferable to repeated use of numerical methods. On the other hand, performing a detailed analytical study in an one-off situation is unreasonable, if not beyond the scope of many practitioners. This could happen in a typical analysis problem in which the statistician might have a restricted time for the analysis. This is where the need of a general purpose numerical integration strategy is called for. In addition, as Smith et al. (1985) point out, all forms of analytic approximations require empirical validation in specific areas of application and at present we need the numerical approaches even for the purpose of judging the power of analytic approximations.

1.1.5 Numerical methods

Numerical methods have recently been introduced to cope with the analytic intractability which often occurs in Bayesian statistics. Reilly (1976) gave an early example of numerical integration in Bayesian analysis. He used a large number of grid points to evaluate the function and approximated the integral needed as a summation of point values. This crude method needs a very large number of function evaluations, especially as the number of parameters increases. Moreover, the choice of location and size of grid is a subjective process which may involve a considerable amount of labour and computation.

Dagenais and Liem (1981) have described a procedure in which univariate marginal densities can be approximated using successive transformations of $\underline{\theta}$, together with results of the assumption of asymptotic normality and additional transformations proposed by Johnson (1949). Subsequently, inferences about the original parameter set are made by inverting all the transformations used.

Naylor and Smith (1982) introduced the first general purpose strategy for numerical integration in Bayesian analysis using Gaussian quadrature to evaluate efficiently integrals of the form (1.4) for a wide range of problems. The method has been refined considerably (see Smith *et al* (1985), Smith *et al* (1987)) since strategies coping with multi-dimensional problems and graphical displays have been presented. The method has been implemented by the Nottingham Statistics group and a computer package called 'BAYESFOUR'. See Naylor and Shaw (1985), (1988) for details. Section 1.2 is concerned with a brief

review of the method.

1.2 The iterative algorithm of Naylor and Smith

A major breakthrough towards the routine implementation of the Bayesian paradigm has been the iterative algorithm suggested by Naylor and Smith (1982). The second chapter of this thesis contains a detailed description of the algorithm, but a brief summary is included here.

The iterative quadrature strategy exploits the asymptotic normal form of the likelihood and employs a transformation of the parameter vector to a parameter which has, at least approximately, zero mean and identity variance-covariance matrix. This transformation is achieved through estimates of first and second moments which in turn have been derived from previous iterations. In each iteration, approximations to posterior moments and normalising constants can be calculated using Gaussian product quadrature formulae. The iterative algorithm is continued until stable answers are obtained for the required integrals between successive iterations.

The method of Naylor and Smith opened a new field in the area of Bayesian analysis making possible the calculation and reconstruction of posterior joint or marginal densities with realistic statistical models and prior distributions. It provided the user of the Bayesian paradigm with the necessary tools to exploit the wide range of options offered to him, such as model sensitivity and robust inference. However, the above iteration scheme imposes several restrictions; it

requires an initial transformation of the parameters to normality, in cases where the asymptotic assumptions are not valid. Moreover, when the dimensions are high, (typically 5 or more), the Gaussian quadrature formulae require an enormous number of function evaluations.

1.3 Other numerical integration strategies

1.3.1 Spherical rules

Naylor and Smith (1982, 1988b) have mentioned the possible use of spherical rules in high dimensions where product rules are too expensive. These rules are more economical than product rules in the sense that, for given a number of points, they are of greater precision than product rules (see Naylor and Smith (1988b) table 1).

The spherical rules are based on the observation that, if we make the transformation θ to ψ where

$$\theta_1 = r\cos\psi_{k-1}\cos\psi_{k-2}\ldots\cos\psi_2\cos\psi_1 ,$$

$$\theta_2 = r\cos\psi_{k-1}\cos\psi_{k-2}\ldots\cos\psi_2\sin\psi_1 ,$$

$$\theta_3 = r\cos\psi_{k-1}\cos\psi_{k-2}\ldots\sin\psi_2 ,$$

$$\vdots$$

$$\theta_k = r\sin\psi_{k-1} ,$$

then we can write the integral (4) as a product of integrals of the form

$$\int_{-\pi}^{\pi} \cos(\psi_i)^{a_i} (\sin\psi_i)^{b_i} d\psi_i \quad , i=1,\ldots,k-1$$

and

$$\int_{0}^{\infty} (r^2)^{c-1} \exp(-r^2)\, dr$$

for some a_i, b_i, c.

These rules, described in detail in Stroud (1971, sections 2.6,2.7), have been embodied in 'BAYESFOUR' (see Naylor and Shaw (1985)) for $k \geq 4$ in (1.4).

Intuitively, we might expect these spherical rules which reflect the spherical symmetry of the posterior density function to perform better than product rules. Unfortunately, there are certain limitations to the use of spherical rules. There are few high precision rules with positive weights. For instance, only rules up to degree seven are used in BAYESFOUR (See Stroud (1971), page 317-319, rules E_n :5-3 and E_n :7-2). Another disadvantage of the spherical rules is that marginal or joint posterior densities cannot readily be derived. A way of overcoming this, is to mix integration stategies as described in Naylor and Smith (1989b, sect. 6). Hence, some parameters can be treated using product rules while the other (nuisance) parameters are dealt with by spherical rules. Of course, the choice of 'interesting' and 'nuisance' parameters can vary between iterations so marginal summarisations can be derived for a large number of parameters.

1.3.2 Sampling based methods

Another way to economize on the number of function evaluations when dealing with high dimensions in (1.4) is to adopt a Monte Carlo approach. This method has been used extensively in the context of Bayesian analysis over the last two decades. See Stewart and Johnson (1972), Stewart (1979, 1983), Kloeck and van Dijk (1978), and van Dijk and Kloek (1980, 1984), Shaw (1988). The general approach proceeds as follows. Suppose it is possible to generate an i.i.d. sequence of random variables $\{\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2, \ldots, \underset{\sim}{\theta}_m\}$ having common density $h(\underset{\sim}{\theta}) > 0$. The Monte-Carlo approach then approximates $S_I(q(\underset{\sim}{\theta}))$ in (1.4) by

$$\hat{S}(q(\underset{\sim}{\theta})) = \sum_{i=1}^{m} q(\underset{\sim}{\theta}_i) w(\underset{\sim}{\theta}_i) ,$$

where $w(\underset{\sim}{\theta}_i) = \ell(\underset{\sim}{x}; \underset{\sim}{\theta}_i).p(\underset{\sim}{\theta}_i)/h(\underset{\sim}{\theta}_i)$. The density $h(\underset{\sim}{\theta})$ is called the importance function and the process of generating $\underset{\sim}{\theta}_i$ according to h is called importance sampling. The efficiency of the method depends clearly on the choice of a suitable importance function. Ideally, $h(\underset{\sim}{\theta})$ should be chosen to resemble $q(\underset{\sim}{\theta})$ and to allow the θ_i's to be generated easily. See Hammersley and Handscomb (1964), and Rubinstein (1981) for more details. In practice, it is possible to estimate simultaneously many $S(q(\underset{\sim}{\theta}))$ for different $q(\underset{\sim}{\theta})$ using the same sample $\{\theta_1, \theta_2, \ldots, \theta_m\}$. This is achieved by choosing $h(\underset{\sim}{\theta})$ to have a similar shape with $q(\underset{\sim}{\theta})$ but with heavier tails, followed by the use of a suitably 'uniform' configuration on points in the k-dimensional hypercube.

A class of univariate distributions which provide a flexible set of distributional shapes, and hence a family of 'suitable' densities $h(\theta)$ is given by Shaw (1986a). The problem of specification of 'uniform' configurations of points in the k-dimensional hypercube has been studied by Shaw (1988). Smith *et al.* (1987) and Naylor and Smith (1988b) describe an iterative importance sampling strategy which has been embodied in BAYESFOUR; this is recommended for use on all problems with 9 or more parameters. See Naylor and Shaw (1985) for more details on the use of Monte Carlo methods in BAYESFOUR.

Considerable objections are made to Monte Carlo methods on the grounds that they add random variation and ignore information such as the position of the generated nodes. See Bacon-Shone's comments in the discussion to van Dijk and Kloek (1985) and O'Hagan (1987).

Very recently, Gelfand and Smith (1988) described sampling-based approaches to calculating marginal densities and Gelfand *et al.* (1989) described how the Gibbs sampler can be used effectively to obtain inference summaries in a range of normal data models. Chapter 6 of this thesis is devoted to applications of Gibbs sampling in the large area of Generalised linear models. This includes, among other things, comparisons with the numerical integration approaches. We present here the basic steps of the Gibbs sampling approach.

The Gibbs sampler is a Markovian updating scheme for the convergence of a density, introduced by Geman and Geman (1984). Given a joint posterior density $p(\underline{\theta}|\underline{x})$, functional forms of the k univariate conditional densities can be readily written down, at least up to proportionality. If these densities are denoted by

$$p(\theta_1|\theta_2,\theta_3,\ldots,\theta_k)$$

$$p(\theta_2|\theta_1,\theta_3,\ldots,\theta_k)$$

$$\ldots \qquad (1.5)$$

$$p(\theta_k|\theta_1,\theta_2,\ldots,\theta_k).$$

then the Gibbs sampling algorithm proceeds as follows: Choose initial values for $\theta_2^{(0)},\theta_3^{(0)},\ldots,\theta_k^{(0)}$ and generate a value $\theta_1^{(1)}$ from the conditional density

$$p(\theta_1|\theta_2^{(0)},\theta_3^{(0)},\ldots,\theta_k^{(0)})$$

Similarly, generate a value $\theta_2^{(1)}$ from the conditional density

$$p(\theta_2|\theta_1^{(1)},\theta_3^{(0)},\ldots,\theta_k^{(0)})$$

and continue up to the value $\theta_k^{(1)}$ from the conditional

$$p(\theta_k|\theta_1^{(1)},\theta_2^{(1)},\ldots,\theta_{k-1}^{(1)}).$$

Then, the new realisation of $\underline{\theta}$ given by $\underline{\theta}^{(1)}$ can be utilised and the above process repeated, say m times, producing $\underline{\theta}^{(m)}$. Following Geman and Geman (1984), under mild conditions,

$$\theta_i^{(m)} \xrightarrow{d} \theta_i \sim p(\theta_i)$$

and therefore, for m large enough, $\theta_i^{(m)}$ can be regarded as a simulated observation from $p(\theta_i)$, the marginal distribution of θ_i.

Replication of the above process t times produces t sets of parameter values, $(\underline{\theta}_j^{(m)}, j=1,t)$, and thus for each element of $\underline{\theta}$ we obtain a simulated sample with size t from its marginal density. Values for this marginal density can be calculated using either a kernel density estimate (see Silverman (1986)) or by averaging over the conditional density:

$$p(\theta_i) \approx \frac{1}{t} \sum_{j=1}^{t} p(\theta_i | \underline{\theta}_j'^{(m)}),$$

where $\underline{\theta}_i'$ denotes the complement of θ_i in $\underline{\theta}$.

Proponents of the above iterative algorithm do not generally claim that it competes with other methods in terms of efficiency, but, it provides a method which is simple to implement and which exploits structural information given by (1.5). It has been demonstrated that the method can be used successfully in otherwise numerically (and analytically) intractable problems (see for example Gelfand *et al.* (1989)). At the time of writing, it seems that potential improvements are expected towards the direction of a general purpose numerical routines for the implementation of the Bayesian paradigm using the Gibbs sampler (see also Hills (1989),Smith and Gelfand (1990)).

1.4 Imbedded sequences of integration rules

This thesis describes how imbedded sequences of positive interpolatory integration rules (PIIR's), obtained from Gauss-Hermite product rules,

can be applied in Bayesian analysis. These imbedded sequences are very promising for two major reasons. Firstly, they provide a rich class of spatially distributed rules which are particularly useful in high dimensions. Secondly, they provide a way of producing more efficient integration strategies by enabling approximations to be updated sequentially, by the addition of new nodes at each step rather than by changing to a completely new set of nodes. Moreover, as points are added successive rules change naturally from spatially distributed non-product rules to product rules. This feature is particularly attractive when the rules are used for the evaluation of marginal posterior densities. The basic theory of these rules is described in chapters 3 and 4, and a suggested integration strategy is proposed in chapter 5.

1.5. Discussion

...It is therefore surely no longer acceptable, neither from an intellectual nor a public relations perspective, simply to proclaim and demonstrate, in the theoretical domain, the inevitability or desirability of the Bayesian position without following the enterprise through to provide the appropriate tools in the practical domain.

A.F.M. Smith (1988)

Vast amounts of research over the last decade have been directed torwards the development of a general purpose software package for Bayesian analysis which can be used routinely by data analysts. Efficient methods of numerical integration and approximation have been

developed and these have been reviewed in this chapter. The technical implementation clearly depends on the ability to calculate the forms of integrals as in (1.4) for any given likelihood and prior specifications.

These methods, however, all have their drawbacks; restricted model choices for analytic approximations; lack of high-degree spherical rules; choice of appropriate importance functions and sampling rules for Monte Carlo methods; expensive and often impracticable high-dimensional product integration. It was therefore intended that the research program should follow the avenue of developing numerical methods for Bayesian analysis which overcome some of these drawbacks. In the following chapters, an extensive review of the relevant literature in numerical integration is carried out, and a new quadrature strategy, based on imbedded sequences of positive interpolatory integration rules (see section 1.4) is suggested.

Chapter 2: Applications of the quadrature strategy of Naylor and Smith

2.1 Introduction

Widespread applications of the adaptive integration strategy introduced by Naylor and Smith (1982) have been documented in the literature, see Table 2.1. These illustrate the rich variety of applications wich have been dealt with and the practical importance of the strategy. Here we describe this strategy and review real applications highlighting some of their important aspects.

In section 2.2 we give a detailed account of the adaptive integration strategy. The following section then gives a simple application of the iterative strategy using BAYESFOUR. This example introduced by Reilly (1976) and was reanalysed by Naylor and Smith (1982).

Although it has been noted in the literature that the numerical strategy can be applied in conjuction with analytic integration over a subset of the parameter space, there is lack of published examples in which this approach is used. In section 2.4 we illustrate using real data a two parameter problem analysed using analytical and numerical integration.

Finally, in section 2.5 of this chapter, we describe an experimental examination of various properties of the numerical strategy of Naylor and Smith.

Table 2.1

Publications with applications of the method by Naylor and Smith (1982)

Reference		Application area	Dimensions
Naylor and Smith	(1982)	Leukaemia data	3
		Stanford Heart Transplant data	3
		Regression with censored data	3
	(1983)	Clinical Chemistry	3
	(1988a)	Archeological data	5
	(1988b)	Haavelmo's consumption model	4
		Directional disequilibrium model	8
Smith and Naylor	(1987)	3 parameter Weibull distribution	3
Smith *et.al.*	(1985)	Probit analysis	2
		Non-linear regression	3
Grieve	(1987)	Regression with propor. hazards	5
Naylor	(1987)	Bayesian alternatives to t-tests	3
Lee	(1987)	Analysis of variance	3
Marriot	(1987)	Box Jenkins models	2
Shaw	(1987a)	Spline logistic regression model	4

2.2 The iterative quadrature strategy of Naylor and Smith (1982)

2.2.1 Genaral formulation

The iterative quadrature strategy of Naylor and Smith (1982) exploits the asymptotic normal form of the likelihood using, where appropriate, parameter transformations to improve the asymptotic normality; for example, a variance σ^2 can be reparameterised to $\log\sigma^2$; a proportion p ($0<p<1$) to $\log(p/(1-p))$. See Hills (1989) for more details. The asymptotic theory enables the argument of (1.4) to be expressed in terms of the product

$$h(\underline{\theta}).n(\underline{\theta};\underline{\varphi},\underline{\Sigma}) \qquad (2.1)$$

where $n(\underline{\theta};\underline{\varphi},\underline{\Sigma})$ is the d-dimensional normal density with mean $\underline{\varphi}$ and covariance matrix $\underline{\Sigma}$ and where under appropriate conditions $h(\underline{\theta})$ is a suitable well behaved function. For assumed $\underline{\varphi}$ and $\underline{\Sigma}$ a linear transformation of $\underline{\theta}$ to a d-dimensional vector $\underline{x}$ leads to the standardized form

$$S(q(\theta)) = If = \int f(\underline{x})e^{-\underline{x}^T\underline{x}}d\underline{x} .$$

This integral is amenable to approximation by standard quadrature formulae of the form

$$Q_m f = \sum_{i=1}^{m} w_i f(\underline{x}_i) \qquad (2.2)$$

where w_i are the weights and the $\underline{x}_i$'s are the nodes of the formula Q_m.

degree of precision

It is common in the numerical analysis context to test the power of an integration formula such as (2.2) by referring to the 'degree' of the formula or of the integration rule. Hence, before we proceed further we state here the definition of the degree of the rule:

A rule of the form (2.2) is said to be a degree p (or precision p or degree of exactness p) if is exact for all monomials of degree p or less (i.e. if it is exact for all monomials $\prod_{i=1}^{d} x_i^{a_i}$ *with* $\sum_{i=1}^{d} a_i \leqslant p$*) and there is at least one monomial of degree p+1 for which it is not exact.*

We mention here that comparison of integration rules merely in terms of their precision, as for example in Naylor and Smith (1988b, table 1), can be misleading. Rabinowitz and Richter (1969) note that product rules can be equal or even superior to non-product rules with higher degree. This happens because product rules of degree p can can integrate all monomials with terms $\prod_{i=1}^{d} x_i^{a_i}$ with $a_i \leqslant p$ and not only these monomials for which $\sum_{i=1}^{n} a_i \leqslant p$. For more details, see Mantel and Rabinowitz (1977) who give definitions of 'optimal' or 'minimal' rules. Having given a brief outline of the general structure of the

procedure we proceed to look at the one dimensional case in detail. This is then followed by a detailed description of the generalisation to d dimensions.

2.2.1.1. 1-dimensional case

Using Gauss-Hermite quadrature rules, we can approximate univariate integrals of the type

$$\int_{-\infty}^{\infty} e^{-t^2} f(t)\, dt \qquad (2.3)$$

by

$$\sum_{i=1}^{n} w_i f(t_i) \qquad (2.4)$$

The error in this is given by

$$E_n = \frac{n!\sqrt{\pi}}{2^n(2n)!} f^{(2n)}(\xi).$$

Thus the rule is exact if f(t) is a polynomial of degree 2n+1 or less. Moreover, the error will be small whenever the high order derivatives of f(t) are sufficiently small. The nodes t_i and the weights w_i of the rule (2.4) can be found in books or can derived using published programs. See Stroud and Secrest (1966) for a list of nodes and weights as well as a FORTRAN program to derive them.

In Bayesian analysis we seek to evaluate integrals of the form

$$S(q(\theta)) = \int_{-\infty}^{\infty} q(\theta)\ell(\theta/\text{data})p(\theta)d\theta$$

For the purposes of our explanation it is convenient to write the above integral in the form

$$q(\theta)\ [\ p(\theta)\ell(\theta/\text{data})\exp((\theta-\mu)^2/2\sigma^2)\ ]\ \exp(-(\theta-\mu)^2/2\sigma^2)$$

where μ and σ^2 represent the posterior mean and variance of θ. Asymptotic theory predicts that, under suitable conditions, $p(\theta)\ell(\theta/\text{data})$ is proportional to a normal density with mean μ and variance σ^2. Thus, the expression within [] should, under suitable conditions, be a slowly varying function of θ. Indeed, for exact normality, the term within [] will be constant.

Transformation to an integral of the form (2.3) by putting $t=(\theta-\mu)/\sqrt{(2\sigma^2)}$ followed by application of (2.4) yields the approximation

$$\sum_{i=1}^{n} m_i q(\theta_i) \qquad (2.5)$$

where

$$\left.\begin{aligned} m_i &= w_i \exp(t_i^2)\sqrt{2}\sigma p(\theta_i)\ell(\theta_i/\text{data}) \\ \theta_i &= \mu + \sqrt{2}\sigma t_i \end{aligned}\right] \qquad (2.6)$$

The m_i's, which involve perhaps time consuming evaluations of the likelihood need only be evaluated once and can be applied to approximate S(q(.)) for various q(.).

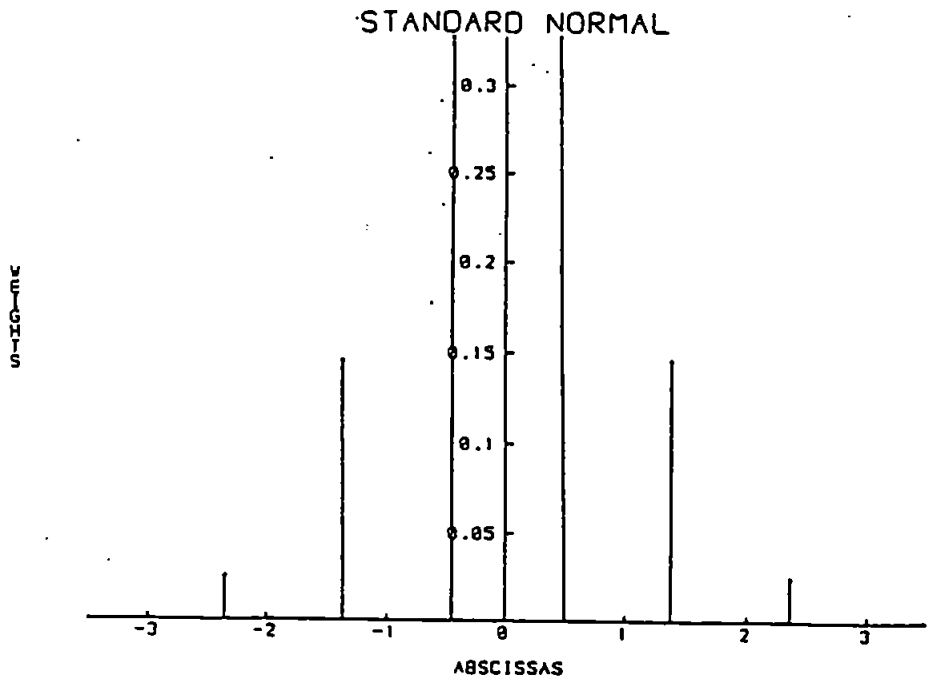
Note that (2.5) yields the approximations to posterior expectations of the form

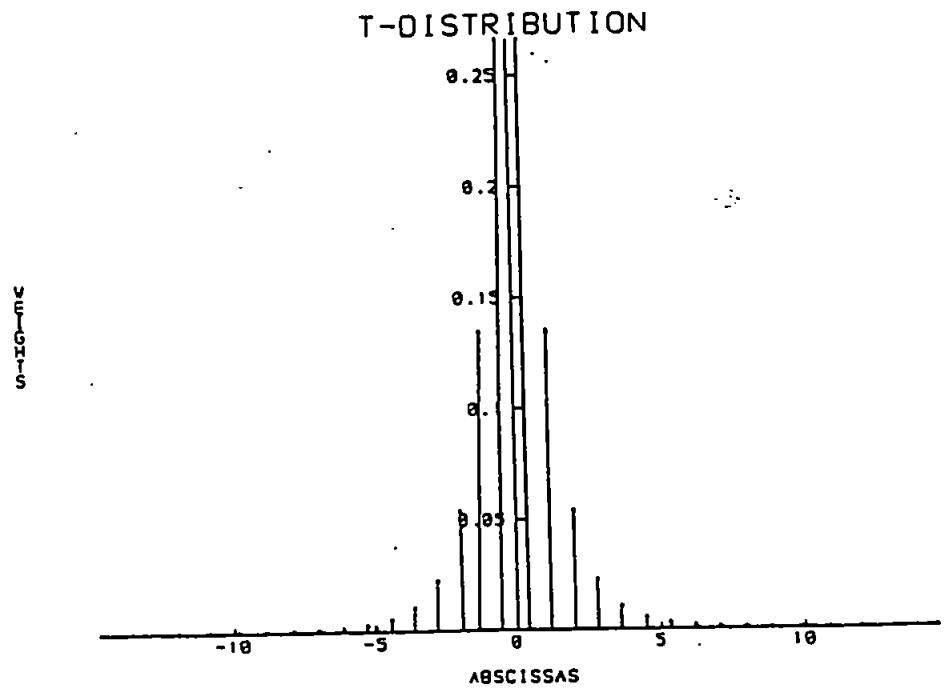
$$E[q(\theta)] = \frac{S_I(q(\underline{\theta}))}{S_I(1)} = \sum_i p_i q(\underline{\theta}_i), \qquad \sum_i p_i = 1$$

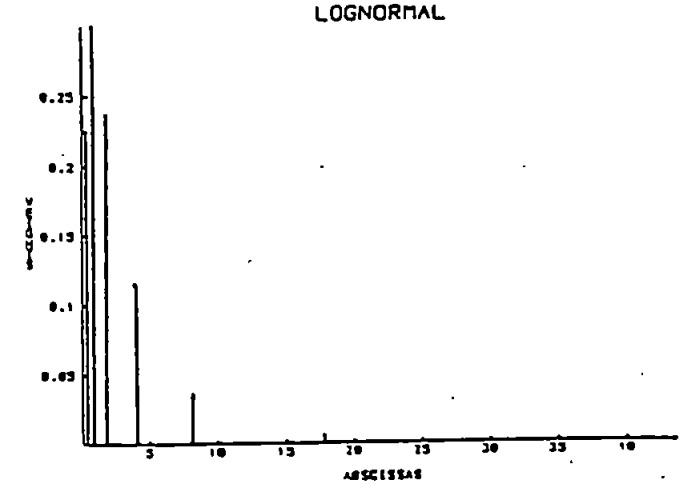
where $p_i = m_i / \Sigma m_i$. The approximation of the expectations can be regarded as an approximation which replaces continous posterior distributions by a discrete distribution with probabilities $p_1, p_2, \ldots, p_n$ at points $\theta_1, \theta_2, \ldots, \theta_n$.

For illustrative purposes, we consider three posterior distributions: (i) a standard normal distribution, (ii) a t-distribution with 3 degrees of freedom and (iii) a lognormal distribution with scale and shape parameters equal to 1. Expectations are then approximated by expectations with respect to the discretised posterior distribution (2.5). For the purpose of these illustrations the true values of the mean and variance were used for μ and σ^2. Figure 2.2 illustrates the discrete distributions obtained for each of the three examples as derived from (2.6). An 8-point Gauss-Hermite formula was used for the two former distributions and a 32-point formula for the latter. Note that for the integration of the lognormal density function a log-transformation of the parameter was applied to achieve normality. Thus, the nodes and weights in (2.6) were replaced by

FIGURE 2.2


STANDARD NORMAL
WEIGHTS
0.3
0.25
0.2
0.15
0.1
0.05
-3
-2
-1
0
1
2
3
ABSCISSAS


T-DISTRIBUTION
WEIGHTS
0.25
0.2
0.15
0.05
-10
-5
0
5
10
ABSCISSAS


LOGNORMAL
0.25
0.2
0.1
5
10
15
20
25
30
35
40
ABSCISSAS

$$m_i^* = m_i \exp(t_i^2 + \mu + /2\sigma t_i)$$

$$\theta_i^* = \exp(\theta_i)$$

Iteration over μ and σ

To use (2.5) we must specify μ and σ in (2.6). The most commonly used method is to give starting values to μ and σ so that the normal density is 'close' to g(t). These values may be the maximum likelihood estimators as given in section 1.1.4, or may be any other reasonable initial estimates. Then we can iterate on (2.5), substituting into (2.6) estimates of μ and σ^2 taken as the posterior mean and variance respectively, as given by (i) and (ii) :

$$\text{(i)} \quad \mu = S(\theta)/S(1)$$

$$\text{(ii)} \quad \sigma^2 = S((\theta-\mu)^2)/S(1)$$

and obtained using (2.5) based on previous values of m_i and θ_i.

In practice, the iterative process begins by iterating within a coarse grid of points and proceeds to the next step by changing μ and σ^2 and/or using a finer grid. The iteration ends when the convergence is satisfactory within and between each grid size(s).

2.2.1.2 Multidimensional case

In cases where higher dimensions need to be considered, the method can be extended using a Gauss-Hermite product rule,

$$\int \cdots \int q(\theta_1,\theta_2,\ldots,\theta_k)\ell(\theta_1,\theta_2,\ldots,\theta_k|\text{data})p(\theta_1,\theta_2,\ldots,\theta_k)d\theta_1\theta_2\ldots\theta_k \simeq$$

$$\simeq \sum_{ik}^{(k)} m_{ik}^{(k)} \cdots \sum_{i2}^{(2)} m_{i2}^{(2)} \sum_{i1}^{(k)} m_{i1}^{(k)} q(\theta_{i1}^{(1)},\theta_{i2}^{(2)},\ldots,\theta_{ik}^{(1)}) \qquad (2.7)$$

where $m_{ij}^{(j)}$, $\theta_{ij}^{(i)}$ can be found using (2.6), substituting the marginal posterior mean and variance of θ_j for μ and σ^2. The number of nodes can be different for different components of $\underline{\theta}$. A further assumption implicitly involved here is that of posterior independence. In problems where this assumption is not reasonable, a transformation of the parameters in $\underline{\theta}$ to a new, approximately orthogonal set of parameters can be applied. This is achieved by applying the transformation of the form

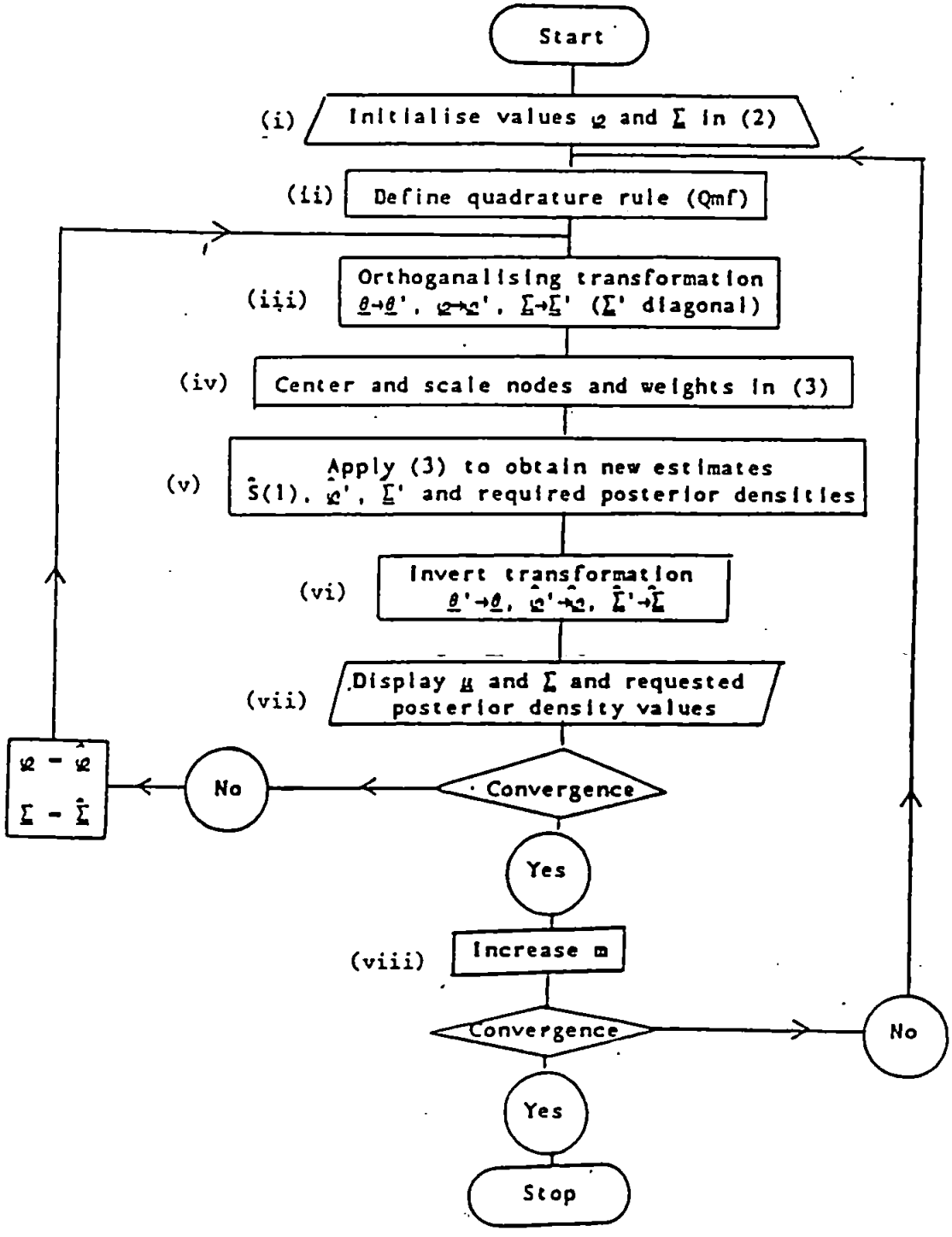
$$\psi_1 = \theta_1$$

$$\psi_i = \theta_i + \sum_{j=1}^{i-1} \beta_{ij}\,\psi_j \ , \quad i=2,\ldots,k$$

with

$$\beta ij = -\text{Cov}\,(\theta_i,\psi_j|x) \ / \ \text{Var}(\psi_j|x)$$

FIGURE 2.1: Flowchart showing the numerical integration strategy of Naylor and Smith (1982)



The flowchart in figure 2.1 shows the essential steps in the iterative strategy. At each stage of the algorithm the transformation is recalculated and the updated posterior moments are used as a guide to the convergence. A key feature of the approach is that the same nodes and weights can be used to calculate normilising constants, marginal density values and moments. This leads to considerable gains in efficiency.

Naturally, the above approach works most efficiently for posterior densities which are very nearly normal. The regularity conditions (Johnson (1970)) indicate a class of problems for which the basic assumption (2.1) is valid. In addition, these conditions are not considered to be the 'weakest' assumptions under which we can achieve such an expansion, and the method can have a theoretical justification that produces satisfatory results for a wide range of problems. Moreover, in cases where no consistent convergence is recorded between iterations, we can conclude the approximation (2.1) is inadequate. Thus, the method provides a 'fail safe' system in that problems outside the class for which the method is appropriate will not produce satisfactory results.

For more than 5 or 6 parameters, the computational power needed to apply the rule (2.5) is enormous. This happens because the number of function evaluations increases rapidly in the use of Gauss-Hermite product rules. In addition, these rules spend all previous function evaluations when procceeding from one rule to another one with higher precision. BAYESFOUR uses Gauss-Hermite rules up to dimension 6. In higher dimensions, other facilities are used to overcome the problem such as fixing some variables or mixing integration rules (see Naylor

and Smith (1988b)). These rules, however, require a very experienced user and add more uncertainty when testing the convergence.

2.3. A three dimensional example

We have already described, in section 1.1.5, the simple approach of Reilly (1976). Here we use one of his examples to demonstrate the Gauss-Hermite rules in higher dimensions.

The data in table (2.2) were generated by Reilly (1976, table 1) using the non-linear regression model

$$\ln y_i = \ln(\alpha+\beta x_i) + \epsilon_i$$

Table 2.2

x_i	y_i
0	4.11
1	6.32
2	8.21
3	10.43
4	14.29
5	16.78

with $\alpha=5$, $\beta=2$ and errors ϵ_i = (i=1,2,...,6) being normally and independently distributed with mean zero and variance σ^2.

Reilly considered the two parameter problem with known variance ($\sigma^2=0.1398^2$). Naylor and Smith (1982) used this example to compare their strategy with Reilly's method, considering also the case where σ^2 is unknown. In the latter case, the three parameter problem with

$$\underline{\theta} = (\alpha,\beta,\sigma^2)$$

has a likelihood of the form

$$\ell(\underline{x};\underline{\theta}) = (2\pi\sigma^2)^{-3}\exp\Big[\frac{-1}{2\sigma^2}\sum_{i=1}^{6}\Big\{\ln\Big(\frac{y_i}{\alpha+\beta x_i}\Big)\Big\}^2\Big]$$

Using an improper prior and a log-transformation of the variance σ^2 we applied the iterative strategy illustrated in figure (2.1). The analysis was carried out using BAYESFOUR with maximum likelihood estimators to obtain initial initial posterior means and covariance matrix. In order to demonstrate the efficiency of the strategy the aggregate measure Δ, described in Naylor and Shaw (1985), was used as an objective measure of convergence. For a d dimensional parameter $\underline{\theta}$ let $p(\underline{x})$, μ_i, σ_i and ρ_{ij} denote respectively the normalising constant, the posterior mean and the posterior standard deviation of the i^{th} component of $\underline{\theta}$, and the correlation between the i^{th} and j^{th} component. The relative change in these quantities from the previous to the current iteration is represented by

$$\Delta p(x) = \frac{p^{(1)}(x) - p^{(0)}(x)}{p^{(0)}(x)},$$

$$\Delta\mu_i = \frac{\mu_i^{(1)} - \mu_i^{(0)}}{\mu_i^{(0)}},$$

$$\Delta\sigma_i = \frac{\sigma_i^{(1)} - \sigma_i^{(0)}}{\sigma_i^{(0)}} ,$$

and $$\Delta\rho_{ij} = \rho_{ij}^{(1)} - \rho_{ij}^{(0)}$$

in which the superscripts (0) and (1) denote the previous and current iterations. The aggregate measure Δ is then given by

$$\Delta = |\Delta p(x)| + \frac{1}{m} \sum_{i=1}^{d} |\Delta\mu_i| + \frac{1}{m} \sum_{i=1}^{d} |\Delta\sigma_i| + \frac{1}{m^2} \sum_{i=2}^{d} \sum_{j=1}^{i-1} |\Delta\rho_{ij}|$$

where m is the number of elements of $\underline{\theta}$ for which moments were calculated.

The convergence between steps being considered satisfactory when an overall change Δ of less than 5% had been achieved. The analysis summary from BAYESFOUR, given below, illustrates the convergence of the posterior moments and the overall change (in percentage) in each iteration. The analysis started with a grid of 4×4×4 points and convergence criterion was satisfied after 4 iterations. This was followed by further 2 iterations with a 5^3 rule, 2 with a 6^3 rule, 2 with a 7^3 rule and finally a single iteration using a 8^3 rule.

In each iteration, BAYESFOUR gives the current estimates of the posterior means and variances of the three parameters, together with their posterior correlations and the normalising constant, which is denoted by p(x). The overall change Δ is given at the end of each iteration. The full picture of the iterative algorithm is integrated with the clear demonstration of the grid changes and the form of the linear transformation that were used.

```
                                                    TUE, 21 MAR 1989 @ 16:08:36
        +-------------------------------+
        ! BAYES FOUR : ANALYSIS SUMMARY !
        +-------------------------------+
                                                     Output summary file = t$2
                                                    Problem data file = re.dat
-------------------------------------------------------------------------------
                                                  moment input file = re.ma::lik
                                                       Starting point supplied:
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  3.99630      2.37920     -5.96940
Posterior Standard Deviations
 0.179815     0.106457     0.577321
Posterior Correlations
         alpha    beta
beta      -0.5769
sigma*2    0.0000   0.0000
Linear transformation used :
Operational parameters are :
 alpha       beta___1    sigma*22
-------------------------------------------------------------------------------
Overflow control: estimate log(p::) =   7.443931
-------------------------------------------------------------------------------
          Parameter space is partitioned into sub-spaces :
          Cartesian product rule on              3 dimensions
          Spherical rule on                      0      "
          Monte-Carlo rule on                    0      "
          giving integration over                3      "
          for a problem with                     3 parameters
          Cartesian product grid size is :   4 X   4 X   4
          A linear transformation is applied to   3-1 parameters

*******************************************************************************
Iteration Number   1
-------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta___1    sigma*22
-------------------------------------------------------------------------------
p(::) =   2.3409376 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.00682      2.37946     -5.47310
Posterior Standard Deviations
 0.225120     0.133315     0.613208
Posterior Correlations
         alpha    beta
beta      -0.5762
sigma*2    0.0053   0.0006

*****  65.4%*****************************************************************
Iteration Number   2
-------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta___1    sigma*22
-------------------------------------------------------------------------------
p(::) =   2.4719310 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.01072      2.37956     -5.36149
Posterior Standard Deviations
```

```
 0.252372     0.149377     0.663818
Posterior Correlations
         alpha    beta
beta      -0.5757
sigma*2    0.0230   0.0011

*****  16.2%**********************************************************************
Iteration Number   3
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta__1    sigma*22
------------------------------------------------------------------------------
p(x) =   2.4875274 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.01159      2.37962     -5.31892
Posterior Standard Deviations
 0.266304     0.157555     0.674465
Posterior Correlations
         alpha    beta
beta      -0.5753
sigma*2    0.0362   0.0011

*****   5.9%**********************************************************************
Iteration Number   4
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta__1    sigma*22
------------------------------------------------------------------------------
p(x) =   2.4751959 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.01178      2.37965     -5.29894
Posterior Standard Deviations
 0.273310     0.161671     0.675368
Posterior Correlations
         alpha    beta
beta      -0.5751
sigma*2    0.0433   0.0010

*****   2.8%**********************************************************************
Iteration Number   5
                                                                    Changes:
                                                                    New grid
          Parameter space is partitioned into sub-spaces :
          Cartesian product rule on              3 dimensions
          Spherical rule on                      0     "
          Monte-Carlo rule on                    0     "
          giving integration over                3     "
          for a problem with                     3 parameters
          Cartesian product grid size is :   5 X   5 X   5
          A linear transformation is applied to   3-1 parameters
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta__1    sigma*22
------------------------------------------------------------------------------
p(x) =   2.6795301 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.01870      2.37948     -5.36683
```

```
Posterior Standard Deviations
 0.266093     0.157258     0.758834
Posterior Correlations
         alpha    beta
beta      -0.5757
sigma*2    0.0313   0.0020

*****  12.6%**********************************************************************
Iteration Number   6
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta__1    sigma*22
------------------------------------------------------------------------------
p(::) =   2.6677575 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.01869      2.37948     -5.36838
Posterior Standard Deviations
 0.266084     0.157192     0.774210
Posterior Correlations
         alpha    beta
beta      -0.5760
sigma*2    0.0263   0.0019

*****   1.3%**********************************************************************
Iteration Number   7
                                                                      Changes:
                                                                      New grid

          Parameter space is partitioned into sub-spaces :
          Cartesian product rule on          -  3 dimensions
          Spherical rule on                     0     "
          Monte-Carlo rule on                   0     "
          giving integration over               3     "
          for a problem with                    3 parameters
          Cartesian product grid size is :   6 X   6 X   6
          A linear transformation is applied to   3-1 parameters
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta__1    sigma*22
------------------------------------------------------------------------------
p(::) =   2.5731978 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.01672      2.37956     -5.29693
Posterior Standard Deviations
 0.282426     0.166778     0.738570
Posterior Correlations
         alpha    beta
beta      -0.5750
sigma*2    0.0502   0.0015

*****   8.6%**********************************************************************
Iteration Number   8
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta__1    sigma*22
------------------------------------------------------------------------------
p(::) =   2.5506539 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
```

```
  4.01704      2.37959     -5.28310
Posterior Standard Deviations
 0.287806     0.169905     0.743151
Posterior Correlations
         alpha    beta
beta      -0.5747
sigma*2    0.0557   0.0013

*****   2.1%**********************************************************************
Iteration Number   9
                                                                        Changes:
                                                                        New grid
          Parameter space is partitioned into sub-spaces :
          Cartesian product rule on              3 dimensions
          Spherical rule on                      0     "
          Monte-Carlo rule on                    0     "
          giving integration over                3     "
          for a problem with                     3 parameters
          Cartesian product grid size is :   7 X   7 X   7
          A linear transformation is applied to   3-1 parameters
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta___1    sigma*22
------------------------------------------------------------------------------
p(::) =   2.6581454 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.02128      2.37951     -5.32822
Posterior Standard Deviations
 0.284116     0.167621     0.781747
Posterior Correlations
         alpha    beta
beta      -0.5749
sigma*2    0.0479   0.0020

*****   6.4%**********************************************************************
Iteration Number  10
------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta___1    sigma*22
------------------------------------------------------------------------------
p(::) =   2.6542192 E    3
Parameter Set
 alpha        beta         sigma*2
Posterior Means
  4.02131      2.37951     -5.32808
Posterior Standard Deviations
 0.293898     0.167460     0.782388
Posterior Correlations
         alpha    beta
beta      -0.5750
sigma*2    0.0463   0.0020

*****   0.2%**********************************************************************
Iteration Number  11
                                                                        Changes:
                                                                        New grid
          Parameter space is partitioned into sub-spaces :
          Cartesian product rule on              3 dimensions
          Spherical rule on                      0     "
          Monte-Carlo rule on                    0     "
          giving integration over                3     "
          for a problem with                     3 parameters
```

```
          Cartesian product grid size is :    8 X    8 X    8
          A linear transformation is applied to    3-1 parameters
-------------------------------------------------------------------------------
Linear transformation used :
Operational parameters are :
 alpha       beta___1    sigma*22
-------------------------------------------------------------------------------
p(x) =    2.5999045 E     3
Parameter Set
 alpha         beta          sigma*2
Posterior Means
  4.01953       2.37955      -5.29503
Posterior Standard Deviations
 0.292821      0.172676      0.772798
Posterior Correlations
          alpha     beta
beta       -0.5742
sigma*2     0.0640   0.0017

*****   4.2%*************************************************************
Iteration Number  12
                                                                      QUIT
                                    Total time       Elapsed  (min:sec)  30:11.00
                                                        CPU (sec)       2.00
                                    Integration time Elapsed  (min:sec)   1: 0.00
                                                        CPU (sec)       0.00
                                                 TUE, 21 MAR 1989 @ 16:11:14
```

2.4. Numerical prediction for the two parameter Weibull distribution

2.4.1 Introduction

In order to demonstrate how the numerical integration strategy can be applied in conjuction with analytic integration over a component of the parameter space we consider as an example the two parameter Weibull distribution. Over recent years, this distribution has become one of the most widely used lifetime models. Moreover, the Bayesian approach to problems involving Weibull distributions has been applied in many recent papers. Soland (1969), Cavanos and Tsokos (1973), Papadopoulos and Tsokos (1976) and Evans and Nigm (1980) all used Bayesian methods in their analysis. More recently, Chen, Hill and Greenhouse (1985) use Bayesian methods in their analysis of survival data on Cancer patients. The papers by Achcar (1984) and Achcar, Brookmeyer and Hunter (1985) describe a Bayesian analysis, based on a Weibull model, applied to medical follow-up studies. Smith and Naylor (1987) compare maximum likelihood estimation with Bayesian estimates for the parameters of the Weibull distribution.

We consider here the two parameter Weibull distribution with p.d.f.

$$p(x/\theta) = \theta_1\theta_2 x^{\theta_2-1} e^{-\theta_1 x} \quad , \; x > 0 \; .$$

Given lifetest data, $\underset{\sim}{x}$, on n items with r failures at times $x_1, x_2, \ldots, x_r$ and n-r right censored observations at times $x_{r+1}, x_{r+2}, \ldots, x_n$ the likelihood function is given by

$$\ell(\underset{\sim}{\theta}/\underset{\sim}{x}) \propto \theta_1^{r}\theta_2^{r}\left[\prod_{i=1}^{r} x_i\right]^{\theta_2} e^{-\theta_1\left[\sum_{i=1}^{n} x_i^{\theta_2}\right]}, \quad \theta_1,\theta_2>0 \; . \qquad (2.8)$$

This form of likelihood is valid providing the censoring mechanism is independent and non-informative (see Kalbfleisch and Prentice (1980), Ch.5). Given a prior distribution on $\underset{\sim}{\theta} = (\theta_1,\theta_2)$ we consider the numerical evaluation of posterior expectations required in a wide range of practical applications. In particular in section 2.4.3 we consider the numerical evaluation of prediction bounds for future lifetimes, whilst in section 4 we consider the posterior distribution of median lifetime. These posterior expectations are more meaningful to practitioners than the moments of θ_1 and θ_2.

Details of the numerical method are presented in section 2.4.2. This method assumes that integration over the scale parameter, θ_1, can be performed analytically. Numerical integration of the shape parameter, θ_2, then leads to a convenient representation of posterior expectations in terms of expectations with respect to a discrete distribution over θ_2. This discrete distribution is obtained by application of the method of Naylor and Smith (see section 2.3).

2.4.2 Evaluation of Posterior Expectations

In general suppose that we need the posterior expectation of $q(\underset{\sim}{\theta})$

$$E[q(\underset{\sim}{\theta})/\underset{\approx}{x}] = \frac{\int_0^\infty d\theta_2 \int_0^\infty \ell(\underset{\sim}{\theta}/\underset{\approx}{x})p(\underset{\sim}{\theta})q(\underset{\sim}{\theta})d\theta_1}{\int_0^\infty d\theta_2 \int_0^\infty \ell(\underset{\sim}{\theta}/\underset{\approx}{x}).p(\underset{\sim}{\theta}).d\theta_1} . \qquad (2.9)$$

For example, with

$$q(\underset{\sim}{\theta}) = e^{-M\theta_1 y^{\theta_2}} \qquad (2.10)$$

we have the predictive probability that the minimum of M future lifetimes exceeds y. With

$$q(\underset{\sim}{\theta}) = (\ell n 2/\theta_1)^{1/\theta_2} \qquad (2.11)$$

we have the posterior density of median lifetime.

For a useful range of priors many of the posterior expectations (2.9) required in practical applications can be expressed in the form

$$E[q(\underset{\sim}{\theta})/\underset{\approx}{x}] = \frac{\int_0^\infty E[q(\underset{\sim}{\theta})/\theta_2, \underset{\approx}{x}].\ell^*(\theta_2/\underset{\approx}{x}).p(\theta_2)d\theta_2}{\int_0^\infty \ell^*(\theta_2/\underset{\approx}{x})p(\theta_2)d\theta_2} \qquad (2.12)$$

where the conditional posterior expectation

$$E[q(\underset{\sim}{\theta})/\theta_2, \underset{\approx}{x}] = \int_0^\infty E[q(\underset{\sim}{\theta})/\theta_1, \theta_2].p(\theta_1/\theta_2, x)d\theta_1 \qquad (2.13)$$

and the integrated *likelihood* x *prior*

$$\ell^*(\theta_2/\underset{\sim}{x}) = \int_0^\infty \ell(\underline{\theta}/\underset{\sim}{x})p(\theta_1/\theta_2)d\theta_1 \qquad (2.14)$$

can be obtained analytically.

The expression (2.12) can then be treated numerically as a posterior expectation in one dimension. Indeed $\ell^*(\theta_2/\underset{\sim}{x})$, being an integrated *likelihood* x *prior*, will be asymptotically normal in form which means that the Gauss-Hermite formulae should provide efficient approximations.

Soland (1969) effectively uses the representation (2.12) with a discretisation of $p(\theta_2)$ to approximate the integral. This method was applied by Evans and Nigm (1980) for the approximation of lower prediction bounds on the minimum of M future lifetimes. The difficulty with Soland's method is the specification of an appropriate discretisation. The search for a suitable choice may be very time consuming. An attractive feature of the method, however, is that, like the application of Gauss-Hermite rules, it leads to an expectation with respect to a discrete approximation to the posterior distribution of θ_2, so that (2.12) is approximated by the sum

$$\sum_{i=1}^{S} P_i \; E[q(\underline{\theta})/\theta_2 = \theta_{2i}, \underset{\sim}{x}] \qquad (2.15)$$

where $P_i > 0$ $(i=1,2,\ldots,5)$ and $\sum_{i=1}^{S} P_i = 1$.

The same abscissas, θ_{2i}, and weights, P_i, can then be used repeatedly for different $q(\underset{\sim}{\theta})$'s.

Our approach, is to apply the method of Naylor and Smith after a log transformation to improve the approximate normality and avoid wastage caused by negative points. Putting $\varphi = \ell n\theta$ in (2.12), yields

$$E[q(\underset{\sim}{\theta})/\underset{\sim}{x}] = \frac{\int_{\infty}^{\infty} E[q(\underset{\sim}{\theta})/\theta_2 = e^{\varphi}, \underset{\sim}{x}] \cdot \ell^*(\theta_2 = e^{\varphi}/\underset{\sim}{x}) \cdot P(\theta_2 = e^{\varphi}) e^{\varphi} d\varphi}{\int_{\infty}^{\infty} e^*(\theta_2 = e^{\varphi}/\underset{\sim}{x}) P(\theta_2 = e^{\varphi}) e^{\varphi} d\varphi} \qquad (2.16)$$

Application of the n-point Gauss-Hermite formulae to (2.16) then yields an approximation of the form (2.15) with

$$\theta_{2i} = e^{\mu + \sqrt{2}.\sigma.t_i}$$

and

$$P_i = \frac{\sqrt{2}.\sigma.w_i \; e^{t_i^2 + \theta_{2i}} . \ell^*(\theta_{2i}/\underset{\sim}{x}) . p(\theta_{2i})}{\sum_{i=1}^{n} \sqrt{2}.\sigma.w_i \; e^{t_i^2 + \theta_{2i}} . \ell^*(\theta_{2i}/\underset{\sim}{x}) . p(\theta_{2i})}$$

where t_i and w_i are the abscissas and weights of the Gauss-Hermite formulae. As with Soland's method, the θ_{2i}'s and P_i's can be used

repeatedly with different $q(\underset{\sim}{\theta})$'s.

2.4.3 Prediction Bounds for Future Lifetimes

In this section we consider the problem of computing prediction bounds for future lifetimes. Suppose we have lifetest data, $\underset{\sim}{x}$, giving a likelihood of the form (2.8) and that there are M future lifetimes, $Y_1,\ldots,Y_M$, following Evans and Nigm (1980), we consider the evaluation of the lower $100\gamma\%$ prediction bound for the shortest lifetime $Y_{(1)}$. In reliability analysis this problem occurs with series system of M identical components with lifetimes $Y_1,\ldots,Y_M$. $Y_{(1)}$ then represents the life of the system. To illustrate the methodology we consider the three examples in Table 2.3; originally from Lawless (1973), which featured in the paper of Evans and Nigm (1980).

Table 2.3

The three examples considered by Lawless (1973).

Example	Future batch size (M)	γ	Test batch size	Failure times	Censoring times
1	40	0.9	10	50.5, 71.3, 84.6, 98.7, 103.8	103.8
2	100	0.9	23	17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.48, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40	-
3	500	0.8	3	45.952, 54.143, 65.440	-

In these examples predictions are made dangerously far beyond the domain of previous experience. In the most extreme case, example 3, predictions about the minimum of 500 lifetimes are made on the basis of a sample of size 3! As far as the application of the Gauss-Hermite formula is concerned, these examples are ill-conditioned in the sense that the integrands required are dominated by $q(\underline{\theta}/\theta_2, \underset{\sim}{x})$ rather than the integrated likelihood x prior $\ell^*(\theta_2/\underset{\sim}{x})$. This difficulty is exacerbated by the restrictions on the range of the prior distribution of θ_2. Thus, although these examples are rather unrealistic and involve very extreme predictions, they should serve as a good test of the Gauss-Hermite formulae as applied here.

Let y_γ be the lower $\gamma 100\%$ prediction bound for the minimum $y_{(1)}$ of M future lifetimes. The problem of evaluating y_γ is equivalent with solving the equation

$$P[y > y_\gamma/\underset{\sim}{x}] = \gamma \qquad (2.17)$$

Following Evans and Nigm (1980) and taking a prior of the form

$$P(\theta_1/\theta_2) \propto \theta_1^{-1}$$

we have

$$P(\theta_1/\theta_2, \underset{\sim}{x}) = \theta_1^{r-1} \exp\left[-\theta_1 \sum_{i=1}^{n} x_i^{\theta_2}\right] \left[\sum_{i=1}^{n} x_i^{\theta_2}\right]^r / \Gamma(r)$$

Also

$$E[P[y>y_\gamma/\underset{\sim}{x},\theta_2]] = \int_0^\infty P[y>y_\gamma/\theta_1,\theta_2].p[\theta_1/\theta_2,\underset{\sim}{x}]d\theta_1$$

$$= \left[\frac{\Sigma x_i^{\theta_2}}{\Sigma x_i^{\theta_2}+My_\gamma^{\theta_2}}\right]^r . \qquad (2.18)$$

The approximate solution to (2.17) can be found iteratively by taking the expectation of (2.18) with respect to the discrete approximation obtained using the Gauss-Hermite formula. Priors (i) and (ii), below, of Evans and Nigm (1980) were used.

(i) $P(\theta_2) = \frac{1}{25}$, $0 < \theta_2 < 25$

(ii) $P(\theta_2) \alpha \theta_2^9 e^{-\theta_2}$, $0 < \theta_2 < 25$

The maximum likelihood estimate of θ_2 and its asymptotic variance were used as starting values for the adaptive integration strategy. Although in practice the 'normal' procedure would be to apply the strategy to approximate the posterior mean and variance and then solve the equation (2.17) using the final approximation, we present the solution to (2.17) at each stage in the strategy. In particular we applied a four point rule and continued updating μ and σ to convergence and then solved 2.12 iteratively. The results are presented in table 2.4.

Table 2.4

Approximate lower prediction bounds and overall changes (Δ) for the three examples of Lawless (1974)

Example 1

	Prior (i)		Prior (ii)	
POINTS	PREDICTION BOUND	Δ	PREDICTION BOUND	Δ
4	0.228263E+02	3.68E-02	0.395505E+02	5.36E-03
5	0.225832E+02	4.86E-02	0.395046E+02	2.07E-02
6	0.225008E+02	6.65E-03	0.394911E+02	1.01E-03
7	0.225004E+02	1.81E-04	0.394911E+02	1.23E-03
8	0.224676E+02	2.91E-03	0.394872E+02	2.95E-04
9	0.224674E+02	2.12E-05	"	
10	0.224624E+02	9.14E-04	"	
11	0.224624E+02	1.68E-06	"	
12	0.224610E+02	1.07E-04	"	
13	0.224610E+02	2.17E-08	"	
14	0.224606E+02	6.67E-05	"	
15	0.224606E+02	2.89E-09	"	
16	0.224605E+02	7.89E-06	"	
17	0.224605E+02	4.47E-10	"	
18	0.224605E+02	2.81E-14	"	
19	0.224605E+02	2.57E-14	"	
20	0.224604E+02	1.64E-06	"	
SOLAND'S	0.22462E+02		0.39283E+02	(25 points)
EXACT	0.22460E+02		0.39487E+02	

Example 2

	Prior (i)		Prior (ii)	
POINTS	PREDICTION BOUND	Δ	PREDICTION BOUND	Δ
4	0.25759E+0	2.80E-03	0.43851E+01	1.41E-03
5	0.25687E+01	1.91E-02	0.43790E+01	1.31E-02
6	0.25668E+01	8.16E-04	0.43775E+01	4.03E-04
7	0.25668E+01	1.09E-05	0.43775E+01	3.40E-06
8	0.25662E+01	7.31E-04	0.43771E+01	3.46E-04
SOLAND'S	0.2357E+01		0.5313E+01	(25 points)
EXACT	0.2566E+01		0.4377E+01	
			(0.4357E+01)	

Example 3

	Prior (i)		Prior (ii)	
POINTS	PREDICTION BOUND	Δ	PREDICTION BOUND	Δ
4	0.192434E+02	5.25E-02	0.222142E+02	2.72E-02
5	0.180045E+02	1.20E-02	0.220889E+02	2.38E-02
6	0.188687E+02	4.27E-02	0.221707E+02	1.41E-03
8	0.185277E+02	2.87E-02	0.221707E+02	5.94E-04
10	0.183622E+02	7.20E-02	0.221357E+02	6.43E-05
12	-	0.123020	0.221352E+02	1.68E-05
12	0.183999E+02	5.30E-03		
14	0.184148E+02	9.34E-03	0.221356E+02	3.08E-04
16	0.184266E+02	1.67E-02	"	
20	0.184148E+02	5.20E-02	"	
24	0.184429E+02	5.71E-02	"	
25	0.184448E+02	4.67E-03	"	
SOLAND'S	0.18433E+02		0.22120E+02	(with 25 points)
EXACT	0.18433E+02		0.22136E+02	

The aggregate measure Δ has been defined in section 2.3 and it is given at the end of each iteration of BAYESFOUR. The exact results given by Evans and Nigm were verified using the Legendre method with 64 points - in example 2 (prior (ii)) a typing error must have occured (.4357 instead of .4377) - the results indicate that the Naylor-Smith method applied here is faster and more efficient than Soland's method, except in the case of the third example (with prior (i)) where Soland's method performs slightly better than the iterative algorithm of Naylor and Smith.

We recall here that the above algorithm requires satisfactory convergence of the normalising constant and the posterior mean and variance, expressed by Δ, within each grid size, and subsequent convergence between the grid sizes before the completion of the iterative process and the calculation of the prediction bounds. With only one exception, in example 3, we always moved to to the next grid size, as suggested from the small size of Δ. In section 2.5.3.2 we

will re-examine the same Weibull examples and we will argue that to update of the maximum likelihood estimates as applied in this section is probably not always the best strategy to be adopted.

2.4.4. Posterior Distribution of Median Lifetime

In many practical situations it is difficult to attach a meaningful interpretation to the parameters θ_1 and θ_2 of the Weibull distribution and for practical purposes it is often desirable to focus on some function of θ_1 and θ_2 which has a meaningful interpretation. Quantiles of the distribution and, in particular, the median are analytically convenient. Achcar (1984), in a medical application, focuses on the median lifetime in a study of survival data on 38 cancer patients. These data are given in Table 2.5.

Table 2.5

Survival times (days) of 38 patients.

FAILURES								CENSORED			
182	81	64	216	374	216	227	237	799	786	754	723
229	264	97	53	361	214	158	75	661	600	561	527
62	147	146	130	67	87	169	201				
510	543	38	18	15	193						

Achcar (1984) uses a log Normal approximation to the posterior distribution of median survival time. The shape parameter of the Weibull distribution being obtained by taking the mode from its

marginal posterior density. We use the Gauss-Hermite method to derive the posterior distribution of the median survival time assuming a non-informative prior for the shape parameter. The two different approximations are compared graphically.

Following Achcar (1984) and taking a locally uniform prior for θ_1, the conditional posterior distribution of θ_1 given θ_2 is

$$P(\theta_1/\theta_2,\underset{\sim}{x}) = \frac{\theta_1^{r}\exp(-\theta_1 \sum_{i=1}^{n} x_i^{\theta_2}).(\sum_{i=1}^{n} x_i^{\theta_2})^{r+1}}{\Gamma(r+1)} .$$

The median survival time m, take

$$m = (\ell n 2/\theta_1)^{1/\theta_2} ,$$

then has the conditional posterior distribution

$$p(m/\theta_2,\underset{\sim}{x}) = S^{r+1}\,\frac{\theta 2}{m}\,.\,\exp(-S)\,/\,r!$$

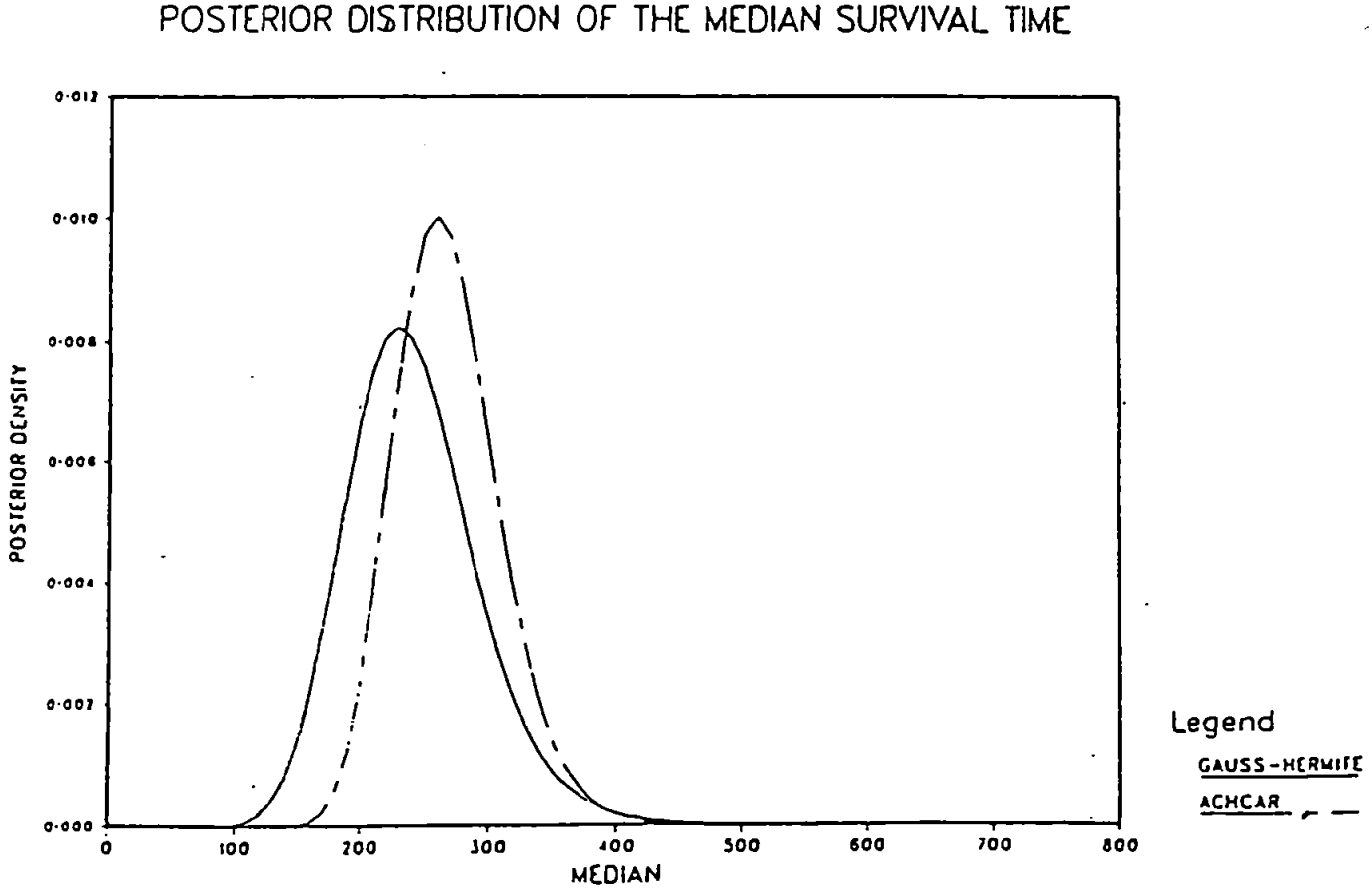
where $S = \ell n 2 \,.\, \sum_{i=1}^{n} \left[\frac{x_i}{m}\right]^{\theta_2}$

Taking the prior $p(\theta_2) \propto 1/\theta_2$ the posterior distribution of median m is approximated by the sum (2.14). This posterior density, together with the approximation given in Achcar (1984) is shown in Figure 2.3.

We can see that the approximation with Gauss-Hermite method gives more "pessimistic" results, in the sense that the lower tail area of Gauss-Hermite approximated density is heavier; for example, using the

Gauss-Hermite formulae the tail area below 200 is 0.21, using the lognormal approximation this tail area is 0.003.

FIGURE 2.3



2.5. Performance of the Gauss-Hermite integration rule

2.5.1 Introduction

It is common to assess numerical methods according to three criteria:

(i) reliability: how often is the method successful

(ii) efficiency: how much effort (computer time) was required to produce the result.

(iii) accuracy: how close is the computed answer to the true answer

In this section, we will try to assess the strategy of Naylor and Smith using the above criteria. In particular, the reliability will be connected with the sensitivity to kurtosis and to the robustness of pertubations in centering and scaling; the efficiency to the wastage of function evaluations and to the effect of the mis$\overset{s}{\underset{\wedge}{}}$pecification in the mean and variance.

2.5.2 Reliability

2.5.2.1 Theoretical background

Laurie (1985) notes that to define reliability, one requires a definition of success and a measurable set of problems. Let us assume that the Naylor and Smith algorithm is succesful if the posterior kernel can be approximated by a polynomial x Normal. This, in turn

implies that convergence is achieved in a m-point quadrature rule Q_m in (2.1), hopefully with m not very large.

Before we proceed to the set of problems to test reliability, we need to mention the vital role of the initial transformation which can make the posterior kernel to be close to Normality. See Smith *et al.* (1985,1987) Shaw (1988, chapter 10) and Hills (1989). Thus, a 'badly behaved' example can be transformed to a 'well behaved' one using a suitable transformation.

Keeping in mind the previous point, we shall try to sketch the set of problems which can indicate whether we may or may not expect success in the Naylor and Smith method. We may say that a posterior kernel is 'well behaved' if it is unimodal, not too skew, continous, with light tails. On the other hand, the posterior kernel is 'badly behaved' if the parameter space is restricted and this leads to 'wastage' of integration points in regions of zero posterior density; or if the posterior density is multimodal. This may occur, for example, in mixture models (see Titterington *et. al.* (1985)) or when the prior and the likelihood are in conflict; or if any other awkward situations occur, for example if the posterior variance does not exist.

Of course, none of the above criteria alone guarantee 'good' or 'bad' behaviour of the kernel. However, they serve as possible kernel features which might indicate possible success or failure of the method.

We proceed by describing the work presented in a Ph.D. thesis by Shaw (1988) concerning the convergence of the Naylor and Smith adaptive

strategy. Shaw illustrated this strategy using a number of case studies but also experimentised with some artificial examples to draw conclusions about its behaviour. His main results are outlined below.

For a given posterior density $p(\underline{\theta})$ where $\underline{\theta}$ denotes a d-dimensional vector of parameters, the steps (iii) to (v) in figure 2.1 define an iterated map T: $X \to X$ where $X \subseteq R^d \times (0,\infty)^d \times (-1,1)^{d(d-1)/2}$ corresponding to d means, d variances and d(d-1)/2 correlations estimated in (v). If the map T for an n-point Gauss quadrature rule Q_m in (2.1) is denoted by T_n, the Naylor and Smith strategy suggests to iterate using T_n up to the point where $T_n(\underline{\mu},\underline{\Sigma}) = (\underline{\mu},\underline{\Sigma})$. Then, we proceed to another map T_m, $m>n$, and check whether $T_n(\underline{\mu},\underline{\Sigma}) = (\underline{\mu},\underline{\Sigma}) = T_m(\underline{\mu},\underline{\Sigma})$.

Shaw notes that a fixed point $(\underline{\mu}_0,\underline{\Sigma}_0)$ can be stable if for all points $(\underline{\mu},\underline{\Sigma})$ in some neighborhood of $(\underline{\mu}_0,\underline{\Sigma}_0)$ $\lim T^k(\underline{\mu},\underline{\Sigma}) = (\underline{\mu}_0,\underline{\Sigma}_0)$ as $k \to \infty$. Otherwise $(\underline{\mu}_0,\underline{\Sigma}_0)$ is unstable. A point $(\underline{\mu},\underline{\Sigma})$ is periodic if $T^k(\underline{\mu},\underline{\Sigma}) = (\underline{\mu},\underline{\Sigma})$ for some integer $k>0$.

In practice, for a Gauss quadrature rule $Q_m f$, we may have any one of the above situations (stable, unstable or periodic point) or any combination of them. There is also the chance of having a chaotic behaviour without any fixed points. Thus, the Naylor and Smith algorithm can converge, diverge, or converge to a limit cycle rather to a point. However, by increasing m, we expect to have a single fixed point which will converge to the true value of $(\underline{\mu},\underline{\Sigma})$.

As far as the one-dimensional case is concerned, Shaw shows that if the posterior density $p(\underline{\theta})$ is log-concave and proper, that is if

$\log p(\alpha\theta_1+(1-\alpha)\theta_2) > \alpha\log p(\theta_1) + (1-\alpha)\log p(\theta_2)$ for all $\theta_1,\theta_2 \epsilon R$, $\alpha\epsilon(0,1)$

and

$$\int_{-\infty}^{\infty} p(\theta)d\theta < \infty$$

then the map $T^3(\mu,\sigma)$ has exactly one fixed point, which in turn implies that we expect good behaviour in unimodal densities with moderately light tails.

2.5.2.2 Sensitivity to Kurtosis

Given a distribution with mean μ and variance σ^2, the standard fourth moment coefficient of kurtosis is given by

$$\mu_4 = \frac{E(x-\mu)^2}{\sigma^4} - 3 \ .$$

This is often regarded as a measure of tail heaviness of a distribution relative to that of a normal distribution or as a measure of peakedness near the centre of a distribution.

A class of exponential power distributions can be written in the general form

$$p(x|\theta,\varphi,\beta) = k\varphi^{-1}\exp\left[-\tfrac{1}{2} \left| \frac{x-\theta}{\varphi} \right|^{2/(1+\beta)} \right] , \ -\infty < x < \infty \quad (2.19)$$

where

$$k^{-1} = \Gamma\left[1 + \frac{1+\beta}{2}\right] 2^{1+\frac{1}{2}(1+\beta)} \quad , \varphi > 0, \ -m < \theta < \infty, \ -1 < \beta \leq 1 \ .$$

This is a member of a class of symmetric distributions which includes Normal (β=0) together with other distributions with various values of the coefficient of kurtosis. (See Box and Tiao (1973), p156-160). The coefficient of kurtosis for a given value of β is given by

$$\mu_4 = \frac{\Gamma\left[\frac{5}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(1+\beta)\right]}{\left\{\Gamma\left[\frac{3}{2}(1+\beta)\right]\right\}^2} - 3 \qquad (2.20)$$

and can take values from -1.2 (β=-1) to 3 (β=1).

The class (2.19) was used here to investigate the performance of 1-D Gauss-Hermite integration rules as described in section 2.2.1, over a range of values of kurtosis. Varying the values of β in (2.20) we integrated (2.19) using Gauss-Hermite method with 4, 6 and 8 points. Given the exact value of the integrand, c, and the approximation for each method, a, the relative error

$$r = \frac{c-a}{c} \qquad (2.21)$$

of the normalising constant and the posterior variance were calculated and plotted against the coefficient of kurtosis as given in (2.20).

The results are illustrated in figures 2.4 and 2.5. In each case, the correct initial values were given to the mean and variance in (2.17). The results indicate that with adequately large grid the method

FIGURE 2.4

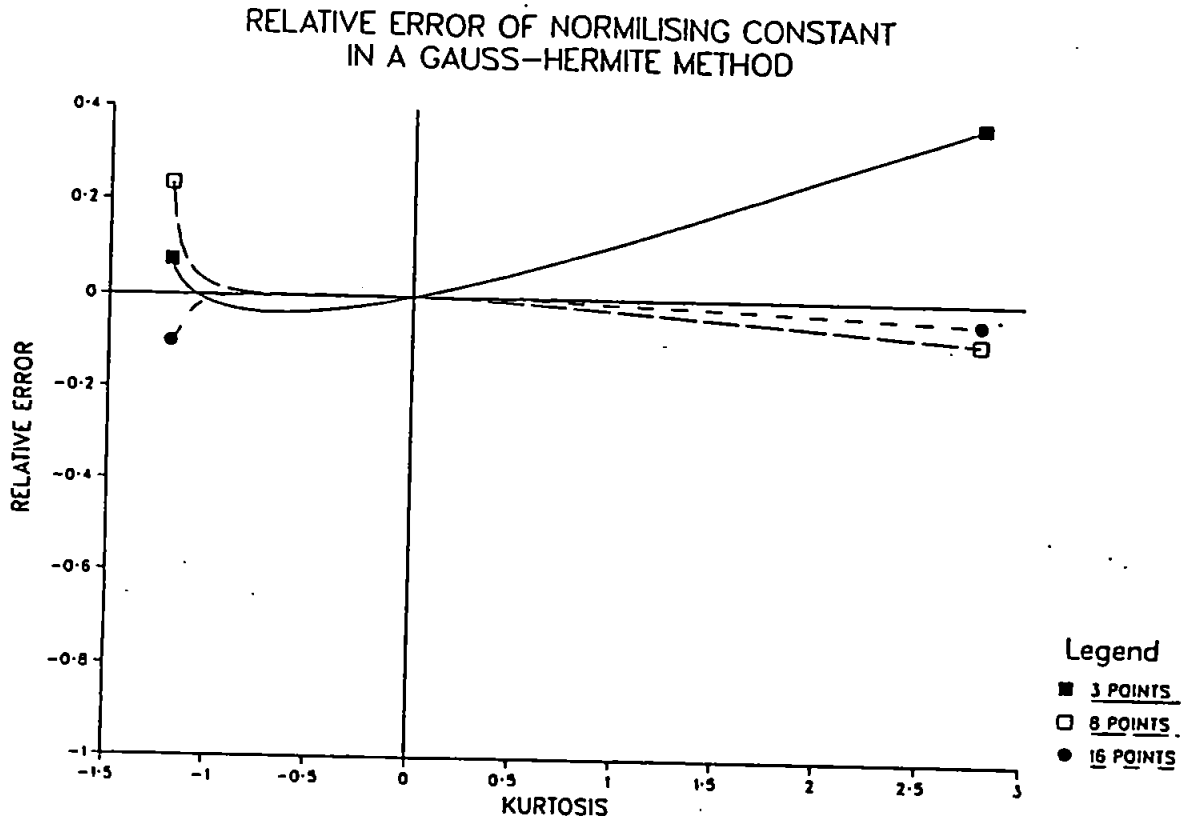
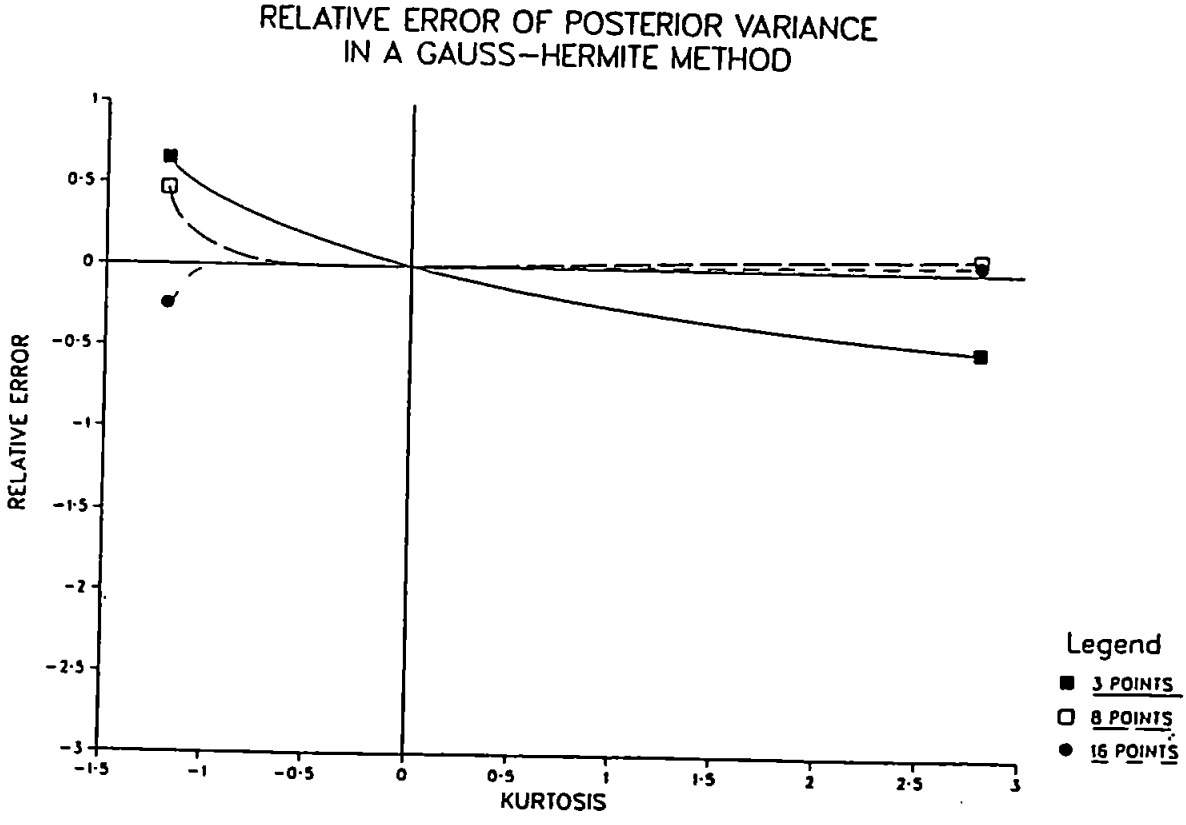


FIGURE 2.5



performs very well over a quite large range of coefficients of kurtosis. It is worth noting the sensitivity of the method when using a 3-point grid, where with $\mu_4=-1$ produces a very good approximation to the normalising constant, much better than the 8-point formula. This is of course justified because three points are too few especially for a density extremely flat around its center. This fact might also be the reason that Naylor and Shaw (1985) suggest an initial grid with 4 points in BAYESFOUR.

2.5.3 Efficiency

2.5.3.1 Choice of scaling and number of nodes

The efficiency of the quadrature is measured in terms of the number of function evaluations required. Occasionally the processing time is used instead, and, as Davis and Rabinowitz (1984,p.423) comment, each measurement has its disadvantage; the former does not take into account the auxiliary computations included in the program and the latter is machine dependent.

The method introduced by Naylor and Smith (1982) discards all previous function evaluations when moving from one iteration to another. This, in general is considered a serious drawback, especially recently where the adaptive integrators have become very popular. See C.de Boor (1971) Piessens *et. al.*(1983) and Elhay and Kautsky (1987).

Let us now assume that for a map T^n there exists a stable point. This

stable point is not the true value of the vector (μ,Σ) in as much as $Q_n f$ is only an approximation to If. In particular, in section 2.5.3.3 we shall examine an example in which the map $T^{32}(\mu,\sigma)$ was initialised with the correct value (0,3) to converge to the wrong value (0,2.55). Therefore, the 'small grid' first iterations may move the vector (μ,Σ) away from its true value, and in addition all computer labour is spent without any feedback. We feel that the important question to be set is whether or not scaling is that important to 'deserve' all this labour.

In the discussion of the paper by Kass *et.al.* (1988) Naylor points out that *Succesful use of Gauss-Hermite quadrature rules depends more on choice of scaling than on number of points.* In the same discussion, however, Shaw notes that *Gauss Hermite integration is actually quite robust to pertubations in centering and scaling.*

In general, this problem depends on the nature of the integrand. Our experience indicates that , especially when the variance is overestimated the Naylor and Smith method is quite robust. We also believe that iterations within small grids decrease the efficiency of the algorithm. We note that the decrease in efficiency occurs, not only due to the wasted function evaluations, but also because of the user's time wasted through moving through different menus to alter the quadrature rule.

In chapter 5 we propose a strategy which we believe increases the efficiency of the Naylor and Smith strategy and at the same time keeps a more flexible option as far as the need of scaling is concerned.

2.5.3.2 The Weibull example revisited

In section 2.4 we considered the two parameter Weibull distribution to demonstrate the application of the numerical integration strategy of Naylor and Smith (1982) in conjuction with analytic integration over the scale parameter θ_1 in (2.8). In relation to our comments in section 2.5.3.1 concerning the question of the use of rescaling and recentering, we re-analyse here the same examples keeping the mean and the variance constant over all grid sizes. These values were taken as the maximum likelihood estimates. Table 2.6 presents the results using the same priors as in 2.4.3

Table 2.6

Approximate lower prediction bounds for the three examples of Lawless (1977)

Example 1

METHOD	PRIOR (i)	PRIOR (ii)
G-H with 2 points	0.224416E+02	0.397863E+02
G-H with 3 points	0.227451E+02	0.395346E+02
G-H with 4 points	0.224783E+02	0.395025E+02
G-H with 5 points	0.224932E+02	0.394930E+02
G-H with 6 points	0.224653E+02	0.394872E+02
G-H with 7 points	0.224653E+02	"
G-H with 8 points	0.224642E+02	"
G-H with 9 points	0.224642E+02	"
G-H with 10 points	0.224614E+02	"
G-H with 11 points	0.224614E+02	"
G-H with 12 points	0.224605E+02	"
G-H with 13 points	0.224605E+02	"
G-H with 14 points	0.224604E+02	"
SOLAND'S with 25 points	0.22462E+02	0.39283E+02
EXACT	0.22460E+02	0.39487E+02

Example 2

METHOD	PRIOR (1)	PRIOR (2)
G-H with 2 points	0.26668E+01	0.44850E+01
G-H with 3 points	0.25789E+01	0.43890E+01
G-H with 4 points	0.25718E+01	0.43820E+01
G-H with 5 points	0.25674E+01	0.43781E+01
G-H with 6 points	0.25664E+01	0.43773E+01
SOLAND'S with 25 points	0.2357E+01	0.5313E+01
EXACT	0.2566E+01	0.4377E+01
		(0.4357E+01)

Example 3

METHOD	PRIOR (1)	PRIOR (2)
G-H with 2 points	0.164749E+02	0.218449E+02
G-H with 3 points	0.203372E+02	0.223968E+02
G-H with 4 points	0.176519E+02	0.219863E+02
G-H with 5 points	0.192492E+02	0.222143E+02
G-H with 6 points	0.182016E+02	0.221005E+02
G-H with 8 points	0.184074E+02	0.221287E+02
G-H with 10 points	0.185011E+02	0.221352E+02
G-H with 12 points	0.185060E+02	0.221363E+02
G-H with 14 points	0.184856E+02	"
G-H with 16 points	0.184549E+02	"
G-H with 20 points	0.184536E+02	"
G-H with 24 points	0.184380E+02	"
G-H with 25 points	0.184380E+02	"
G-H with 64 points	0.184417E+02	"
SOLAND'S with 25 points	0.18433E+02	0.22120E+02
EXACT	0.18433E+02	0.22136E+02

Comparing the results with them of table 2.4, we can see that rescaling and recentering does not increase the efficiency of the Gauss Hermite integration rule. In fact, the results in table 2.6 are slightly better in some cases. Of course, these one dimensional examples cannot provide a definite conclusion, but, they can serve as an indication to support the robustness of the integration rules to perbutation of mean and variance. Our belief is, and it will be emphasised again in chapter 5, that in 'well-behaved' kernels the maximum likelihood estimates provide good choices for use in Gauss-Hermite integration rules and repeated change of their values results in loss of efficiency.

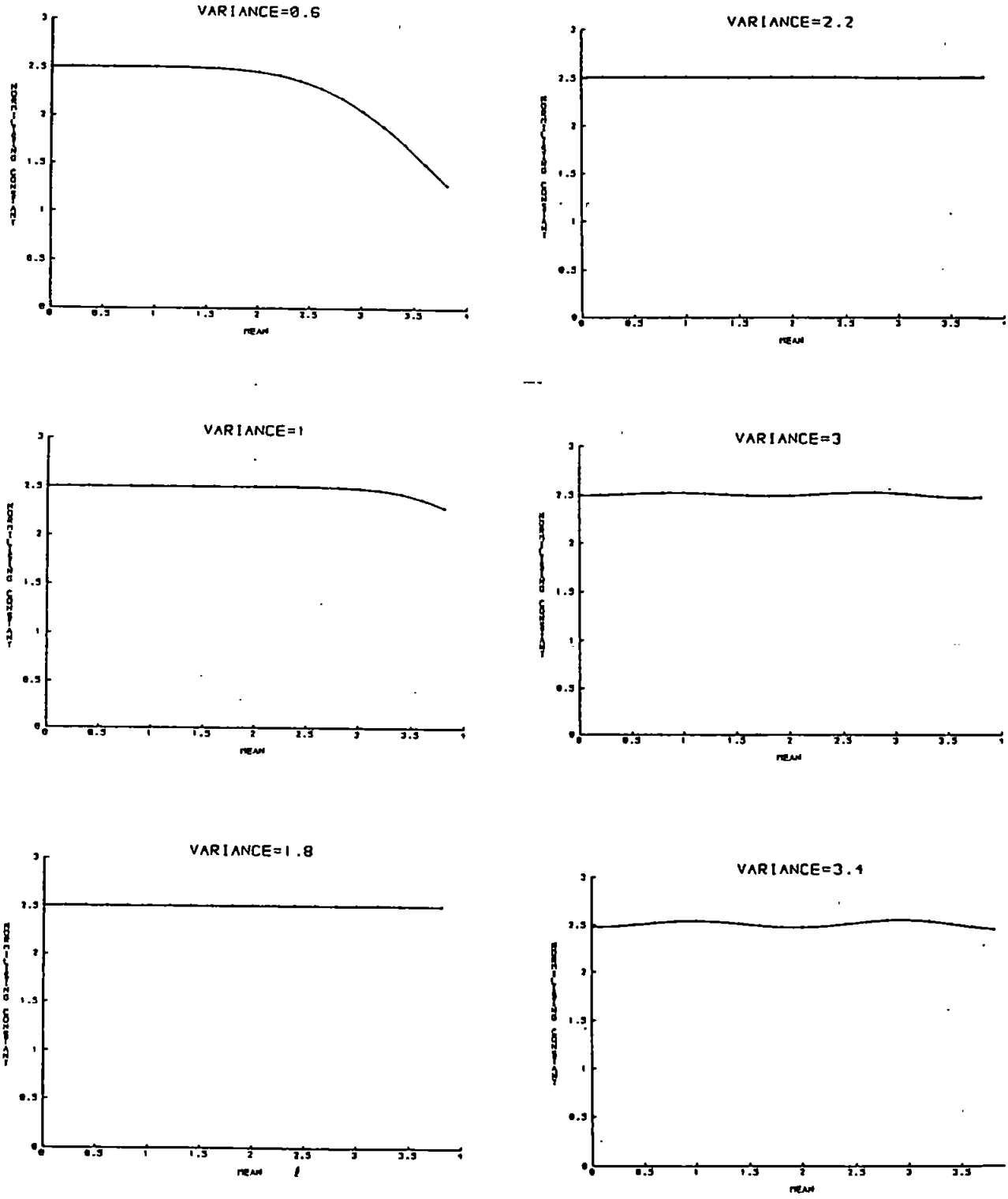
2.5.3.3 Misspecification of mean and variance

As pointed out in section 2.2.1, one of the key features of the iterative strategy is the initial specification of the mean and the variance in (2.6). Moreover, if a linear transformation is to be applied in higher dimensional cases, a covariance matrix could be used initially for the stage (iii) in figure 2.1. In cases where the maximum likelihood estimates are available, they offer very good initial starting values. See section 2.4 for an one-dimensional example where the initial maximum likelihood estimates achieve very good approximations to normality. There are however cases where the maximum likelihood estimates are not readily derived and we need to start the iteration strategy with some first crude approximations. We performed some experiments to test the sensitivity of the approach in such cases.

Initially, we tried integrating $g(x)=\exp(-x^2/2)$ with an 8-point Gauss Hermite rule varying the mean and the variance. When the correct value of the mean and variance are used, the proper value of the integral, ($\sqrt{2\pi}=2.5066$) must be approximated exactly. In figure 2.6 the different graphs illustrate the behaviour of the normalising constant for a given variance when the mean is misspecified. It is evident that in cases of uncertainty about the variance, a larger value is preferable. This remark has been also made by Naylor (1982).

A typical example of heavy tailed distribution is a t-distribution with three degrees of freedom. In our next artificial example, we tested

FIGURE 2.6: Mispecification of mean and variance



True values: Normalising constant 2.5066
Mean 0
Variance 1

Iterative method with 8-point formula

VARIANCE	POS.VARIANCE	NOR.CONSTANT	POS.MEAN
0.3000000000D+01	0.2441655153D+01	0.2605811122D+01	-0.8215650382D-14
0.2441655153D+01	0.2290335991D+01	0.2647776824D+01	-0.1243449788D-13
0.2290335991D+01	0.2248590485D+01	0.2657023085D+01	-0.1199040867D-13
0.2248590485D+01	0.2236960980D+01	0.2659388240D+01	-0.1731947918D-13
0.2236960980D+01	0.2233710846D+01	0.2660032227D+01	-0.4440892099D-15
0.2233710846D+01	0.2232801672D+01	0.2660211029D+01	-0.7549516567D-14
0.2232801672D+01	0.2232547278D+01	0.2660260954D+01	-0.6217248938D-14
0.2232547278D+01	0.2232476092D+01	0.2660274916D+01	-0.1132427485D-13
0.2232476092D+01	0.2232456171D+01	0.2660278823D+01	-0.1154631946D-13
0.2232456171D+01	0.2232450597D+01	0.2660279916D+01	-0.1554312234D-13
0.2232450597D+01	0.2232449037D+01	0.2660280222D+01	-0.2176037128D-13
0.2232449037D+01	0.2232448600D+01	0.2660280307D+01	-0.1265654248D-13
0.2232448600D+01	0.2232448478D+01	0.2660280331D+01	-0.9769962617D-14
0.2232448478D+01	0.2232448444D+01	0.2660280338D+01	-0.1532107774D-13
0.2232448444D+01	0.2232448434D+01	0.2660280340D+01	-0.1421085472D-13
0.2232448434D+01	0.2232448431D+01	0.2660280341D+01	-0.7105427358D-14
0.2232448431D+01	0.2232448431D+01	0.2660280341D+01	-0.1021405183D-13

Iterative method with 32-point formula

VARIANCE	POS.VARIANCE	NOR.CONSTANT	POS.MEAN
0.3000000000D+01	0.2647782880D+01	0.2719028110D+01	-0.1136590821D-13
0.2647782880D+01	0.2624702691D+01	0.2719206587D+01	-0.2423061751D-13
0.2624702691D+01	0.2623048790D+01	0.2719210173D+01	-0.2343958361D-13
0.2623048790D+01	0.2622929515D+01	0.2719210393D+01	-0.1637578961D-13
0.2622929515D+01	0.2622920909D+01	0.2719210409D+01	-0.2667310817D-13
0.2622920909D+01	0.2622920288D+01	0.2719210410D+01	-0.2524369602D-13
0.2622920288D+01	0.2622920244D+01	0.2719210410D+01	-0.2178812686D-13
0.2622920244D+01	0.2622920240D+01	0.2719210410D+01	-0.1905420266D-13
0.2622920240D+01	0.2622920240D+01	0.2719210410D+01	-0.1600108934D-13

TABLE 2.7: Mispecification of mean and variance

True mean: 0
True variance: 3

the iterative strategy of Naylor and Smith (1982) to $g(t)=[1+(x^2/3)]^{-2}$ starting with the correct mean 0 and variance 3. Even though the convergence of the normalising constant was close to the correct value (2.7206989) it is extraordinary that the iteration convergences to the wrong values of the variance even though we started with the proper values and despite the fact that we used a high precision rule of 32 points. The results are shown in table 2.7. Thus, we have an example where a very high precision rule converges to a value substantially different from the actual value. With the added complexity of working in high dimensional cases, a quite experienced user might be needed for the judge of whether and how we should proceed in the different steps of figure 2.1.

2.5.4 Accuracy

The accuracy of Naylor and Smith algorithm is assessed when checking for convergence between two different quadrature rules Q_n and Q_l with $n<l$ in (ix) in the flowchart of figure 2.1. Implicitly, by moving from a map T^n to a map T^l we use the properties of Gauss-Hermite rules

$$\lim_{m\to\infty} Q_m f = If \qquad (2.22)$$

which holds if, for all sufficiently large values $|x|$, $f(x)$ satisfies the inaquality

$$|f(x)| \leq \exp(x^2)/|x|^{1+\rho} \quad \text{for some } \rho>0$$

see Davis and Rabinowitz (1984, p.227) and Uspensky (1928). To make

use of (2.22) we also assume that there is a unique stable point in map T^n and no unstable or periodic points. Naturally, we also assume that if this holds for for the map T^n it also holds for the map T^1.

The way to assess the accuracy of a quadrature rule is usually achieved by checking the absolute or relative error between two successive estimations. In BAYESFOUR this can be done by using the aggregate measure Δ,(see section 2.3), or by assessing the convergence of each of the elements of vector $(\underline{\mu},\underline{\Sigma})$ separately.

In practice, the accuracy of Naylor and Smith algorithm is closely related to its reliability. If the method is successful, we expect good accuracy and we can normally detect good accuracy through a rapid convergence.

Chapter 3: Imbedded integration rules

3.1 Introduction

We consider integrals of the form

$$I(kf) = \int_a^b k(x)f(x)dx \ , \qquad -\infty \leqslant a < b \leqslant \infty \qquad (3.1)$$

where $k(x)$ is such that $I(kf)$ exists for a family of functions which includes P_n, the set of all polynomials of degree $\leqslant n$. We are interested in approximating $I(kf)$ by interpolatory integration rules (IIR's) of the form

$$Q_{nf} = \sum_{j=1}^{n} w_{jn} f(x_{jn}) \ . \qquad (3.2)$$

where the set of points $X_n = \{ x_{jn}: j=1,\ldots,n \}$ is specified in advance and the weights w_{jn}, $j=1,\ldots,n$ are interpolatory. If the weights w_{jn} are positive, such rules are called positive interpolatory integration rules (PIIR's).

The degree of such rules is n-1, in the sense that it can integrate exactly all the mononomials of degree n-1 or less. If the set X_n is optimally chosen, the rule (3.2) can be of degree 2n-1. Such rules are called integration rules of Gauss type. For example, by putting $k(x) = \exp(-x^2)$, $a = -\infty$, $b = \infty$ in (3.1), and using X_n the zeros of the

Hermite polunomials of degree n (see Davis and Rabinowitz (1984, p.34), the rule (3.2) is the Gauss-Hermite rule (2.4) used in the Naylor and Smith (1982) numerical integration strategy. The weights w_{jn} and points x_{jn} for the Gauss-type integration rules can be found in books, for example Stroud and Secrest (1966).

In the sequence $\{Q_{nf}\}$ of Gauss-formulae, for different values of n, the respective node sets $X_n = \{x_j, j=1,\ldots,n\}$ do not have any points in common, except the mid-point, which is a node when n is odd. This is a serious drawback of the approach, because proceeding from a computation of Q_{nf} to Q_{mf} with $m>n$ almost all previous function evaluations are discarded.

Sequences, in which the nodes of a given rule form a subset of the nodes of its successor, overcome this drawback and are the particular concern of this chapter. Essentially, there are two ways in which such sequences can be obtained: either by the addition of nodes to an existing rule to form an extended rule, or by taking subsets of nodes from an existing rule to form an imbedded rule. If, in the above procedures, the derived sequences have as a highest degree rule a Gauss-type rule, the integration rules obtained are called Gauss-based integration rules (GBIR's).

Historically, the subject of this chapter emanates from Kronrod (1964). Motivated by a desire to estimate economically the error in the classical Gaussian quadrature formula, he proposed to extend the n-point Gauss quadrature rule to a (2n+1)-point quadrature rule by inserting n+1 additional points and making the extended quadrature rule have maximum degree of exactness. This early work of Kronrod has

led to a vast amount literature in the field of numerical analysis, extending and refining the GBIR's.

An important feature of these rules is that they provide a sequence of approximations in which previous function evaluations are exploited when proceeding from one approximation to another. They also provide a means of measuring an estimation error in the quadrature formula. It is not surprising, therefore, that such rules are used in nearly all well known automatic quadrature routines.

This chapter concentrates on the applications of the one-dimensional imbedded sequences of rules. Their properties and potential applications in Bayesian analysis are the subject of section 3.2. We shall adopt a similar notation to that of Rabinowitz *et al.* (1987), our main reference in this area. Other relevant references in the numerical analysis field are Davis and Rabinowitz (1984, p.106-109, 426), Atkinson(1978, p.243-248), and the reviews given by Monegato (1979) and Gautschi (1988).

In section 3.3, we present a recent development in this area namely the imbedded sequences of positive interpolatory rules. We believe that these rules are particularly promising in Bayesian analysis. Artificial and real examples are used for the illustration of these methods in section 3.4.

3.2 Gauss-based sequences of interpolatory integration rules

3.2.1 Patterson rules

Patterson (1968a) proposed the use of sequences of interpolating integration rules based on subset of certain Gauss and Lobatto integration points. In these sequences, each rule is an extension of the previous rule in that it uses all the points of the previous rule, or equivalently, each rule is imbedded in its successor. Such a sequence is also called an imbedded sequence.

In his paper, Patterson started with a n-point rule, $n = 2^r+1$, with the points denoted by x_j, $j=1,\ldots,n$, $x_n < x_{n-1} < \ldots < x_1$. He then formed the subsets

$$S_i = \{x_{2^i(j-1)+1} : j=1,2,\ldots,2^{r-i}+1\} , \qquad i=1,2,\ldots,r \qquad (3.3)$$

by successively deleting alternative points from the previous subset. Therefore, the points in S_i-S_{i+1} interlace those of S_i+1, i.e. between any two points of S_{i+1} there is a point of S_i-S_{i+1}. Figure 3.1 represents the subsets S_i, $i=1,2,\ldots,6$ for the case of $n=65$. For each of the subsets S_i, Patterson computed the weights w_{ji}, $j=1,2,\ldots,i$ needed for the calculation of Q_{if} in (3.2) using numerical integration of the Lagrangian interpolating coefficients, given by

$$L_k(x) = \prod_{\substack{j=0 \\ j\neq k}}^{i} \frac{x-x_j}{x_k-x_j} , \qquad x_j \in S_i \qquad (3.4)$$

of sufficiently high order. If, for example, $k(x) = e^{-x^2}$ $a=-\infty$ and $b=\infty$ in (1), then the weights needed for the calculation of Q_{nf} in (3.2) using a pre-assigned set of nodes S_n as in (3.3) would be obtained as described below:

The Lagrangian interpolating polynomial $p(x)$ of degree n-1 for a function $f(x)$ given at the points x_i, $i = 1,\ldots,n$ is given by

$$f(x) \approx p(x) = \sum_{i=1}^{n} L_i(x)\ f(x_i),$$

where $L_i(x)$ is given in (3.4), and is the unique member of the set of polynomials of degree $\leq n$ with this property. Therefore, we have

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)\ dx \approx \int_{-\infty}^{\infty} e^{-x^2} p(x)\ dx =$$

$$= \int_{-\infty}^{\infty} e^{-x^2} \sum_{i=1}^{n} L_i(x)\ f(x_i)\ dx = \sum_{i=1}^{n} w_{in}\ f(x)\ ,$$

with

$$w_{in} = \int_{-\infty}^{\infty} e^{-x^2} Li(x)\ dx \qquad (3.5)$$

and thus the weights are given by (3.5) using a Gauss-Hermite formula with n/2 points when n is even and (n+1)/2 points when n is odd.

In his paper, Patterson (1968a) used a Gauss-Legendre rule with a=-1,

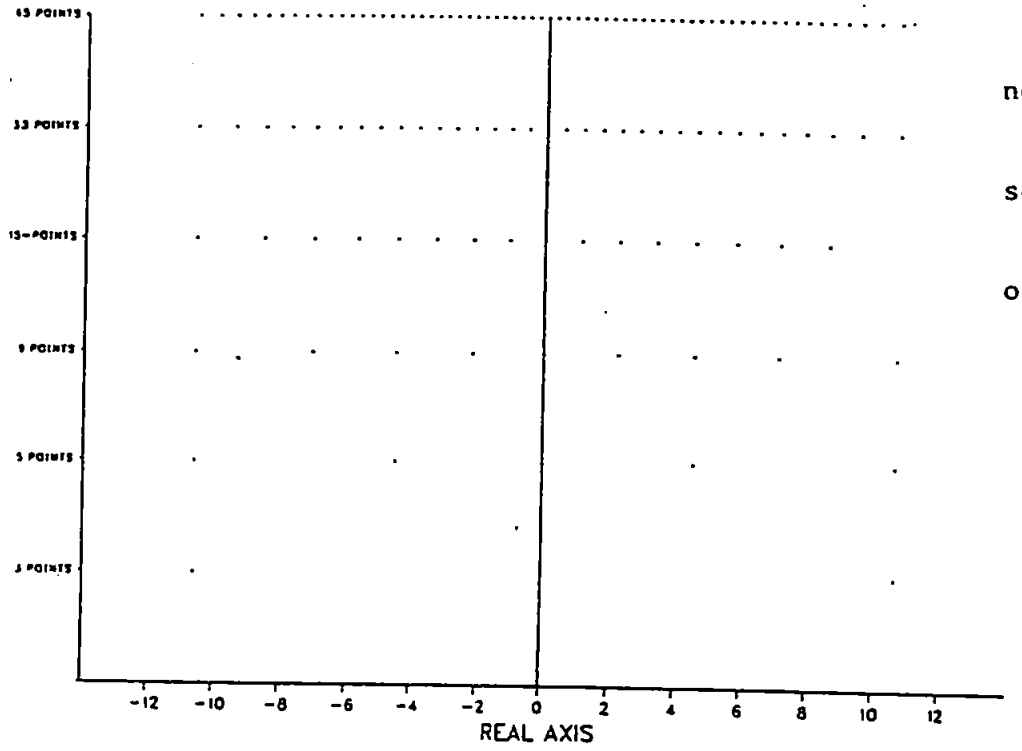
b=1, k(x)=1 in (3.1). It turned out that all the weights are positive, ie. the resulting rules are all positive interpolatory integration rules (PIIR's). We shall be mainly interested in these rules, because negative weights are un-attractive from the Statistical viewpoint. Shaw (1987b) notes that the negative sign in the weights might cause the embarrassing possibility of estimating the normalising constant to be negative, that general rounding error can occur in the calculation of the posterior expectations, and that theoretical bounds on the error of the approximation of (3.1) and (3.2) often involve the expression $\Sigma|w_{in}|$, which can be large if the weights w_{in} are not all the same sign. Additionally, the interpretation of approximations to expectations with respect to a discrete distribution breaks down if the weights are negative.

Our conjecture was that any Gauss rule would be a positive integration rule, and this was encouraged by Davis and Rabinowitz (1984, pp109) where it has been commented that 'experience has shown these weights to be nonnegative'. However, as it can be seen from tables 3.1, 3.2 and 3.3 in the Gauss-Hermite case, at least two (symmetric) negative weights appeared in the new subsets on the n-weights set, where n = 17, 33 or 65. These weights were calculated using (3.5), where the nodes were calculated using a FORTRAN program by Stroud and Secrest (1966). The possibility that these negative weights resulted from rounding error during their computation was investigated. This involved the application of a quadruple precision program.

A further difficulty with Patterson's method is that it uses extraordinarily widely dispersed nodes for small numbers of points. For example, if we apply a 65 point final precision formula to

FIGURE 3.1

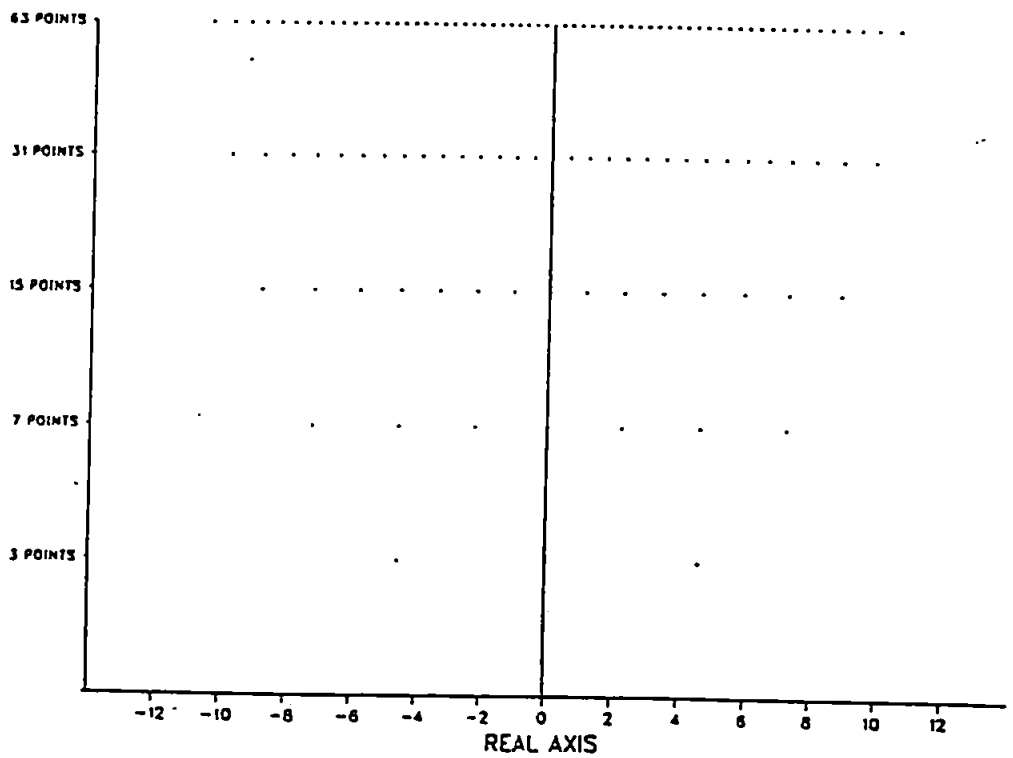
GRAPHICAL DISPLAY OF PATTERSON METHOD



nodes in sequences of rules

FIGURE 3.2

GRAPHICAL DISPLAY OF PSEUDO-PATTERSON METHOD



nodes in sequences of rules

TABLE 3.1: Weights derived using a 17-point final precision Patterson formula

3 POINTS	5 POINTS	9 POINTS	17 POINTS
0.186731305564D-01	-0.316766609740D-02	-0.164498505396D-04	0.458057886809D-10
0.173510758978D+01	0.109709912356D+00	0.404626472724D-03	0.497707898162D-07
0.186731305564D-01	0.155936935838D+01	0.714117170146D-02	0.711228913945D-05
	0.109709912356D+00	0.355432689142D+00	0.298643286695D-03
	-0.316766609740D-02	0.104652977597D+01	0.506734995761D-02
		0.355432689142D+00	0.409200341481D-01
		0.714117170146D-02	0.172648297670D+00
		0.404626472724D-03	0.401826469469D+00
		-0.164498505396D-04	0.530917937623D+00
			0.401826469469D+00
			0.172648297670D+00
			0.409200341481D-01
			0.506734995761D-02
			0.298643286695D-03
			0.711228913945D-05
			0.497707898162D-07
			0.458057886809D-10

TABLE 3.2: Weights derived using a 33-point final precision Patterson formula

3 POINTS	5 POINTS	9 POINTS	17 POINTS	33 POINTS
0.842126987299D-02	-0.166093905068D-02	-0.296080124389D-04	-0.438312899004D-09	0.115331621300D-22
0.175561131082D+01	0.534721947511D-01	0.968539517657D-03	0.265414894992D-07	0.165709470130D-18
0.842126987299D-02	0.166883133917D+01	-0.110409328811D-01	-0.477456671487D-06	0.240778567955D-15
	0.534721947511D-01	0.222542224208D+00	0.449398476850D-05	0.943481415879D-13
	-0.166093905068D-02	0.134757340490D+01	0.158976380585D-04	0.147398093694D-10
		0.222542224208D+00	0.353114659978D-02	0.112892224711D-08
		-0.110409328811D-01	0.717323056234D-01	0.480774561784D-07
		0.968539517657D-03	0.427420359341D+00	0.123769336655D-05
		-0.296080124389D-04	0.767046346899D+00	0.204236840508D-04
			0.427420359341D+00	0.225442770595D-03
			0.717323056234D-01	0.171845463776D-02
			0.353114659978D-02	0.926568997066D-02
			0.158976380589D-04	0.359879823185D-01
			0.449398476856D-05	0.102069079846D+00
			-0.477456671477D-06	0.213493931133D+00
			0.265414894998D-07	0.331552000749D+00
			-0.438312898995D-09	0.383785266519D+00
				0.331552000749D+00
				0.213493931133D+00
				0.102069079846D+00
				0.359879823185D-01
				0.926568997066D-02
				0.171845463776D-02
				0.225442770595D-03
				0.204236840508D-04
				0.123769336655D-05
				0.480774561784D-07
				0.112892224711D-08
				0.147398093694D-10
				0.943481415879D-13
				0.240778567955D-15
				0.165709470130D-18
				0.115331621300D-22

TABLE 3.3: Weights derived using a 65-point final precision Patterson formula

3 POINTS	5 POINTS	9 POINTS	17 POINTS	33 POINTS	65 POINTS
0.393179843004D-02	-0.802647213858D-03	-0.200776103569D-04	-0.338388971102D-08	-0.315755284154D-18	0.825161079872D-49
0.176459025403D+01	0.262170145604D-01	0.773239839374D-03	0.284224940751D-06	0.703386781762D-16	0.270767584008D-43
0.393179843004D-02	0.172162511619D+01	-0.951019572255D-02	-0.595636116226D-05	-0.373128562687D-14	0.589628446545D-39
	0.262170145604D-01	0.122954438632D+00	0.636588761375D-04	0.947502660032D-13	0.285418486390D-35
	-0.802647213858D-03	0.154405904061D+01	-0.439490530254D-03	-0.146520593496D-11	0.495258625501D-32
		0.122954438632D+00	0.223540741335D-02	0.155581766488D-10	0.396328698479D-29
		-0.951019572255D-02	-0.114850506654D-03	-0.121727456221D-09	0.170591158107D-26
		0.773239839374D-03	0.352180245895D+00	0.736980458233D-09	0.437697419487D-24
		-0.200776103569D-04	0.106461525963D+01	-0.265597439529D-08	0.720161078912D-22
			0.352180245895D+00	0.136865658910D-06	0.802221870040D-20
			-0.114850506653D-03	0.806055687314D-05	0.630789104097D-18
			0.223540741335D-02	0.260282846315D-03	0.361819961854D-16
			-0.439490530254D-03	0.423970301316D-02	0.155466357219D-14
			0.636588761376D-04	0.361393508077D-01	0.511391748167D-13
			-0.595636116226D-05	0.164599570881D+00	0.131125161063D-11
			0.284224940751D-06	0.406502135599D+00	0.266086534778D-10
			-0.338388971101D-08	0.548955373800D+00	0.432865615344D-09
				0.406502135599D+00	0.570758293276D-08
				0.164599570881D+00	0.615779622143D-07
				0.361393508077D-01	0.548045603500D-06
				0.423970301316D-02	0.405224939101D-05
				0.260282846316D-03	0.250453426281D-04
				0.806055687324D-05	0.130082915729D-03
				0.136865658968D-06	0.5703989665370-03
				-0.265597437118D-08	0.211998163069D-02
				0.736980460674D-09	0.670140453658D-02
				-0.121727455554D-09	0.180694331115D-01
				0.155581767074D-10	0.416611087617D-01
				-0.146520592590D-11	0.823001633692D-01
				0.947502665298D-13	0.139526139482D+00
				-0.373128558299D-14	0.203250574154D+00
				0.703386787944D-16	0.254628811852D+00
				-0.315755280790D-18	0.274478226558D+00
					0.254628811852D+00
					0.203250574154D+00
					0.139526139482D+00
					0.823001633692D-01
					0.416611087617D-01
					0.180694331115D-01
					0.670140453658D-02
					0.211998163069D-02
					0.5703989665370-03
					0.130082915729D-03
					0.250453426281D-04
					0.405224939101D-05
					0.548045603500D-06
					0.615779622143D-07
					0.570758293276D-08
					0.432865615344D-09
					0.266086534778D-10
					0.131125161063D-11
					0.511391748167D-13
					0.155466357219D-14
					0.361819961854D-16
					0.630789104097D-18
					0.802221870040D-20
					0.720161078912D-22
					0.437697419487D-24
					0.170591158107D-26
					0.396328698479D-29
					0.495258625501D-32
					0.285418486390D-35
					0.589628446545D-39
					0.270767584008D-43
					0.825161079872D-49

integrate a normal density, then a 3 point formula gives one node at the mean and two nodes well over 10 standard deviations away from the mean! Clearly, though the formulae will integrate a normal density exactly, it will be influenced to a great extent by the behaviour of the distribution in the extreme tails. For example, the formula gives grossly incorrect values when applied to a heavy tailed distribution. In a t-distribution with 3 degrees of freedom, a 3 point formula gives a normalising constant equal to $0.33 \cdot 10^{43}$ (this should be compared with the correct value of 2.72!).

Rabinowitz *et al.* (1987) in slightly different context make this comment and point out that the nodes in the subsets are far away from the sets corresponding to the Gauss rules with the same number of points. To overcome the above difficulty and to eliminate the negative weight, we developed a pseudo-Patterson method. The method was constructed from a final high-accuracy rule of $n = 2^i-1$ points, in a similar way to Patterson method. Starting now with $n = 2^i-1$ points new subsets are created by successively striking out every second point starting from the first points. The new subsets of a $n = 2^i-1$ point formula will then have the form

$$S_i = \{x_{2i_j} \quad j=1,2,\ldots,2^{r-1}-1\} , \qquad i=1,2,\ldots,r-1$$

Figure 3.2 represents the subsets of a 63-point formula. Comparing it with Figure 3.1 (Patterson's original method) we note that the new method is built with nodes more concentrated around the mid-point, but using one step less than Patterson's method. The weights were calculated for subsets of nodes from the 15 point and 31 point final precision rules. In the former case all weights were positive but,

TABLE 3.4

Weights of the subsets of 15-point final precision Patterson-type (pseudo-Patterson) formula

3-POINTS	7-POINTS	15-POINTS
0.819210342477D-01	0.228636956632D-03	0.152247580425D-08
0.160861178240D+01	0.247879730213D-02	0.105911554771D-05
0.819210342477D-01	0.330523582941D+00	0.100004441232D-03
	0.110599181650D+01	0.277806884290D-02
	0.330523582941D+00	0.307800338724D-01
	0.247879730213D-02	0.158488915795D+00
	0.228636956632D-03	0.412028687498D+00
		0.564100308725D+00
		0.412028687498D+00
		0.158488915795D+00
		0.307800338724D-01
		0.277806884290D-02
		0.100004441232D-03
		0.105911554771D-05
		0.152247580425D-08

in the latter case some negative weights occurred. See table 3.4.

3.2.2 Experiments with Patterson type rules

Having found the weights for some Patterson and pseudo-Patterson sequences, we used the class of exponential power distributions described in section 2.5.1 to test their behaviour over a range of distributions with different coefficients of kurtosis. The Patterson sequence based on a 17-point final precision formula and the pseudo-Patterson based on a 15 point final precision formula were examined using the graphical display of figure 3.3-3.6. The plots represent the relative error aqainst the coefficient of kurtosis (see (2.20) and (2.21)).

The figures 3.3 and 3.4 confirm our remarks in previous section and verify that the pseudo-Patterson method is clearly much better than Patterson's method and achieves small relative errors for distributions which are 'close' to the Normal (kurtosis close to 0). The small number of steps in two methods however, does not permit convergence checks, and therefore prevents any further clear conclusions being drawn.

3.2.3 Gauss-Kronrod-Patterson Rules

The imbedded Gauss based sequences of Patterson (1968a) attracted little interest in the numerical analysis literature, perhaps because simultaneously and in the same journal, Patterson published another paper (1968b) in which he introduced the Gauss-Kronrod-Patterson (G-K-P) sequence of integration rules, which received a considerable amount of attention in subsequent years. These rules, are more

FIGURE 3.3

RELATIVE ERROR OF NORMILISING CONSTANT
IN A PSEUDO–PATTERSON METHOD

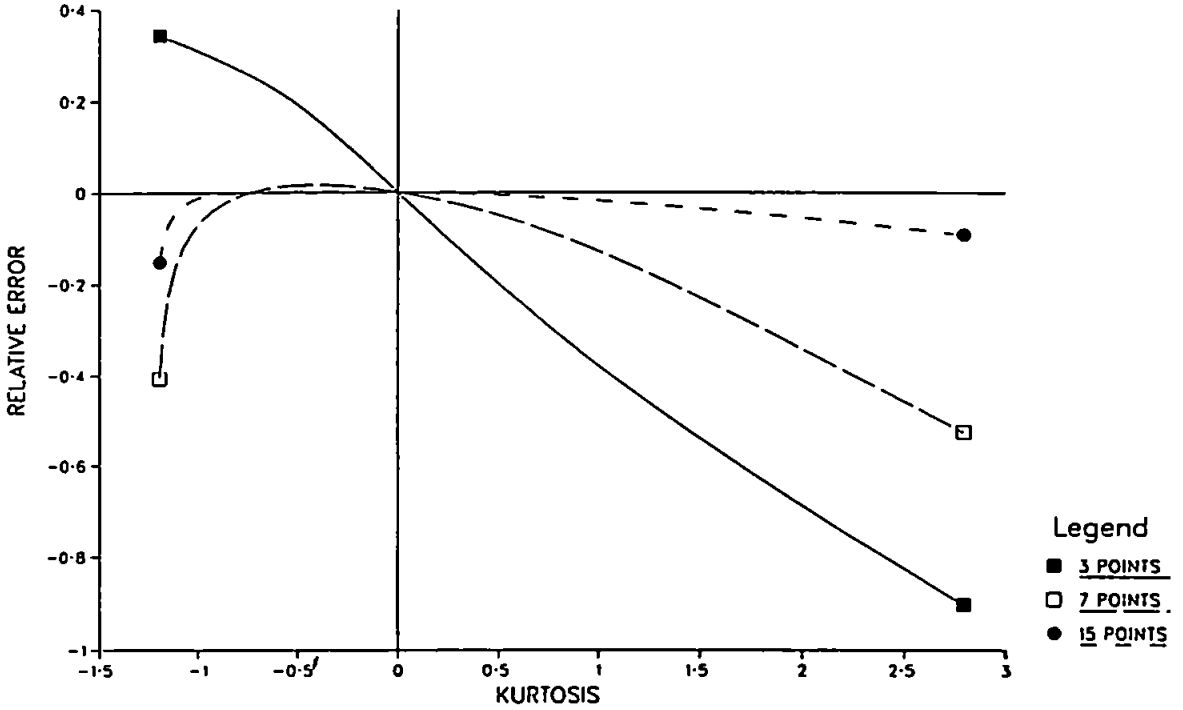


FIGURE 3.4

RELATIVE ERROR OF NORMILISING CONSTANT
IN A PATTERSON METHOD

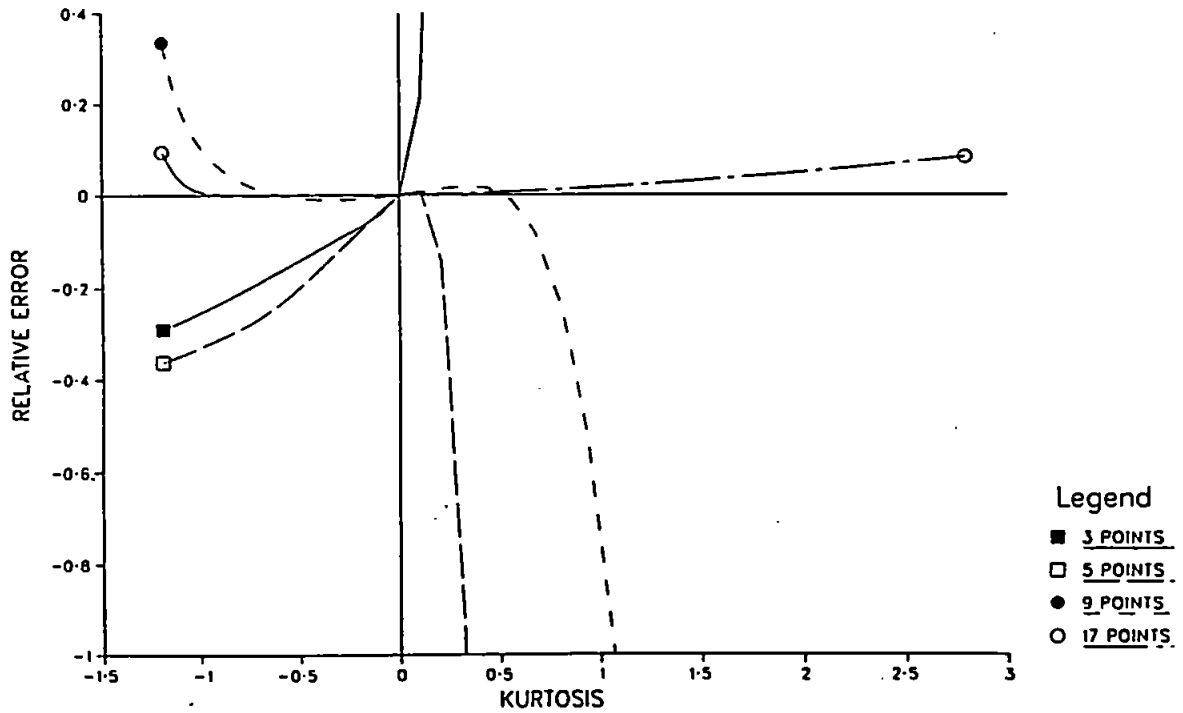


FIGURE 3.5

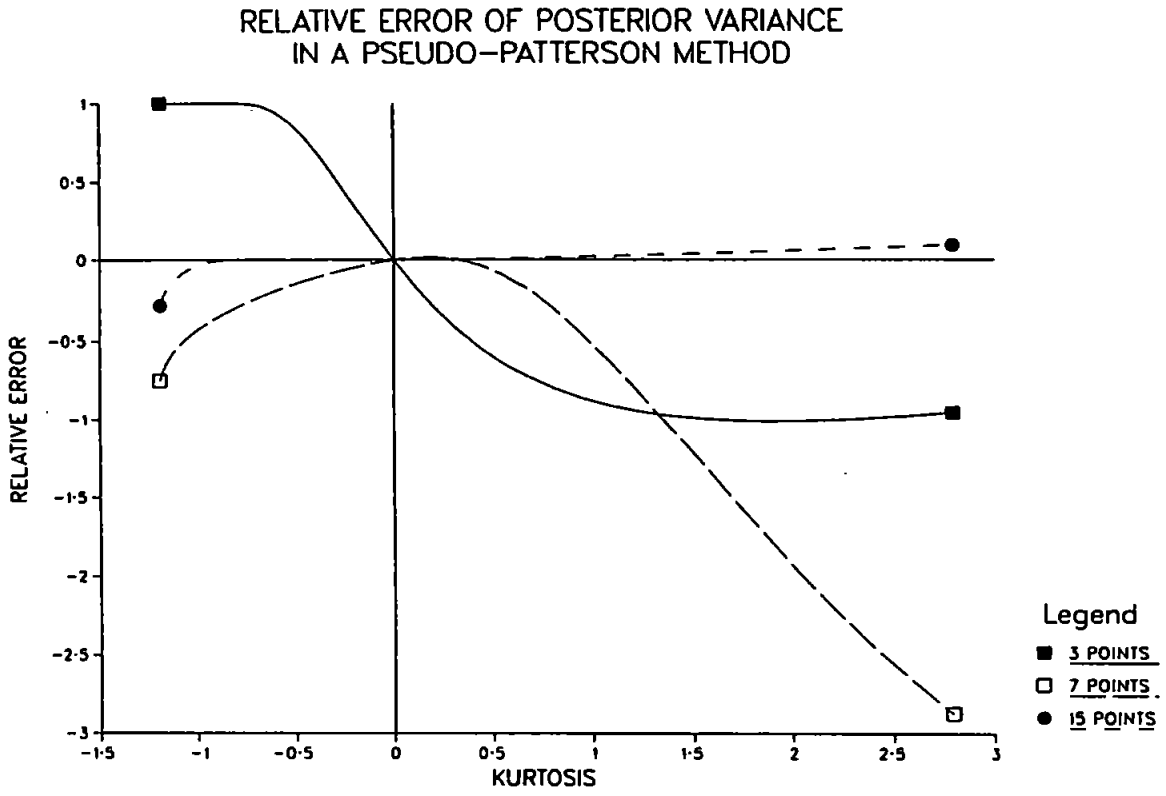
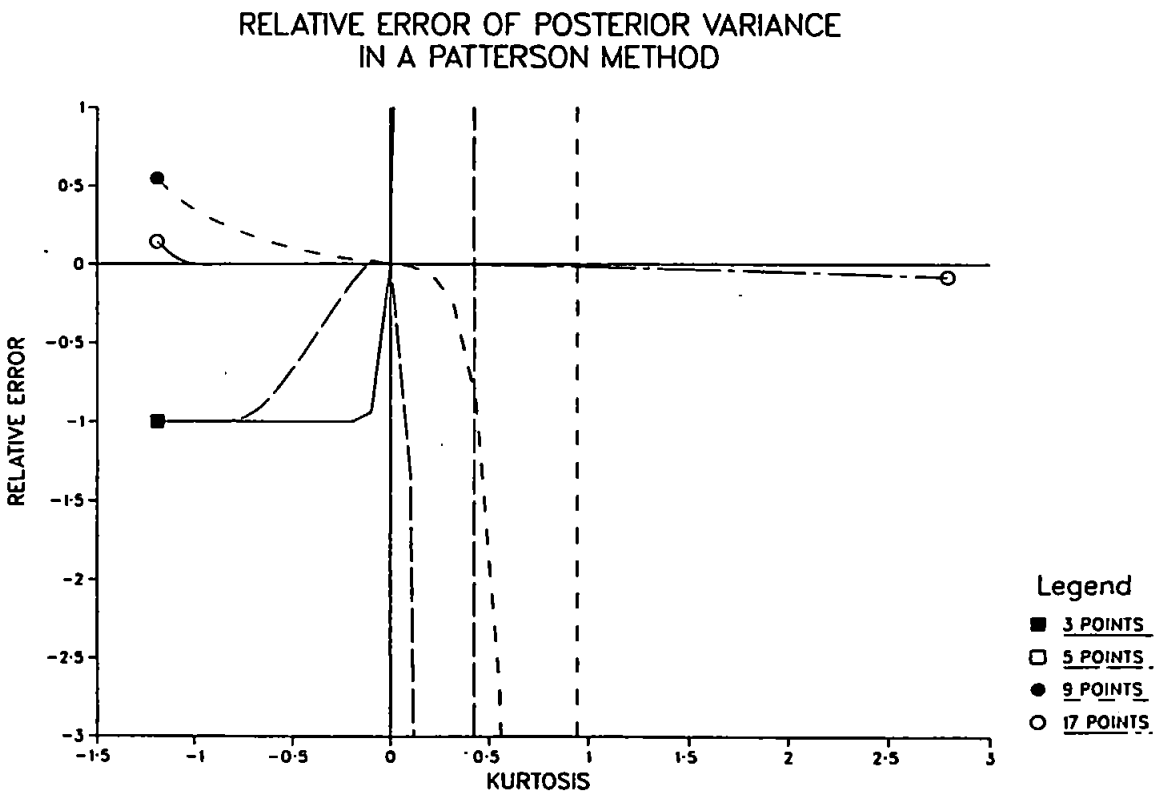


FIGURE 3.6



accurate than GBIR rules and in addition are open-ended, whereas in the GBIR sequences one has to specify the base (final) rule in advance. This explains the neglect for the GBIR rules, and the concentration on developments of G-K-P rules (eg. Piessens and Branders (1974), Monegato (1978)) and automatic algorithms (eg. Piessens (1973), Patterson (1973)) were published. The basic theory and development of these (G-K-P) rules are outlined below -a recent survey can be found in Gautschi (1987).

The original idea behind the G-K-P rules came from Kronrod (1965), who considered (3.1) for the case a=-1, b=1. He showed how an n-point Gaussian quadrature formulae may be augmented by a set of n+1 nodes to yield quadrature formulae of degree 3n+1 if n is even and 3n+2 if n is odd. The problem can be expressed then as an approximation of (3.1) by

$$\sum_{i=1}^{n} a_i \, f(x_i) + \sum_{i=1}^{n+1} \beta_i \, f(\xi_i) \qquad (3.6)$$

where the xi's are the nodes of an n-point Gaussian quadrature formula.

We want to determine the additional nodes ξ_i and the weights α_k and β_k so that the degree of precision of (3.6) is maximal. It is known that the nodes ξ_i must be the zeros of the polynomial $\varphi_{n+1}(x)$ which satisfies

$$\int_a^b p_n(x)\varphi_{n+1}(x)x^k dx = 0 \qquad k=0,1,\ldots,n \qquad (3.7)$$

where $p_n(x)$ is the Legendre polynomial of degree n. Kronrod (1965) gave a simple method for the computation of $\varphi_{n+1}(x)$. Patterson (1968b) expanded $\varphi_{n+1}(x)$ in terms of Legendre polynomials and derived a sequence of quadrature formulas by iterating the process described in (3.6) and (3.7). He considered the Gauss-Legendre, rule and starting with a 3-point rule reached a 127-point rule.

Unfortunately, the zeros of $\varphi_{n+1}(x)$ in (3.7) are not necessarily real, and the construction of the G-K-P sequence is not always possible. Nor is the positivity of the weights ensured. For the particular case of $k(x) = \exp(-x^2)$ with $a=-\infty$, $b=\infty$ in (3.1), (Gauss-Hermite case), Ramsky (1974) notes that extensions of the type (3.6) exist only for $n=1,2,4$. Monegato (1976) confirms these results experimentally, and proves them in a paper two years later (1978). Finally, Elhay and Kautsky (1984) compute a new G-K-P type sequence of imbedded quadratures for Gauss-Hermite case. The sequence consists of formulae with 2, 5, 9 and 17 nodes with two (symmetric) negative weights appearing in the 17-point formula. It, therefore, seems that the G-K-P sequences are unsuitable for use in statistical analysis.

3.3. Imbedded sequences of positive interpolating integration rules

Nearly twenty years after the first GBIR sequence was constructed by Patterson (1968a), Rabinowitz *et al.* (1987) published a paper which considered GBIR rules with positive weights. These rules are preferable to the original GBIR rules, primarily because they contain

positive integration rules in Gauss Hermite case, and also because when progressing from one rule to its successor fewer points are added. This is the case because the number of points increases arithmetically rather than geometrically, which was the case in GBIR rules. This latter property gives the user the flexibility to use a larger final-precision number of points, compared with GBIR rules, and to monitor the convergence of the integral more often.

We consider again integrals of the form (3.1) which can be approximated by interpolating integration rules (IIR's) of the form (3.2), where the set of points $X_n = \{x_{jn}: j=1,\ldots,n\}$ is specified in advance and therefore this IIR is said to be based on X_n. We recall here that a IIR is called positive IIR (PIIR) if all weights r_{jn} in (3.2) are positive, and that a rule Qmf is imbedded in Qnf if $X_m \subset X_n$. We state now theorem 1 as given in Rabinowitz et al (1987):

Theorem 1: Given any PIIR Q_nf based on a set X_n, there exists a finite sequence of PIIR's, $\{Qn_kf;\ k=1,\ldots,m\leq k\}$ such that $Qn_1f = Qnf$ and such that Qn_kf is of precision n_k-2 for $k=1,\ldots,m-1$.

The theorem guarantees that, if we are dealing with a symmetric situation, as for example in Gauss-Hermite case, where the weight function $k(x) = e^{-x^2}$ in (3.1) is symmetric about 0, we can construct a sequence of PIIR's starting with a final high-precision set of points X_n and dropping each time two (symmetric) points. The interpolating weights for each subset of X_n can then be calculated with the same method which used by Patterson (1968a) and described earlier in this section. The FORTRAN package IQPACK is specially written to evaluate weights of interpolatory quadratures with prescribed nodes. See

Kautsky and Elhay (1982) and Elhay and Kautsky (1987) for more details. These weights need be calculated only once, stored, and used each time, as indicated by Naylor and Smith (1982). Assuming that we start using 2 nodes when n is even and 3 nodes when n is odd, we need to store for each n a set of (n+1)/2 nodes (due to symmetry) and (n-1)/2 sets of weights, corresponding to each increase in he number of nodes by one. For example, with a 16-point final precision base rule we need to store 8 (symmetric) nodes and 8 sets of weights, and for a 17-point final precision sequence we need to store 9 nodes (8 symmetric and the mid-point) and 8 sets of weights.

Rabinowitz *et.al.* (1987) tested different sequences of PIIR's in various situations. We describe here the more interesting (experimental) results concerning the convergence of the sequences of PIIR's. Their suggestion is to stop and accept the current approximation if the results of two (or more conservatively three) successive approximations agree to within desired accuracy. This convergence is not to the true value of the integrand but to the final precision based sequence. It is however generally true that when rapid convergence occurs, it will be to the true value of the integral. Thus, a general strategy would be to start with a high accuracy based rule . In their paper, Rabinowitz *et.al.* start with a 36-point Gauss-Hermite rule to demonstrate the convergence of the imbedded sequence of a PIIR.

The problem of the false convergence is, of course, inevitable. This phenomenon is well known to numerical analysts, see for example Lyness (1983), Davis and Rabinowitz (1984, p.421-424), Laurie (1985), and therefore, it is worthwhile to keep this in mind in the context of the

application of these methods in Bayesian analysis. We will discuss this matter extensively in chapter 5.

3.4 Applications of imbedded sequences of PIIR's in 1 dimension

3.4.1 Reanalysis of the Weibull example of section 2.4

A PIIR sequence as described in chapter 4 for the Gauss-Hermite case was constructed and applied to the Weibull example of (section 2.4). We chose a 16 point final precision formula, and the PIIR sequence was found as described in Rabinowitz *et al.* (1987). Having then chosen the initial rule, we dropped a point each time and we computed the interpolatory weights for the resulting set of points using (3.5). The indices of the points in the order in which they were removed are

$$8,6,7,5,4,3,2,1. \qquad (3.8)$$

Note that reading from right to left the sequence gives the indices of the points in the order of building up a (k+1)-point PIR from a k-point PIR.

The lower 90% prediction bound was calculated using the method described in (2.4.3), and for a better demonstration of the method, was plotted against the number of function evaluations rather than the number of points -see figures 3.7-3.12. Having applied the Gauss-Hermite method with 4, 6 and 8 points it is therefore assumed that when an 8-point formula is used, 4+6+8=18 function evaluations

FIGURE 3.7

EXAMPLE 1 WITH PRIOR(1)

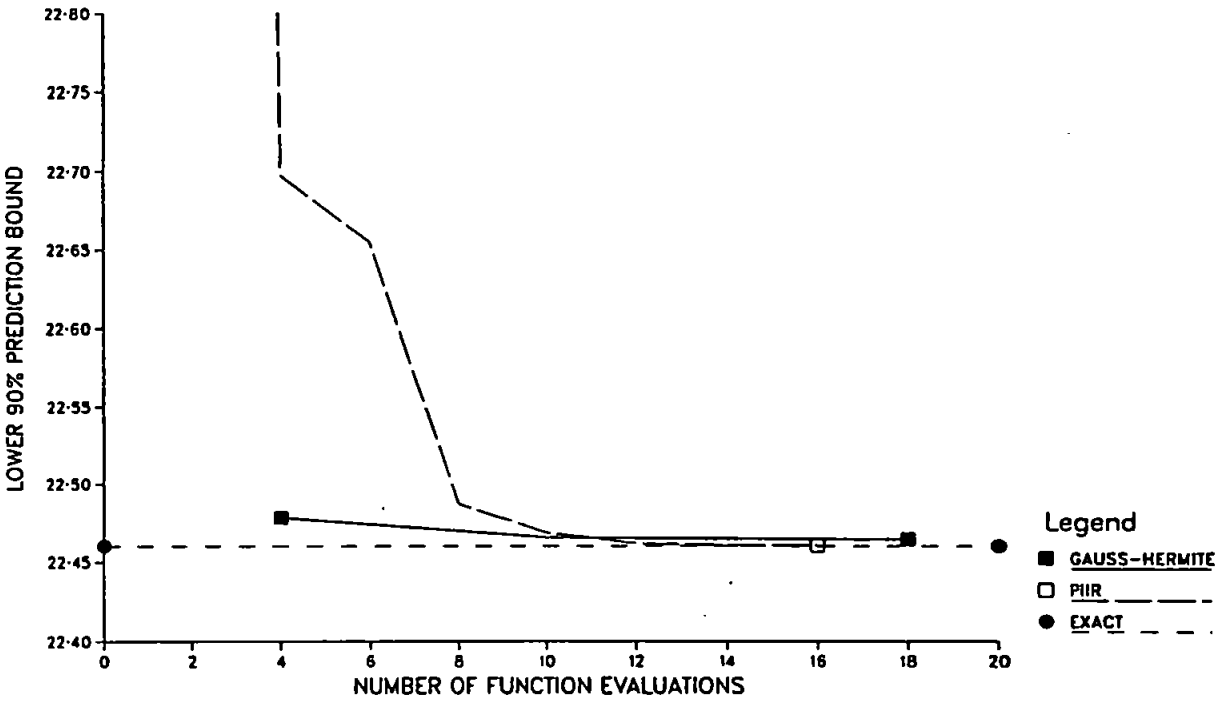


FIGURE 3.8

EXAMPLE 1 WITH PRIOR(2)

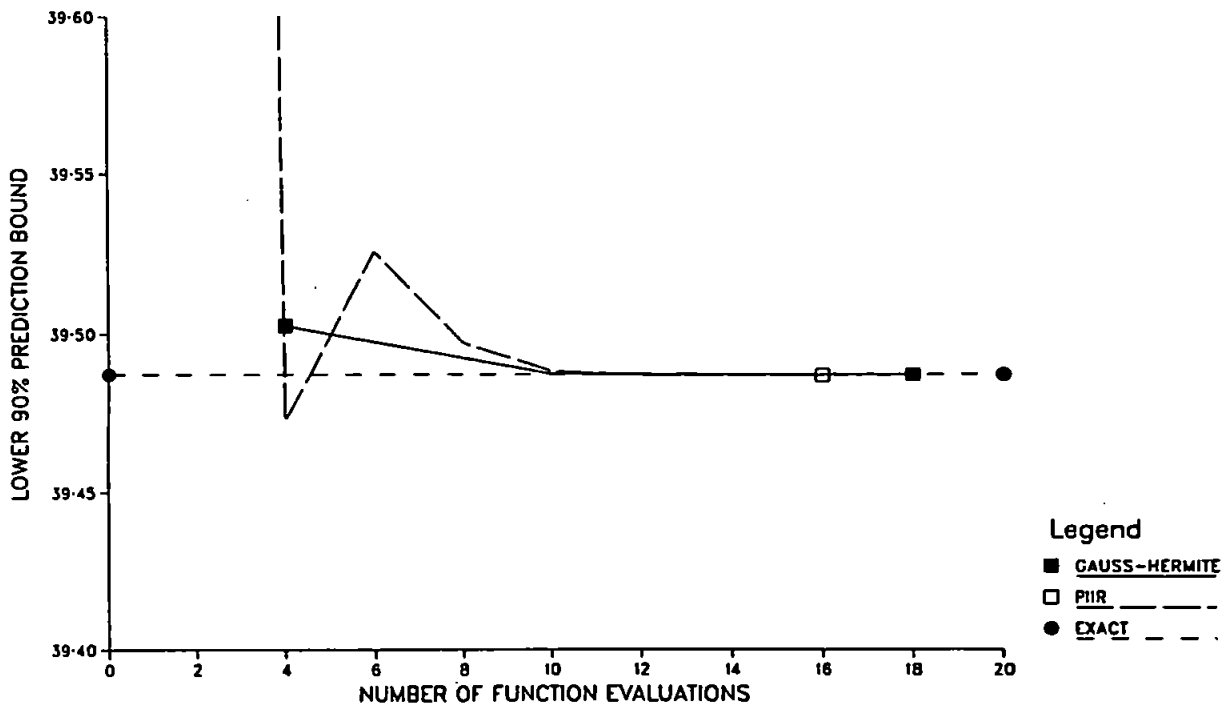


FIGURE 3.9

EXAMPLE 2 WITH PRIOR(1)

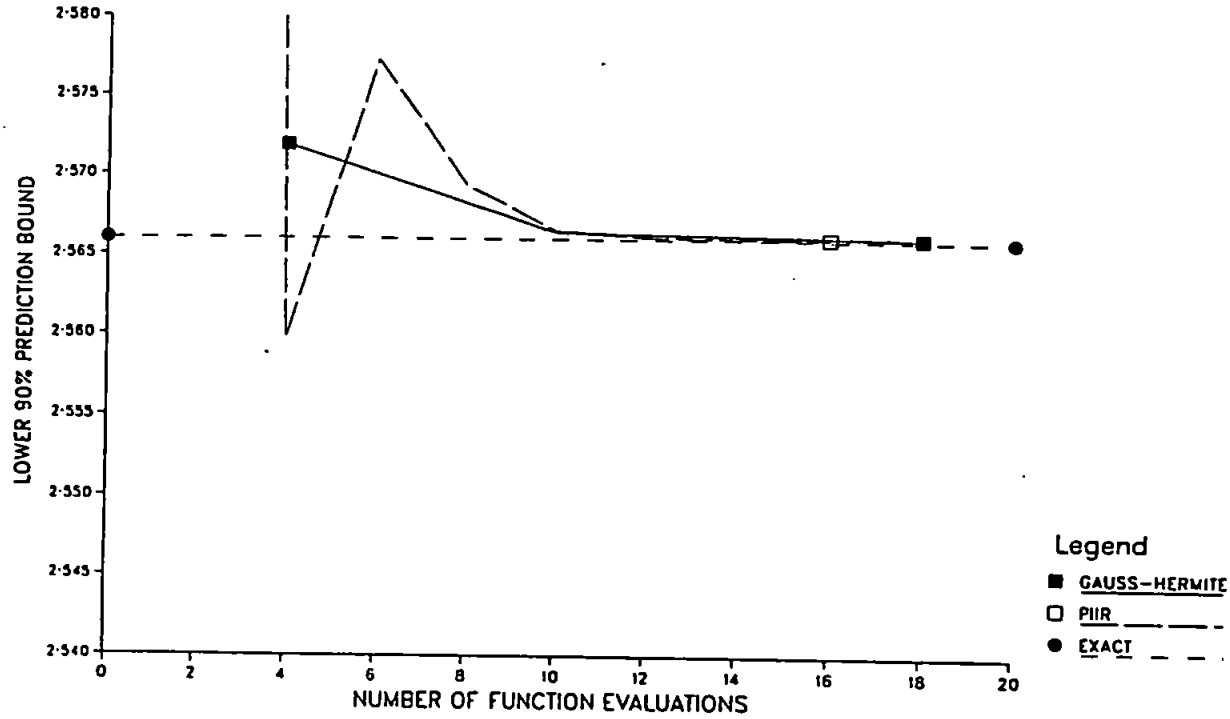


FIGURE 3.10

EXAMPLE 2 WITH PRIOR(2)

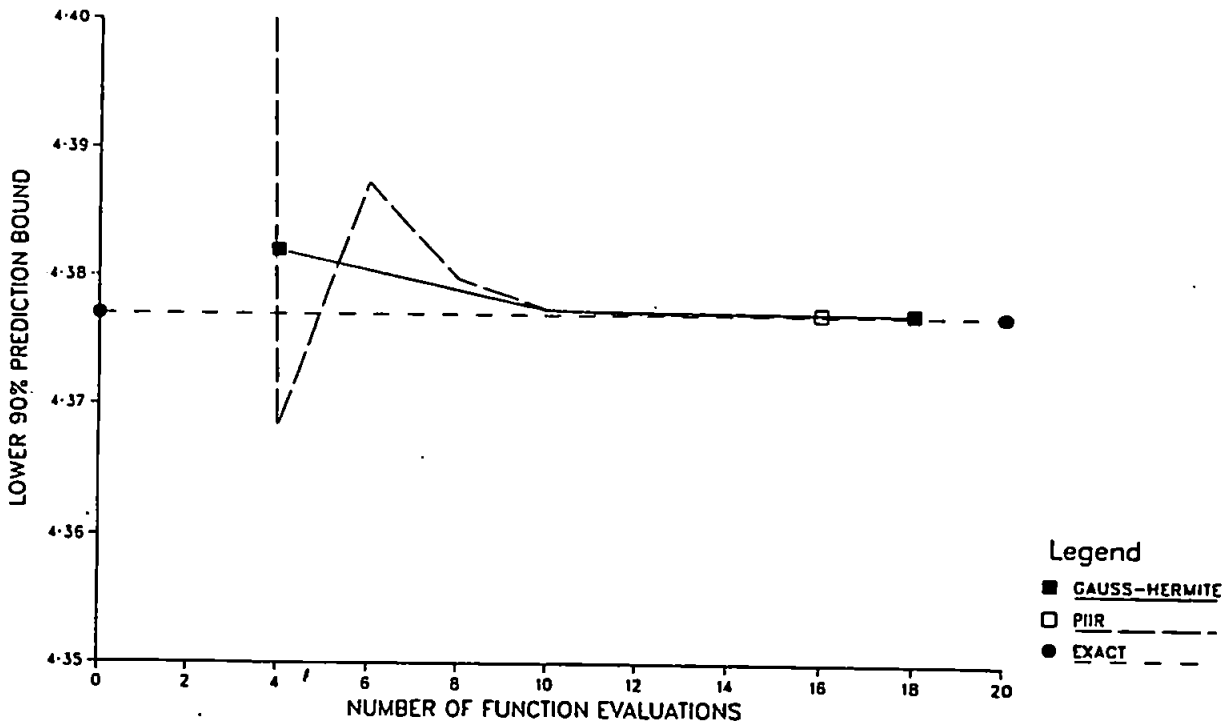


FIGURE 3.11

EXAMPLE 3 WITH PRIOR(1)

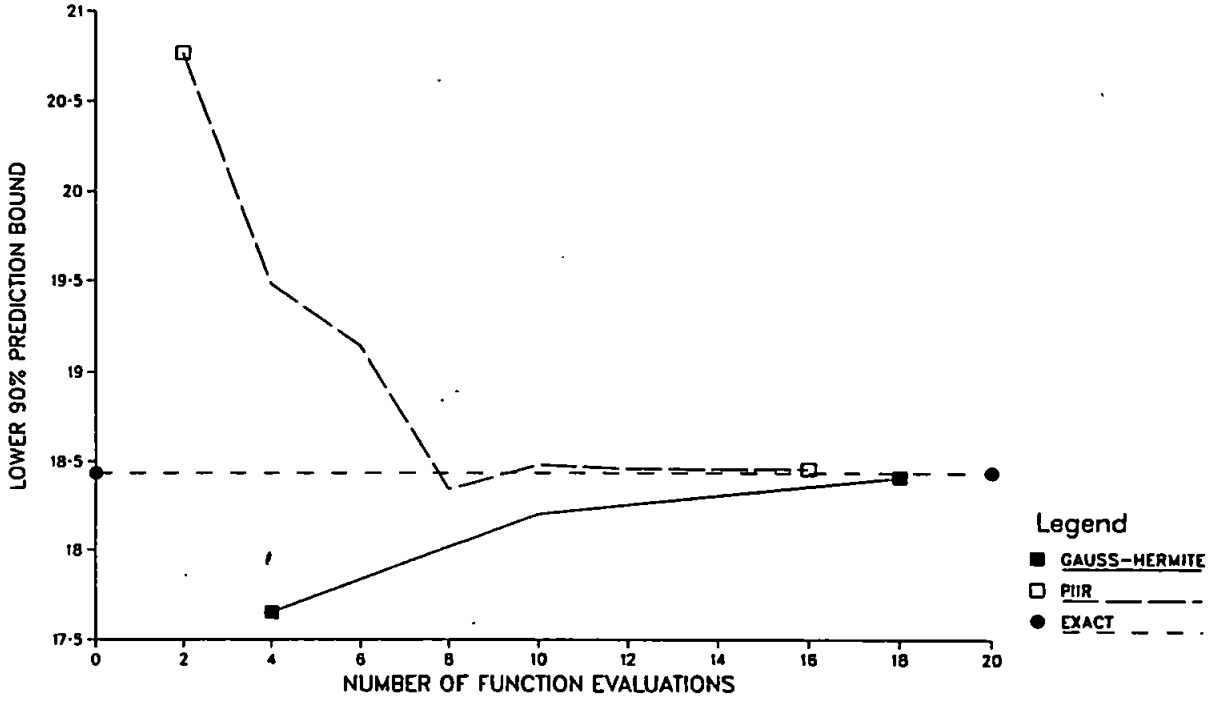
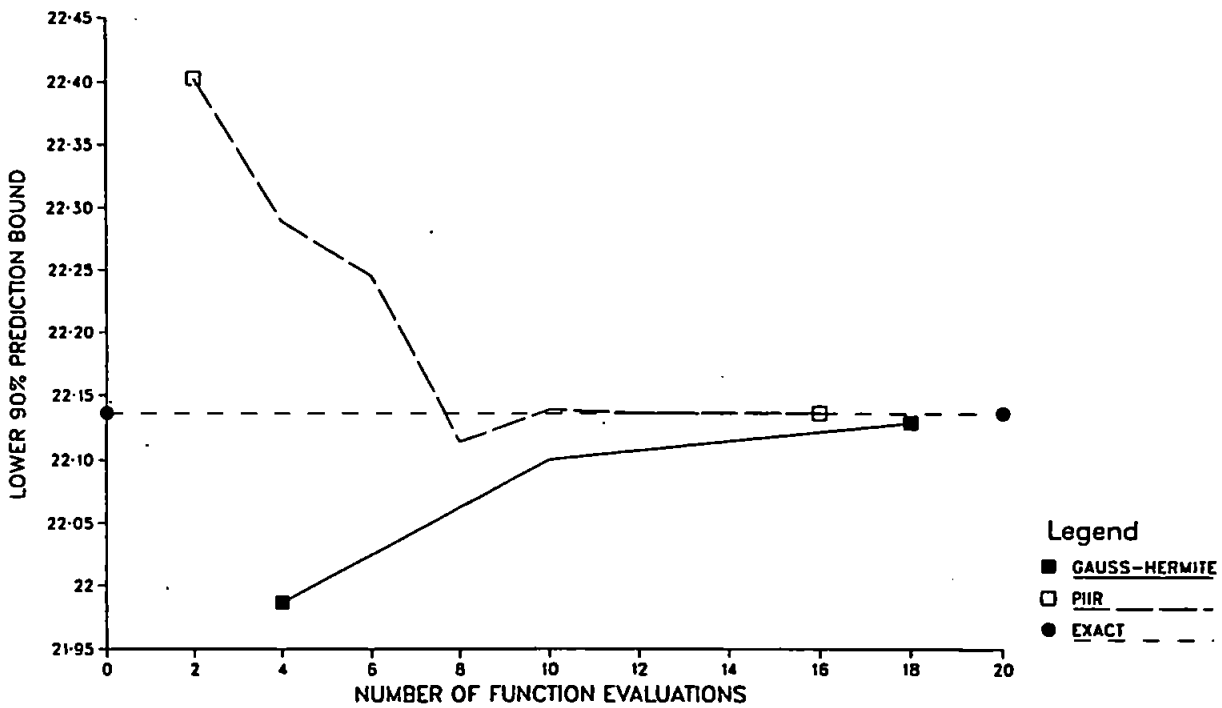


FIGURE 3.12

EXAMPLE 3 WITH PRIOR(2)



have been performed.

The results indicate that the PIIR's give very poor initial approximation, but converge reasonably fast. The advantage of PIIR's over the pseudo-Patterson (and of course the Patterson) method is clear in that convergence can be monitored more easily due to the larger number of steps. We note, that the sequence of PIIR's used for these examples is only one of many sequences which can be derived from a 16 point base rule, and that the optimal (or near-optimal) sequence is difficult to define. In general the choice is dependent mainly on the nature of the integrand.

3.4.2 An artificial example involving one dimension

Naylor (1982, Section 3.3.2) presented an artificial example for the demonstration of the adaptive iterative scheme as described by Naylor and Smith (1982). He considered an exponential distribution with p.d.f.

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

which, for a sample of size $n = 5$ with sample mean 1 gives a likelihood function of the form

$$\ell(\underline{x}|\theta) = \theta^{-5}e^{-5/\theta} .$$

The prior density was taken as

$$p(\theta) \propto \theta^{-1} .$$

Naylor (1982) applied the adaptive iterative scheme of Naylor and Smith (1982) after a log transformation. Starting values for the iteration were obtained using maximum likelihood estimates. In order to demonstrate the effectiveness of PIIR sequences we applied two such sequences obtained from a 16 point Gauss-Hermite rule and a 36 point Gauss-Hermite rule.

With the 16 point rule a sequence of imbedded rules was obtained by successively deleting, in size order, symmetric nodes 8, 6, 7, 5, 4, 3 and 2. With the 64 point rule a sequence of imbedded rules was obtained by successively deleting symmetric nodes 18, 14, 17, 16, 12, 15, 10, 13, 8, 11, 9, 7, 6, 5, 4, 3 and 2. The results together with those taken from Naylor (1982) are given in Table 3.5.

From Table 3.5 it can be seen, that in this particular example, the imbedded integration rules perform extremely well. The approximations from the two sequences of PIIR's show superior convergence after 16 function evaluations than the adaptive integration rule does after 126 function evaluations. Of course this is a particularly well behaved example, but it does illustrate that in some cases where good initial estimates are available sequences PIIR's provide an attractive alternative to the adaptive integration rules of Naylor and Smith (1982).

TABLE 3.5

Comparison of Adaptive Integration Strategy with PIIR sequences

(a) Results from Naylor (1982)

Iteration	Grid Size	Cum. no. of function evaluations	Normalizing constant x 10^3	Mean	Variance
0	6	6	7.700		
1	6	12	7.674		
2	6	18	7.674		
3	6	24	7.674	1.2478	0.4831
4	7	31	7.681		
5	7	38	7.682		
6	7	45	7.682	1.2478	0.4977
.	.	.			
.	.	.			
9	8	61	7.679	1.2492	0.5061
.	.	.			
.	.	.			
12	9	96	7.680	1.2496	0.5111
.	.	.			
.	.	.			
15	10	126	7.680	1.2497	0.5144

(b) Results obtained by applying a sequence of PIIR's based on a 16 point rule

No. of Generators	Cum. no. of function evaluations	Normalizing constant x 10^3	Mean	Variance
1	2	8.908	1.0242	0.0455
2	4	7.863	1.2356	0.4178
3	6	7.744	1.2358	0.4015
4	8	7.671	1.2455	0.4760
5	10	7.680	1.2491	0.5021
6	12	7.680	1.2498	0.5148
7	14	7.680	1.2500	0.5194
8	16	7.680	1.2500	0.5207

(c) Results obtained by applying a sequence of PIIR's based on a 32 point rule

No. of Generators	Cum. no. of function evaluations	Normalizing constant x 10^3	Mean	Variance
1	2	9.095	1.0105	0.0204
2	4	8.310	1.2447	0.5823
3	6	7.695	1.2361	0.4710
4	8	7.695	1.2363	0.4693
5	10	7.670	1.2503	0.4974
6	12	7.692	1.2470	0.5044
7	14	7.677	1.2501	0.5152
8	16	7.680	1.2498	0.5167
9	18	7.680	1.2500	0.5190
10	20	7.680	1.2500	0.5202
11	22	7.680	1.2500	0.5200
12	24	7.680	1.2500	0.5204
13	26	7.680	1.2500	0.5207
14	28	7.680	1.2500	0.5208
15	30	7.680	1.2500	0.5208
16	32	7.680	1.2500	0.5208
17	34	7.680	1.2500	0.5208
18	36	7.680	1.2500	0.5208

True posterior mean = 1.2500 and variance = 0.5208

4.1 Introduction

In chapter 2 we described the Gaussian product formulae which have been used by Naylor and Smith (1982) for the approximation of d-dimensional integrals. In this chapter, we consider the construction and application of imbedded sequences of multidimensional PIIR's which can be derived in the same manner as the one-dimensional sequences of PIIR's described in chapter 3.

The application of conventional numerical methods to Bayesian analysis is unfortunately not an easy task, especially in our context of multidimensional integration over the high dimensional space exploiting properties of product rules. The reason is simple. Traditionally, the numerical analysts focussed on efficient integration rules, and made objective comparisons using the degree of precision of each rule (see section 2.2.1). Thus, rules which achieve maximum precision with least number of function evaluations (sometimes called optimal rules) attracted much of the research interests and therefore the (sub-optimal!) Gauss-product rules were disregarded. There is no paper that we are aware of which contains any discussion of the product rules in connection with the special properties which make them fully symmetric integration rules. Yet this is the basic property which we exploit to produce imbedded sequences of rules.

This chapter reviews the basis properties of multidimensional fully symmetric integration rules. Adopting a similar notation to our main

reference paper by Rabinowitz *et al* (1987), we present, in the next section, some basic definitions concerning integration rules and, in particular, fully symmetric integration rules. The exploitation of the special properties of these fully symmetric rules will enable the construction of imbbeded sequences of multidimensional PIIR's in section 4.3, and, their application on some high dimensional Bayesian analysis in section 4.4.

More details concerning the connection between the numerical analysis and the Bayesian statistics will be given in chapter 5.

4.2 Fully Symmetric Integration Rules

The imbedded sequences of integration rules dealt with in this chapter are obtained from Gauss-Hermite product rules with the same number of points in each dimension. These rules are appropriate because they are fully symmetric integration rules. The general class of fully symmetric integration rules have received a considerable amount of attention in the numerical analysis literature; see, for example, Lyness (1965), McNamee and Stenger (1967), Rabinowitz and Richter (1969), Mantel and Rabinowitz (1977), Keast and Lyness (1979) and Genz (1986). This section gives the basic definitions of fully symmetric integration rules and some of their important properties. We begin with definitions of fully symmetric points, sets and functions.

Two points $\underline{x}$ and $\underline{y}$ of the n-dimensional Euclidean space E^n are said to be fully symmetric, denoted $\underline{x} \sim \underline{y}$, if $\underline{y}$ can be obtained from $\underline{x}$ by permutation and/or changes in sign of the coordinates of $\underline{x}$. In

passing we note that the relation '$\sim$' is an equivalence relation and in any equivalence class there are at most $n!2^n$ points. A subset of points $F_n \subseteq E^n$, is called a fully symmetric set if $\underset{\sim}{x} \in F_n$ and $\underset{\sim}{x} \sim \underset{\sim}{y}$ implies $\underset{\sim}{y} \in F_n$. A function g is said to be fully symmetric if $g(\underset{\sim}{x}) = g(\underset{\sim}{y})$ whenever $\underset{\sim}{x} \sim \underset{\sim}{y}$. Examples of fully symmetric sets are the n-dimensional unit sphere, the n-dimensional unit cube and the entire space. We consider integration rules of the form

$$If = \int_{R_n} \cdots \int k(\underset{\sim}{x}) f(\underset{\sim}{x}) d\underset{\sim}{x} \simeq \sum_{i=1}^{m} w_i f(x_{i1}, x_{i2}, \ldots, x_{in}) = Q_m f \qquad (4.1)$$

in which the set of nodes are fully symmetric and the weight function comprises a fully symmetric function defined on the set of nodes. Such fully symmetric integration rules can be completely specified by a set of M generators $\underset{\sim}{y}_i$, i=1,2,...,M, (the unique representatives of an equivalence class) and the corresponding weights $w_i^{(M)}$ (i=1,2,...,M). Each of the M generators $\underset{\sim}{y}_i$ defines a fully symmetric set of nodes $(\underset{\sim}{x}:\underset{\sim}{x} \sim \underset{\sim}{y}_i)$ with the same weight $w_i^{(M)}$. Thus, a fully symmetric integration rule can be written in the form

$$Q_M f = \sum_{i=1} w_i^{(M)} \sum_{FS} f(\underset{\sim}{y}_i) \qquad (4.2)$$

where $\sum_{FS}$ denotes the sum over all fully symmetric points which can be obtained from $\underset{\sim}{y}_i$.

Table 1 gives the types of three dimensional generators and their corresponding number of points in their representative FS set. Thus, in three dimensions, Qmf will have the form

TABLE 4.1

Types of three-dimensional generators.

Types	Generators		Number of points in a set
$(0,0,0)$	$(0,0,0)$		1
$(\alpha,0,0)$	$(a_i,0,0)$	$i = 1,\ldots,k_1$	6
$(\beta,\beta,0)$	$(\beta_i,\beta_i,0)$	$i = 1,\ldots,k_2$	12
$(\gamma,\delta,0)$	$(\gamma_i,\delta_i,0)$	$i = 1,\ldots,k_3$	24
$(\epsilon,\epsilon,\epsilon)$	$(\epsilon_i,\epsilon_i,\epsilon_i)$	$i = 1,\ldots,k_4$	8
(ζ,ζ,n)	(ζ_i,ζ_i,n_i)	$i = 1,\ldots,k_5$	24
(θ,μ,λ)	(θ_i,μ,λ_i)	$i = 1,\ldots,k_6$	48

$$Qmf = \sum_{i=1}^{k_0} w_{i0} f(0,0,0) +$$

$$+ \sum_{i=1}^{k_1} w_{i1} \sum_{FS} f(a_i,0,0) + \sum_{i=1}^{k_2} w_{i2} \sum_{FS} f(\beta_i,\beta_i,0)$$

$$+ \sum_{i=1}^{k_3} w_{i3} \sum_{FS} f(\gamma_i,\delta_i,0) + \sum_{i=1}^{k_4} w_{i4} \sum_{FS} f(\epsilon_i,\epsilon_i,\epsilon_i)$$

$$+ \sum_{i=1}^{k_5} w_{i5} \sum_{FS} f(\zeta_i,\zeta_i,n_i) + \sum_{i=1}^{k_6} w_{i6} \sum_{FS} f(\theta_i,\mu_i,\lambda_i)$$

where $\sum_{j=0}^{6} k_j = m$.

It is notable that Gauss-Hermite product rules with the same number of points in each dimension are fully symmetric. The number of generators in such a rule can readily be derived using a combinatorial argument. For a d dimensional rule with n non-negative points in each dimension, the number of generators is given by $M = {}^{n+d-1}C_d$. Thus, for example, a 5^5 product rule can be expressed as a fully symmetric rule with $M = {}^{3+5-1}C_5 = 21$ generators.

Two particularly important properties of the fully symmetric integration Qmf rule are as follows.

(i) If f is a monomial containing an odd power of a coordinate variable then $Q_M f = If = 0$.

(ii) If f is a monomial with only even powered coordinates then If and Qmf depend only on the exponents and not on the ordering of the coordinates.

It is clear from (i) that if a fully symmetric integration rule is exact for all monomials up to degree 2k it is exact to degree 2k+1. Moreover, (ii) provides a means of deriving the weights from a set of m generators specified in advance. Given m generators we obtain weights $w_i^{(m)}$ (i=1,2,...,m) which integrate exactly m monomials of the form

$$\underline{x}^{\underline{j}} = x_1^{j_1} x_2^{j_2} \ldots x_d^{j_d} \qquad j_1 \geq j_2 \geq \ldots \geq j_d \geq 0. \quad (j_i \text{ even})$$

Following Rabinowitz et al (1987) we adopt an ordering of the monomials in which $\underline{x}^j$ precedes $\underline{x}^k$ if the degree of $\underline{x}^j$ is less than $\underline{x}^k$ or in cases where $\underline{x}^j$ and $\underline{x}^k$ are of the same degree $\underline{x}^j$precedes $\underline{x}^k$ if for the first p(p=1,2,...) on which j_p and k_p differ $j_p < k_p$. For example, in three dimensions this results in the ordering 1, x_1^2, $x_1^2x_2^2$, x_1^4, $x_1^2x_2^2x_3^2$, $x_1^4x_2^2$, x_1^6, ...

Given a set of $m \leq M$ generators (which without loss of generality we denote by $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_m$) defining as imbedded rule we obtain weights $w_i^{(m)}$ by solving the system

$$\sum_{i=1}^{m} w_i^{(m)} \sum_{\underline{x} \sim \underline{y}_i} I\underline{x}^{je} = I\underline{x}^{je}, \qquad e = 1,2,\ldots,m', \qquad (4.3)$$

and making the rule exact for the first m' monomials. These m' monomials being the first which give a unique set of weights. Note here that $m' > m$ if the first m monomials lead to a dependent system of equations.

4.3 Construction of Imbedded Integration Rules

Rabinowitz et al (1987) gave the following theorem.

Theorem 2: Given an integration rule with M generators and positive weights which is exact for the first M monomials, then there exists a rule with M-1 generators and non-negative weights which is exact for the first M-1 monomials.

Theorem 2 guarantees the existence of at least one sequence of M imbedded rules. Furthermore, (4.3) provides the basis for an algorithm for obtaining sequences of imbedded rules by working from an M-generator base rule and creating sequentially M-1, M-2, ..., 1 generator imbedded rules. We will use this theorem to construct an imbedded sequence of PIIR's based on a product Gauss-Hermite rule. As we remarked in the previous section, this rule satisfies all conditions to be converted from (4.1) to (4.2), but the solution of (4.3) in this case is a quite complicated problem and its implementation requires the adoption of a specific strategy.

The weights in a Gauss-Hermite product rule can be readily obtained by multipling the corresponding one dimensional weights. However, we must construct a system of linear equations of the form (4.3) in order to derive the imbedded sequence of integration rules. The solution of (4.3) with all points of the product rule serves as a check on the algorithm before we proceed creating imbedded rules. The construction and solution of (4.3) with the full set of generators in the base rule is the first step of our strategy.

We noted in chapter 2 that d-dimensional Gauss-Hermite product rules with n points in each dimension can integrate all monomials with terms $\prod_{i=1}^{d} x_i^{a_i}$ with $a_i \leq p$ $(i=1,2,\ldots,d)$. We also remarked in the previous section that the number of generators for that rule is ${}^{n+d-1}C_d = M$, where n denotes the number of non-negative points in each dimension. It is simple to show that the number of monomials that this rule can integrate is ${}^{2n+d-1}C_d > M$. Our task is to find the first M monomials which make the matrix of coefficients of the system (4.3)

non-singular. In addition, it is known (see for example Davis and Rabinowitz (1984)) that this system has a unique solution. Therefore, these first M monomials give a unique set of weights.

When this is accomplished, the next step is to drop one generator (or, equivalently, one column of the matrix of coefficients of the system (4.3)) and try to solve the system of M-1 equations. Following Rabinowitz et al (1987) our strategy for obtaining an M-1 generator imbedded rule from an M generator rule is to drop the most expensive generator (ie. the one which generates most points) which leads to a positive rule. In this case, care must be taken when we try to find the set of the M-1 independent equations which make the matrix nonsingular: we do not want to integrate monomials with higher degree and omit monomials with less degree. Our strategy therefore is to find the <u>first</u> M-1 monomials which make the matrix of coefficients of the system nonsingular. Similarly, we proceed for the construction of M-2,M-3,...,2 generator imbedded rules.

A numerical algorithm to implement the above is as follows:

(i) Create an ordered list of monomials that the full product rule can integrate exactly using the ordering described above.

(ii) Derive the RHS of (3) using the formula

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{2i_1}x_2^{2i_2}\cdots x_n^{2i_n}\exp(-x_1^2+x_2^2+\ldots+x_n^2)\,dx_1dx_2\ldots dx_n =$$

$$= \Gamma(i_1+1/2)\,\Gamma(i_2+1/2)\ldots\Gamma(i_n+1/2)$$

(iii) Derive a system (4.3) produced by the fist M monomials as they are ordered in (i). Start solving the system using total pivoting with scaling to avoid round-off errors. (For more details in these methods see Steinber (1974) or Atkinson (1978)). If the matrix is singular, the total pivoting will stop when the first k_1 diagonal elements are non-zero and the last (bottom) $M-k_1$ rows zero.

(iv) Remove the last $M-k_1$ equations of the system and replace them from the equations derived of the $M+1, M+2, \ldots, M+k_1$th monomials, as they are ordered in (i). Continue the total pivoting starting from the $k+1$th row.

(v) Repeat step (iv), say i times, until $k_i=M$. Then, solve the system and check whether the solution produces the Gauss-product weights for verification of the method.

For the $M-1, M-2, \ldots, 2$ generator rule, we remove each time from the system of equations one generator starting from the most expensive and try to solve the $(M-1)\times(M-1), (M-2)\times(M-2), \ldots, 2\times2$ system of equations. If the solution produces a set of positive weights we stop and proceed to the next rule. Otherwise we try to remove the next most expensive generator.

One important factor of the strategy is that we are able to record in

each step the exponents of the last monomial integrated exactly by the rule. We can therefore 'judge' the power of each imbedded rule because we know that at least all monomials ordered before that can be integrated exactly by the rule. Of course, there are more monomials integrated exactly which have been ordered after the last one and have been removed during the above numerical algorithm. In the next section we will specify exactly which are these monomials.

Even though the fully symmetric rules were used in the past to produce integration rules with specified accuracy, (or degree of precision), the above remarks indicate that sequences of PIIR's can be used to produce rules which lie between two rules of precision, say, p-2 and p. In the context of numerical analysis, certainly the reason for constructing rules of specified accuracy is concerned with the comparison of different methods. However, we believe that these rules are potentially useful, especially in a Bayesian analysis, since they fill a gap by providing a rich class of positive integration rules over the d-dimensional space.

We need to mention that in contrast with the linear algebra context, the matter of 'zero elements' in a numerical solution of a large system of linear equations requires considerable amount of attention. We specified a 'tolerance', which is a adequately small number (say 0.1e-6) which could serve as the smallest positive number. The sensitivity of our solution was tested repeatedly and we found that the algorithm is very stable in the sense that it produces same solution over a wide range of tolerance.

An alternative way to work out steps (iii) and (iv) above is to use

procedures (see, for example NAG (1987)) which can calculate the number of non-zero singular values and thus the number of independent equations in the system (or, equivalently, the rank of the system). The way to proceed with this method is to start with two rows and add one row at a time, calculating the new rank at each stage. If the rank does not increase by one, then that row is not needed. This method of course has the disadvantage that it is not theoretically founded: Theorem 2 guarantees the existence of an imbedded sequence of PIIR's starting from the large rule and dropping one generator at a time. The reverse procedure could create the unfortunate problem of having to stop at a point where the sequence does not produce a positive rule. However, this method can serve as a way to derive some PIIR's in cases where the whole sequence is not needed.

Using the above algorithm , we have produced imbedded sequences of PIIR's based on $-5^3, 9^3, 13^3, 5^5, 5^7, 5^9$ Gauss-Hermite product rules. In the next section we illustrate how these sequences can produce efficient results within the Bayesian framework.

4.4 Properties of the imbedded sequences

4.4.1 Related results from Numerical analysis theory

The idea of dropping a node from an one-dimensional Gauss-quadrature formula was suggested by Berntsen and Espelid (1984) and Laurie (1985). This idea was applied to circular symmetric planar regions by Cool and Haegemans (1987) and to the cube by Bernten and Espelid

(1988). Rabinowitz *et al.* (1987) investigated the existence of such rules with certain optimal properties, and recently Cools and Haegemans (1989) and Cools (1989) have explored further properties and presented a method for the construction of multidimensional imbedded rules, based on invariant theory and ideal theory.

Even though the theoretical work by Cools and Haegemans (1989) is a theoretical generalisation of the work we presented in section 4.3, it is worthwile describing here some properties of the imbedded sequences of PIIR's obtained from the Gauss-Hermite product rules. We have mentioned in section 2.2.1 that the product rules are relatively 'strong' in the sense that they can integrate many more monomials than other rules with the same degree. This is a very important property which we generalise here for the whole sequence of the imbedded rules derived in section 4.3.

Suppose we have an n-dimensional Gauss-Hermite product rule of degree 2m-1. Then the nodes are based on the zeros of the Hermite polynomials of degree m, say $H_m(x)$ (see David and Rabinowitz (1984). Let $H_m(x_j)$ be the Hermite polynomials on x_j, $j=1,2,\ldots,n$ and X a polynomial of degree $\prec$ m. Then, the following result holds:

Result 1: If an integration rule based on the roots of $H_m(x_j)$ is exact for all polynomials of degree $\leq$ deg(X) + m-2, then the rule is also exact for Xx_j^m.

Proof: $H_m(x_j) = x_j^m + \Phi(x_j)$, where Φ is a polynomial of degree $\leq$ m-2. Let If be the integral of the function f as defined in (4.1). Then,

$$IXH_m(x_j) = 0 \quad \text{for each } X \quad \text{(see Engels (1980), p. 239)}$$

Thus,

$$IXx_j^m = -IX\Phi(x_j)$$

Let Q be an integration rule based on the Hermite-points, then

$$QXH_m(x_j) = 0 \Rightarrow QXx_j^m = -QX\Phi(x_j)$$

Consequently, if the integration rule is exact for all polyomials of degree $\leq \deg(X) + \deg(\Phi)$, then it is also exact for Xx_j^m because

$$QXx_j^m = -QX\Phi(x_j) = -IX\Phi(x_j) = IXx_j^m$$

As an example, consider the case n=2:

Choose X=1, then we have that if the formula is exact for all monomials of degree $\leq m-2$, then it is also exact for x_1^m and x_2^m.
Choose $X=x_1^i x_2^j$, $i+j<m$. Then if the formula is exact for all polynomials of degree $\leq m-2+i+j$, then it is also exact for $x_1^{m+i}x_2^j$ and $x_1^i x_2^{m+j}$.

Result 1 is useful because it enables us to obtain a precise picture of the power of each rule by examining the monomials that are integrated exactly by it.

4.4.2 Practical error estimation

In the second edition of the book by Davis and Rabinowitz (1984) there is a section on practical error estimation containing references to several papers on the topic. The error analysis associated with quadrature rules is usually based on a study of the derivative of the argument. See for example Engels (1980), chapter 3. A more practical approach is to compare approximations obtained by different rules. Imbedded sequences of integration rules have an important role to play in this area because of their attractive property to 'overlap' on the set of nodes they use.

Let If and $Q_M f$ denote the integral and an M-quadrature integration rule of the function f as given in (4.1) and (4.2) respectively. We are interested in producing estimates of the error

$$Ef = If - Qf.$$

Normally an estimate for Ef is produced by applying two quadrature rules Q_M and Q_L, with $M>L$, and an estimate of the error in the approximation given by Q_M is given by

$$E_M f = |Q_M f - Q_L f| \qquad (4.4)$$

The hope is that

$$|If - Q_M f| \leq E_M f \qquad (4.5)$$

which in turn guarantees that the quadrature routine is reliable.

However, because the approximation (4.2) and the error estimate (4.4) are based on only a finite number of function values, they are not completely reliable. In practice one does not know that $Q_M f$ is more accurate than $Q_L f$, even if Q_M has a higher degree than Q_L.

The assumption behind the expression (4.4), see De Boor (1971), is that

$$E_M f \approx |If - Q_L f| > |If - Q_M| \qquad (4.6)$$

If (4.6) is satisfied, then (4.5) would be satisfied.

Another important concept in the discussion of error estimation was introduced by Lyness (1965). He defined Null rules of degree L as

$$N_L = Q_M - Q_L \qquad (4.7)$$

If we apply N_L on any polynomial of degree less or equal to L, the result will be zero. Very often the the error estimate (4.4) is scaled by some factor λ in order to balance reasonably between efficiency and reliability. It is then easy to prove that λEf is also equal to the difference of two integration rules of the same degree as the original ones:

$$\lambda Ef = \lambda Q_M f - \lambda Q_L f = Q_M f - (\lambda Q_M f + (1-\lambda) Q_L f) = Q_M f - Q'_M f$$

where $Q'_M f$ is a linear combination of $Q_M f$ and $Q_L f$ and therefore of degree equal to the degree of Q_L.

The question which is of great importance, relates to the choice of a suitable λ, or, different values of λ that will provide a set of Null rules. In section 5.6 we will dicuss the importance of the integration of such rules in the proposal of our integration scheme for Bayesian analysis.

4.5 Illustrative examples

4.5.1 A 5-dimensional imbedded sequence of PIIR's

The data in Table 4.2 were analysed by Grieve (1987) who applied a Bayesian analysis using a Weibull regression model with proportional hazards. Using the same notation as Grieve (1987) the time to tumour, t, has p.d.f.

$$p(t/\underline{\beta},p) = pt^{p-1}e^{\underline{z}\underline{\beta}} \exp[-tPe^{\underline{z}\underline{\beta}}] , \qquad t > 0, \qquad (4.8)$$

where p is the shape parameter of the Weibull distribution, $\underline{z}$ a row vector of covariates and β a column vector of regression coefficients. For the analysis of the data in Table 2 $\underline{z} = (z_0, z_1, z_2, z_3)$ is defined as follows:

$z_0 = 1$ for all mice;

$z_1 = 1$ for mice in the vehicle control group and 0 otherwise;

$z_2 = 1$ for mice in the test substance group and 0 otherwise;

$z_3 = 1$ for mice in the positive control group and 0 otherwise.

TABLE 4.2

Photocarcinogenicity data from Grieve (1987)

I Irradiated control			II Vehicle control			III 8-MOP			IV Positive control		
Mouse no.	Week of death (censoring time)	Week of tumour	Mouse no.	Week of death (censoring time)	Week of tumour	Mouse no.	Week of death (censoring time)	Week of tumour	Mouse no.	Week of death (censoring time)	Week of tumour
1		12	1		32	1		22	1		27
2		17	2		27	2		26	2		18
3		21	3		23	3	10		3		22
4		25	4		12	4		28	4		13
5		11	5		18	5		19	5		18
6		26	6	40		6		15	6		29
7		27	7	40		7		12	7		28
8		30	8		38	8		35	8	20	
9		13	9		29	9		35	9		16
10		12	10		30	10		10	10		22
11		21	11	40		11		22	11		26
12		20	12		32	12		18	12		19
13		23	13	40		13	24		13	29	
14		25	14	40		14		12	14	10	
15		23	15	40		15	40		15		17
16		29	16	40		16	40		16		28
17		35	17		25	17		31	17		26
18	40		18		30	18		24	18		12
19		31	19		37	19		37	19		17
20		36	20		27	20		29	20		26

Following Grieve (1987), and considering the $n+m = 80$ times in Table 2 as ordered in such a way that the first $n=65$ times $t_1, t_2, \ldots, t_n$ are uncensored and the last $m=15$ times, $t_{n+1}, t_{n+2}, \ldots, t_{n+m}$, are censored (ie. corresponding to deaths) the likelihood function can be written

$$\ell(\underline{\beta}, p/\text{data}) = \left[\prod_{j=1}^{n} p t_j^{p-1} e^{\underline{z}_j \underline{\beta}} \right] \left[\prod_{j=1}^{n+m} \exp[-t_j^p e^{\underline{z}_j \underline{\beta}}] \right]$$

where $\underline{z}_j$ denotes the vector of covariates for the j^{th} mouse.

An initial imbedded PIIR sequence based on a 5^5 product rule has been derived (see table 4.3) and applied to the above example starting with maximum likelihood estimates and the associated asymptotic covariance matrix. The results obtained from the full PIIR sequence are shown for illustrative purposes in figures 4.1-4.11. They indicate that the sequence converges rapidly, and as a result it can be used to save a considerable number of function evaluations. In next chapter we shall discuss how a proposed numerical integration strategy based on this sequence of PIIR's could be applied in this example.

TABLE 4.3: Imbedded sequence of PIIR's obtained from a 5^5 Gauss-Hermite based rule

Step	No of Gener's	Generators					No of points	Sum of points	Exponents of last monomial
1	2	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	1	1	
		0.0000000	0.0000000	0.0000000	0.0000000	2.0201829	10	11	2 0 0 0 0
2	1	0.0000000	0.9585725	0.9585725	0.9585725	2.0201829	320	331	2 2 0 0 0
3	1	0.0000000	0.0000000	0.9585725	0.9585725	0.9585725	80	411	4 0 0 0 0
4	1	0.0000000	0.0000000	0.0000000	0.0000000	0.9585725	10	421	2 2 2 0 0
5	1	0.0000000	0.0000000	0.0000000	2.0201829	2.0201829	40	461	4 2 0 0 0
6	1	0.9585725	0.9585725	0.9585725	0.9585725	0.9585725	32	493	2 2 2 2 0
7	1	2.0201829	2.0201829	2.0201829	2.0201829	2.0201829	32	525	4 2 2 0 0
8	1	0.0000000	0.0000000	0.0000000	0.9585725	0.9585725	40	565	4 4 0 0 0
9	1	0.0000000	0.0000000	0.0000000	0.9585725	2.0201829	80	645	2 2 2 2 2
10	1	0.0000000	0.0000000	2.0201829	2.0201829	2.0201829	80	725	4 2 2 2 0
11	1	0.0000000	0.9585725	0.9585725	0.9585725	0.9585725	80	805	4 4 2 0 0
12	1	0.9585725	0.9585725	0.9585725	0.9585725	2.0201829	160	965	4 2 2 2 2
13	1	0.0000000	2.0201829	2.0201829	2.0201829	2.0201829	80	1045	4 4 2 2 0
14	1	0.0000000	0.0000000	0.9585725	0.9585725	2.0201829	240	1285	4 4 2 2 2
15	1	0.0000000	0.9585725	0.9585725	2.0201829	2.0201829	480	1765	4 4 2 2 2
16	1	0.9585725	0.9585725	0.9585725	2.0201829	2.0201829	320	2085	4 4 4 2 0
17	1	0.0000000	0.0000000	0.9585725	2.0201829	2.0201829	240	2325	4 4 4 2 2
18	1	0.0000000	0.9585725	2.0201829	2.0201829	2.0201829	320	2645	4 4 4 4 0
19	1	0.9585725	0.9585725	2.0201829	2.0201829	2.0201829	320	2965	4 4 4 4 2
20	1	0.9585725	2.0201829	2.0201829	2.0201829	2.0201829	160	3125	4 4 4 4 4

FIGURE 4.1

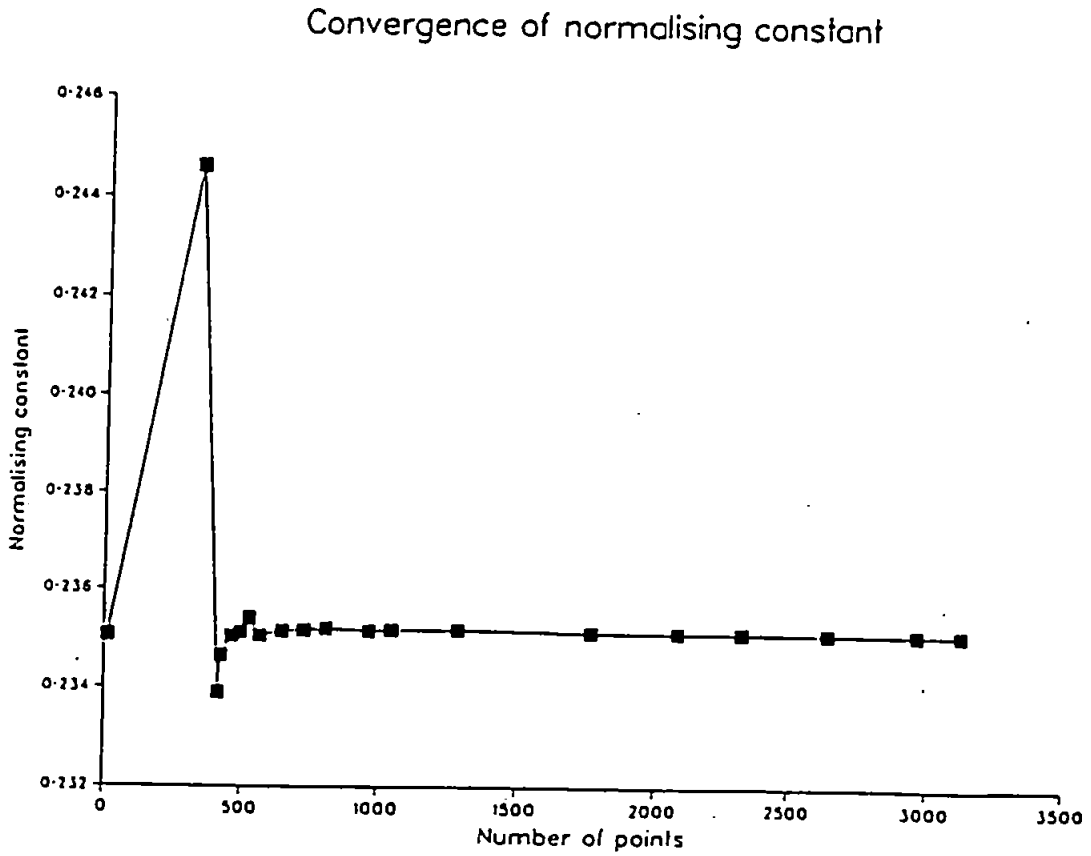


FIGURE 4.2

Convergence of posterior mean of b_0

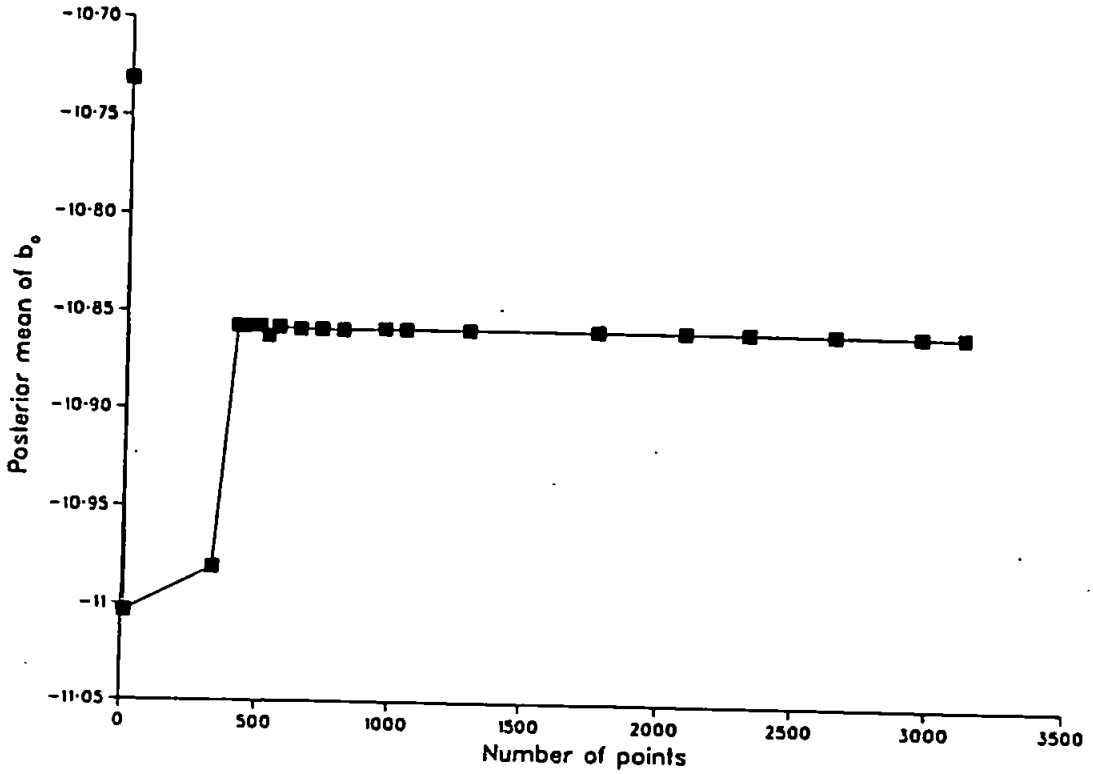


FIGURE 4.3

Convergence of posterior variance of b_0

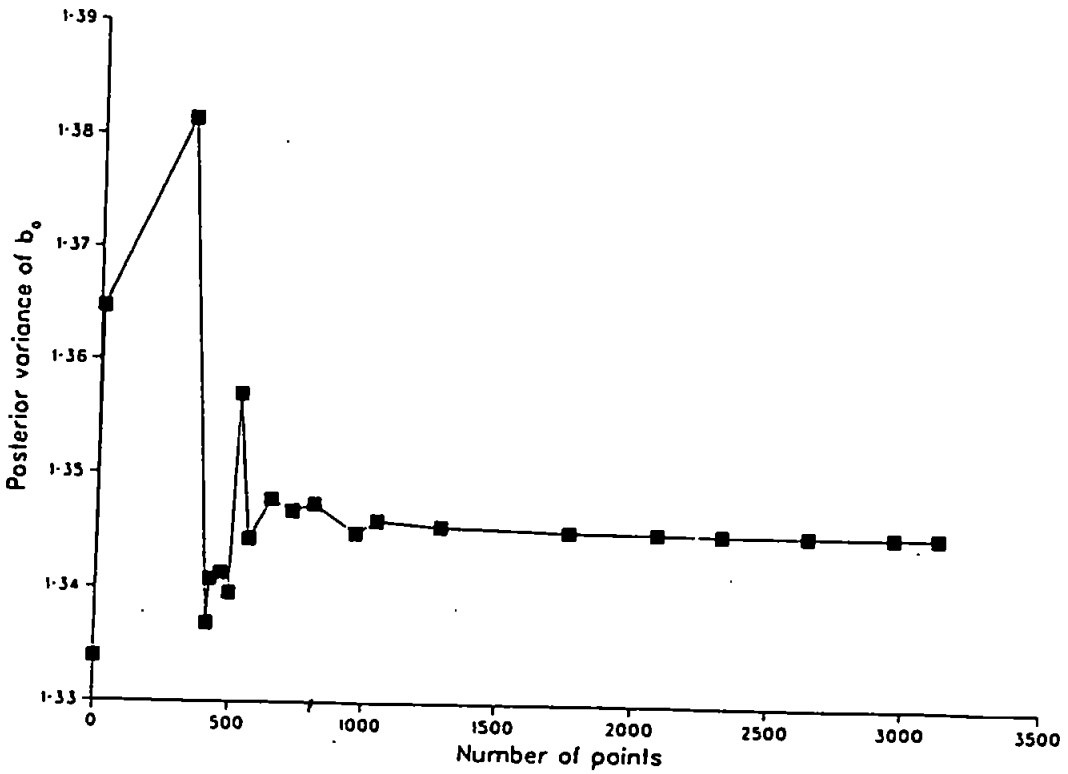


FIGURE 4.4

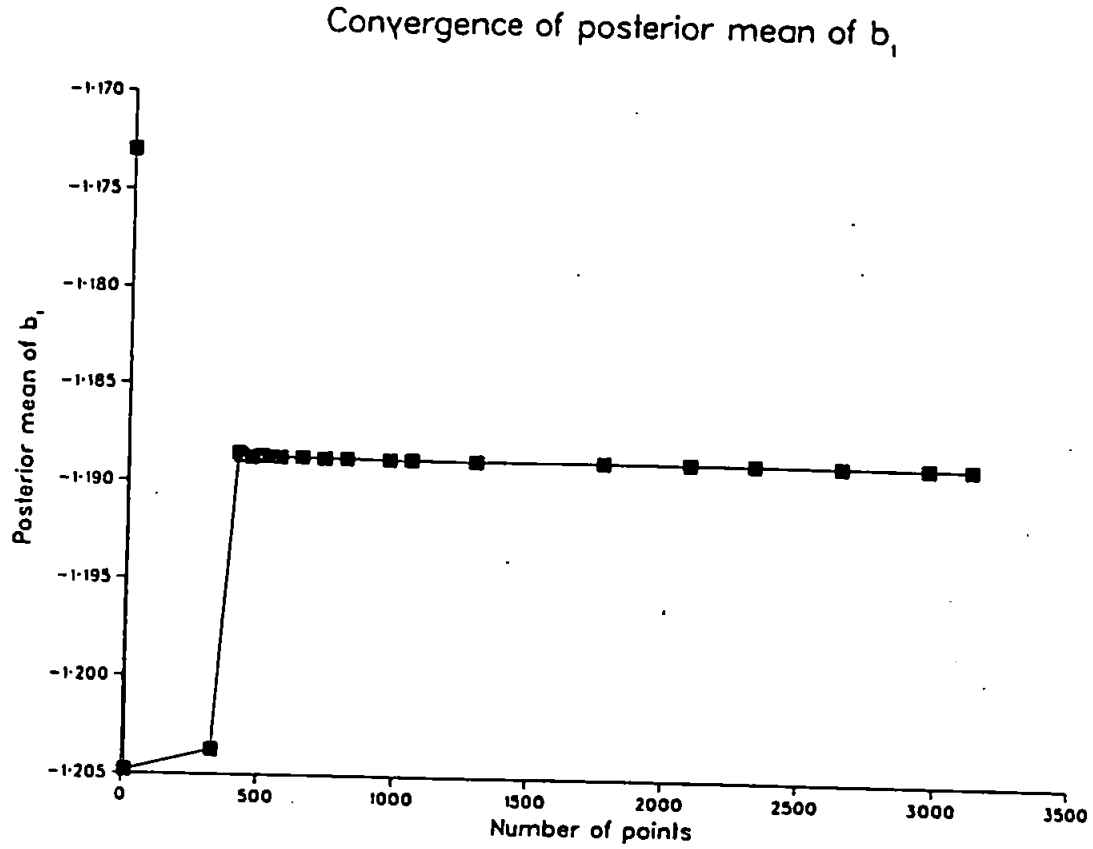


FIGURE 4.5

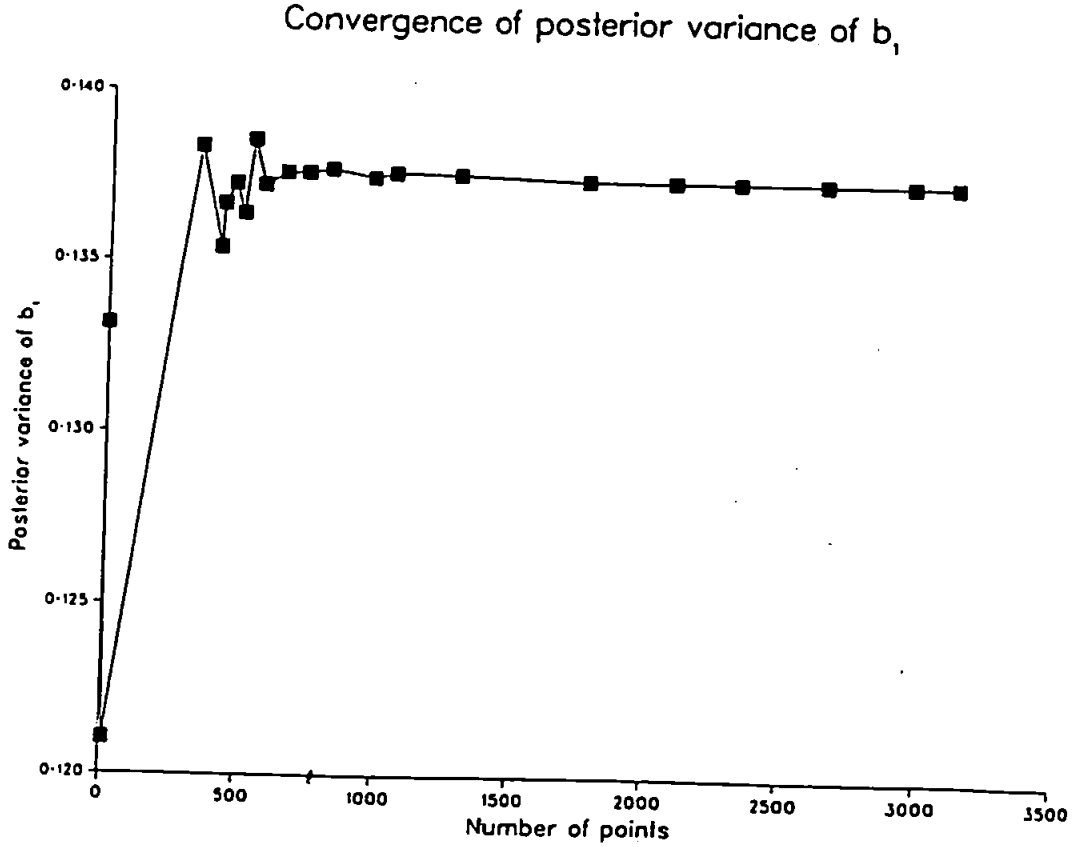


FIGURE 4.6

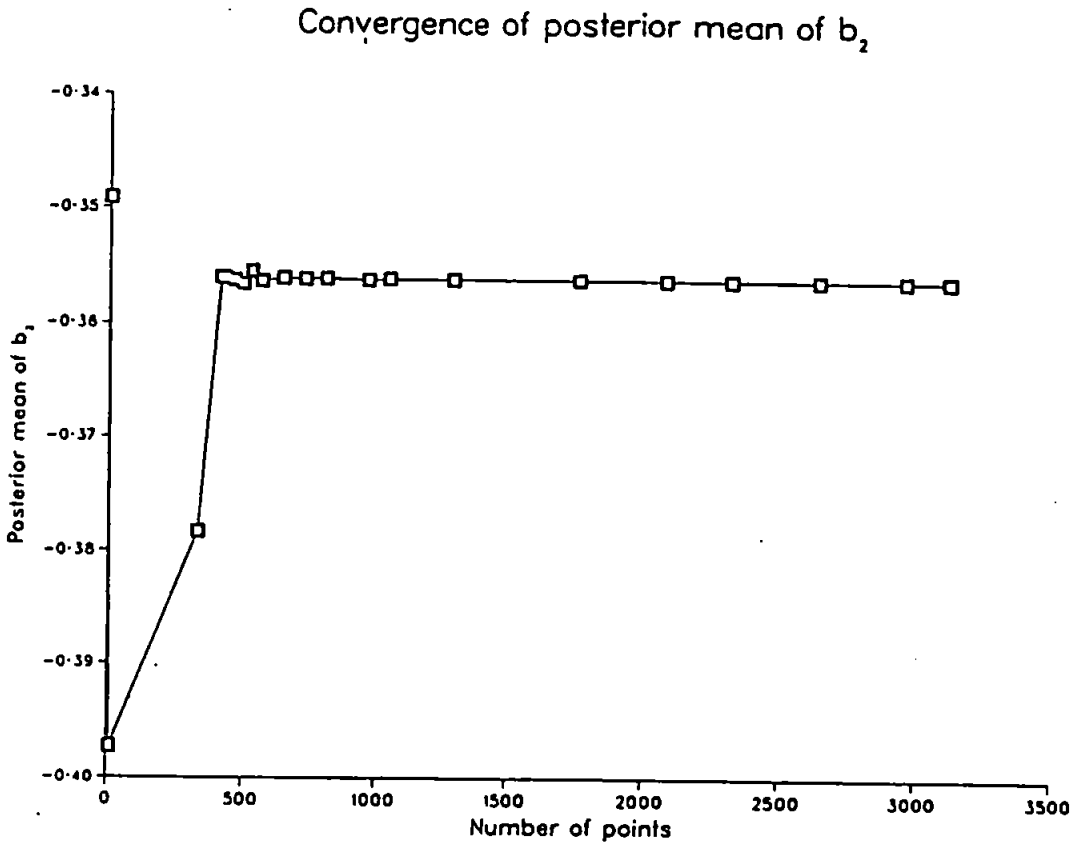


FIGURE 4.7

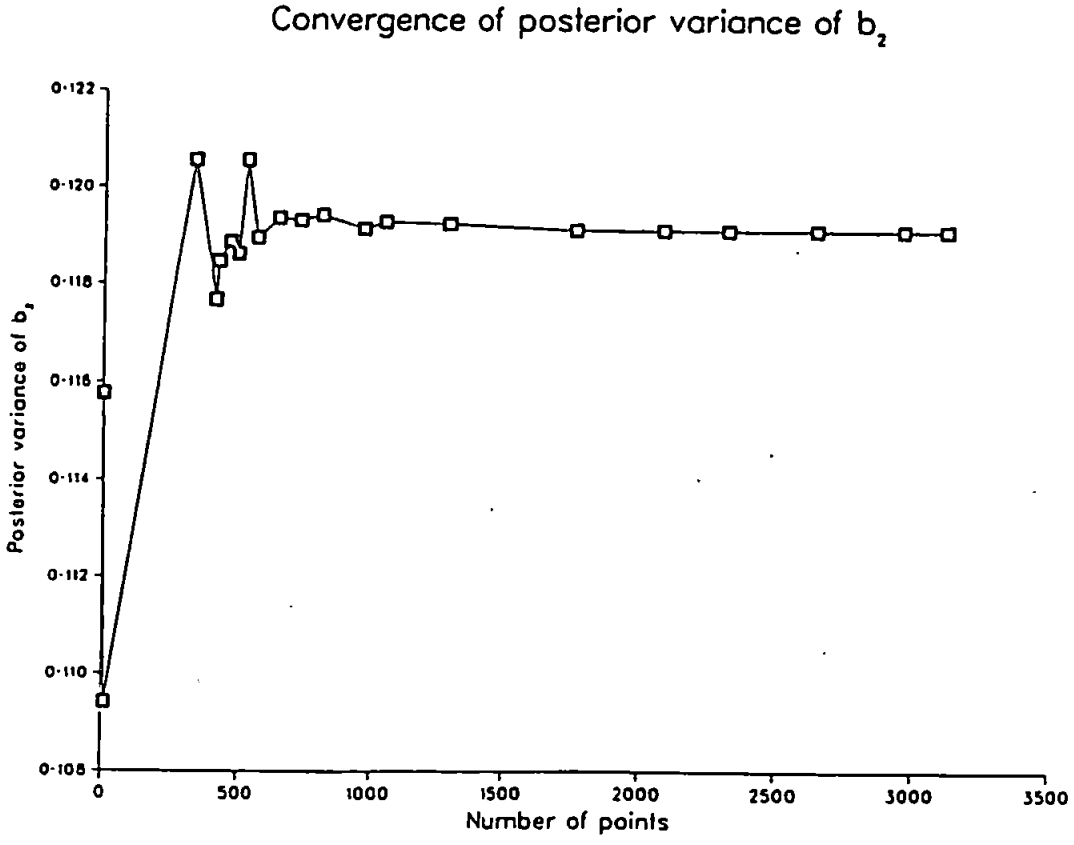


FIGURE 4.8

Convergence of posterior mean of b_3

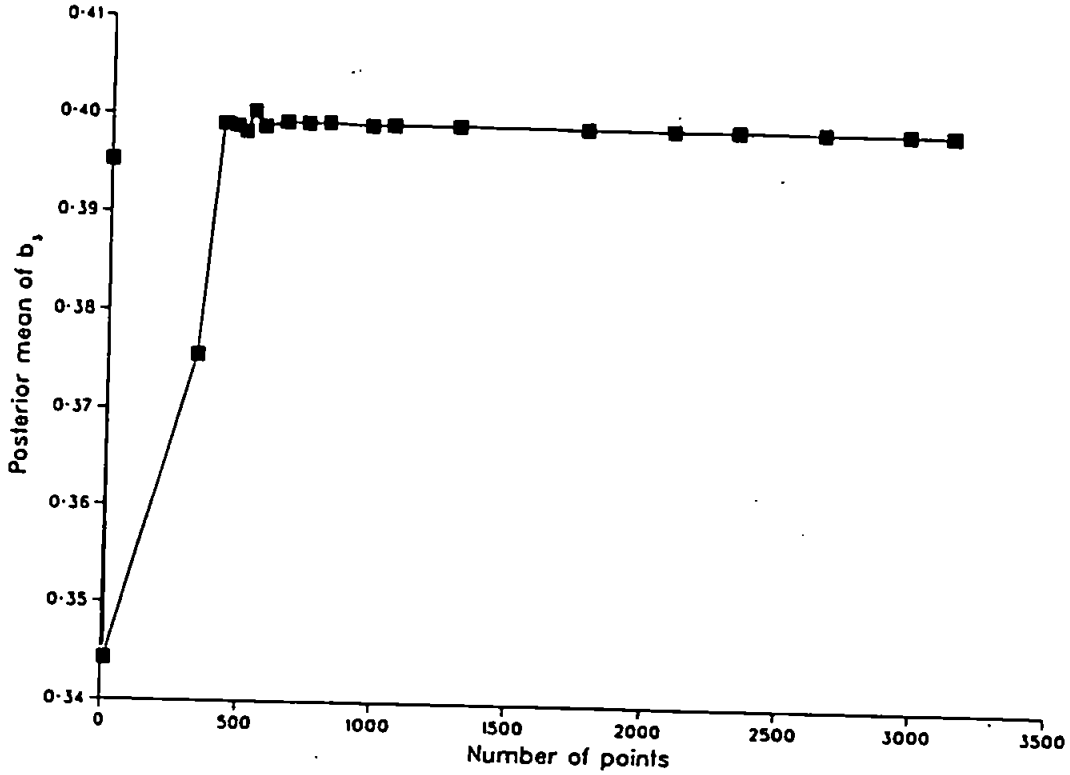


FIGURE 4.9

Convergence of posterior variance of b_3

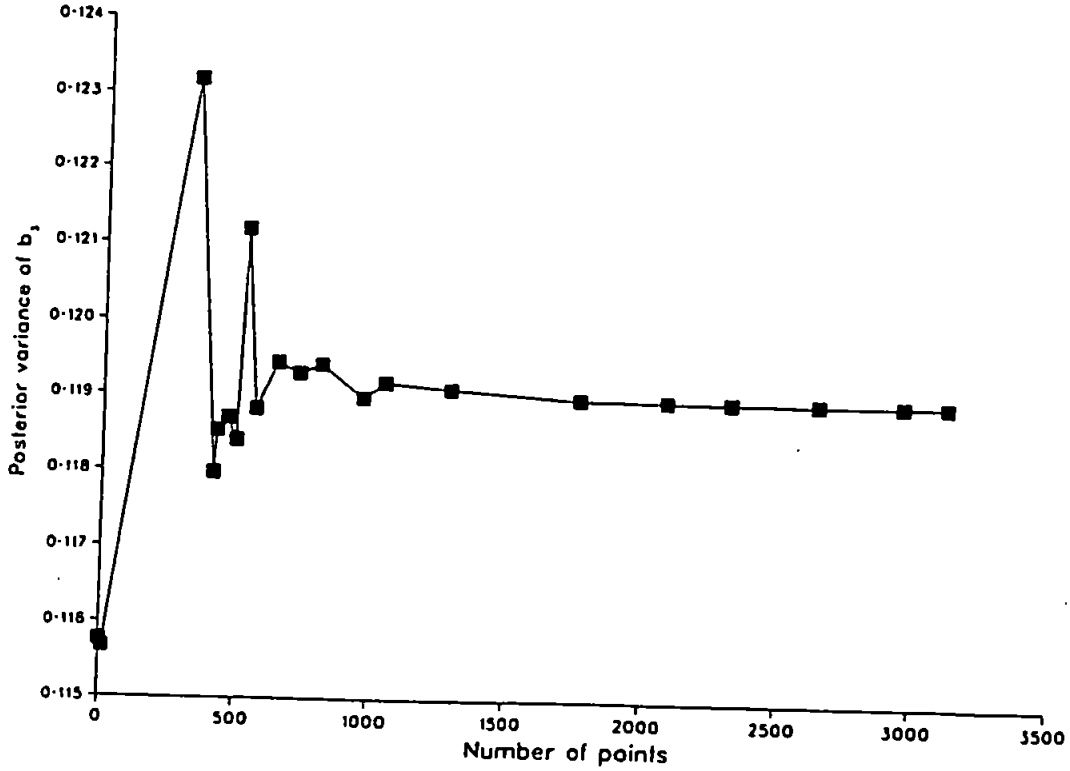


FIGURE 4.10

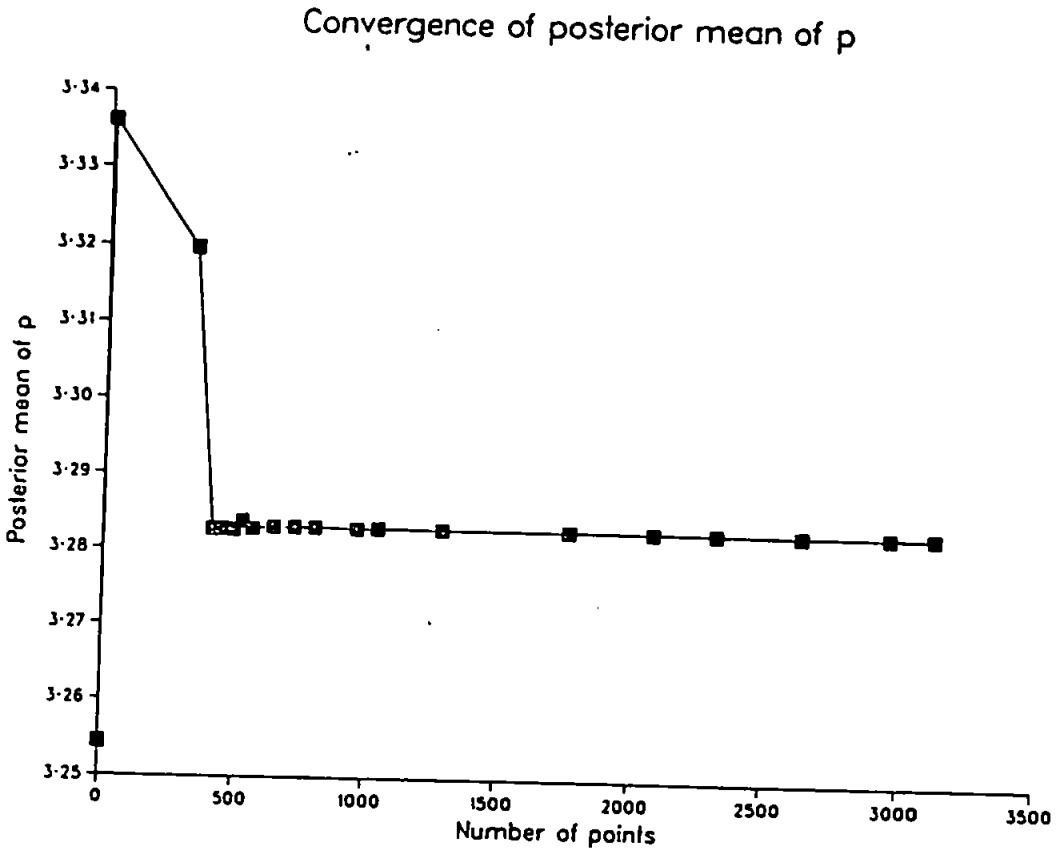
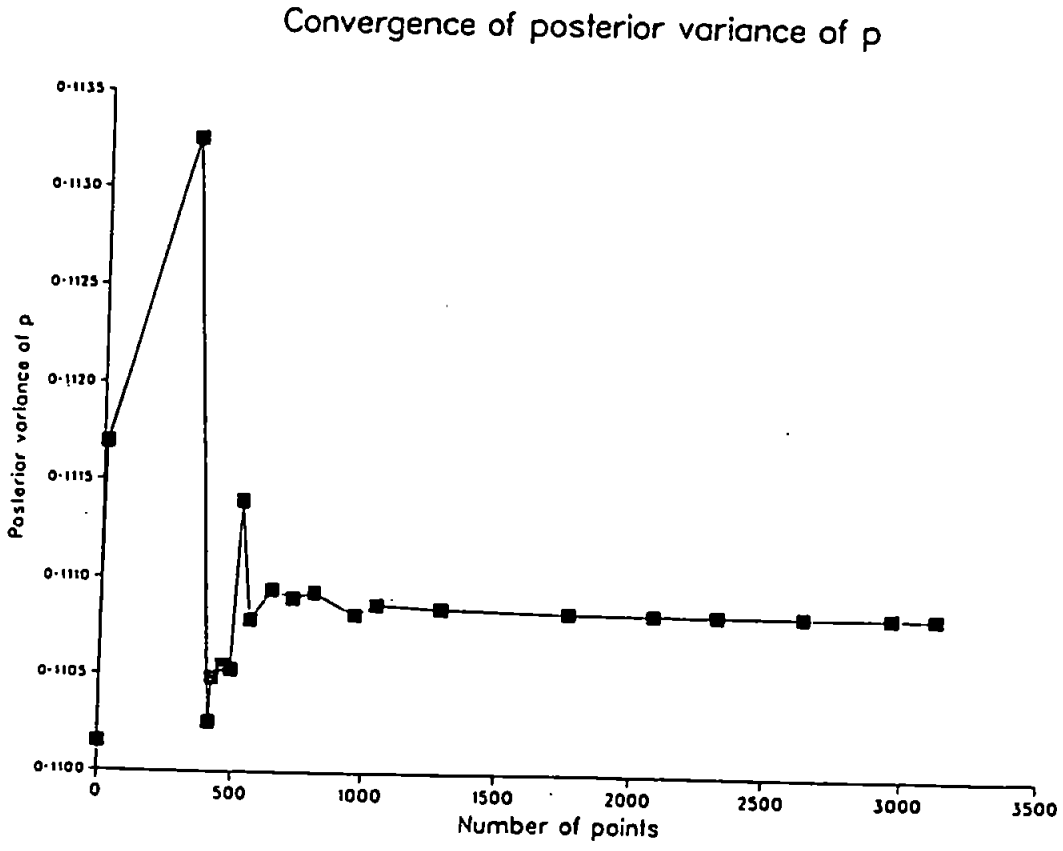


FIGURE 4.11



4.5.2 A 7-dimensional imbedded sequence of PIIR's

Lawless (1982, p.337) presented a set of data which is reproduced here in table 4.4. This consists of survival times in months and regressor variables for 65 multiple myeloma patients and it is a subset from a more comprehensive set given by Krall *et al.* (1975). The problem is to relate survival times for multiple myeloma patients to a number of prognostic variables. These prognostic variables are:

x_1 Logarithm of a blood urea nitrogen measurement at diagnosis

x_2 Hemoglobin measurement at diagnosis

x_3 Age at diagnosis

x_4 Sex : 0, male; 1, female

x_5 Serum calcium measurement at diagnosis

Asterisks denote censoring times.

We used the model (4.8) to analyse these data, the 6-dimensional vector $\underline{z}$ being defined in this case as follows:

$z_0 = 1$ for all patients

$z_1 = x_1 - \bar{x}_1$

$z_2 = x_2 - \bar{x}_2$

$z_3 = x_3 - \bar{x}_3$

$z_4 = x_4$

$z_5 = x_5 - \bar{x}_5$

We used the maximum likelihood values as our initial estimates and we applied the sequence of PIIR's based on a 5^7 Gauss Hermite product rule. The convergence of the posterior mean and variance vectors are illustrated in figures 4.12-4.25. Comparatively with the previous example the convergence is more rapid and the saving in computer labour can be really worthwile. As with the previous example, it sufficed to illustrate in this section the efficiency of the sequence of PIIR's. We shall examine the same data in the next chapter, following the description of our proposed strategy.

TABLE 4.4: Survival times and regressor variables

for multiple myeloma patients

t	x_1	x_2	x_3	x_4	x_5	t	x_1	x_2	x_3	x_4	x_5
1	2.218	9.4	67	0	10	26	1.230	11.2	49	1	11
1	1.940	12.0	38	0	18	32	1.322	10.6	46	0	9
2	1.519	9.8	81	0	15	35	1.114	7.0	48	0	10
2	1.748	11.3	75	0	12	37	1.602	11.0	63	0	9
2	1.301	5.1	57	0	9	41	1.000	10.2	69	0	10
3	1.544	6.7	46	1	10	42	1.146	5.0	70	1	9
5	2.236	10.1	50	1	9	51	1.568	7.7	74	0	13
5	1.681	6.5	74	0	9	52	1.000	10.1	60	1	10
6	1.362	9.0	77	0	8	54	1.255	9.0	49	0	10
6	2.114	10.2	70	1	8	58	1.204	12.1	42	1	10
6	1.114	9.7	60	0	10	66	1.447	6.6	59	0	9
6	1.415	10.4	67	1	8	67	1.322	12.8	52	0	10
7	1.978	9.5	48	0	10	88	1.176	10.6	47	1	9
7	1.041	5.1	61	1	10	89	1.322	14.0	63	0	9
7	1.176	11.4	53	1	13	92	1.431	11.0	58	1	11
9	1.724	8.2	55	0	12	4*	1.945	10.2	59	0	10
11	1.114	14.0	61	0	10	4*	1.924	10.0	49	1	13
11	1.230	12.0	43	0	9	7*	1.114	12.4	48	1	10
11	1.301	13.2	65	0	10	7*	1.532	10.2	81	0	11
11	1.508	7.5	70	0	12	8*	1.079	9.9	57	1	8
11	1.079	9.6	51	1	9	12*	1.146	11.6	46	1	7
13	0.778	5.5	60	1	10	11*	1.613	14.0	60	0	9
14	1.398	14.6	66	0	10	12*	1.398	8.8	66	1	9
15	1.602	10.6	70	0	11	13*	1.663	4.9	71	1	9
16	1.342	9.0	48	0	10	16*	1.146	13.0	55	0	9
16	1.322	8.8	62	1	10	19*	1.322	13.0	59	1	10
17	1.230	10.0	53	0	9	19*	1.322	10.8	69	1	10
17	1.591	11.2	68	0	10	28*	1.230	7.3	82	1	9
18	1.447	7.5	65	1	8	41*	1.756	12.8	72	0	9
19	1.079	14.4	51	0	15	53*	1.114	12.0	66	0	11
19	1.255	7.5	60	1	9	57*	1.255	12.5	66	0	11
24	1.301	14.6	56	1	9	77*	1.079	14.0	60	0	12
25	1.000	12.4	67	0	10						

FIGURE 4.12

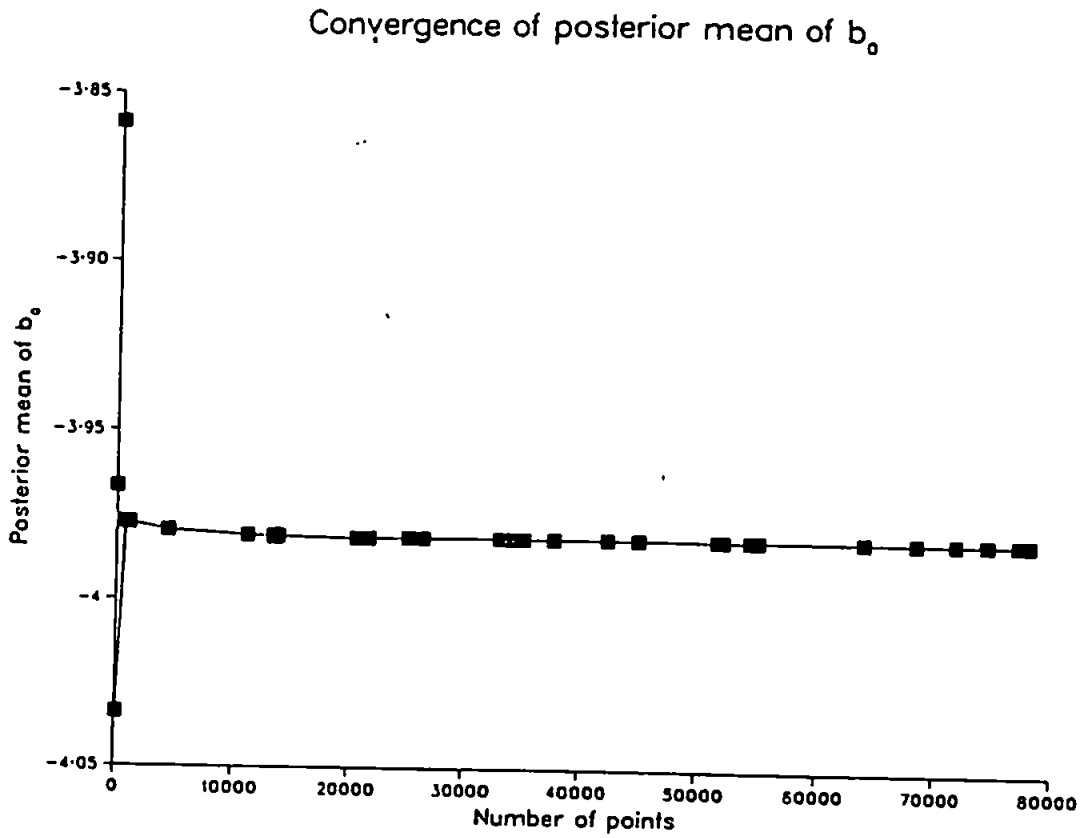


FIGURE 4.13

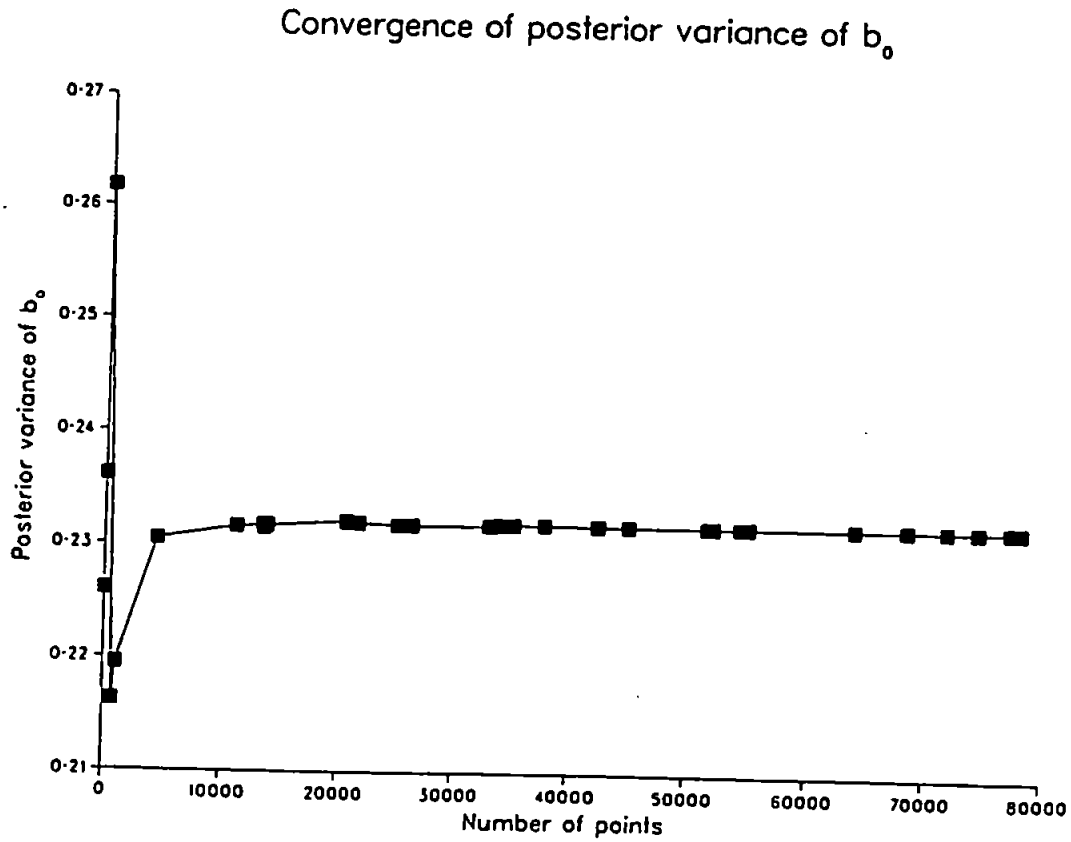


FIGURE 4.14

Convergence of posterior mean of b_1

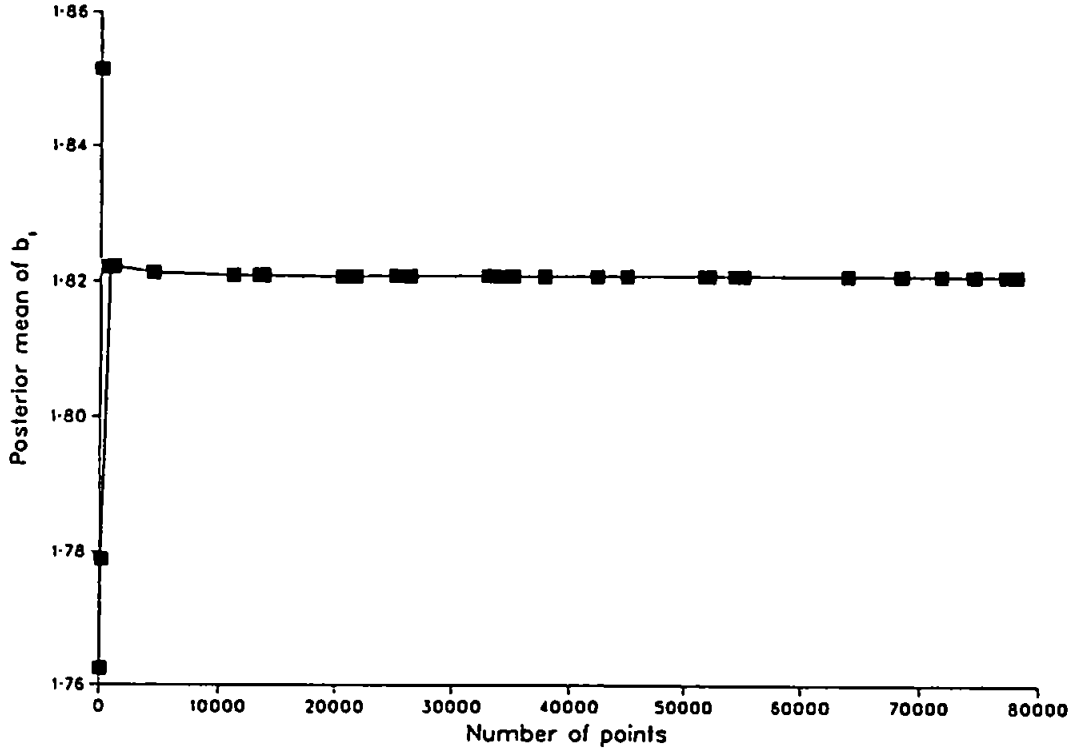


FIGURE 4.15

Convergence of posterior variance of b_1

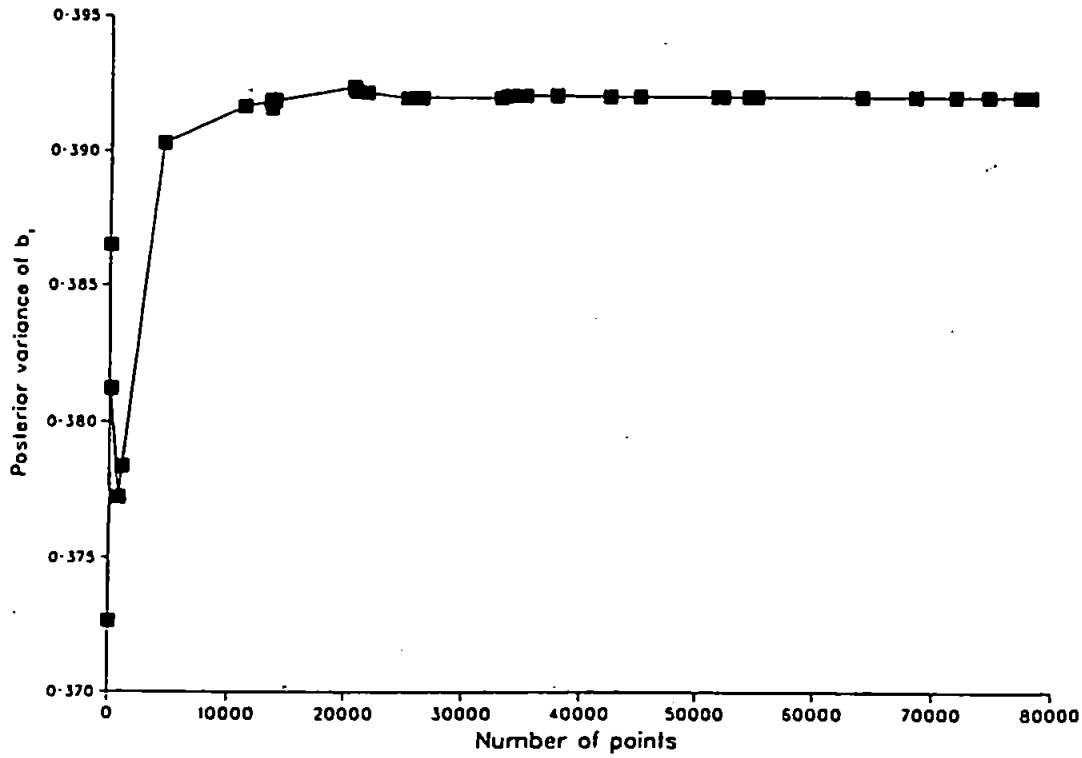


FIGURE 4.16

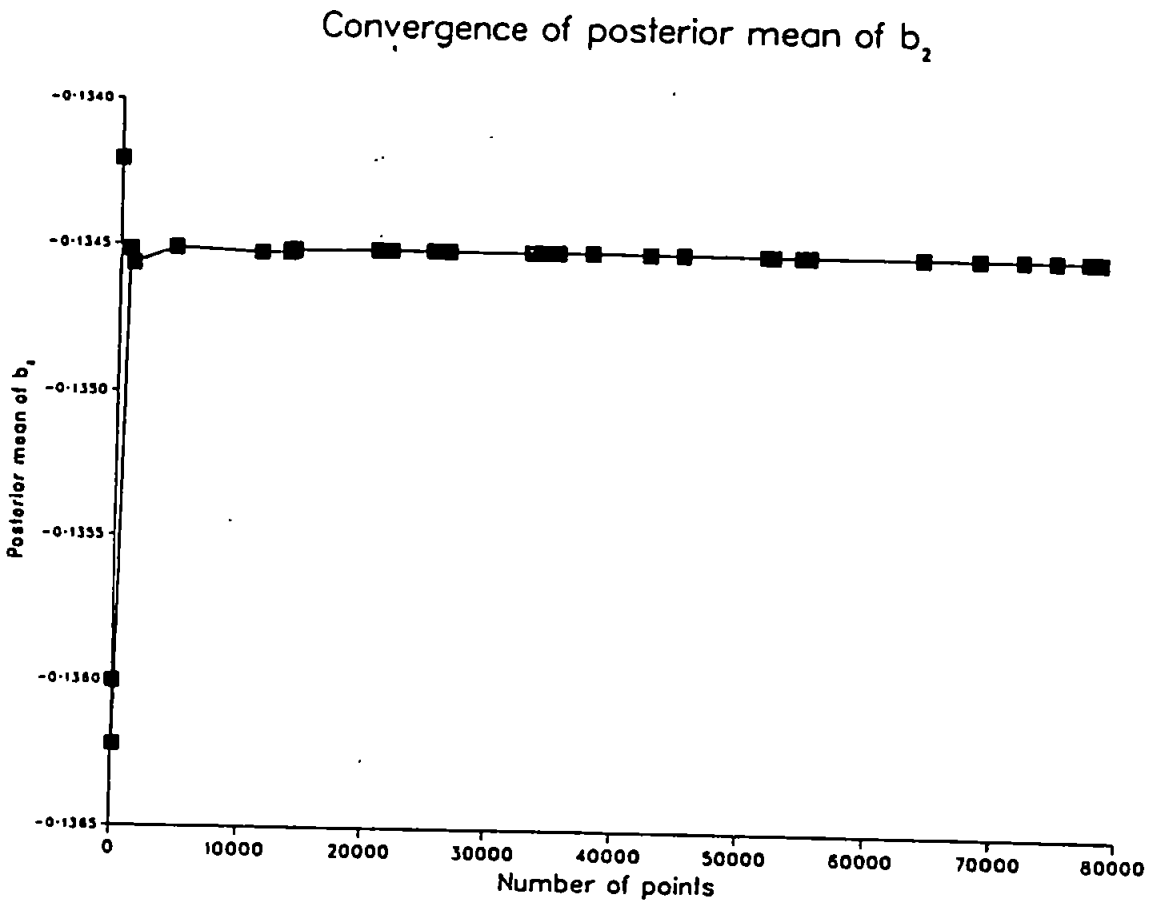


FIGURE 4.17

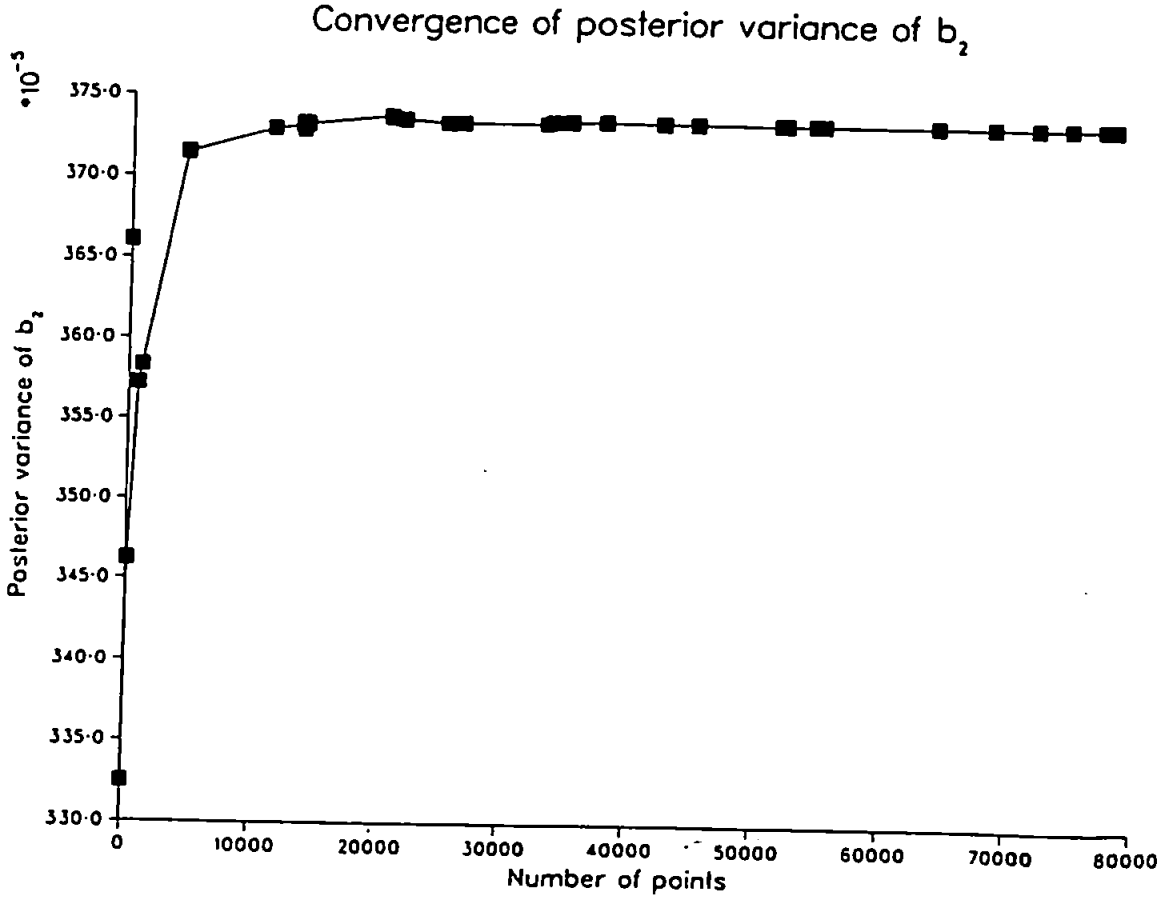


FIGURE 4.18

Convergence of posterior mean of b_3

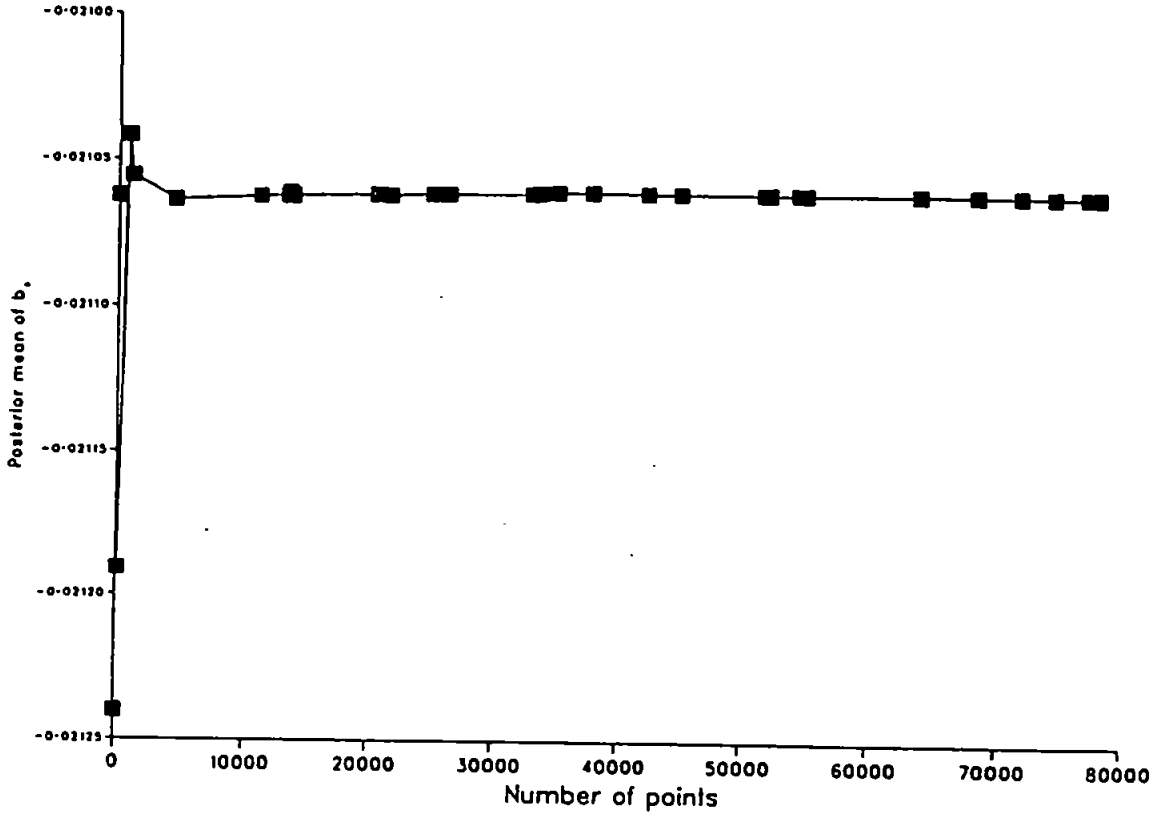


FIGURE 4.19

Convergence of posterior variance of b_3

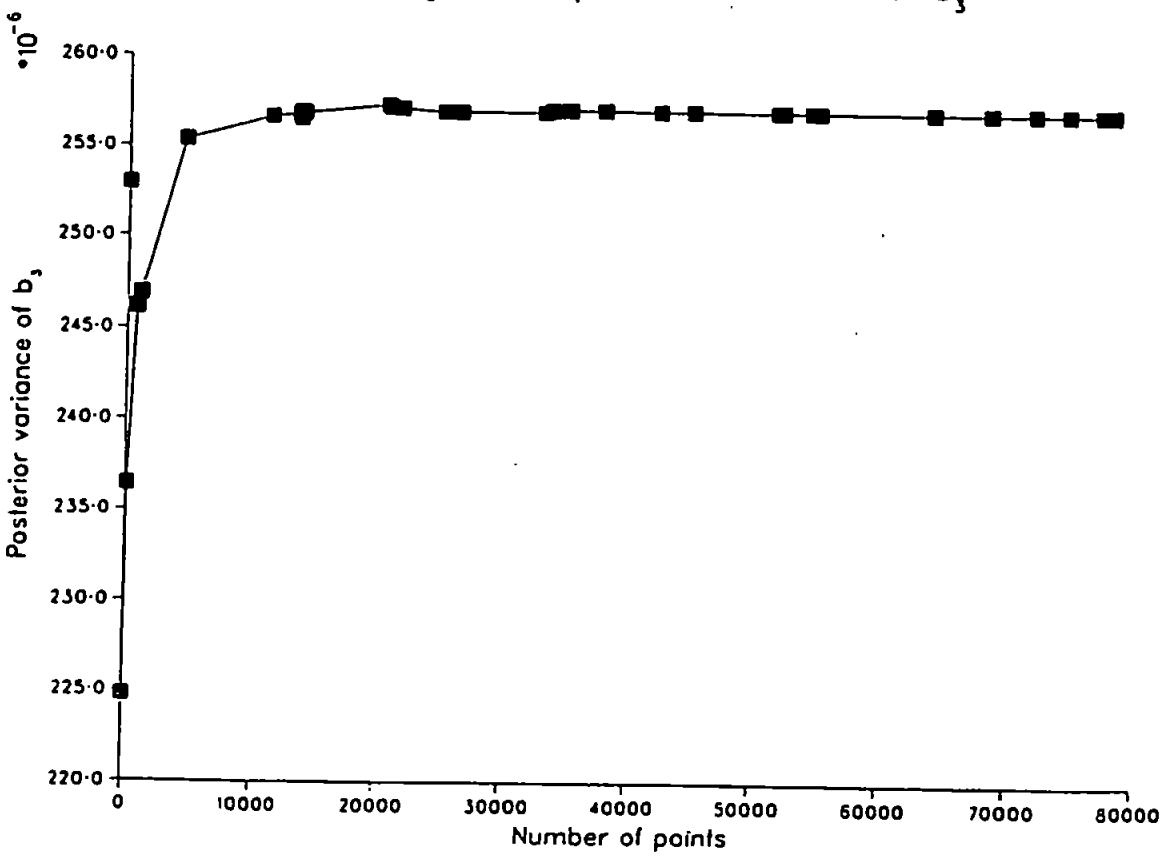


FIGURE 4.20

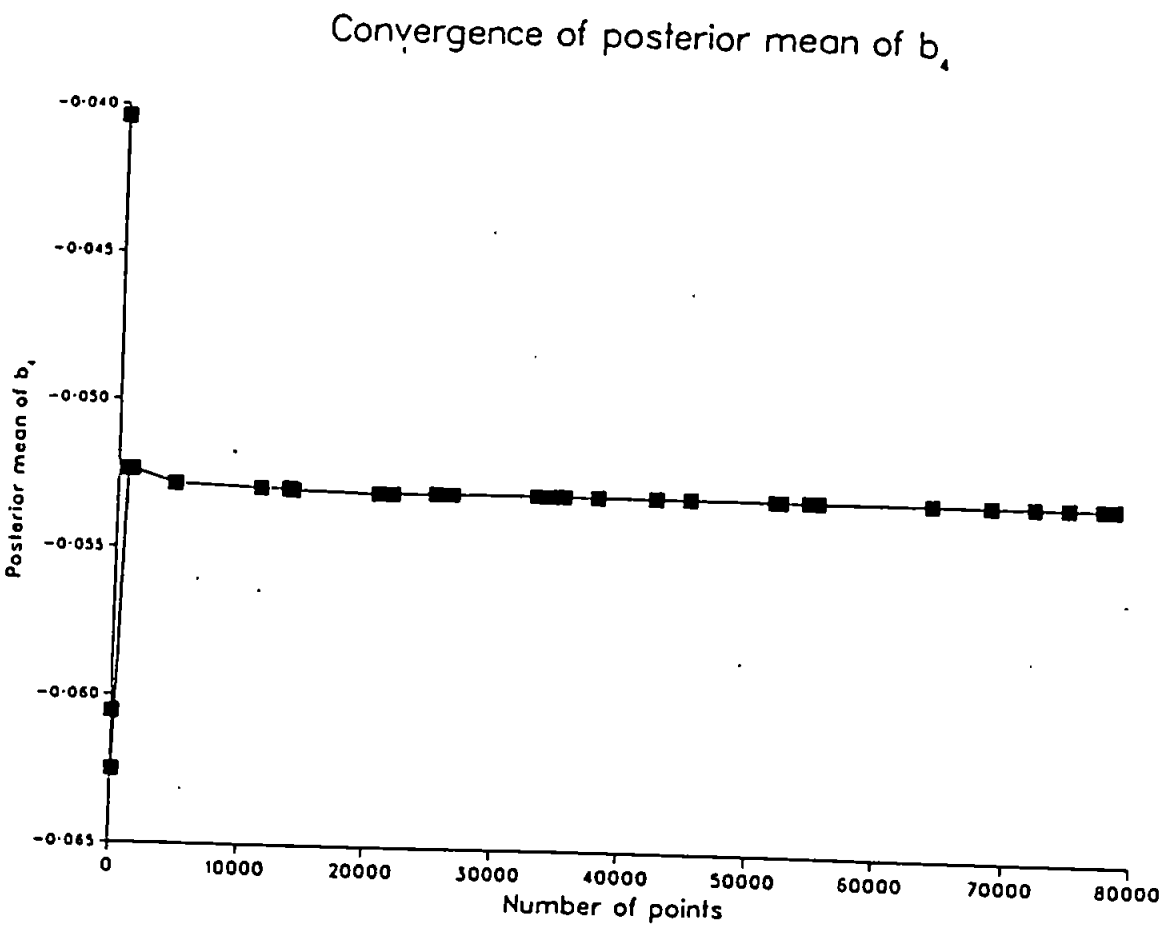


FIGURE 4.21

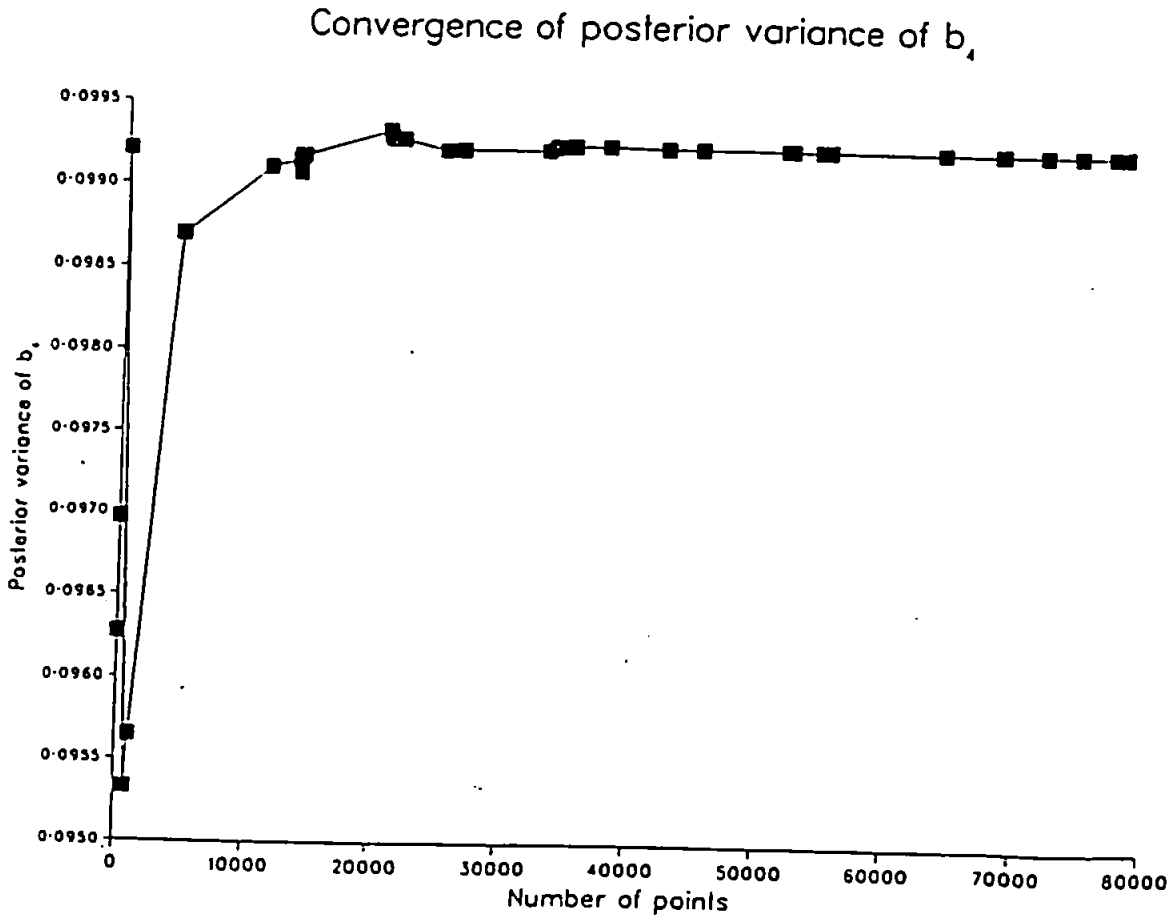


FIGURE 4.22

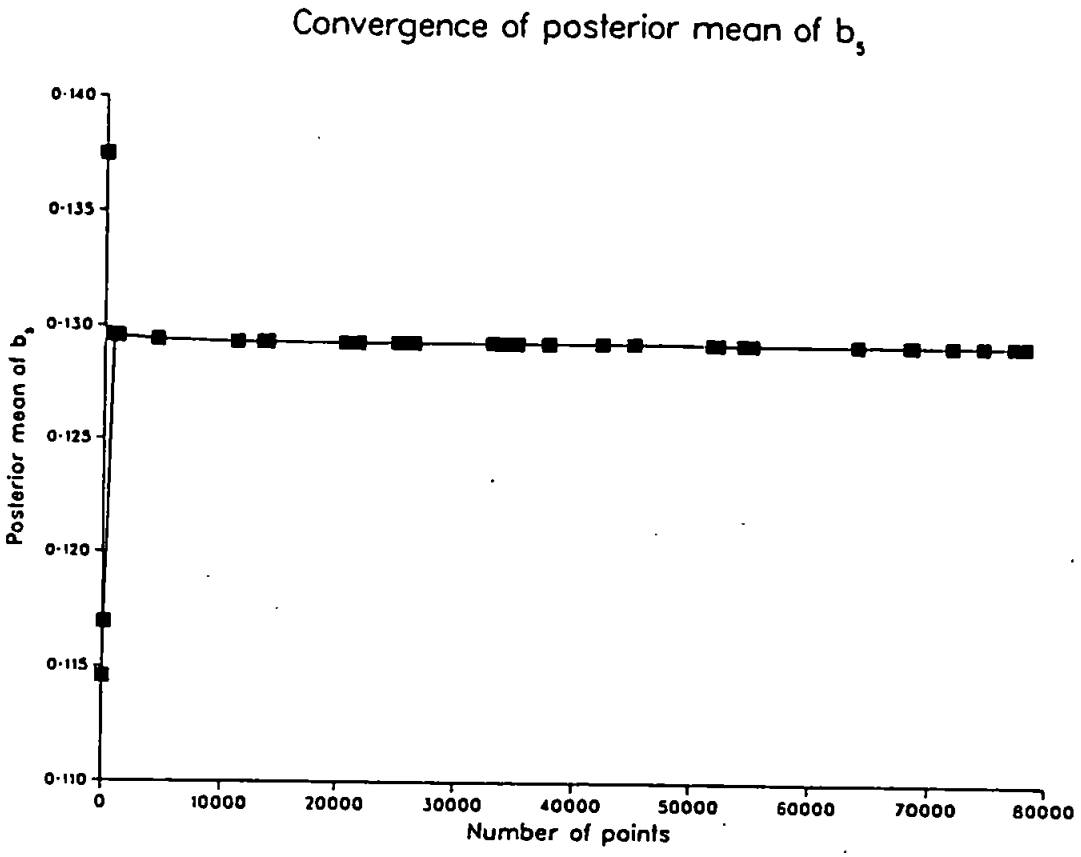


FIGURE 4.23

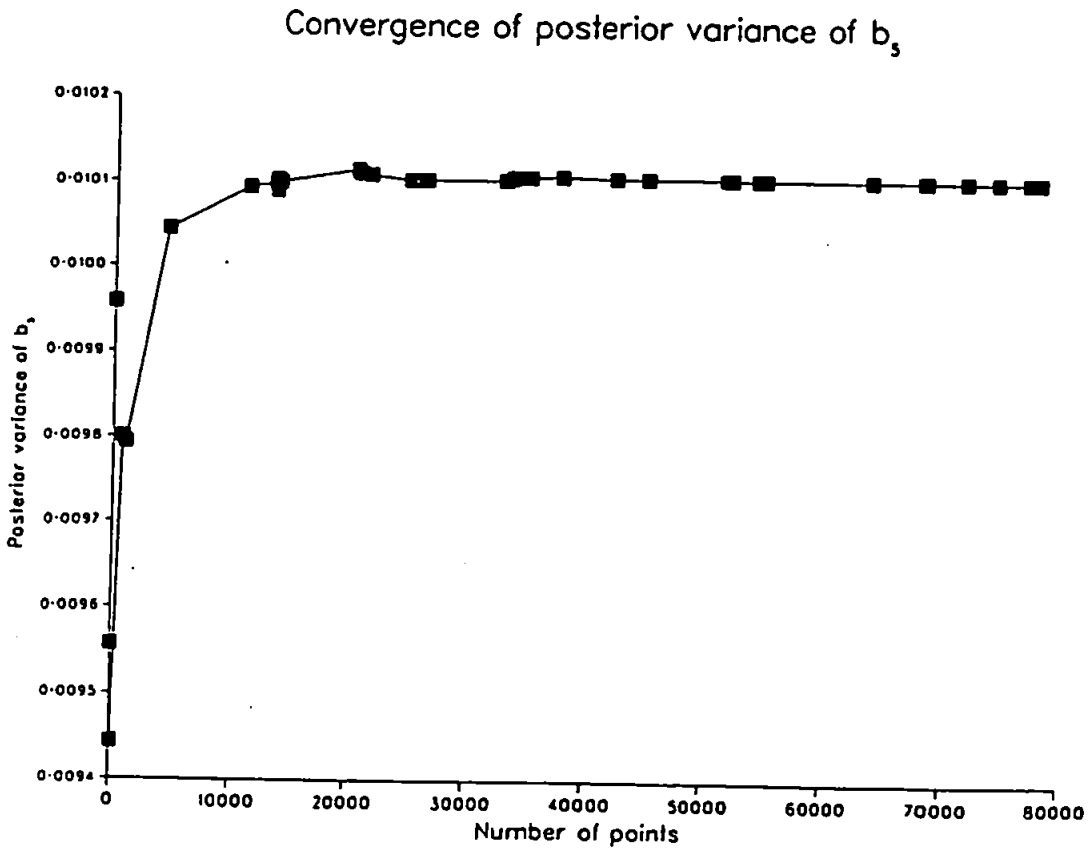


FIGURE 4.24

Convergence of posterior mean of p

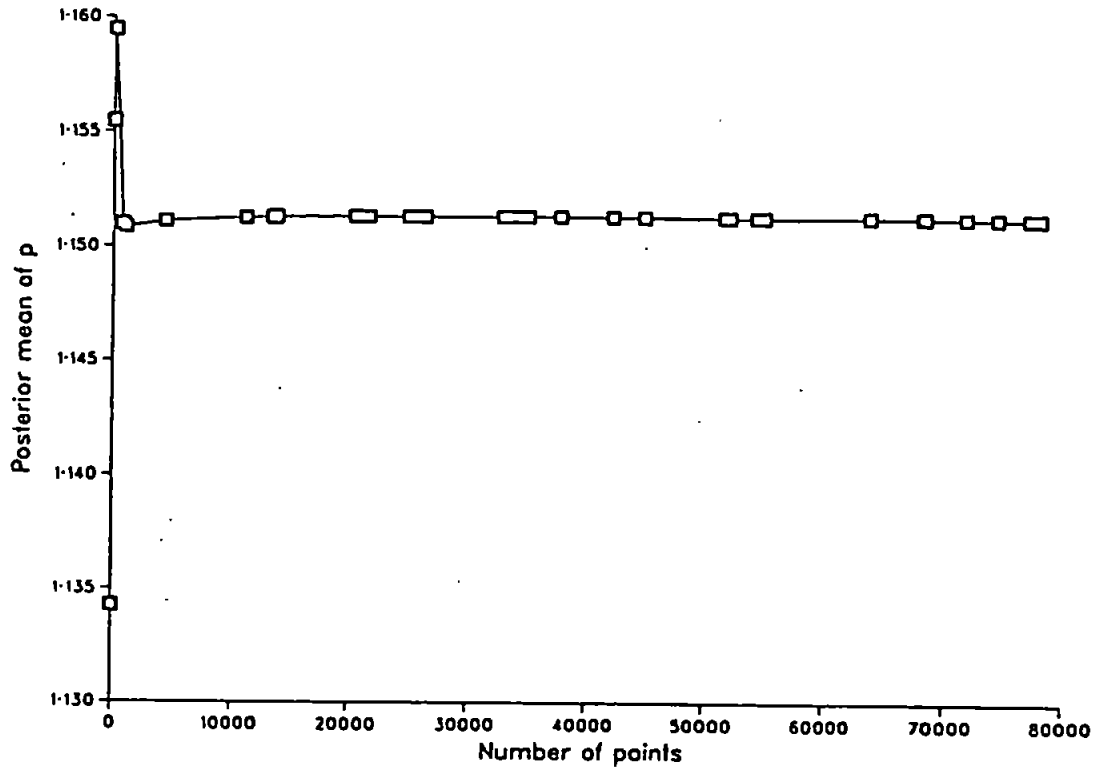
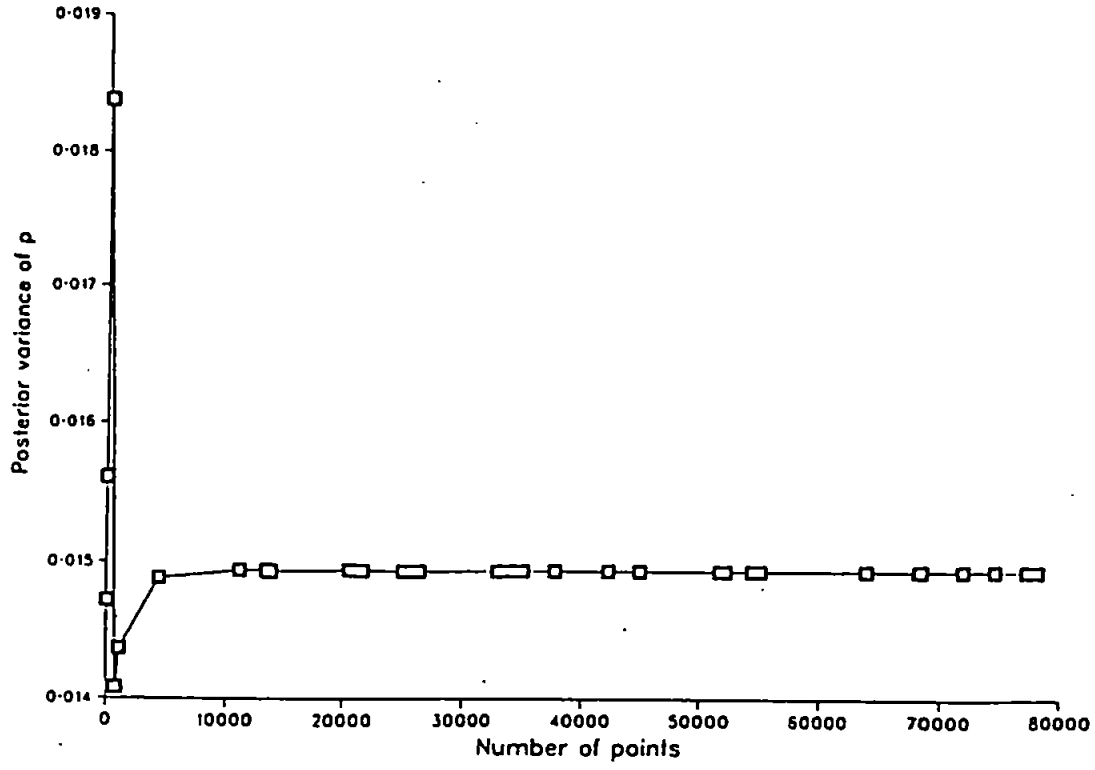


FIGURE 4.25

Convergence of posterior variance of p



5. Applications of imbedded sequences of PIIR's in Bayesian Analysis

5.1 A numerical integration strategy

In this chapter we describe the use of imbedded sequences of PIIR's for the improvement of the adaptive integration strategy of Naylor and Smith (1982). Two main threads are discussed.

First, the use of imbedded sequences of PIIR's to provide a useful source of spatially distributed positive integration rules for use in five dimensions and upwards. These rules can be incorporated into the existing Naylor and Smith adaptive strategy to fill the gaps between the low precision spherical rules and the very expensive Gauss-Hermite product rules.

Secondly, the use of imbedded sequences of PIIR's to improve the adaptive integration strategy of Naylor and Smith. This involves incorporating a facility to work through a sequence of imbedded rules monitoring convergence after at each stage and changing to a sequence with a different location and spread only when it is deemed necessary. We propose a strategy based on the following steps. Of course, the initial parameter transformation and the possible orthogonalising transformation should also be incorporated within this strategy.

(i) Start the iterative strategy by selecting an appropriate base rule.. In general, this will be larger than that used in the adaptive integration strategy of Naylor and Smith (1982) -typically 5 or more points in each dimension.

(ii) Apply an imbedded sequence of PIIR's and check the convergence after each approximation. If in the early steps there is evidence of considerable mispecification of the vector $(\underline{\mu},\underline{\Sigma})$ stop and iterate again starting with the updated elements of $(\underline{\mu},\underline{\Sigma})$.

(iii) If the convergence in the sequence of PIIR's is rapid, then there is good indication that convergence is to the proper value of the integrand. This may well happen when the initial value of $(\underline{\mu},\underline{\Sigma})$ is not close to the posterior vector $(\underline{\mu}',\underline{\Sigma}')$ but the integrand is "well behaved"; see section 2.5.4. In such cases we suggest the completion of the sequence and the derivation of the marginal densities. Of course, the option of verification using new approximations to $(\underline{\mu},\underline{\Sigma})$, possibly with a larger base rule is available, especially when the initial rule is of low precision. Our experience though, coincides with that of Rabinowitz *et. al.* (1987) who reported that their experiments (in one dimension) show that if rapid convergence is achieved with a PIIR sequence then in general this will be to the true value of the integrand.

(iv) If the convergence is slow, we suggest that there are two possible causes for this: mis$\overset{\varsigma}{\underset{\wedge}{}}$pecification of the vector $(\underline{\mu},\underline{\Sigma})$ or a "badly behaved" posterior kernel. Therefore, updating the vector $(\underline{\mu},\underline{\Sigma})$ is recommended. If the iteration does not improve the convergence, then we may come to the conclusion that the assumptions are invalid. An increase in the size of base rule may overcome the problem, but a question arises -provided that we started with a large enough number of base rule- whether it is useful to proceed with an doubtful and expensive procedure or to use existing information from

the posterior distribution (for example, characteristics such as kurtosis or skewness) and use an importance sampling integration, see Shaw(1988a,b). An interesting issue arises here, whether it is possible to use existing information to make a (possibly better than the initial) transformation of the parameter space. This remains a matter for future research.

We note that the strategy proposed above is in essence an extension of the iterative strategy embodied in BAYESFOUR, at least as far as product rules are concerned. The important point is that we apply the product rule by proceeding through a sequence of imbedded rules. The information in the sequence of approximations is used to enable us to stop early or diagnose convergence problems. In the standard BAYESFOUR strategy the full product rule is used, effectively moving to the end of the sequence without exploiting any information about the development of the approximations.

In the remainder of this chapter, we shall demonstrate the efficiency of the above strategy via real examples. When it is possible, comparisons with the currently available method of section 2.2 will be made. However, we feel that use of an interactive adaptive algorithm is very much subjective, and numerical illustrations could be misleading. In most examples, it suffices to illustrate the option of stopping relatively early without using all generators, and how the convergence behaviour can provide information on whether an increase in size of base rule is desirable.

5.2 The 1-dimensional examples of section 2.4 revisited

In section 2.4, we demonstrated how the Gauss-Hermite rule can be more efficient than Soland's method. In our examples, we used the maximum likelihood estimates as initial values for the vector (μ,σ), and illustrated the efficiency of the integration rule by comparing the number of function evaluations needed to up to the convergence to the true value. Following the Naylor and Smith (1982) method, the way to approach this particular problem would be to iterate between and within grid size(s). We argue in this section that an imbedded sequence of PIIR's can produce useful information more efficiently.

We chose a 36-point Gauss-Hermite final precision rule as our initial PIIR. We hope in advance that this rule is large enough to provide efficient approximation to the integrand (2.12), maybe after some transformation of (μ,Σ).

The imbedded sequence of section 3.4.2 was used to produce successive estimates of the posterior vectors (μ,Σ), and consequently of the prediction bounds for future lifetimes. We make the implicit assumption that convergence of the vector (μ,Σ) implies convergence of the probability bound. Tables 5.1-5.3 contain the results of the PIIR sequence and figures 5.1-5.6 illustrate them graphically.

In the first two examples rapid convergence is achieved within the sequence of the PIIR. The convergence is to the true value, in accordance with our remarks in section 5.1. Example 3 is a particularly badly behaved example, for the reasons already mentioned in section 2.4. The convergence is slower in this example, but

TABLE 5.1

Convergence of lower prediction bounds: Example 1

No of points	prior (i)	prior (ii)
4	22.6202149149	39.1725502631
6	22.4752570666	39.5017764965
8	22.4794575144	39.5022994633
10	22.4726103142	39.4891424601
12	22.4642702381	39.4872090199
14	22.4608467189	39.4868499414
16	22.4610449498	39.4867726786
18	22.4603842710	39.4867307658
20	22.4604278882	39.4867234200
22	22.4604881856	39.4867256654
24	22.4604614396	39.4867244697
26	22.4604472323	39.4867241635
28	22.4604433675	39.4867241102
30	22.4604423201	39.4867241018
32	22.4604420800	39.4867241012
34	22.4604420355	39.4867241011
36	22.4604420298	39.4867241011

TABLE 5.2

Convergence of lower prediction bounds: Example 2

No of points	prior (i)	prior (ii)
4	2.46584237628	4.26778796217
6	2.57160207566	4.38178302767
8	2.57170578173	4.38188694312
10	2.56672456904	4.37757288453
12	2.56639244308	4.37731686234
14	2.56615339468	4.37715551223
16	2.56612633119	4.37713940334
18	2.56611390018	4.37713314447
20	2.56610894168	4.37713090598
22	2.56610954492	4.37713115046
24	2.56610901586	4.37713096930
26	2.56610889736	4.37713093549
28	2.56610887500	4.37713093030
30	2.56610887135	4.37713092960
32	2.56610887096	4.37713092961
34	2.56610887093	4.37713092962
36	2.56610887093	4.37713092962

TABLE 5.3

Convergence of lower prediction bounds: Example 3

No of points	prior (i)	prior (ii)
4	20.0054998079	22.2793217019
6	17.6540087948	21.9913700886
8	17.6938657226	22.0004909801
10	18.6938237240	22.1593960625
12	18.2481950220	22.1135958655
14	18.5049346964	22.1407018109
16	18.4381989262	22.1357805291
18	18.4512340637	22.1367613822
20	18.4396385034	22.1361378323
22	18.4384731948	22.1360813988
24	18.4377465701	22.1360588810
26	18.4377293443	22.1360602004
28	18.4377015500	22.1360601156
30	18.4376920991	22.1360601126
32	18.4376888776	22.1360601120
34	18.4376879361	22.1360601119
36	18.4376877217	22.1360601119

TABLE 5.4

Prediction bounds for example 3 with prior (i)

Results updating the posterior mean and variance

No of points	First iteration	Second iteration
4	20.0569529782	19.5590237432
6	19.1622926720	19.1681122196
8	19.1411405873	19.1473957968
10	18.2781875060	18.3072448533
12	18.7450233636	18.7254772579
14	18.3845533347	18.3964856778
16	18.4541753325	18.4544927342
18	18.4339520166	18.4362926335
20	18.4424482458	18.4419137941
22	18.4438783529	18.4431172032
24	18.4435278094	18.4424717722
26	18.4431554803	18.4420349541
28	18.4430067931	18.4418497428
30	18.4429459792	18.4417730099
32	18.4429234778	18.4417440343
34	18.4429161691	18.4417344005
36	18.4429142665	18.4417318173

FIGURE 5.1

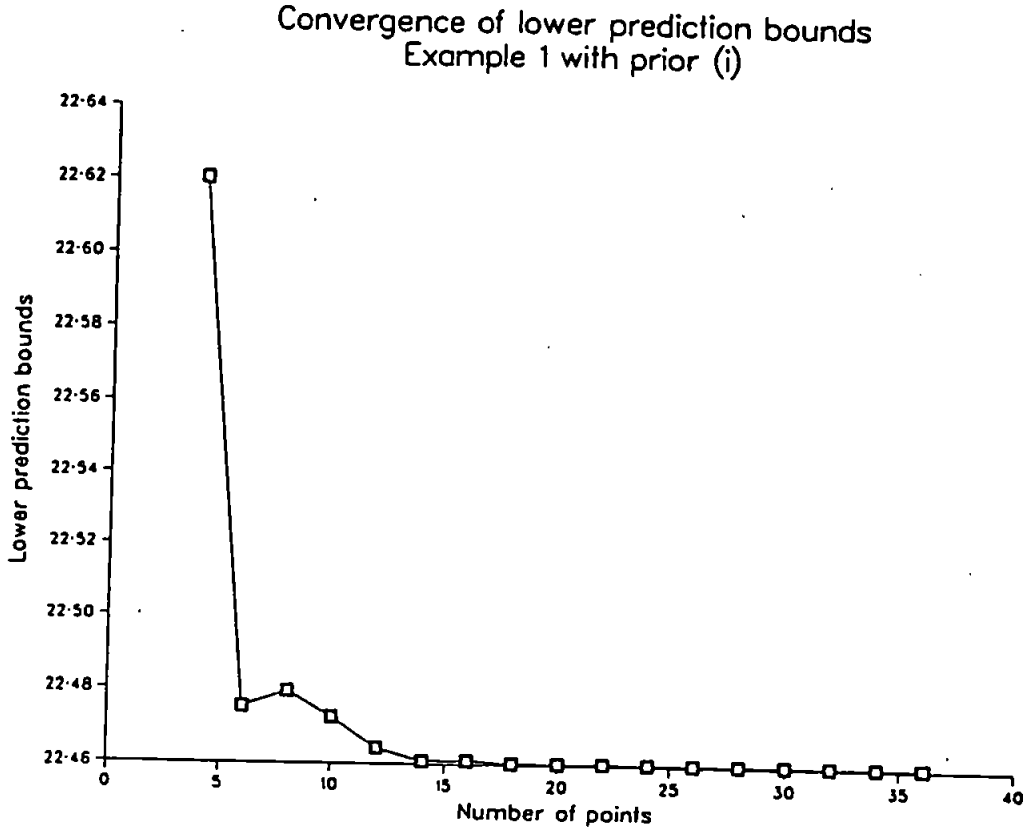


FIGURE 5.2

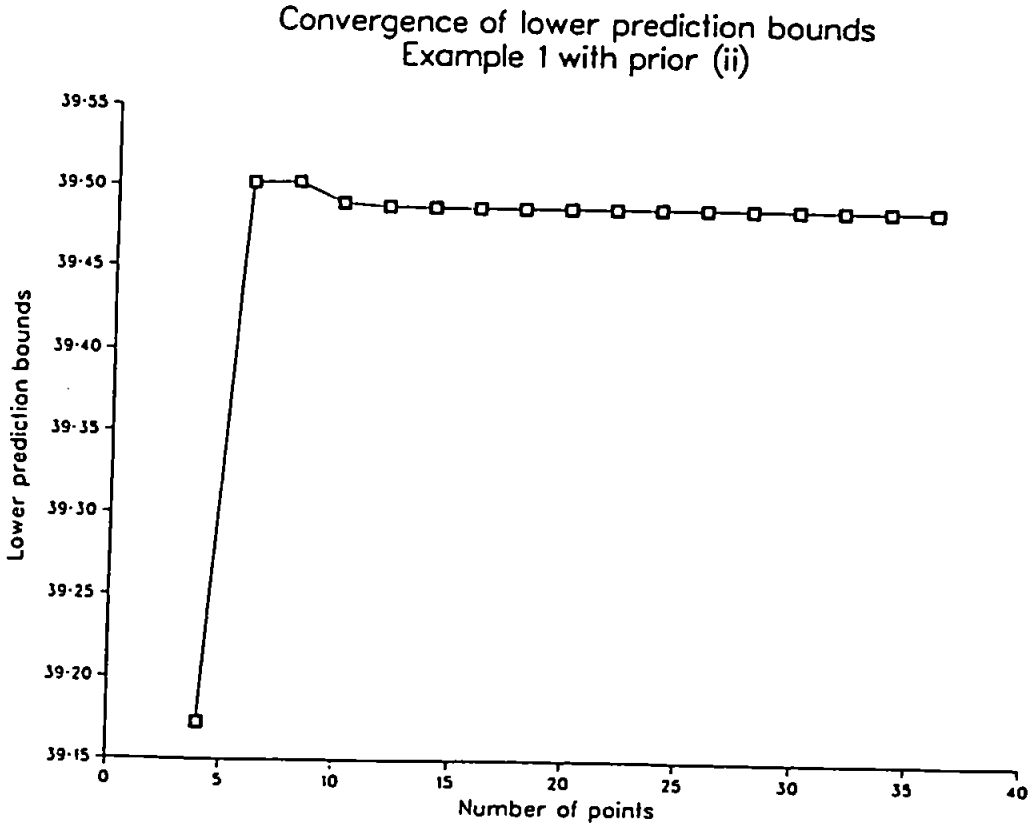


FIGURE 5.3

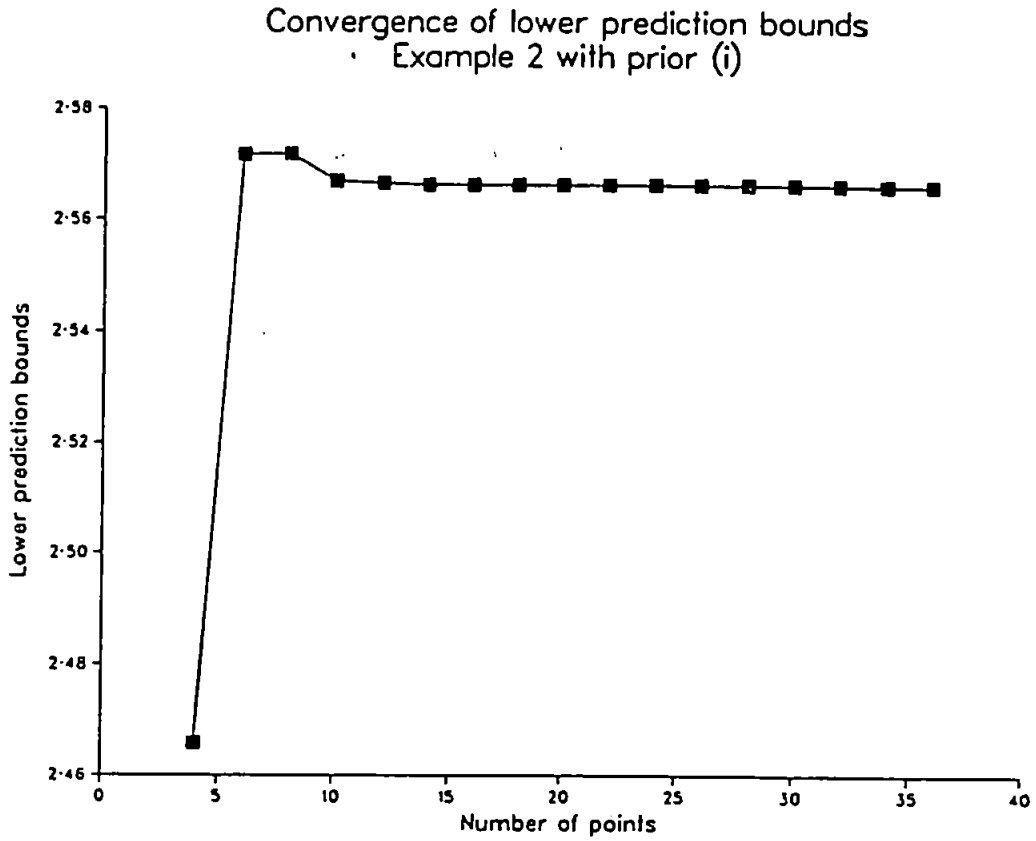


FIGURE 5.4

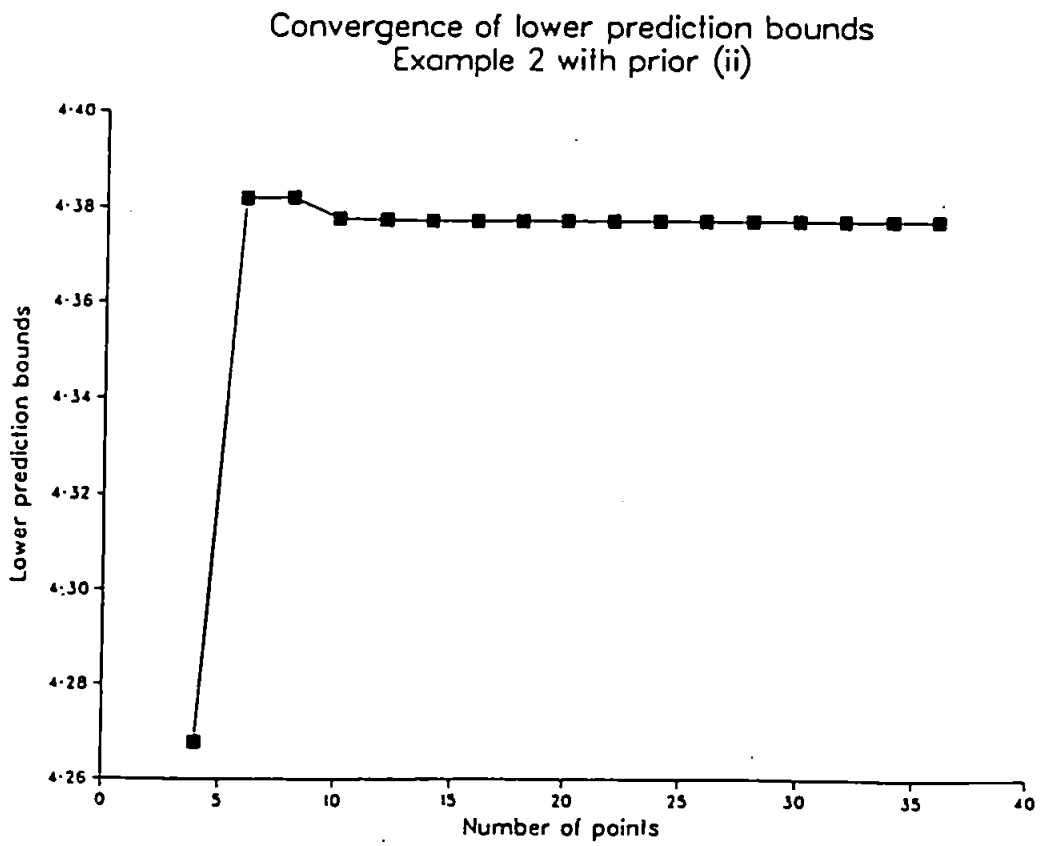


FIGURE 5.5

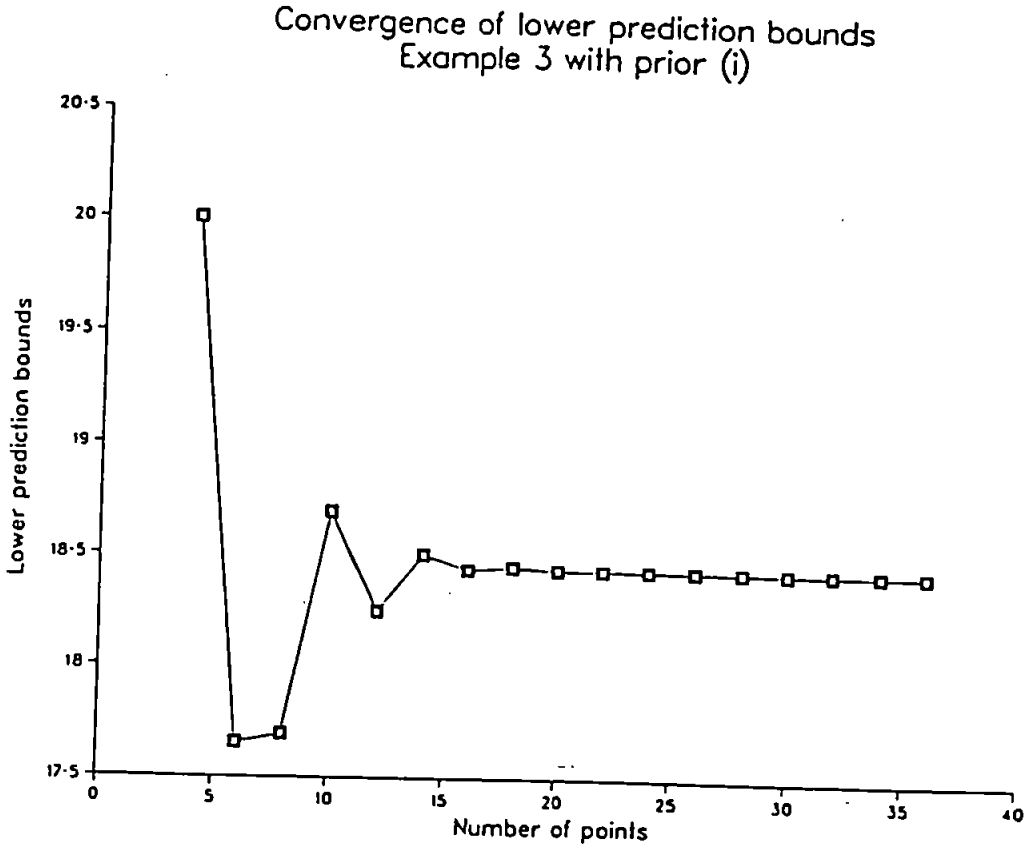
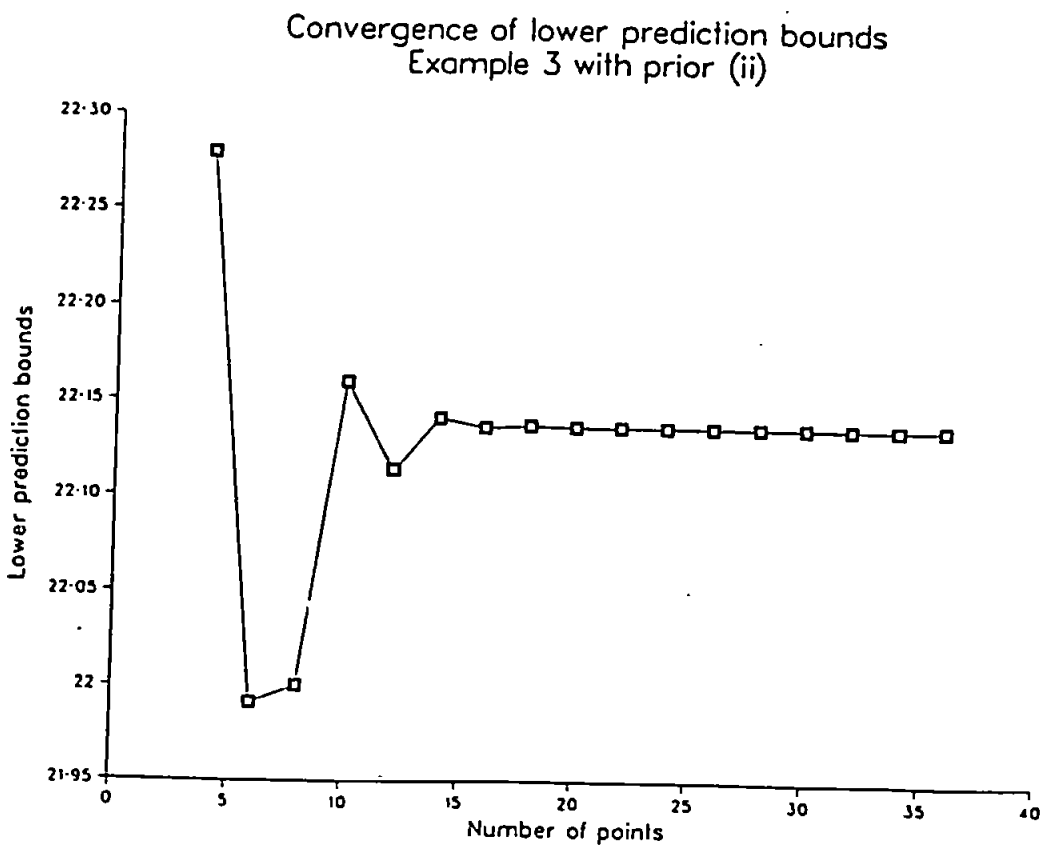


FIGURE 5.6



improves after 14 or 16 function evaluations to produce rapid convergence towards the end of the sequence. Comparatively, the sequence of PIIR's produces the slowest convergence in the third example with prior (i), and this is the only example where the sequence does not converge to the true value of the integrand.

As we mentioned in our proposed strategy in section 5.1, we argue that if the convergence is rapid, an iteration updating the vector (μ,σ) within the same grid size should not be recommended. In fact, the irregularities in the early steps of the PIIR sequence in example 3 only signifies the bad behaviour of the integral. However, for illustrative purposes and reasons of objectivity, we applied two more iterations updating the vector (μ,σ). The results are shown in table 5.4. Indeed, the imbedded sequence produces poorer results compared with the first iteration, and the rate of convergence does not improve.

Thus, these examples illustrated that the imbedded sequence of the 36-point Gauss-Hermite based PIIR produced efficient results, simple to interpret and quite informative as far as the behaviour of the integrand is concerned.

5.3 Two three-dimensional examples

5.3.1 Reanalysis of Stanford Heart Transplant Data.

We consider here a three dimensional parameter model which was used by Turnabull *et. al.* (1974) to describe data from the Stanford heart transplant program and was referred by them as the Pareto model. This model was used by Naylor and Smith (1982), Tierney and Kadane (1986) and Kass *et. al.* (1988) for the demonstration of their method. Furthermore, the latter paper involved some useful comments and questions in the discussion followed, concerning the way Naylor and Smith method should be applied. See section 2.5.4 for details. Therefore, our consideration of the problem here will be in exactly the same way as in the previously cited papers, for ease of comparison.

In the Stanford heart transplant program, out of the 82 patients who accepted in the program, 30 of them did not receive a heart transplant. However, these 30 patients do not form control group since their selection was by circumstances beyond of the control of the experiment, such as early death or recovery. The Pareto model views individual patients in the nontransplant group as having lifetimes following the exponential density

$$p(t/\varphi) = \varphi e^{-\varphi t}$$

The mean φ of the above exponential density is assumed itself to be a random variable drawn independently for each patient from a gamma

distribution with density of the form

$$p(\varphi/\lambda,p) = \frac{\lambda}{\Gamma(p)}\,(\lambda\varphi)^{p-1}\,e^{-\lambda\varphi} \quad ; \; p,\lambda>0.$$

For an individual transplant patient the lifetime distribution is taken to be exponential with mean $\tau\varphi$ in place of φ. Clearly the transplant is effective in prolonging life if $\tau<1$.

The marginal density for the future lifetimes of a candidate if no transplant were performed is given by

$$p(t) = \int_0^\infty p(t/\varphi)p(\varphi/\lambda,p)d\varphi = \frac{p\lambda^p}{(\lambda+t)^{p+1}}$$

In this way the resulting likelihood function of the parameter vector $\underline{\theta} = (\tau,\lambda,p)$ is

$$l(\underline{x};\underline{\theta}) = \prod_{i=1}^{n} \frac{p\lambda^p}{(\lambda+x_i)^{p+1}} \prod_{i=1}^{N} \left(\frac{\lambda}{\lambda+x_i}\right)^p \prod_{j=1}^{m} \frac{\tau p\lambda^p}{(\lambda+y_j+\tau z_j)^{p+1}} \cdot \prod_{j=m+1}^{M} \left(\frac{\lambda}{\lambda+y_j+\tau z_j}\right)^p$$

where the x_i are the survival times in days of the N=30 non-transplant patients, n=26 of whom died, and y_j,z_j are the times to transplant and survival times, respectively, for the M=52 transplant patients, m=34 of whom died.

Naylor and Smith (1982) and Tierney and Kadane (1986) used an improper uniform prior on the parameter vector $\underline{\theta}$ of the form

$$p(\underline{\theta}) = \text{constant} \qquad \text{for } \tau, \lambda, p > 0$$
$$= 0 \qquad \text{otherwise}$$

They also mentioned the possible integration over the p parameter analytically, but for illustrative and comparative purposes they have chosen to work with the full three-parameter likelihood. As has been already mentioned, we shall adopt the above approaches to faciliate the comparison of the methods.

Naylor and Smith (1982) noted that a run over a series of 5^3 grids failed to show satisfactory convergence whereas a final convergence was achieved between 8^3 and 10^3 grids. Their results also suggested that different order of orthogonalising transformations (which correspond to different transformations of the parameter space) produce different convergence rates, so we adopt here the 'optimum' parameter order (p, λ, τ) for our transformations.

The interesting point in the Turnbull *et al.* data is that the use of maximum likelihood based approximations can be misleading, especially for λ. See comments in the paper by Naylor and Smith (1982), and by Tierney and Kadane (1986). It is therefore important to see how an imbedded sequence of a PIIR can handle a situation in which the initial vector $(\underline{\mu}, \underline{\Sigma})$ is mispecified.

Using the algorithm of section 4.3 we produced an imbedded sequence of PIIR based on a 9^3 Gauss-Hermite product rule. This sequence consists of ${}^{5+3-1}C_3 = 35$ generators. We used as initial values for the parameter vector $(\underline{\mu}, \underline{\Sigma})$ the maximum likelihood estimates. (We note

here two misprints in Naylor and Smith (1982) paper, the s.d. of p is 0.1 instead of 1.1 and the correlation of λ and τ is -0.18 instead of -0.46). In figures 5.7-5.13 we illustrate the convergence of the posterior mean and variance of $\underline{\theta}$, in two different situations. The first iteration (black squares in figures 5.7-5.13) denotes the initial sequence based on the maximum likelihood estimates and the associated covariance matrix. We can see that the imbedded sequence gives an early information about the mispecification of the initial estimates. It also provides a sequence of estimates with a lot of 'jumps'. Thus, even if we do not stop early and we we use all generators in the sequence, the behaviour of the convergence indicates that some assumptions might be invalid or that the initial estimates might be far away from the posterior parameter vector.

In the same figures, we illustrate the last iteration sequence which represents the behaviour of the posterior mean and variance vectors when a convergence is achieved within the 9^3 product rule. The rapid convergence of the sequence indicates that the results do not need a further verification: an increase to the 10^3 grid will produce the same results, within a tolerance error (aggregate measure Δ=0.022).

Adopting the suggestion of Rabinowitz *et al.* (1987) and stopping if three successive approximations show convergence, we applied the strategy of section 5.1 using $\Delta<.03$ as our criterion for convergence. The initial imbedded sequence based on the 9^3 Gauss-product rule showed convergence after 277 function evaluations. With three more iterations updating the mean and the covariance matrix, convergence both within the imbedded sequences between sequences with different $(\underline{\mu},\underline{\Sigma})$'s occurred after a total of 4x277= 1108 function evaluations.

FIGURE 5.7

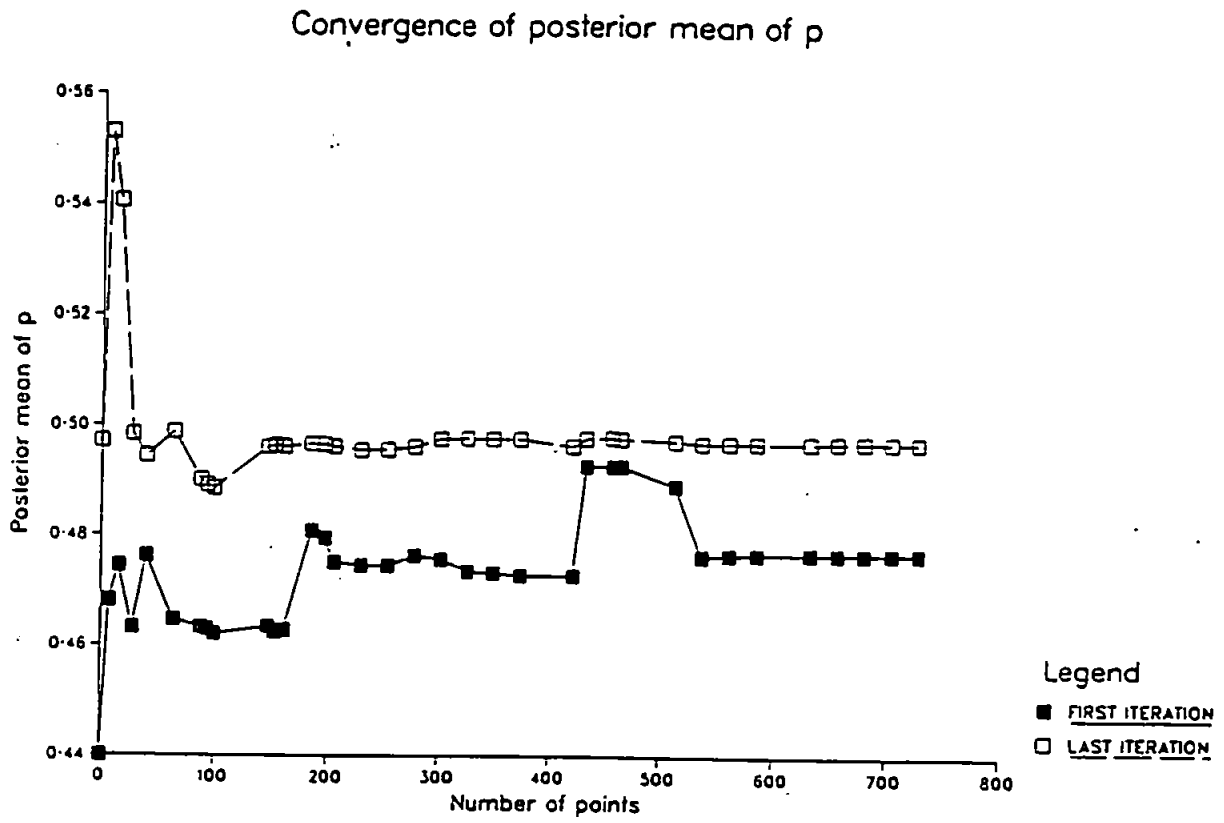


FIGURE 5.8

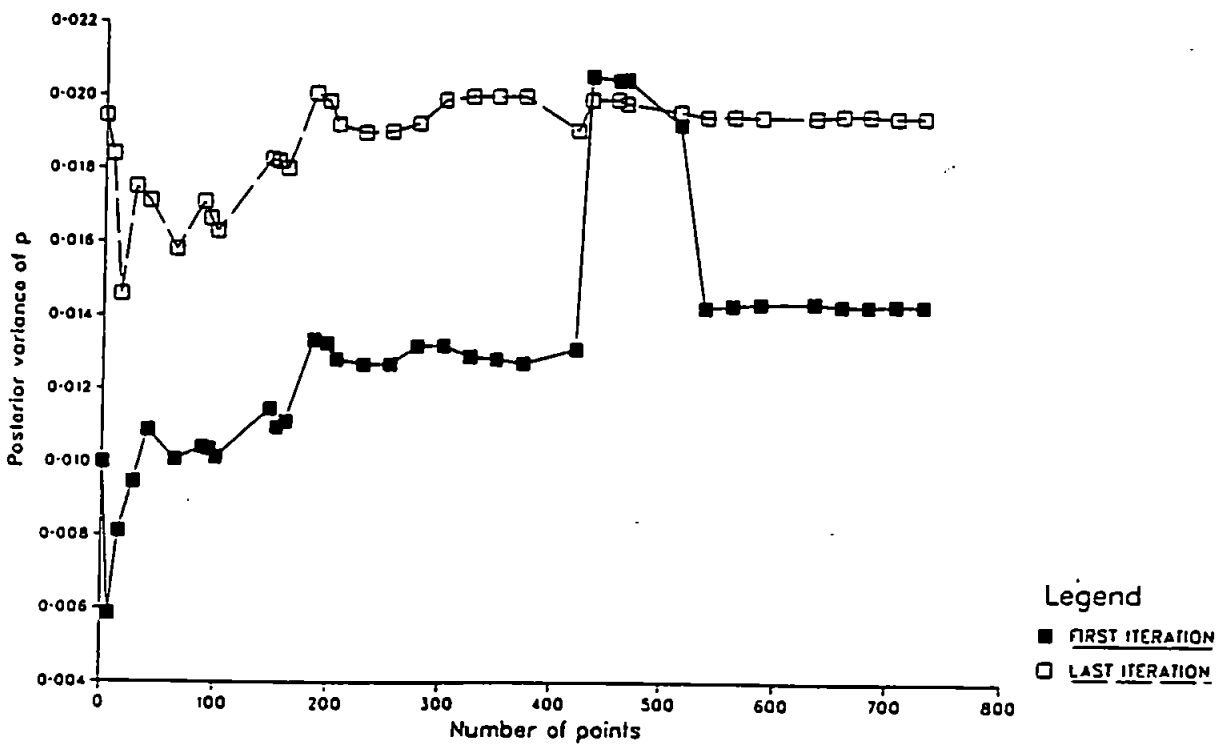


FIGURE 5.9

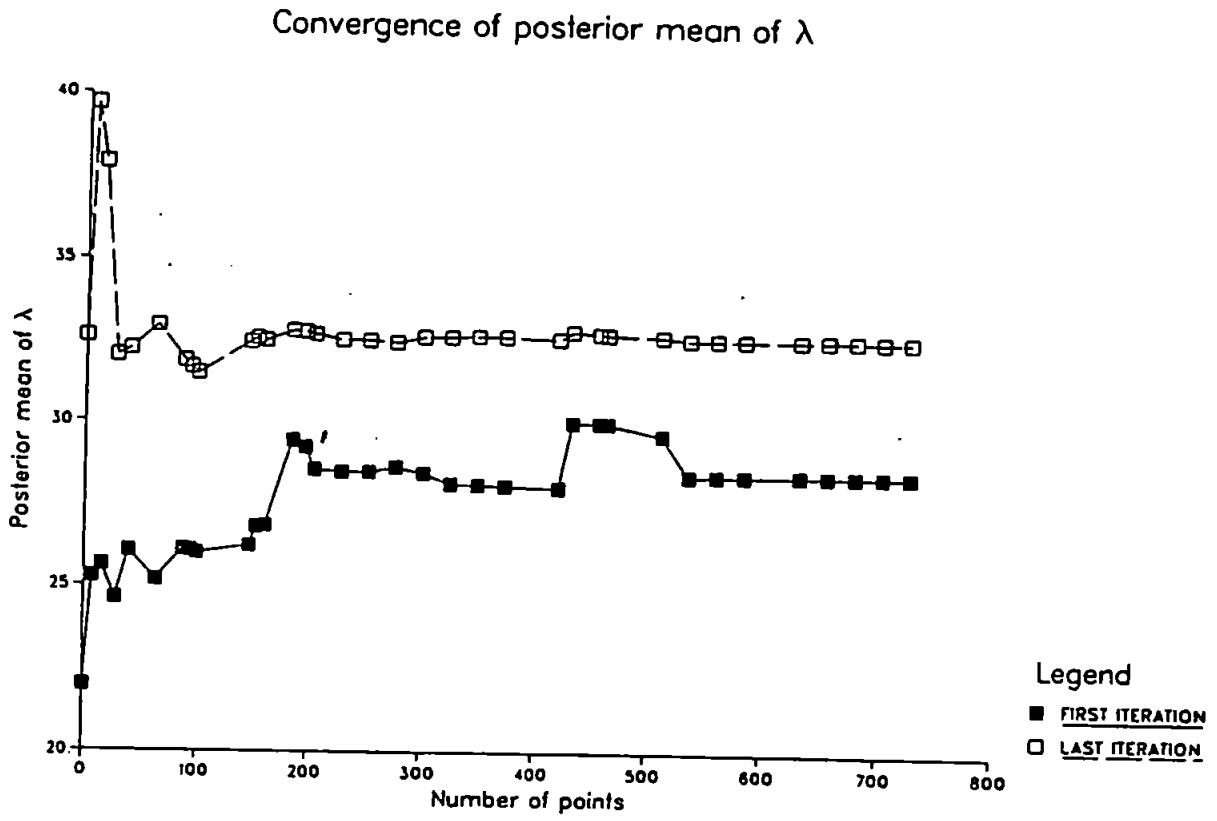


FIGURE 5.10

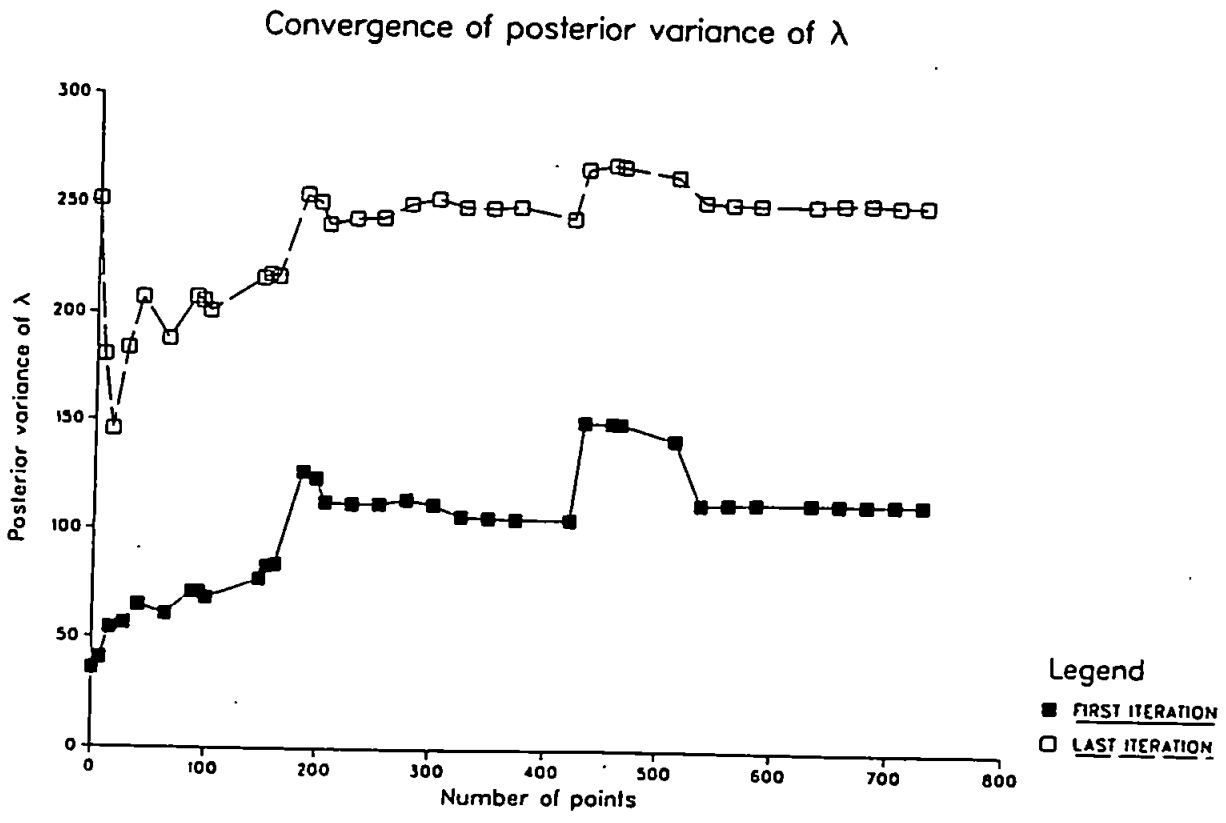


FIGURE 5.11

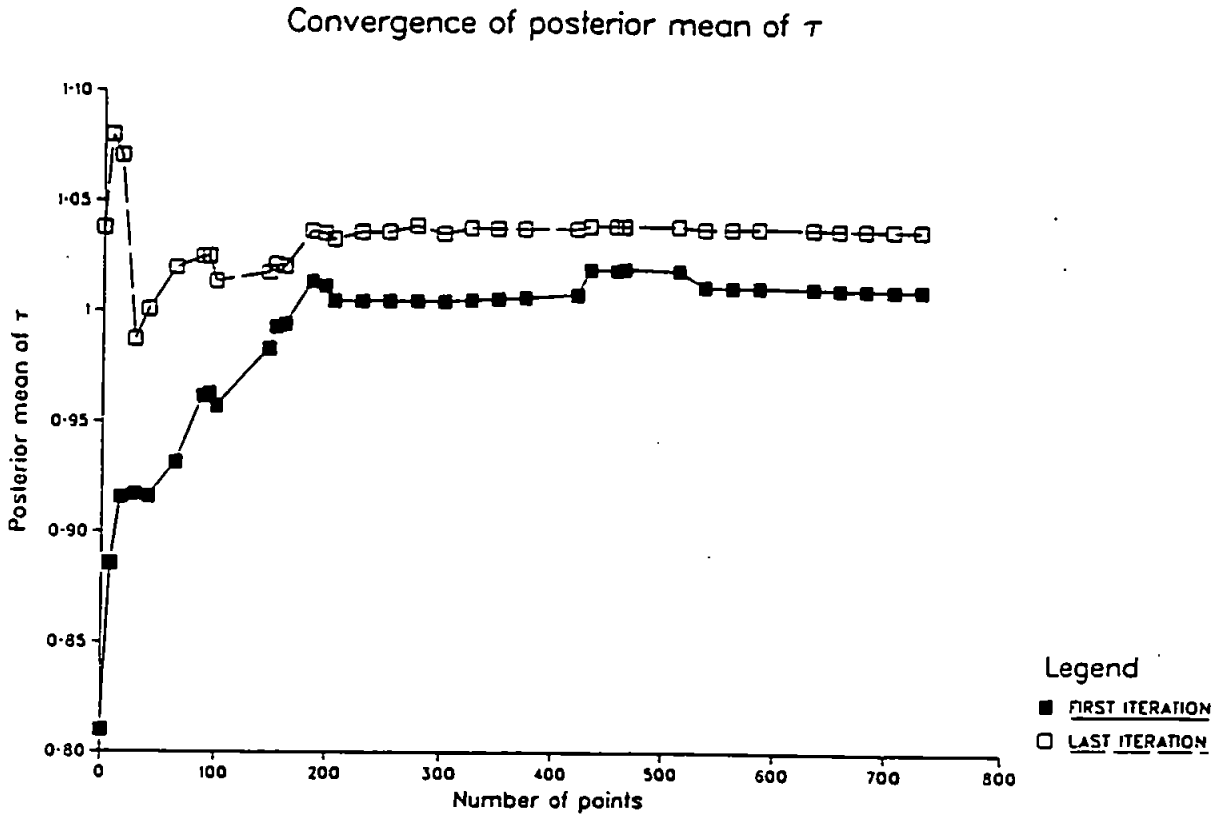


FIGURE 5.12

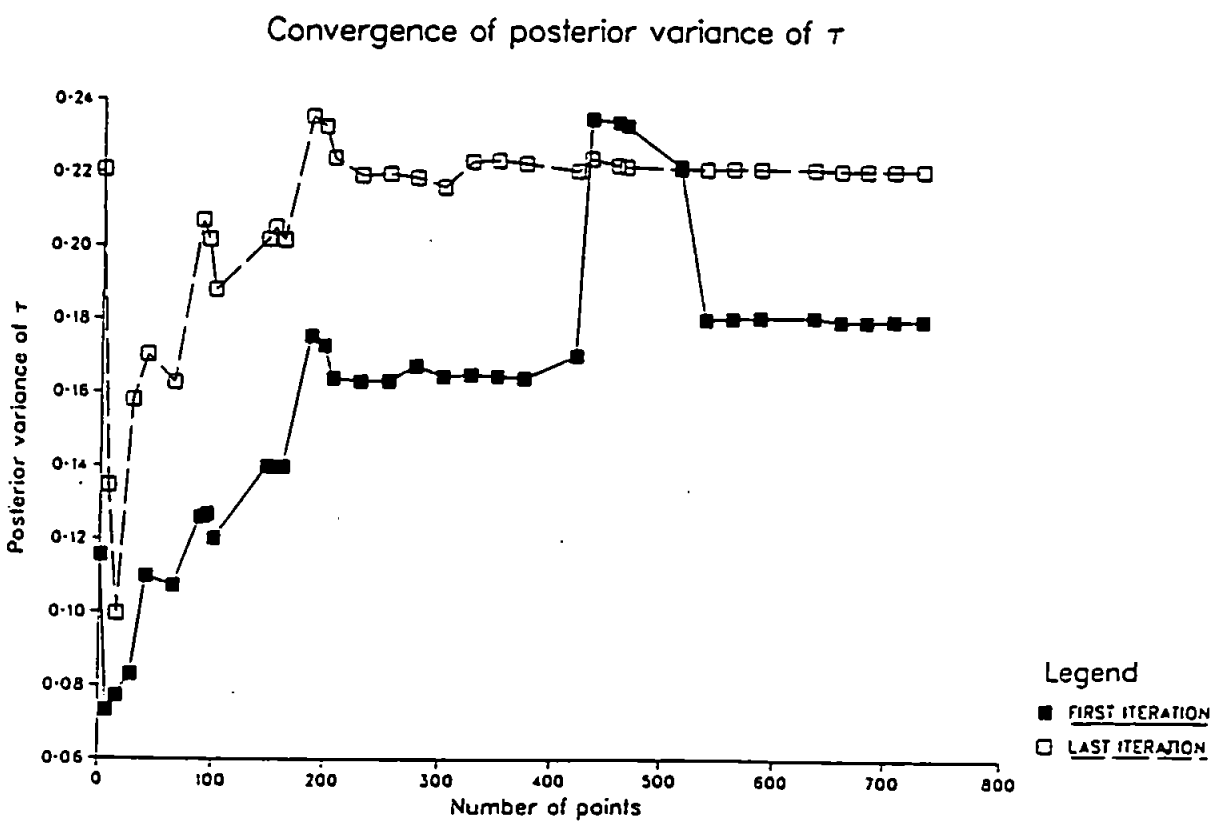
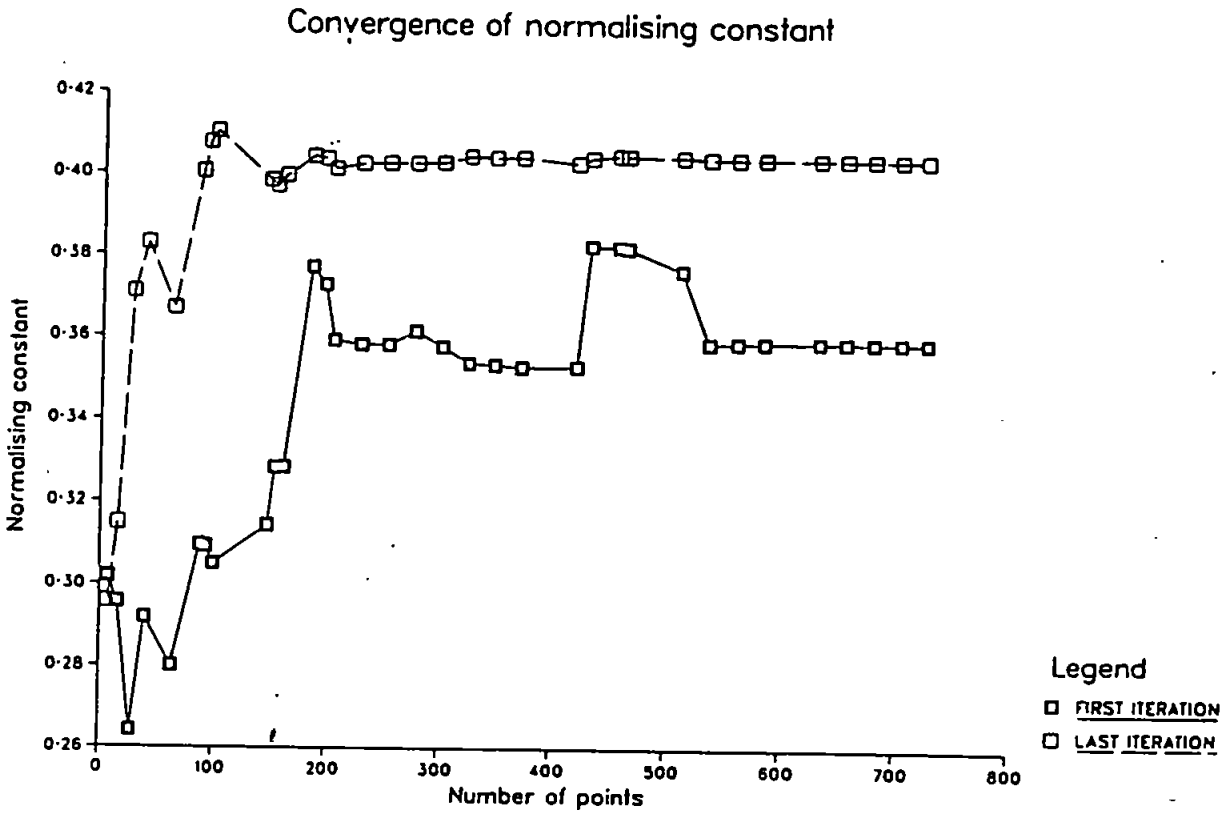


FIGURE 5.13



If marginal densities are needed we could add 277 more points to end up with a full product rule including 1560 function evaluations.

The same convergence criterion ($\Delta<0.03$) was also used for a BAYESFOUR run in the same problem. Starting with a 4^3 Gauss Hermite product rule, 16 iterations were needed ending up in 8^3 product rule and a total of 2730 function evaluations ($6x4^3+4x5^3+3x6^3+2x7^3+1x8^3$). Moreover, if we consider that one more iteration is normally needed in

BAYESFOUR to verify the convergence, the total number of function evaluations would be $2730+9^3-3459$. We remark here that such verification is not needed in the imbedded sequence of PIIR's: The behaviour of the sequence is the factor which guards against false convergence.

We re-iterate that the strategy based on imbedded sequences leads to a 9^3 product rule positioned and centered appropriately after 1560 function evaluations whilst the standard Naylor-Smith strategy leads to an 8^3 product rule after 2730 function evaluations.

5.3.2 The example of section 2.3 revisited

The three dimensional example of section 2.3, Reilly (1976), is a particularly badly behaved example. In the demonstration of BAYESFOUR in section 2.3 we have chosen as our criterion for convergence to be $\Delta<0.05$. Convergence occurred after 11 iterations, but interest lies in the issue of false convergence. Indeed, the increase of grid size decreases Δ, but very slowly. Even after 17^3 grid sizes the posterior moments have not stabilised, Δ not being less than 0.02. In fact, such behaviour is expected given the restrictions on the parameter space imposed by the model (section 2.3), and the small sample size ($n=6$).

It is interesting to explore the behaviour of the imbedded sequence of PIIR's used in such badly-behaved example. The imbedded sequence of section 5.3.1 was applied again and figures 5.14-5.20 describes the behaviour of the integration rules. In a similar way as in figure 5.2,

the black squares denote the values obtained in the first iteration based on the maximum likelihood estimates, whereas the white squares denote the values at the last iteration. An interesting feature in figure 5.14-5.20 is that the jumps in the values of the posterior expectations and the normalising constant do not vanish even in the last iteration. There is therefore information from the imbedded sequence which otherwise would have been lost if a straight Gauss-product rule was used.

In fact, detailed investigation of this particular problem showed that one or more generators with nearly all their nodes making the likelihood zero are added at a particular point. It was initially thought that this happened because of the specific choice of the imbedded sequence of PIIR's: we remarked in section 4.3 that there are many imbedded sequences that can be derived based on a Gauss product rule. Thus, we derived another sequence of PIIR's, changing slightly the algorithm of section 4.3, hoping that a sequence which will include all the badly placed nodes in the early steps will produce a good convergence at the late stage of the sequence. Unfortunately, this was not possible, as the convergence always being influenced by such nodes.

Models with such badly behaved integrands are often considered as being of great interest, and thus the information derived from the sequences of PIIR's is valuable, if, however, not yet specifically determined. It is noted in section 2.5 that badly behaved integrands are in any case difficult to handle by numerical integration techniques, and so the identification of such irregularities can be of considerable use.

FIGURE 5.14

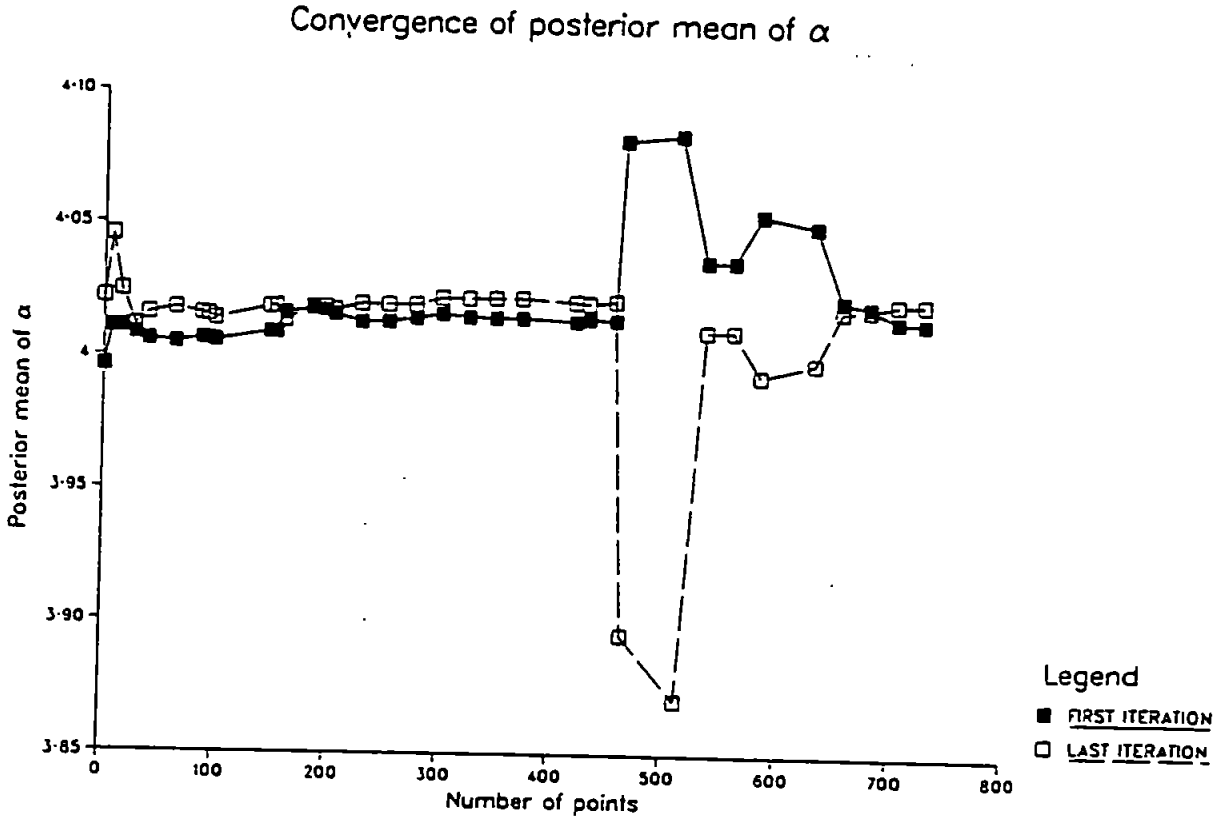


FIGURE 5.15

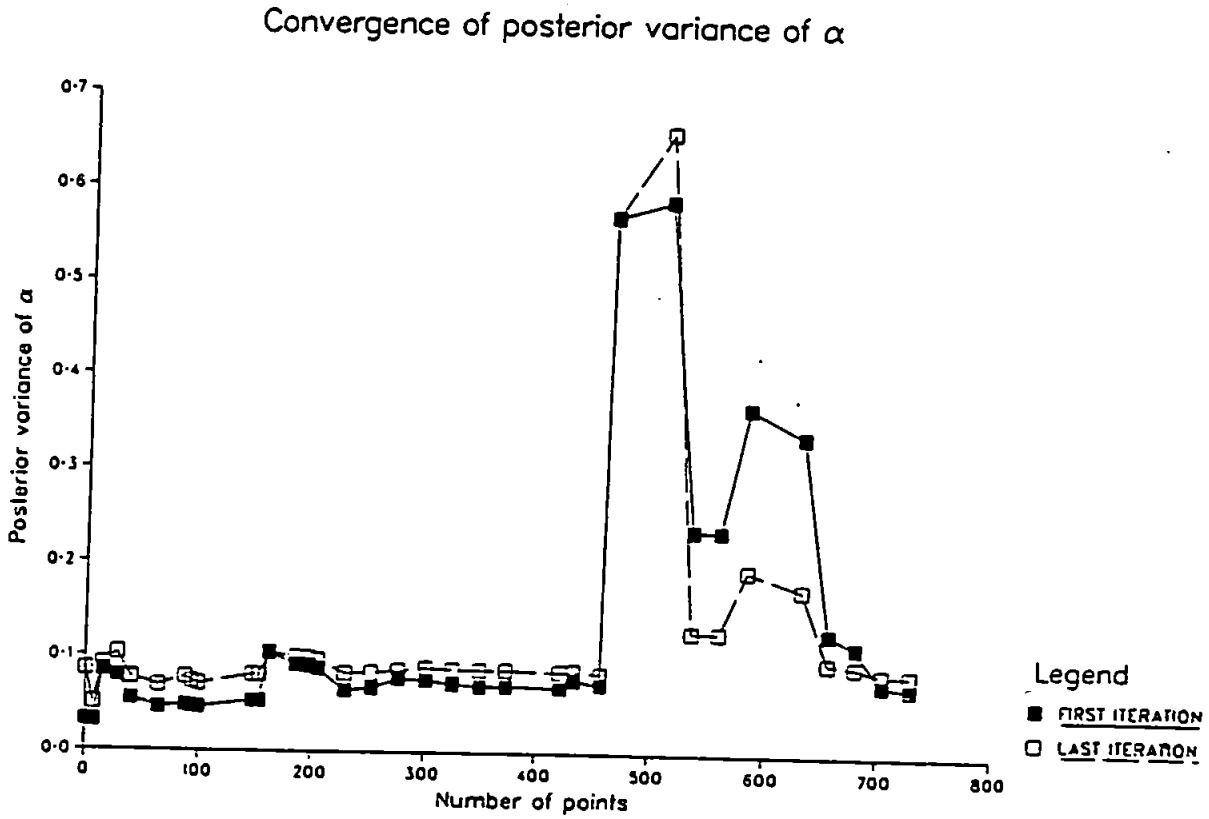


FIGURE 5.16

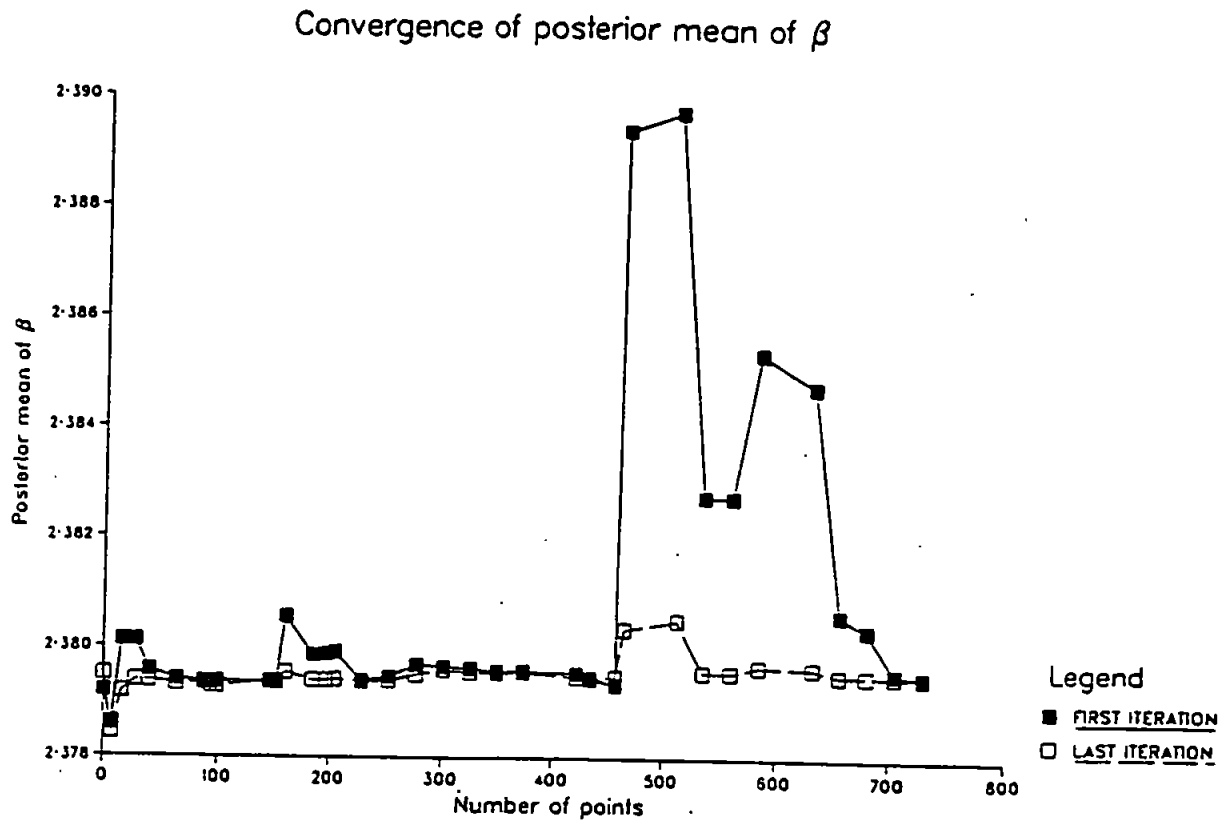


FIGURE 5.17

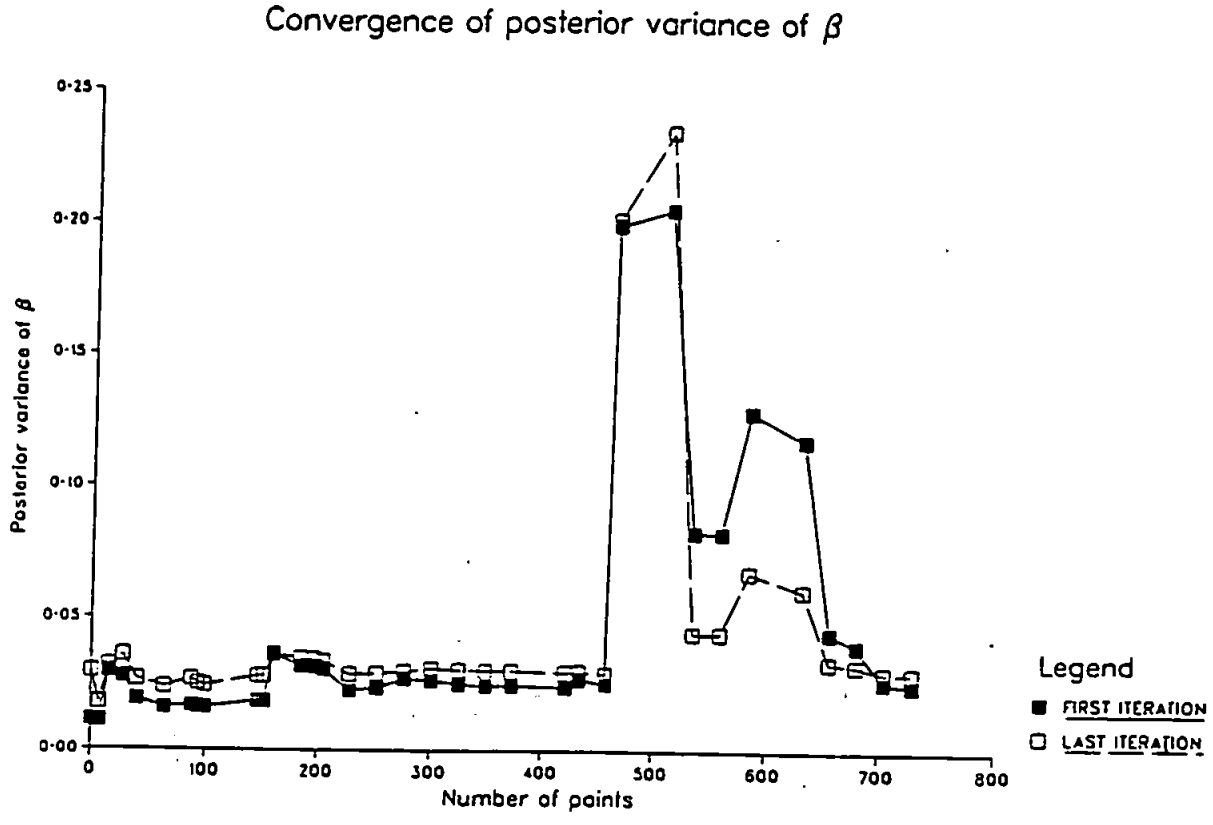


FIGURE 5.18

Convergence of posterior mean of $\log(\sigma^2)$

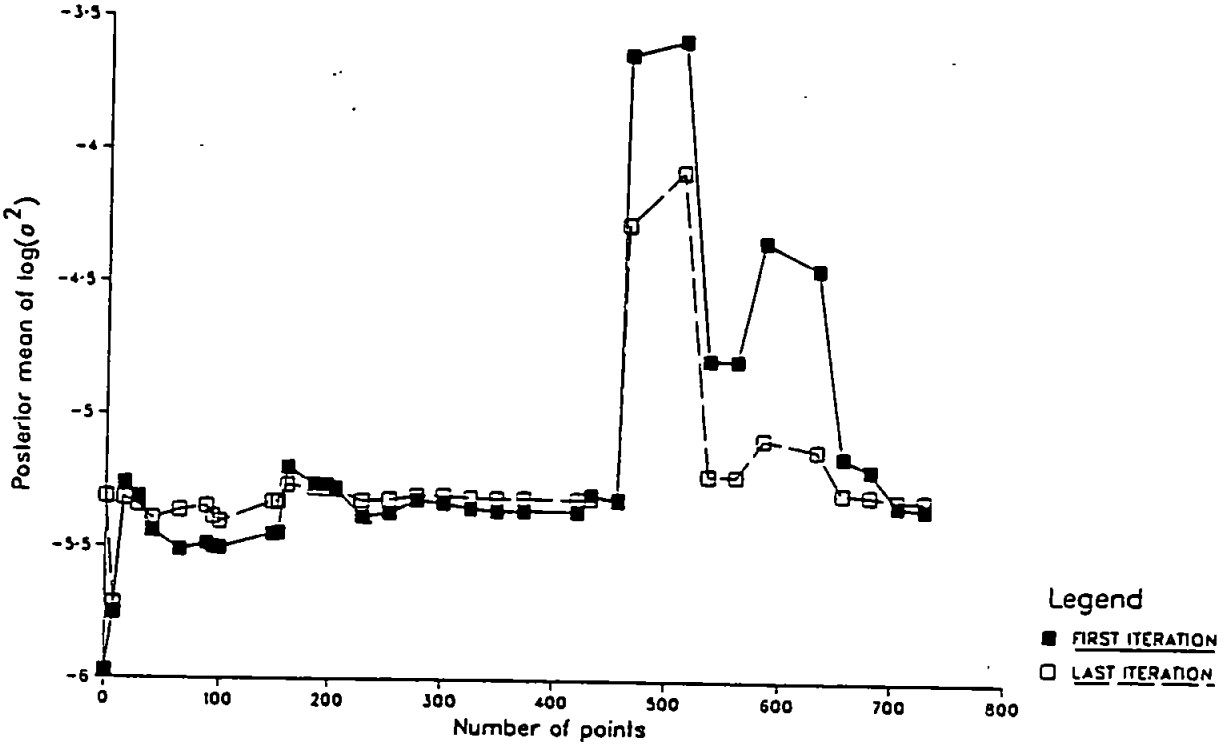


FIGURE 5.19

Convergence of posterior variance of $\log(\sigma^2)$

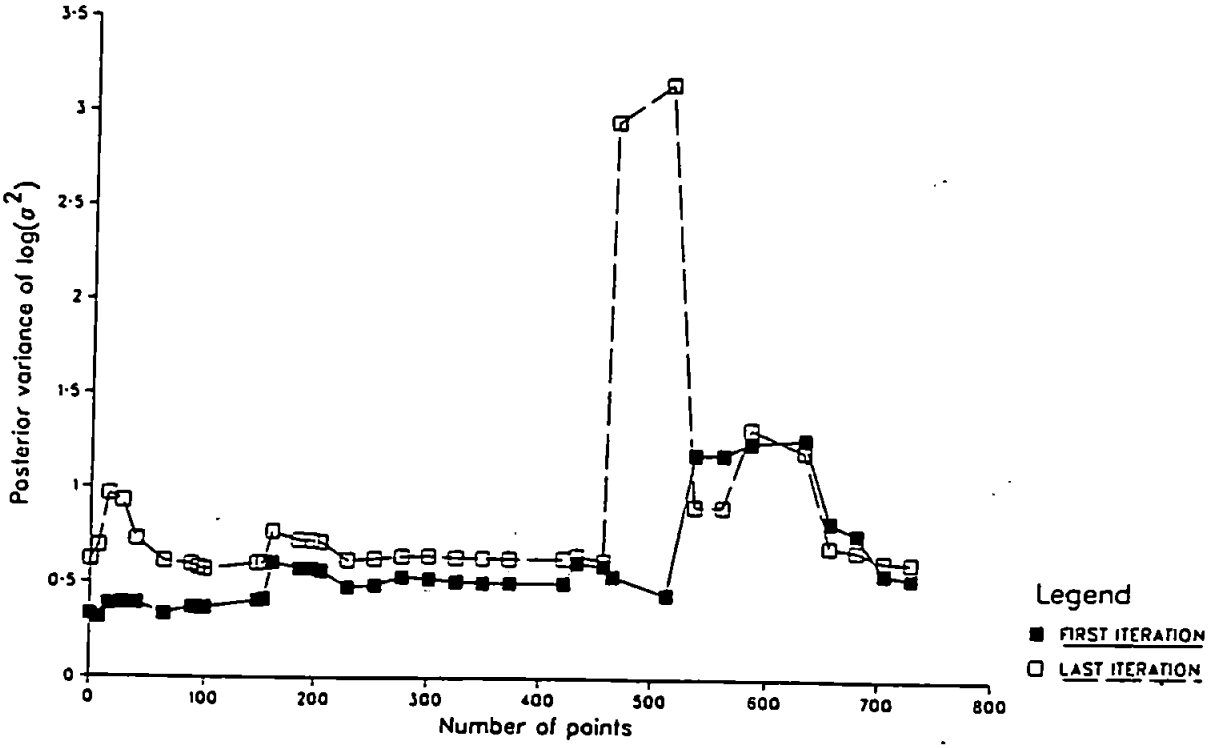
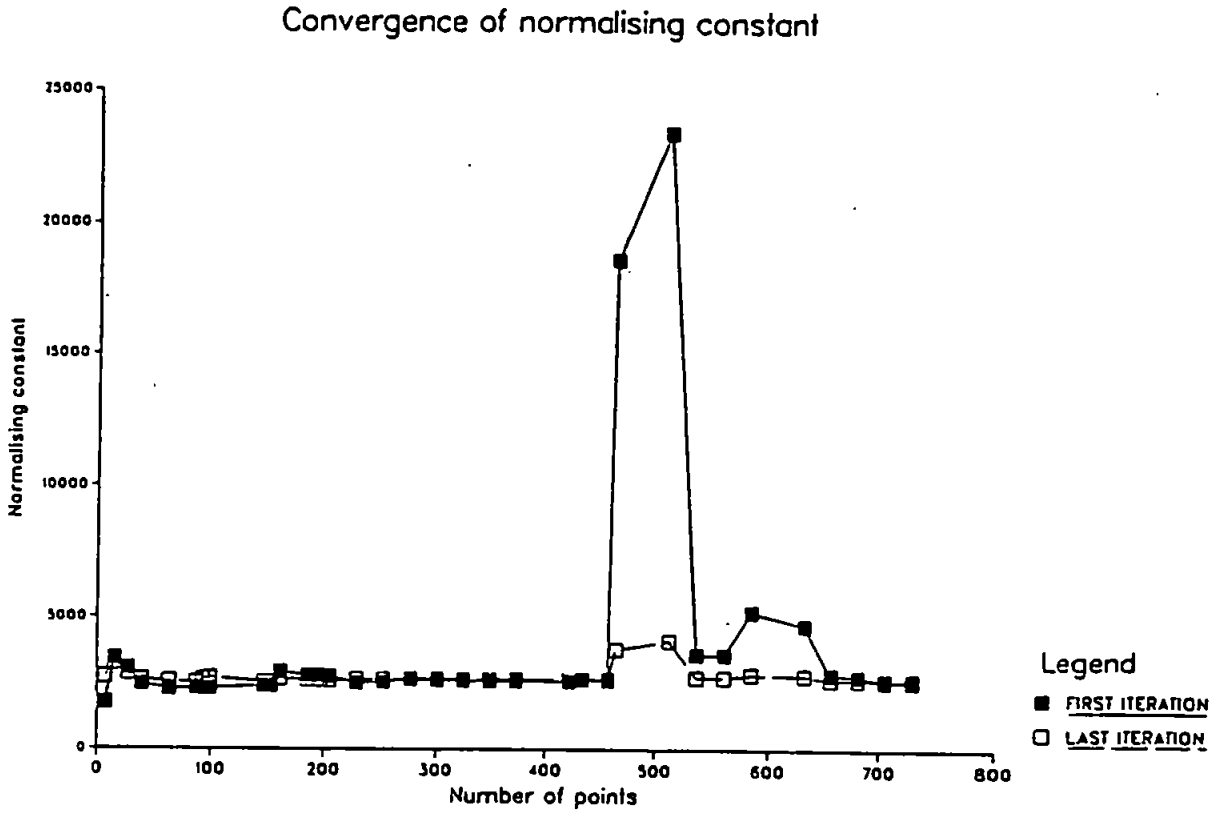


FIGURE 5.20


Convergence of normalising constant
Normalising constant
25000
20000
15000
10000
5000
0
0
100
200
300
400
500
600
700
800
Number of points
Legend
FIRST ITERATION
LAST ITERATION

5.4 A 5-dimensional example

In section 4.4.1, we used a 5-dimensional example to illustrate an imbedded sequence of a PIIR. Grieve (1987), assuming an improper locally uniform prior $p(\underline{\beta},p)$ = constant, applied the adaptive integration strategy of Naylor and Smith (1982) to obtain the marginal posterior densities of p and the regression coefficients. This involved starting with maximum likelihood estimates and the associated asymptotic covariance matrix using a 4^5 grid. Convergence was achieved after 9 iterations ending with a $6^2 \times 5^3$ grid.

Adopting the suggestion of Rabinowitz *et al* (1987) and stopping if three successive approximations show convergence, we applied the strategy of section 3.1 using $\Delta<0.001$ as our criterion for convergence. An initial imbedded PIIR sequence based on a 5^5 product rule converged after 805 function evaluations. This represents a considerable saving on the full product rule with 5^5 = 3125 points. The posterior moments were updated within the same sequence of PIIR's and convergence to the same values occurred with a further 805 points. The convergence is very rapid as can be seen from figures 4.1-4.11, and according to the strategy of section 5.1, we can stop and construct marginal densities. As a matter of interest, we moved to a 6^5 point PIIR where convergence both within the imbedded sequence of the PIIR and between the two PIIR's was achieved after a further 5056 function evaluations. Thus convergence between and within PIIR's occured after a total of 6666 function evaluations compared with the minimum 13294 we estimate were used by Grieve. Note that, if marginal densities were needed, we could add 6 more generators to end up with a full product rule and readily derived marginals, or, when

convergence occured within the sequence, we could create a mixture of integration rules (for example any combination of product,spherical or PIIR) in a manner similar to Naylor and Smith (1988b).

5.5 A 7-dimensional example

The computation labour required for the application of Gauss-Hermite product rules in more than six dimensions is enormous. For this reason, the BAYESFOUR user guide (Naylor and Shaw (1985) suggests that spherical or Monte Carlo rules should be used when the parameter space exceeds six. Indeed, BAYESFOUR does not contain any seven dimensional Gaussian rules.

In figures 4.12-4.25 we illustrated how the 7-dimensional example given in Lawless (1982, p.337) can be handled using an imbedded sequence of PIIR's. That illustration really serves as an example where any of the positive rules based on the 5^7 Gauss-Hermite product rule could be used to fill the gap in the positive integration catalogue: The currently available integration rules for 7 dimensions are the 7-degree spherical rule with 452 points (Stroud (1971), pp 317-319, rule E_n: 7-2) and the 4^7 Gauss Hermite product with 16384 points. Thus, rules taken from the sequence of the imbedded sequence of the 5^7 Gauss-product rule really serve as intermediate integration rules due to their important property to lie between two rules of specific degree (section 4.6). Moreover, even though the degree of these rules is less than 7, Result 1 of section 4.4.1 indicates that these are very powerfull in terms of the number of monomials they can integrate.

Returning to the actual statistical problem, in such experiments interest lies in the marginal densities of the parameter vector $\underline{\beta}$. Figures 4.12-4.25 illustrate that one iteration on a 7^5 sequence enables us to judge the behaviour of the integrand, and, according to the remarks of section 5.1, no further iterations are needed because the convergence is rapid. However, this can be considered as an 'expensive' approach, because the final Gauss-product rule requires 5^7=78125 function evaluations. An alternative approach would be to stop in the middle of the integration rule and, using the current estimates of the posterior moments, construct the marginal densities using mixtures of integration rules. Depending on how expensive or cheap these rules are, the former or latter approach may be more efficient. Other integration rules chosen from an imbedded sequence of PIIR's can be used in such mixtures. For example, rules taken from the 5^5 based imbedded sequence of section 4.5 can be combined together a 4^2 product rule. In this particular example, the first approach was chosen. The marginal densities are illustrated in figures 5.21-5.22.

Figure 5.5 suggests that patients with higher blood urea nitrogen measurement at diagnosis (variable z_1) have longer survival times. It also provides some evidence that exponential distribution could be appropriate instead of the Weibull (shape parameter p).

FIGURE 5.21

Marginal densities

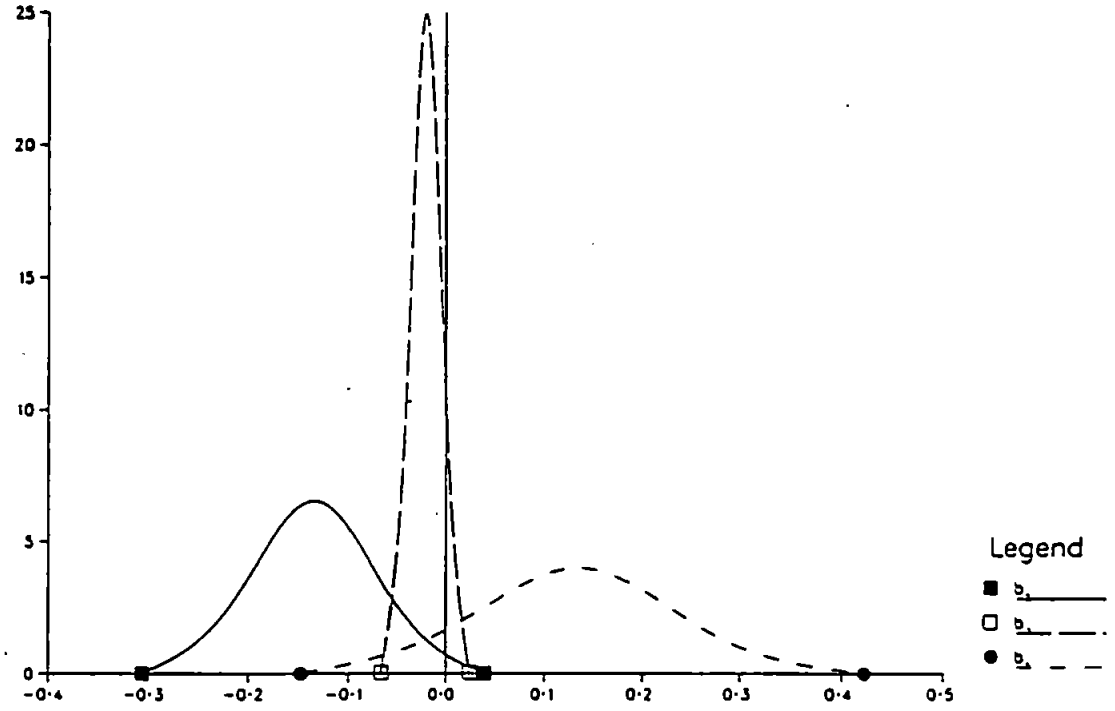
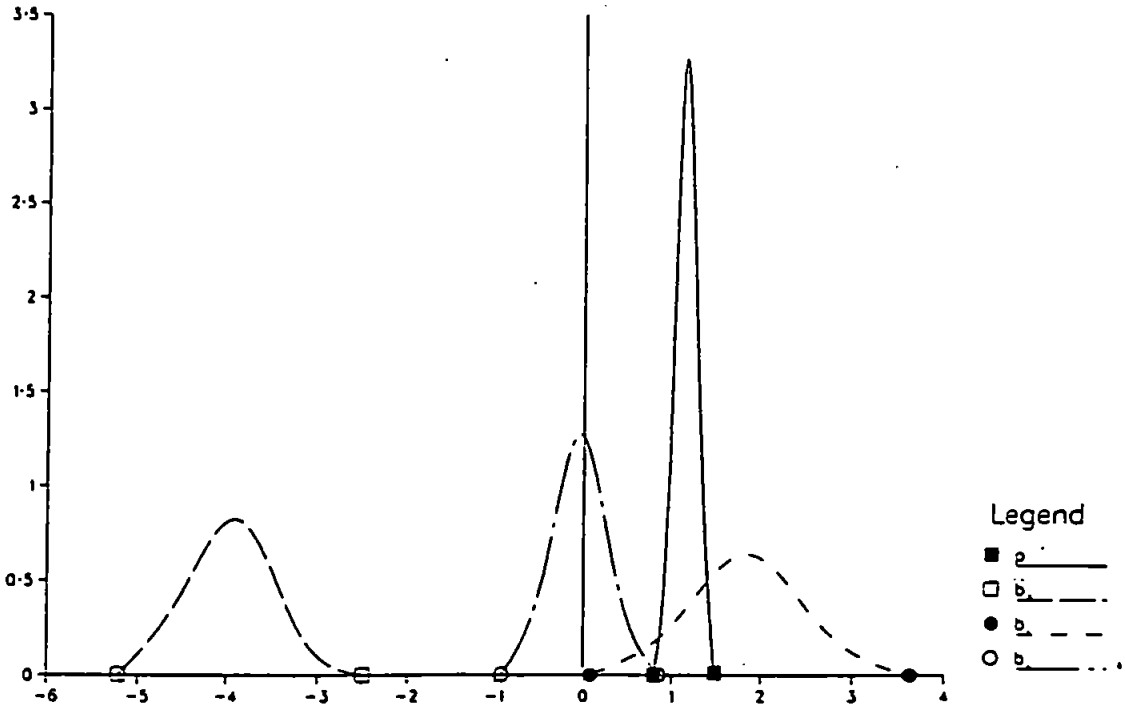


FIGURE 5.22

Marginal densities



6.1: Introduction

In section 1.3.2 we described how the Gibbs sampler can be used as a method for calculating marginal posterior densities, according to the description of Gelfand and Smith (1988). Up to the time of writing, this method has already been succesfully applied by Clayton (1989), Zeger and Karim (1989), Gelfand *et al.* (1989,1990), Racine-Poon *et al.* (1990). Such a remarkable number of published applications over such a short time, reflects the enormous potential of the Gibbs sampler. Any thesis in the general area of the implementation of Bayesian paradigm produced at this time would be incomplete without at least one chapter devoted to Gibbs sampling. Consequently, in this chapter, we will take a close look in the implementation details of the method, we will describe how it can be applied in the large family of Generalised linear models (Nelder and Wedderburn (1972)), and finally we will demonstrate it using the proportional hazards models used in sections 5.4 and 5.5. It is hoped that the importance of the work presented here will compensate for the disjointing effect on the thesis.

In the sequel, we will follow the notation as section 1.3.2.

6.2: Sampling from conditional densities

The Gibbs sampler involves drawing random samples from all full conditional densities of the form

$$p(\theta_i | \theta_j, j \neq i). \qquad (6.1)$$

Often the likelihood and prior forms specified in Bayesian analysis, lead to standard distributions in (6.1), typically normals or gammas. See for example Gelfand and Smith (1988), Gelfand *et al.*(1989). In these cases standard algorithms are available to generate random variates, see for example the books from Ripley (1987) or Devroy (1986). In other cases, a more general purpose random number generating procedure has been suggested (Gelfand and Smith (1988), Gelfand *et al.* (1989,1990)), the ratio of uniforms, see Ripley (1987). However, this method requires at least two, and possibly three, numerical maximisations. Given that a standard maximisation routine requires on average at least seven to eight function evaluations, the ratio of uniforms method can be very inefficient. In general, it is appropriate for badly behaved density functions where alternative sampling techniques are not readily available.

Another general purpose method, which is found to be very usefull in practice, is "rejection sampling", see Ripley (1987). The probability density function we need to sample from, $p(\theta)$ say, needs only to be specified up to a constant of proportionality. To obtain a sample from $p(\theta)$, choose a probability density function $g(\theta)$ and a constant $c \geq 1$ such that

$p(\theta) \leq cg(\theta)$ for every θ in the domain of p

Then, generate two independent random variates θ from g(.) and u from a Uniform (0,1). Let $T=cg(\theta)/p(\theta)$. If $uT \leq 1$ accept θ as a random variate from p(.). Otherwise generate another θ and u and repeat the process until the condition is satisfied.

The above method requires a dominating density g(.), called envelope function, a simple method for generating a random variate from it, and knowledge of the constant c. In general, careful study of p(.) can result in a suitable choice of g(.) and c. In addition, a maximisation of p(.) or of p(.)/g(.), preferably analytical but often numerical, is normally required.

The efficiency of the above rejection sampling procedure can be improved through the "squeezing" method, see for example Ripley (1987). This proceeds as follows:

Choose $g_u \equiv g$ and c as the rejection sampling method above, and also another function g_l such that $g_l \leq p(\theta)$ for all θ in the domain of p. Generate a random variate θ from g_u and independently u from a Uniform (0,1). Then proceed as follows:

```
If u ≤ g_l(θ)/g_u(θ) then accept θ else
    If u ≤ p(θ)/g_u(θ) then
        accept θ
    Else
        reject θ
    Endif
Endif
```

The function $p(\theta)$ is therefore squeezed between $g_l(\theta)$ and $g_u(\theta)$ and the calculation of the ratio $p(\theta)/g_u(\theta)$ is avoided when $u \leq p(\theta)/g_u(\theta)$. The more closely $g_u(\theta)$ and $g_l(\theta)$ 'squeeze' $p(\theta)$, the less often evaluations of $p(\theta)$ are required.

Returning to the application of Gibbs sampling, rejection sampling has often been used for sampling from full conditional densities as in (6.1). Zeger and Karim (1990), suggest choice of $cg(\theta)$ as $c_1 N(\theta^*, c_2\sigma_\theta^*)$, where θ^* and σ_θ^* are the maximum likelihood estimates of the conditional density $p(\theta_i|\theta_j, j\neq i)$, given the current simulated values of $\theta_j, j\neq i$. The constants c_1 and c_2 are chosen so that the modes of the full conditional p and $cg(.)$ are equal, and c_2 is large enough, say $c_2=2$. If the maximum likelihood estimates cannot be found analytically, Zeger and Karim apply a numerical maximisation over the full conditional to determine c_1, and then choose $c_2>1$ "to be certain the approximating Gaussian function covers the true posterior ... over the range in which θ is likely to occur".

Racine-Poon *et al.* (1990) use the same envelope function $c_1 N(\theta^*, c_2\sigma_\theta^*)$. They proceed one step further than Zeger and Karim by specifying c_2 analytically, maximising the function $h(\theta)=p(\theta)/g(\theta)$. However, they note that such an envelope function can only be used when $p(\theta)$ is log-concave. If this is not true, they use an alternative envelope function based on maximisation of $p(\theta)$.

Clayton (1989)and Forster (1990) use a histogram or a polygon as an envelope function.

All the above approaches are based on the study of the form of $p(.)$,

and consequently on the application of various devices to obtain the envelope function g(.) and the constant c. While such variate generation techniques are proved to be efficient, they are ad hoc, and depend on the mathematical background and insight of the designer.

6.3: Rejection sampling from log-concave density functions

An important class of density functions which we shall consider in the remainder of this chapter is the class of log-concave density functions. This class includes many common probability density functions. See Gilks and Wild (1990) or Devroy(1986, p.287) for a list of such densities. We begin with a formal definition of what is meant by log concavity. This is followed by description of a specified rejection sampling method for dealing with log concave density functions.

A function f on R^n is called concave if it is a twice continously differentiable real valued function on an open convex set C in R^n, and its Hessian matrix

$$H_x = (H_{ij}(\theta)), \quad H_{ij}(\theta) = \frac{\partial^2 f}{\partial\theta_i \partial\theta_j}\,(\theta_1,\ldots,\theta_n)$$

is negative semi-definite for every $\theta \epsilon C$. If the Hessian matrix is negative definite, the function f is called strictly concave. A function f on R^n is log-concave if logf is concave on its support.

The log-concavity of a density function enables us to use general

purpose algorithms for the generation of random variates. These methods, in general, require knowledge of the position of the mode. Devroye (1986, p.287-309) presents a clear account of many available methods for sampling from log-concave density functions.

Recently, Gilks and Wild (1990), proposed a method for sampling from any log-concave univariate probability density function, called "Adaptive rejection sampling". Their suggested algorithm is based on the remark that, any concave function, say f, can be bound by a piece-wise-linear upper and lower bounds (hulls), constructed using tangents at, and chords between, evaluated points on the domain of f(.). The detailed algorithm is as follows:

Assume that we need to generate random variates from the probability density function $p(\theta)=\exp(L(\theta))$. Suppose that $L(\theta)$ and $L'(\theta)$ have been evaluated at k ordered points $\theta_1,\theta_2,\ldots,\theta_k$, let $T_k = [\theta_i, i=1,\ldots,k]$, and denote the upper and lower hulls $u_k(\theta)$ and $\ell_k(\theta)$ respectively. Assume also that the mode of $L(\theta)$ lies between θ_1 and θ_k, and that $L(\theta)$ is twice continously differentiable on a real interval (a,b), where a and b can be $-\infty$ or ∞, and the second derivative must be non-positive throughout (a,b). Define

$$S_k(\theta) = \exp(u_k(\theta)) \; / \; \int \exp(u_k(\theta'))d\theta'$$

and proceed according to the following algorithm:

Repeat until desired number of points have been sampled

Sample θ from $S_k(\theta)$ and independently u from Uniform (0,1)

If $u \leq \exp(\varrho_k(\theta)-u_k(\theta))$ then

Accept θ

Else

If $u \leq \exp (L(\theta)-u_k(\theta))$ then

Accept θ

Else

Reject θ

Endif

Add θ to T_k, increament k, relabel the members of T_k

Endif

End Repeat

The Adaptive rejection sampling has two important advantages compared with other existing general purpose methods for generating independent observations from a probability density function. First, numerical maximisation is not needed, so it is more efficient. Second, it is adaptive in the sense that when more points are rejected in the rejection sampling algorithm, the probability of rejection is decreasing for the next random variate sampled from the envelope function. This happens because with the addition of more points the density function is more close to the upper and lower functions used to 'squeeze' it. Moreover, even though the Gibbs sampling normally requires only samples of size one from each conditional density, the adaptive rejection sampling can be utilised in special cases to exploit this second advantage and therefore to offer large gains in efficiency. See section 6.5 for more details.

6.4: Log-concavity and Generalised Linear models

Generalised linear models, introduced by Nelder and Wedderburn (1972), include a large class of useful statistical models. In this section we investigate the potential use of Gibbs sampling as means of making inferences about the parameters in a generalised linear model. In particular, we intend to use the adaptive rejection sampling technique introduced by Gilks and Wild (1990), so our main interest lies on the log-concavity of the likelihood function.

Let the data consist of a vector of responses $\underline{y}$ of length n, and a nxp matrix of regressors $\underline{Z}$ of known constants. The responses $\underline{y}$ are assumed to be a realisation of a vector of random variables $\underline{Y}$ independently distributed with means $\underline{\mu}$. Generalised linear models are characterised by the following structure.

(i) The distribution of the responses is assumed to belong to a natural exponential family

$$f(y|\theta) = \exp[\ (\theta y - b(\theta))\ /\ a(\varphi) + c(y,\varphi)\ ]$$

for some functions a(.),b(.) and c(.), for a natural parameter θ. Many parametric density functions belong to this family, for example Binomial, Normal, Poisson, Gamma.

(ii) The matrix $\underline{Z}$ influences $\underline{y}$ via a linear combination $\underline{\eta}=\underline{Z\beta}$, where $\underline{\beta}$ is a p-dimensional parameter vector and $\underline{\eta}$ is a vector termed the linear predictor.

(iii) The linear predictor $\underline{\eta}$ is related to the mean $\underline{\mu}$ of $\underline{Y}$ by a a link function g, such that $\eta_i=g(\mu_i)$, i=1,...,n. Of special importance are the natural link functions, which occur when $\theta=\underline{Z\beta}$, for the natural parameter θ.

The above family includes some very well known models. For example for the normal distribution and the natural link function $g(\mu)=\mu$ we obtain the classical linear regression model. For the Poisson distribution, the natural link function $g(\mu)=\log(\mu)$ gives rise to the log-linear Poisson model which can be used, for example, in the analysis of multidimensional contigency tables. When the responses follow the binomial distribution with mean π, the link functions $g(\pi)=\text{logit}(\pi)$, $g(\pi)=\Phi^{-1}(\pi)$ and $g(\pi)=\log(-\log(1-\pi))$ yield the logistic, probit and the complementary log-log model respectively.

Now interest lies on the log-likelihood function L of a certain generalised linear model. Maximum likelihood estimators are frequently used to estimate the vector $\underline{\beta}$ of coefficients of the linear combination $\underline{Z\beta}$, see for example McCullagh and Nelder (1989). These methods rely heavily on the asymptotic properties of the maximum likelihood estimators as the sample size n of observations tends to infinity. In particular, certain regularity conditions have been given by different authors which guarantee, at least for natural link functions, weak consistency and asymptotic normality of the maximum likelihood estimators, see for example Haberman (1977) and Fahmeir and Kafmann (1980). Among others, these regularity conditions assume that the Fisher information matrix is positive definite. This assumption is of great importance because, in the case of natural link functions,

the Hessian and the information matrix coincide, see Nelder and McCullagh (1989, p.43), and therefore log-concavity of the log-likelihood is automatically implied.

Unfortunately, the above result cannot be generalised for non-natural link functions. However, Wedderburn (1972) provides a series of special caes in which he proves log-concavity of the likelihood function. Hs results are summarised as follows:

Normal: L is strictly concave only in the case of the natural link function, g=id.

Gamma: Strict log-concavity is attained for $g(\mu)=\log(\mu)$ and $g(\mu)=\mu^x$ $(-1\leq x<0)$. It is assumed here that $y_i\geq 0$ for every i, $i=1,2,\ldots,n$.

Poisson: L is strictly concave for $g(\mu)=\log(\mu)$ and $g(\mu)=\mu^x$ $(0<x<1)$. For the link function $g(\mu)=\mu$ the log-likelihood is strictly concave if $y_i>0$ for every i, and concave for any value of y_i.

Binomial: The logistic, probit and complementary log-log models defined above attain strict log-concavity of the likelihood function. L is also strictly concave for the link functions $g(\mu)=\mu$, and $g(\mu)=\sin^{-1}\sqrt{\mu}$.

An interesting point should be made here. Wedderburn (1972) shows that for the logistic, probit and complementary log-log models, the maximum likelihood estimators are guaranteed to be finite only when $0<y_i<m_i$ for every i, where y_i is the number of positive responses out of m_i trials. In addition, for the last two link functions, $g(\mu)=\mu$ and $g(\mu)=\sin^{-1}\sqrt{\mu}$, the finitiness of the maximum likelihood estimates is not guaranteed. However, in the Bayesian context, the prior distribution should overcome this problem yielding a well behaved

posterior density. Consequently, the difficulties which arise in the maximum likelihood estimation approach do not occur when a Bayesian approach with a suitable prior is adopted.

The application of the adaptive rejection sampling method described in section 6.3 requires the log-cocavity of the full conditional distributions. We have shown that for certain cases of Generalised linear models the likelihood is log-concave. We remark that this statement implies log-concavity of the full conditional likelihood because, from definition, the Hessian matrix is negative semi-definite so all its diagonal elements are non-positive. Before we proceed to investigate the log-concavity of full joint conditional densities, we include in the special cases of Generalised linear models another large family of models.

Weibull, exponential and extreme value: These distributions can be used for modelling censored survival data in which the response variate is the lifetime of a component or the survival time of a patient, see Kay (1977), Aitkin and Clayton (1980). The link function is the same as for log-linear Poisson models, $g(\mu)=\log(\mu)$, except that there is a fixed intercept (offset) included in the linear predictor. It is straightforward to prove that for each of the above distributions used for proportional hazards models, the full conditional likelihood is concave. First, note that the likelihood under the Weibull model is given by

$$L(\underline{\beta},\rho/\text{data}) = \left[\prod_{j=1}^{n} \rho t_j^{\rho-1} e^{\underline{z}_j\underline{\beta}} \right] \left[\prod_{j=1}^{n+m} \exp[-t_j^{\rho} e^{\underline{z}_j\underline{\beta}}] \right] \qquad (6.2)$$

where $t_j, j=1,\ldots,n$ and $t_j, j=n+1,\ldots,m$ denote the uncensored and censored lifetimes respectively, ρ is the shape parameter of the Weibull distribution ($\rho>0$), and $\underline{z}_j, j=1,\ldots,m$ is the vector of covariates for the j^{th} case. Then simple manipulation yields

$$\frac{\partial^2 \ell ogL}{\partial \beta_k{}^2} = -\sum_j z_{kj}{}^2 \mu_j \leqslant 0, \text{ for a parameter } \beta_k, \ 1 \leqslant k \leqslant p, \qquad (6.3)$$

$$\frac{\partial^2 \ell ogL}{\partial \rho^2} = -n/\rho^2 - \Sigma(\log t_j)^2 \mu_j < 0$$

where $\log \mu_j = \rho \log t_j + \underline{z}_j \underline{\beta}$.

Thus (6.3) guarantees log-convavity under the Weibull model. Furthermore, note that (6.2) holds also for the exponential model, being in fact a special case of the Weibull with $\rho=1$. Finally, the transformation $u=e^t$ in (6.2) yields the extreme value distribution, so log-concavity is readily shown for this model too.

According to the Bayesian paradigm, a prior density function is placed on every parameter which, combined with the information from the data obtained through the likelihood function, yields the posterior distribution. Therefore, the full conditional posterior log-density function is derived as a sum of the full conditional log-likelihood and the logarithm of the prior density function. Consequently, if the prior density function is log-concave, the full posterior conditional will be log-concave, as a sum of two log-concave functions (see Rockafellar (1972).

In cases where the prior is not log-concave, a sampling-resampling technique described in Smith and Gelfand (1990) can be adopted, see also Stephens and Smith (1990): Assume that we need to sample from a function $p_1(\theta)$, but a sampling technique is not readily available. Furthermore, suppose that samples from another function $p_2(\theta)$ are available, say $\theta_1,\ldots,\theta_n$. Then calculate $w_i=p_1(\theta_i)/p_2(\theta_i)$ and $q_i=w_i/\Sigma w_i$. Draw a θ^* from the discrete distribution over $[\theta_1,\ldots,\theta_n]$ with probability masses q_i on θ_i. Then θ_i^* is approximately distributed according to p_2.

Thus, the requirements for the adaptive rejection sampling are fulfilled for certain -and in fact, most common- Generalised linear models under any prior density specifications. Gibbs sampling is therefore applicable for making inferences about the parameters of interest, when a Bayesian approach is adopted.

6.5: Optimising Gibbs algorithm

One of the major advantages of the adaptive rejection algorithm introduced by Gilks and Wild (1990), is that it very efficient when samples are drawn repeatedly. In fact, Gilks and Wild report that in general the number of evaluations needed for a sample of size n increases approximately in proportion to the cube root of n. However, for the Gibbs sampling, only samples of size 1 are required in each iteration, and consequently this gain of efficiency can not be utilised. However, in this section we will demonstrate how, in some special cases, we can make use of this property of the adaptive rejection sampling and speed up the Gibbs sampling algorithm.

The model used in section 5.4 is a proportional hazards model of the type (6.2), with all covariates $\underline{z}_j$ being zero or one. Suppose that we wish to apply the Gibbs sampling (Gelfand and Smith (1988)) to make inferences about the parameters ρ and $\beta_i, i=1,\ldots,4$. For the implementation of Gibbs sampling algorithm, independent observations must be available from each full conditional density. First note that the full conditional for the shape parameter of the Weibull distribution ρ is given by

$$p(\rho|\underline{\beta},\text{data}) \propto \rho^{n} \left(\prod_{j=1}^{n} t_j\right)^{\rho} \prod_{j=1}^{n+m} \exp\left(-t_j^{\rho} \prod_{k=1}^{4} \left(e^{\beta_k z_{jk}}\right)^{\rho}\right) \qquad (6.4)$$

and samples from (6.4) are not readily available. The conditional (6.4) is however, according to section 6.4, log-concave, and adaptive rejection sampling can be used to sample from it.

The conditional densities for the parameters in the linear predictor $\underline{Z\beta}$ are given by

$$p(\beta_i|\beta_{\ell,\ell\neq i},\rho,\text{data}) = \prod_{j=1}^{m+n} \left(\alpha_{ij}\psi_i^{z_{ij}}\right)^{w_j} \exp\left(-\gamma_{ij}\psi^{z_{ij}}\right) \qquad (6.5)$$

where $\alpha_{ij} = \rho t_j^{\rho-1} \prod_{k\neq i} \exp(\beta_k z_{jk})$

$\gamma_{ij} = \rho t_j \alpha_{ij}$

$\psi_i = \exp(\beta_i)$

$$w_j = \begin{cases} 1, & j \leqslant n \\ 0, & j > n \end{cases}$$

Noting that the regressors matrix $\underline{Z}$ contains only 0 or 1, (section 4.5.1), (6.5) can be written as

$$p(\beta_i | \beta_{\ell, \ell \neq i}, \rho, \text{data}) \propto \psi_i^A \exp(-B\psi_i) \qquad (6.6)$$

where $$A = \sum_{j=1}^{m+n} z_{ji} w_j$$

$$B = \sum_{j=1}^{m+n} \gamma_{ij} z_{ji}$$

Therefore, samples from (6.6) can be drawn simply sampling from a gamma density Ga(A+1,B), using well known methods, see Ripley (1987), and then trasforming the sampled variates $\psi_i \rightarrow \log\psi_i = \beta_i$.

A closer look of (6.5) reveals an interesting fearure which can be used to speed up the sampling procedure. Suppose that a sample of size 1 was drawn from

$$p(\beta_i | \beta_\ell = b_\ell, \ell \neq i, \rho = p) \qquad (6.7)$$

and, according to the Gibbs sampling algorithm, in a later stage another sample of size 1 needs to be obtained from

$$p(\beta_i | \beta_\ell = b'_\ell, \ell \neq i, \rho = p'). \qquad (6.8)$$

Both (6.7) and (6.8) are of the form (6.6), simply substituting ψ_i by $\exp(\beta_i)$. Let

$$\gamma_{ij} = t_j^p \prod_{k \neq i} \exp(b_k z_{jk})$$

and

$$\gamma'_{ij} = t_j^{p'} \prod_{k \neq i} \exp(b'_k z_{jk})$$

Then, instead of sampling from (6.8), we can generate an independent observation from (6.7), say b_i, and apply the transformation

$$b_i \rightarrow b_i + \log\left(\sum_j \gamma_{ij} z_{ji} \Big/ \sum_j \gamma_{ij} z_{ji} \right) \equiv b'_i$$

Then, simple manipulation shows that b'_i is a indepentend observation drawn from (6.8). Consequently, the above argument allows us, during the Gibbs sampling algorithm, to sample repeatedly from the same conditional density and then simply rescaling the sampled variates according to the updated values of the other parameters. This is a great advantage if a method such the adaptive rejection sampling is to be used, according to our comments at the end of section (6.3). The gain in the efficiency in the above example will be demonstrated in the next section.

6.6: Illustartive examples

6.6.1: A proportional hazards model

In this section we analyse in this section the proportional hazards model (6.2), adopting the Gibbs sampling approach. The full conditional densities are given by (6.4) and (6.5), and cannot be simplified to allow sampling from any known density function. Consequently, we adopt the adaptive rejection sampling introduced by Gilks and Wilk (1990) to sample from the full conditionals. These conditionals are log-concave according to section 6.4, so the requirements for adaptive rejection sampling ar fulfilled.

Gibbs sampling (section 1.3.2) requires initial values for all but one parameter. We gave initial values to the parameteres $\beta_i, i=1,..,4$, taken from the maximum likelihood estimates. These initial values have been also given to the application of the imbedded sequences of PIIR's for the same example, see section 5.5. Thus, comparisons of the two methods can be made on fair grounds.

At each iteration of the Gibbs sampling, adaptive rejection sampling from every conditional density requires (at least) two points which can be used as initial points for the construction of upper, using tangents, and lower, using chords, bounds. These initial points were taken as the sample mean ± one standard deviation, where the sample moments were calculated from the previous iteration of Gibbs sampling. In cases where the two initial points did not lie to each side of the mode of the conditional density, additional points were supplied.

Gibbs sampling converged after 70 iterations, using 500 replications in each iteration. The resulting marginals, closely resemble the marginals derived from the imbedded sequences of PIIR's. In figures 6.1-6.7, we illustrate the marginals derived with the imbedded sequences of PIIR's in the analysis of section 5.5, and the marginals from the Gibbs sampling after 60 and 70 iterations. Note that the marginals do not exactly coincide. In fact, in some cases there are differences even between the two marginals derived with the Gibbs sampling approach.

The differences in the marginals in figures 6.1-6.7 might imply that the Gibbs sampling has not converged after 70 iterations. While the matter of convergence is currently a difficult problem, different ways have suggested to overcome it. Gelfand and Smith (1988) suggest the use of Q-Q plots or graphical comparison of marginals, derived at regural intervals during iterations. Following this avenue, Gelfand *et al*. (1990) suggest comparing marginals derived every 5 iterations. Foster (1990) uses as indicators sample moments. All these methods can provide an informal assesment of the convergence, and in fact, while not being rigorously justified, they should normally be reliable. In our example, we checked the marginals and the sample moments every 10 iterations. We believe that, the discrepancies between the marginals are justified by the fact that they are constructed from a finite sample, and that oscillations of the sample moments around the true posterior moments should be expacted. Moreover, the inferences made in section 5.5 do not change because of these discrepancies, at least for this particular example. It is also noteworthy that the differencies in inferences drawn from Gibbs

FIGURE 6.1

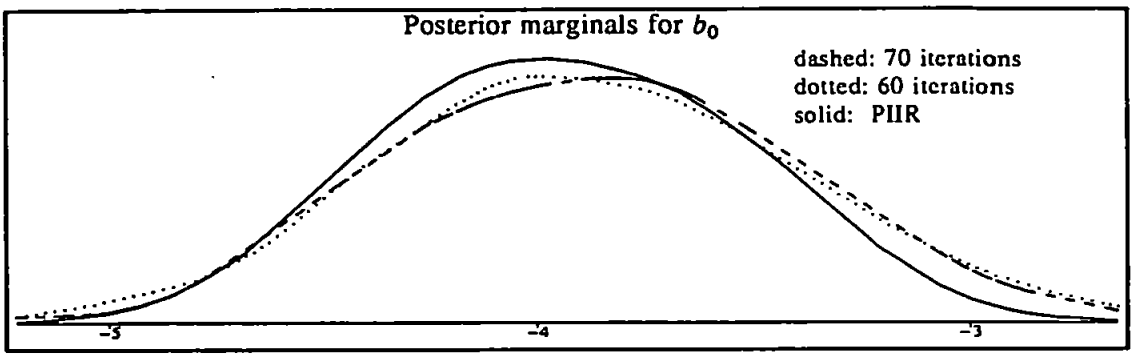


FIGURE 6.2

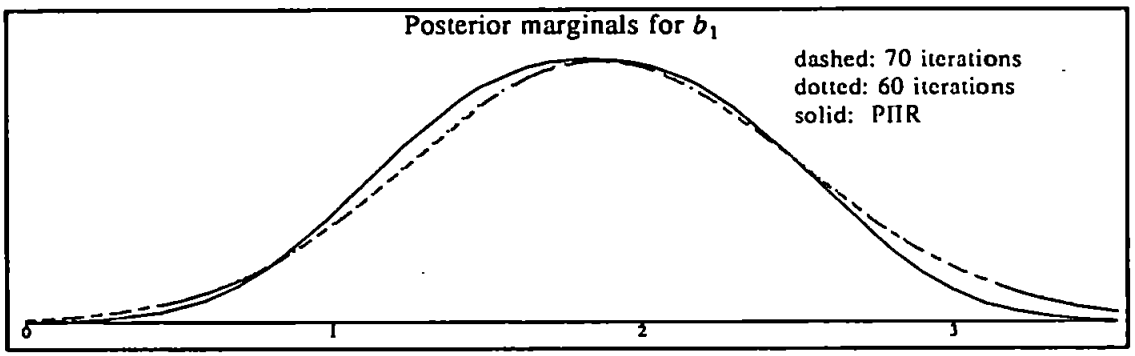


FIGURE 6.3

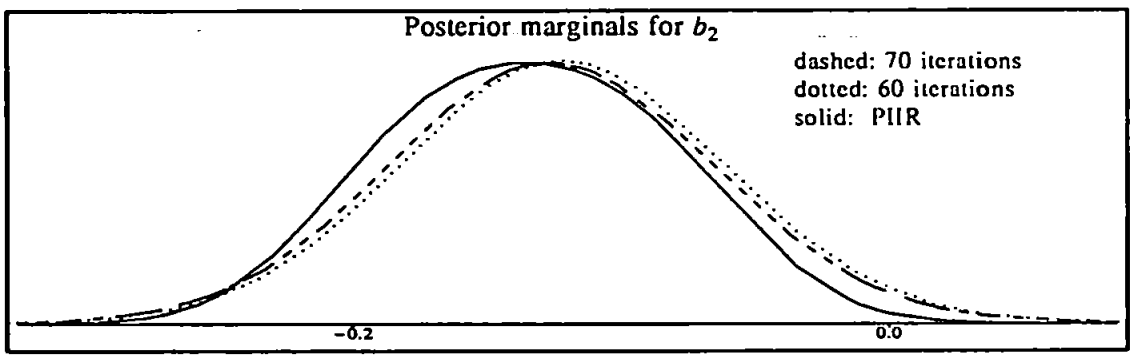


FIGURE 6.4

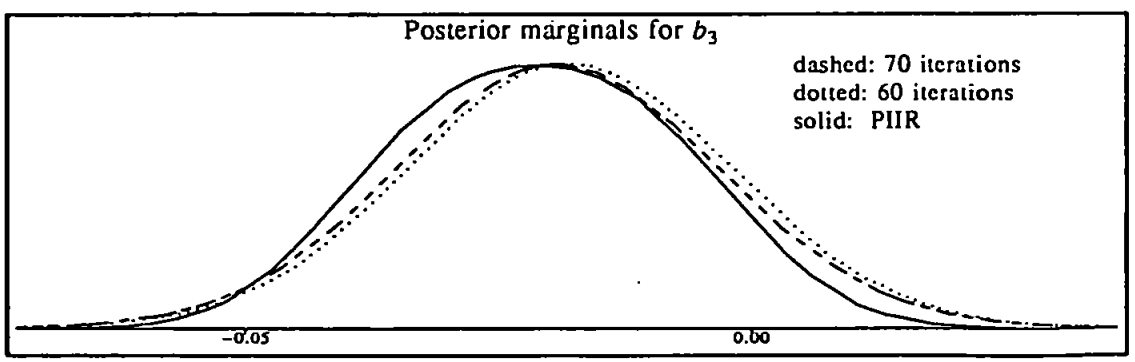


FIGURE 6.5

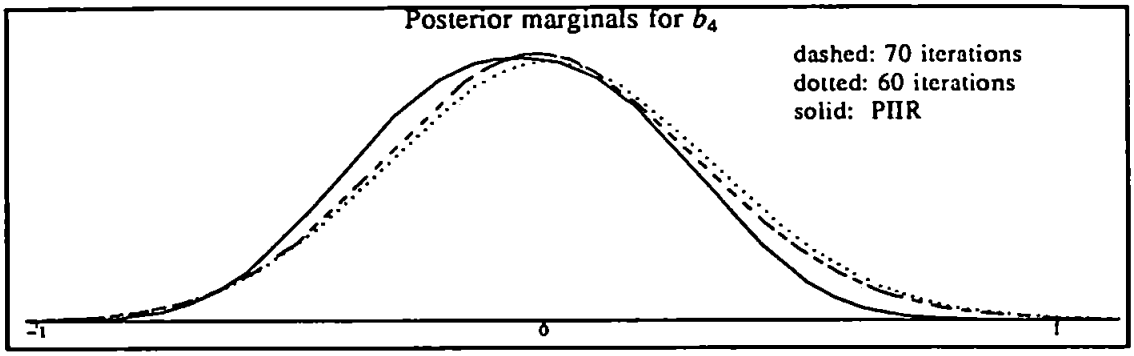


FIGURE 6.6

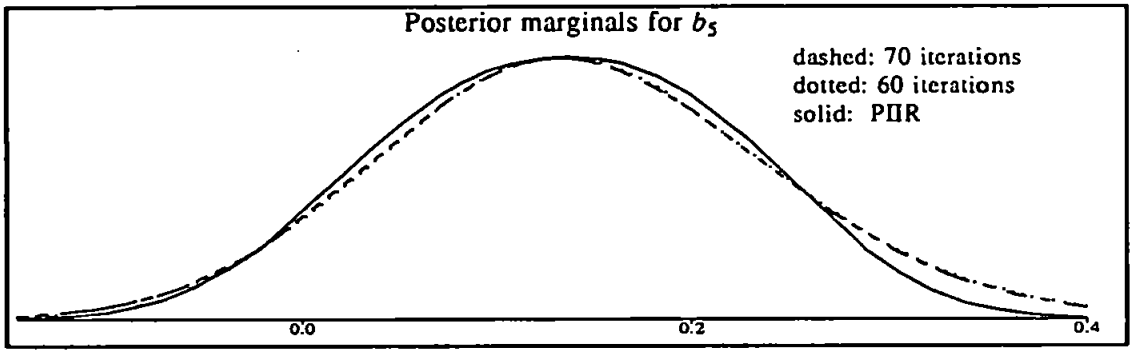
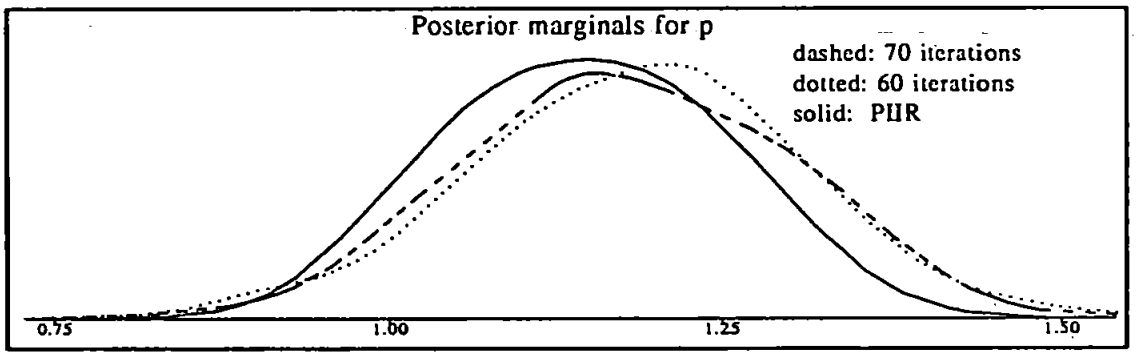


FIGURE 6.7



sampling and from the numerical integration are of minor importance when compared with maximum likelihood estimation inferences.

Another reason for the marginal descrepancies might be the high correlation between the shape parameter ρ and the parameter vector of regressors $\underline{\beta}$. A linear orthogonalising transformation of the type used in numerical integration strategies, see section 2.2.1.2, can be applied, but this still remains a matter of future research. Another policy to overcome this problem has been suggested by Zeger and Karim (1989). They propose multiple samples from highly correlated variables. Thus, in our example, for each sampled variate for $\underline{\beta}$, we might sample 10 variates for ρ.

An average of 3.94 function evaluations were used for the sample of size one from each conditional density. Our initial starting points for the adaptive rejection sampling were very poor, giving average function evaluations for each conditional at the first iteration 4.16. This is a considerable gain in the efficiency compared with other black-box sampling techniques, for example ratio of uniforms, which requires at least two numerical maximisations spending on average for each one 7.5 function evaluations.

Of course, comparisons with numerical integration techniques should not be made only in terms of the efficiency. A more general comparison will be made in the next section.

6.6.2: A special case of proportional hazards model

We discuss in this section the details of the application of the special proportional hazards model described in section 6.5. We recall that during the Gibbs sampling, this special case allows for each of the regressor parameters to be sampled only from one conditional distribution, rather than differrent conditionals. We applied the data of section 4.5.1 which, were analysed using imbedded sequences of PIIR's in section 5.4.

Convergence was achieved after 150 iterations using 500 replications. Comparisons of illustrative marginals and sample moments were made every 50 iterations. The constructed marginal densities are shown in figure 6.8.

A normal application of adaptive rejection sampling for each of the 4 conditional densities of the regressor parameters β_i, i=1,...,4, whould require, as in section 6.6.1, on average 3.94 function evaluations. Consequently, for the whole analysis, and for these particular 4 parameters, an approximate total of 500x150x4x3.94=1182000≈1.2 million function evaluations would be required. Table 6.1 shows the function evaluations required when the algorithm is optimised using the algorithm described in section 6.5.

It is clear that the gain of the efficiency is outstanding, the optimised algorithm requiring only 493 function evaluations, compared with approximately 1.2 million function evaluations needed for a usual application of adaptive rejection sampling.

FIGURE 6.8

Marginal densities

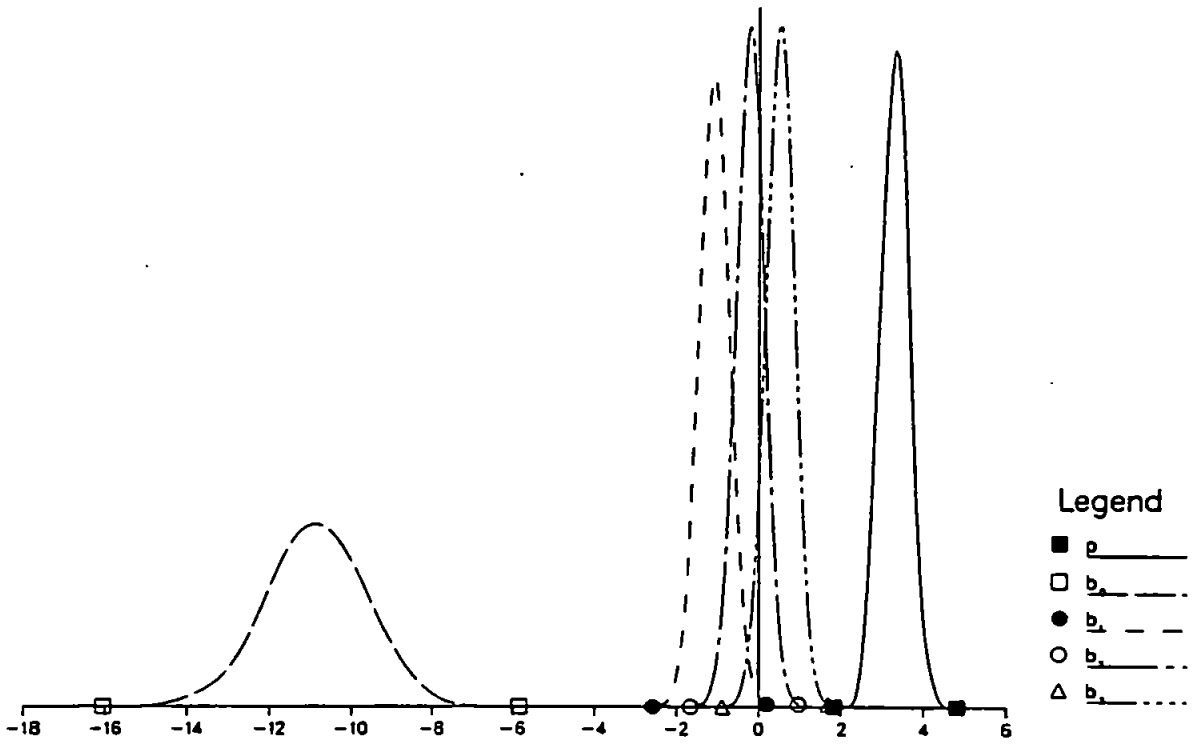


Table 6.1: Function evaluations for 4 regressor parameters

Iteration number	Total function evaluations for 4 parameters
1	99
2	124
3	141
4	156
5	169
10	205
20	255
30	301
40	322
50	346
100	440
150	493

6.7: Discussion

There are at least 4 different approaches for the implementation of the Bayesian paradigm: Numerical integration strategies, analytic approximations, Monte Carlo integration techniques and the Gibbs sampling approach. The main part of this thesis investigated numerical integration strategies, but, the current chapter considered application of Gibbs sampling as an alternative way to make Bayesian inferences. Inevitably, this section will concentrate on comparisons between the two approaches.

Suppose that a user, experienced or unexperienced, faced with an analytically intractable or tedious problem, wishes to choose a general purpose implementation technique to make use of the Bayesian paradigm.

If a numerical integration technique is to be employed, the user has to provide the functional forms of the likelihood and the prior. Then, a close look of the parameters of interest should reveal whether or not parameter transformations are necessary to satisfy assumptions of normality. If these transformations need to be made, the user must supply the Jacobian matrix of these transformations. Then, a chosen numerical integration strategy requires an interactive implementation by the user. This involves mainly choice of integration rules and decision making concerning the convergence of the numerical approximations. Furthermore, choice of the order of orthogonalising parameters must be constantly made, depending on which are the parameters of interest.

Gibbs sampling approach requires the functional form of full conditional density functions, at least up to a constant of proportionality. In addition, the Gibbs sampling user must provide a way to generate independent observations from each of the conditionals. This might involve use of sophisticated sampling techniques which in turn might require numerical maximisations or functional forms of the derivatives of conditionals. Then, the Gibbs sampling updating scheme offers the flexibility of batch run: The user must provide only once the number of replicates and the number of iterations, and, given that these numbers are adequately large, can simply obtain desired posterior marginal or predictive density functions.

The different philosophy of the two methods is evident. The numerical integration strategies, attacking the problem directly, evaluate high dimensional integrals using sophisticated techniques based on certain assumptions. Transformation-choice before the implementation of the problem, must be followed with assumption-checking by means of an interactive running. On the other hand, Gibbs sampling attacks the problem indirectly, in the sense that integrals are not calculated. The asumptions for implementing an inference problem are weaker, and after the initial choice of a sampling technique, the system is fail-safe, in the sense that, convergence will be eventually obtained.

The difference between the two approaches would not be so evident, if, for every possible problem, proper transformations could be obtained and a numerical integration approach was possible. However, at the time of writting, certain models such as hierarchical models or models with multimodal posterior densities seem to be unsolvable with

numerical integration techniques. Therefore, it is important to note that the two techniques do not always overlap at the range of possible applications.

The numerical integration strategies are limited in their applicability in cases where the number of parameters is relatively small. In fact, even with the current speed of computers, numerical integration does not seem able to cope with more than ten parameters. On the other hand, Gibbs sampling, exploiting the ability to be implemented via batch jobs, can provide answers to very high-dimensional problems.

Note however, that Gibbs sampling cannot compete with numerical integration in terms of efficiency, the yardstick taken either the number of function evaluations or the computer running time. Consequently, for low-dimensional problems where the assumptions of approximate normality are satisfied, the numerical integration is the most suitable approach. For example, the proportional hazard models analysed repeatedly in this thesis, are classical examples where a numerical integration stategy provides the fastest and most easily obained results.

Before choosing a strategy to implement the Bayesian paradigm, the potential user of either the above methods, must keep in mind that, at least at the time of writting, numerical integration and Gibbs sampling overlap minimaly in respect to the range of problems in which they should be applied. Dimensionality, need for initial transformations and efficiency must all taken into consideration before a decision is made.

7. Criticisms and future research

This section outlines possible topics for future research together with some extensions of previous sections. The essence of the material presented in the thesis, regarded not only as a topic for theoretical self-gratification, but, also as an aid to increasing the diversity of creative statistical thinking, will be criticised.

Even though research in the applications of numerical integration techniques in Bayesian analysis started one decade ago, it seems that progress in this field has passed its peak (1982-1985) and has slowed down in recent years. Among other things, this could be explained from an apparent inability to provide software to the four potential users (Smith (1988)): The Bayesian research statistician, the non-Bayesian research statistician, the broadly-focussed applied statistician and the student.

A major drawback of the integration techniques described in this thesis is the need for an initial parameter transformation (see section 2.5.4). Progress towards overcoming this drawback has been made by the work of Hills (1989), but much remains to be done. An ideal scheme would be one in which transformations were made automatically. Given that the Naylor and Smith algorithm contains two iterative procedures (between and within grid size), it is questionable why, for a given grid size, the only information updated relates to mean vector and the covariance matrix, and not to other information which might possibly a more appropriate parameter transformation. An interesting idea would be to use prior information from previous function evaluations in order to choose a parameter, λ

say, which will produce a parameterisation $\psi_\lambda(\theta)=\varphi$, for a family of distributions ψ_λ with φ close to normal, or at least more 'symmetric' than θ.

According to our results in sections 2.5.2 and 3.4, Shaw's results described in section 2.5.3, and following the discussion and our proposals in sections 2.5.4 and 5.1 respectively, it is doubtful that updating the mean vector and the covariance matrix is the best policy to achieve maximum efficiency. We believe that, in most cases, the behaviour of the imbedded sequence of PIIR's is a good indication for deciding whether or not rescaling and/or relocation should be performed. However, until this takes the form of a formal (and hopefully, automatic) decision-making criterion, support of one or other opinion becomes problematic.

Similarly, consider the related problem of error estimation. If the error (4.5) can be estimated more accurately, a decision can be made as far as relocation and rescaling are concerned. The problem here is that, by adding some nodes to the integration rule, in a sense we subtract some terms from the error in (4.1). Unfortunately, these terms can not be estimated, and therefore the meaning of the aggregate measure Δ (section 2.3) is more or less unjustified. Appropriate choice of a null rule, or a set of null rules (section 4.4.2) could possibly help in this direction.

All available numerical integration rules ultimately possess the same danger: the situation where convergence is not achieved however much the grid size is increased. A general purpose multidimensional integration package may then possibly be adopted, but its efficiency

will be very low because it would not exploit the asymptotic behaviour of the posterior density. The use of Monte Carlo methods (section 1.3.2) seems the best available policy at the present time.

The above problems, together with the proposed solutions, give a current status of the research and the potential applications of the numerical integration in Bayesian Statistics. It has been noted indirectly in section 2.5.4, and needs highlighting in the epilogue of the thesis, that it seems that at the moment of writing only the first potential user among the 4 mentioned above can use numerical integration techniques in Bayesian analysis. This happens because only an expert user can overcome the above problems using his judgement obtained from his experience. We would like to believe, however, that the solution of the above problems will lead to more general adoption of numerical integration in Bayesian analysis.

The revolution of information technology in 1980's has led to the world of the user-friendly, easy to implement, computer packages. While the advances in this field are being achieved with enormous speed, their influence on the rest of the science is becoming more and more apparent. If the philosophically sound Bayesian framework is to be proclaimed in the world of active statistical thinking, it has certainly to be adjusted in this sociological framework (see Smith (1984),(1987)). The work of this thesis has targeted this area, but the danger remains that if the problems mentioned earlier in this section do not produce a satisfactory answer, the social currents will isolate the potential users of numerical integration techniques to the first amongst the four users mentioned above: The Bayesian research statistician.

REFERENCES

Achcar, J.A. (1984). Use of Bayesian Analysis to Design of Clinical Trials with One Treatment. *Commun. Statist.-Theor. Meth.*, 13(14), p. 1693-1707.

Achcar, J.A. (1987). Transformation of survival data to an extreme value distribution. *The Statistician*, 36, 229-234.

Achcar, J.A., Brookmeyer, R., Hunter, W.G. (1985). An Application of Bayesian Analysis to Medical Follow-up Data. *Statistics in Medicine*, Vol. 4, pp 509-520.

Aitkinson,K.E. (1978). *An Introduction to Numerical Analysis*. New York: Wiley.

Barnett V.D. (1982). *Comparative Statistical Inference*. London: Wiley.

Berntsen J. and Espelid T. (1984). On the use of Gauss-quadrature in adaptive integration schemes. *BIT*, 24, 239-242.

Berntsen J. and Espelid T. (1988). On the construction of higher degree three-dimensional embedded integration rules. *SIAM J. Numer. Anal.*, 25, 222-234.

Box G.E.P. and Tiao G.C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Addison-Wesley.

Cavanos, G.C. and Tsokos, C.P. (1973). Bayesian Estimation of Life Parameters in the Weibull Distribution. *J. Oper. Res.*, 21, pp 755-63.

Chen, W.C., Hill, B.M., Greenhouse, J.B. (1985). Bayesian Analysis of Survival Curves for Cancer Patients Following Treatment. *Bayesian Statistics 2*, pp 299-328. Eslevier Science Publishers B.V. (North Holland).

Clayton D. G. (1989). A Monte Carlo method for Bayesian inference in Frailty models. *Technical report*, University of Leicester, Department of Community Health, Leicester, England.

Cools R. (1989a). The construction of cubature formulae using invariant theory and ideal theory. Ph.D. thesis, department of Computer Science, Katholieke Universiteit Leuven, Belgium.

Cools R. (1989b). Personal communication.

Cools R. and Haegemans A. (1987). Contruction of sequences of embedded cubature formulae for circular symmetris planar regions. In: P Keast and G Fairweather (eds), Numerical integration: Recent developments. *Software and Applications*, pp113-139. NATO ASI SERIES, REIDEL C203.

Cools R. and Haegemans A. (1989). On the construction of multi-dimensional embedded cubature formulae. *Numer. Math.* 55,735-745.

Cox D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman Hall, New York: Halsted Press.

Davis P.J. and Rabinowitz P (1984). *Methods of Numerical Integration* (2nd ed.). Academic Press, Orlando, Florida.

De Boor (1971) CADRE: an algorithm for numerical quadrature. In: J.R.Rice, ED., *Mathematical Software*. Academic Press, NY.

De Boor (1971). On writi ing an automatic integration algorithm. In: J.C.Rice, Ed., *Mathematical Software*. Academic Press, NY.

DeGroot M.H. (1970). *Optimal Statistical Decisions*. McGraw Hill: New York.

Dellaportas P. and Wright D.E. (1988). Numerical Prediction for the Two Parameter Weibull Distribution. *Research Report MSOR-88-02*, Polytechnic South West. Submitted for publication.

Dellaportas P. and Wright D.E. (1989). Positive imbedded integration in Byesian analysis. *Research report MSOR-89-05*, Polytechnic South West. Submitted for publication.

Elhay S. and Kautsky J. (1984). A Method for Computing Quadratures for the Kronrod Patterson Type. *Austral. Comput. Sci. Comm.*, 6, 1, p. 15-1 - 15-9.

Elhay S. and Kautsky J. (1987). Algorithm 655: IQPACK : FORTRAN subroutines for the weights of interpolatory Quadratures. *ACM Transactions on Mathematical Software*, Vol. 13, No 4, pp 319-413.

Engels H. (1986). *Numerical Quadrature and Cubature*. Academic Press, London.

Evans, I.G. and Nigm, A.M. (1980). Bayesian Prediction for Two-Parameter Weibull Lifetime Models. *Commun. Statist.-Theor. Meth.*, A9(6), pp 649-658.

Fahmeir L. and Kaufmann H. (1985). Consistency and Asymptotic normality of the Maximum Likelihood estimator in Generalised Linear Models. *The annals of Statistics*, 13, 1, 342-368.

Foster J.J. (1990). Models and marginal densities for multiway contigency tables. *Ph.D. thesis*, Department of mathematics, University of Nottingham, UK (in preparation).

Gautschi W. (1987). Gauss-Kronrod Quadrature – A survey. In: *Numerical methods and Approximation Theory III* , Ed.: G.V. Milovanovič, Niš, pp 39-66.

Grieve A.P. (1987). Applications of Bayesian Software: two examples. *The Statistician*, 36, pp283-288.

Gelfand A.E. and Smith A.F.M. (1988). Sampling based approaches to calculating marginal densities. *Research report* 08-88, Department of Mathematics, University of Nottingham, Nottingham UK.

Gelfand A.E., Smith A.F.M. and Lee T-m. (1990). Bayesian analysis of constrained parameter and truncated data problems. *Technical report 90-04*, Department of Mathematics, University of Nottingham, UK.

Gelfand A.E., Hills S.E., Racine-Poon A. and Smith A.F.M. (1989). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Rsearch report 01-89*, Department of Mathematics, University of Nottingham, Nottingham, UK.

Geman S. and Geman D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 6,721-741.

Genz A (1986). Fully Symmetric Interpolatory Rules for Multiple Integrals. *SIAM J. Numer. Anal*: 23, 6, pp1273-1283.

Gilks W.R. and Wild P. (1990). Adaptive rejection sampling for Gibbs sampling. *Technical report*, Medical Research Council Biostatistics Unit, Cambridge, UK.

Haberman S. J. (1977). Maximum Likelihood estimates in exponential reponse models. *The annals of Statistics*, 5, 5, 815-841.

Hills S.E. (1989). The parametrisation of statistical models. Unpublished Ph.D. thesis, Department of Mathematics, University of Nottingham, Nottingham, UK.

Johnson R.A. (1970). Asymptotic expansions associated with posterior distributions'. *Annals of Mathematical Statistics*, 41, 851-864.

Johnson N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrica*, 36, 149-176.

Kalbfleisch, Z.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data.* New York: John Wiley.

Kass R.E., Tierney L. and Kadane J.B. (1986). Asymptotics in Bayesian Computation. *Bayesian Statistics* 3, Bernardo J.M., DeGroot M.H., Lindley D.V. and Smith A.F.M (Eds). Oxford Univ. Press.

Kass R.E., Tierney L. and Kadane J.B. (1989). The validity of posterior expansions based on Laplace's method. In: *Essays in Honour of George Barnard*, Eds S.Geisser, J.S.Hodges, S.J.Press and A.Zellner, pp. 473-88. Amsterdam: North Holland.

Kautsky J. and Elhay S. (1982). Calculation of the weights of interpolatory quadratures. *Numer. Math.* 40, pp 407-422.

Keast P and Lyness J.N. (1979). On the Structure of Fully Symmetric Multidimensional Quadrature Rules. *SIAM J. Numer. Anal.* 16, pp11-29.

Kloeck T. and van Dijk H.K. (1978). Bayesian estimates for equation system parameters; an application of integration by Monte Carlo. *Econometrica* 46, 1-19.

Krall J., Uthoff V. and Harley J. (1975). A step up procedure for selecting variables associated with survival. *Biometrics*, 31, p. 49-57.

Kronrod A.S. (1964). Nodes and Weights for Quadrature Formulae. Sixteen-place Tables. *"Nauka"*, Moscow; English Translation:

Consultans Bureau, New York, 1965. MR 32 #597, 598.

Laurie D.P. (1985). Practical error estimation in numerical integration. *Journal of Computational and Applied Mathematics*, 12 and 13, pp 425-431. North Holland.

Lawless, J.F. (1973). On the Estimation of Safe Life when the Underlying Life Distribution is Weibull. *Technometrics*, 15(4), p. 857-65.

Lawless J.F. (1982). *Statistical models and methods for lifetime data*. John Wiley & Sons, Inc., USA.

Lee T.D. (1987). Assesment of inter- and intra-laboratory variances: a Bayesian alternative to BS 5497. *The Statistician*, 36, 161-170.

Lindley D.V. (1980). Approximate Bayesian Methods. In: *Bayesian Statistics*. (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith). University Press: Valencia.

Lyness J.N. (1965). Symmetric Integration Rules for Hypercubes. *Math. Comp.*, 19, pp260-276, 394-407, 625-637.

Lyness J. (1983). When not to use an automatic quadrat routine. *SIAM Rev*. 25 pp 63-85.

Mantel P and Rabinowitz P (1977). The Application of Integer Programming to the Computation of Fully Symmetric Integration Formulas in Two and Three Dimensions. *SIAM J. Numer. Anal.*, 14, pp425-431.

Marriot J. (1987). Bayesian numerical and graphical methods for Box–Jenkins time series. *The Statistician*, 36, 265-268.

McNamee J and Stenger F (1967). Construction of Fully Symmetric Numerical Integration Formulas. *Numer. Math.*, 10, pp327-344.

Monegato G. (1976). A Note on Extended Gaussian Quadrature Rules. *Mathematics of Computation*, 30, 136, pp812-817.

Monegato G. (1978). Some Remarks on the Construction of Extended Gaussian Quadrature Rules. *Mathematics of Computation*, 32, 141, pp247-252.

Monegato G. (1979). An overview of results and questions related to ronrod schemes. *Numerische Integration*, Ed.: G. Hammerlin, ISNM 45, pp 231-240, Birkhauser, Basel.

Naylor J.C. (1982). Some numerical aspects of Bayesian inference. Unpublished PhD thesis. University of Nottingham.

Naylor J.C. (1987). Bayesian Alternatives to t-tests. *The Statistician*, 36, pp241-246.

Naylor J.C. and Shaw J.E.H. (1985). BAYES FOUR–USER GUIDE. Nottingham Statistics group, Department of Mathematics. University of Nottingham.

Naylor J.C. and Smith A.F. (1982). Applications of a Method for the Efficient Computation of Posterior Distributions. *Appl. Statist.*, 31,

3, 214-225.

Naylor J.C. and Smith A.F.M. (1983). A Contamination Model in Clinical Chemistry. *The Statistician*, 32, pp214-225.

Naylor J.C. and Smith A.F.M. (1988a). An Archaeological Inference Problem. *J. Amer. Statist. Assoc.*, 83, No.403, pp 588-595.

Naylor J.C. and Smith A.F.M. (1988b). Econometric Illustrations of Novel Numerical Integration Strategies for Bayesian Inference. *Journal of Econometrics*, 38, 103-126.

O'Hagan A. (1987). Monte Carlo is fundamentally unsound. *The Statistician*, 36, 247-249.

Papadopoulos, A. and Tsokos, C.P. (1976). Bayesian Analysis of the Weibull Failure Model with Unknown Scale and Shape Parameters. *Statistica*, 36, pp 547-60.

Patterson T.N.L. (1968a). The Optimum Addition of Points to Quadrature Formulae. *Math. Comp.*, 22, pp847-856.

Patterson T.N.L. (1968b). On Some Gauss and Lobatto Based Quadrature Formulae. *Math. Comp.*, 22, pp877-881.

Patterson T (1973). Algorithm 468: Algorithm for Numerical Integration over a Finite Interval. *Communications of the ACM*, 16, 11, pp694-699.

Piessens R (1973). An Algorithm for Automatic Integration. *Angewandte Informatik*, 9, pp399-401.

Piessens R., De Doncker-Kapenga E., Uberhuber C. and Kahaner D.K. (1983). QUADPACK: *A subroutine Package for Automatic Integration*. Springer, Berlin.

Piessens R. and Randers M.B. (1974). A Note on the Optimal Addition of Abscissas to Quadrature Formulas of Gauss and Lobatto Type. *Mathematics and Computation*, 28, 125, pp135-139.

Rabinowitz P, Kautsky J, Elhay S and Butcher J.C. (1987). On Sequences of Imbedded Integration Rules. In: P Keast and G Fairweather (eds), Numerical integration: Recent developments. *Software and Applications*, pp113-139. NATO ASI SERIES, REIDEL C203.

Rabinowitz P and Richer N (1969). Perfectly Symmetric Two-dimensional Integration Formulas with Minimal Number of Points. *Math. Comp.*, 23, pp765-780.

Racine-Poon A., Smith A.F.M. and Gelfand A.E. (1990). Bayesian analysis of population models using the Gibbs sampler. *Technical report 90-03*, Department of Mathematics, Univerity of Nottingham, UK.

Reilly P.M. (1976). The Numerical Computation of Posterior Distributions in Bayesian Statistical Inference. *Appl. Statist.*, 25, pp201-209.

Shaw J.E.H. (1986). A class of univariate distributions for use in

Monte Carlo studies. *Research report* 04-86, Nottingham Statistics Group, Dept. of Mathematics, University of Nottingham, UK.

Shaw J.E.H. (1987a). Numerical Bayesian Analysis of some Flexible Regression Models. *The Statistician*, 36, pp147-153.

Shaw J.E.H. (1987b). Aspects of Numerical Integration and Summarisation. *Bayesian Statistics* 3, Bernardo J.M., DeGroot M.H., Lindley D.V. and Smith A.F.M (Eds). Oxford Univ. Press.

Silverman B.W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Skene A.M. (1983). Computing marginal distributions for the dispersion parameters of analysis of variance models. In: *Practical Bayesian Statistics*, Eds. A.P. Dawid and A.F.M.Smith, Logman: Harlow.

Smith A.F.M. (1987). Present position and potential developments: some personal views. Bayesian Statistics. *J.R. Statist. Soc.* A,147,Part 2,245-259.

Smith A.F.M. (1988). What Should be Bayesian about Bayesian Software? *Bayesian Statistics* 3, Bernardo J.M., DeGroot M.H., Lindley D.V. and Smith A.F.M (Eds). Oxford Univ. Press.

Smith A.F.M and Gelfand A.E. (1990). Bayes Theorem from a sampling-resampling perspective. *Research report* 90-01, Department of Mathematics, University of Nottingham, Nottingham, UK.

Smith R.L. and Naylor, J. (1987). A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. *Appl. Statist.* 36, 3, pp 358-369.

Smith A.F.M., Skene A.M., Shaw J.E.H., Naylor J.G., Dransfield M. (1985). The Implementation of the Bayesian Paradigm. *Commun. Statist. Theor. Meth.*, 14(5), pp1079-1102.

Smith A.F.M., Skene A.M., Shaw J.E.H., Naylor J.C. (1987). Progress with Numerical and Graphical Methods for Practical Bayesian Statistics. *The Statistician*, 36, pp75-82.

Smith R.L. and Naylor J.C. (1984). A comparison of maximum likelihood and Bayesian estimators for the three parameter Weibull distribution. *Appl. Statist.* 36, 3, 358-369.

Soland, R.N. (1969). Bayesian Analysis of the Weibull Process with Unknown Scale and Shape Parameters. *IEEE Trans. Rel.*, R-18, pp 181-4.

Steinberg D.I. (1974). *Computational Matrix Algebra*. McGraw-Hill, Inc.

Stephens D.A. and Smith A.F.M. (1990). Sampling-resampling techniques for the computation of posterior densities in normal means problems. *Technical report 90-5*, Department of Mathematics, University of Nottingham, UK.

Stewart L. (1979). Multiparameter univariate Bayesian analysis. *J. Amer. Statis. Assoc.* 74, 684-693.

Stewart L. (1983). Bayesian analysis using Monte Carlo integration - a powerful methodology for handling some difficult problems. *The Statistician* 32, 195-200.

Stewart L. (1987). Hierarchical Bayesian analysis using Monte Carlo integration: computing posterior distributions when there are many possible models. *The Statistician*, 36, 211-219.

Stewart L. and Johnson J.D. (1972). Determining optimum burn-in and replacement times using Bayesian decision theory. *IEEE Transactions on Reliability* R-21, 170-175.

Stroud A.H. and Secrest D (1966). *Gaussian Quadrature Formulas*. USA: Prentice-Hall Inc.

Tierney L. and Kadane J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, 81, 82-86. (Also more detailed as Technical report No. 431, University of Minnessota)

Tierney L., Kass R. and Kadane J.B. (1987). Interactive Bayesian Analysis using accurate asymptotic aproximations. In: *Proc. of the 19th Symp. on the Interface*, Ed. R.M. Heiberger, pp.15-21. Alexandria: ASA.

Titterington D.M., Smith A.F.M. and Makov V.E. (1985). *Statistical Analysis of Finite mixture Distributions*. John Wiley and Sons Ltd., GB.

Uspensky, J.V. (1928). On the convergence of quadrature formulas related to the infinite interval. *Trans. Amer. Math. Soc.* 30, p.542-559.

van Dijk H.K., Hop J.P. and Louter A.S. (1987). An algorithm for the computation of posterior moments and densities using simple importance sampling. *The Statistician*, 36, 83-90.

van Dijk H.K. and Kloeck T. (1980). Further experience in Bayesian analysis using Monte Carlo integration. *J. Econometrics* 14, 307-328.

van Dijk H.K. and Kloeck T. (1984). Experiments with some alternatives for simple importance sampling in Monte Carlo integration. In: *Bayesian Statistics* II, J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds.). North Holland, Amsterdam.

Wedderburn R.W.M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalised linear models. *Biometrica*, 63, 1, 27-32.

Zeger S. L. and Karim M. R. (1989). Generalised linear models with random effects; A Gibbs sampling approach. *Technical report P691*, Department of Biostatistics, The Johns Hopkins University.

Additional Reference

Heyde C.C. and Johnstone I.M. (1979): On Asymptotic posterior normality for Stochastic processes. J. R. Statist. Soc. B, 41, pp 184-189.