Integration of Hierarchical Models

by Valen E. Johnson*

Duke University

Institute of Statistics and Decision Sciences

DP #89-07

Abstract

A technique called quantile integration is proposed for the integration of hierarchical models. If \( n \) denotes the number of points used to evaluate the integral, the error of approximation is \( O(\log(n)/n) \) when the moment generating function exists for each conditional density appearing in the model, and is \( O(n^{-1+1/k}) \) when each conditional density possesses \( k \) moments. Approximations produced by the technique take the form of probability density functions, and the method requires no modification for densities defined on bounded or semi-bounded intervals. In many models, computational requirements increase only linearly with the dimension of the integral, and little analytical effort is needed for implementation.

* Valen E. Johnson is Assistant Professor, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27706. The author thanks John Geweke and Mike West for several helpful comments and Mike West for suggesting the model used in Example 3.
1. INTRODUCTION

A frequently encountered problem arising in the implementation of the Bayesian paradigm is the requirement to evaluate integrals having a form similar to

\[ p(\theta_1) = \int \cdots \int p(\theta_1|\theta_2)p(\theta_2|\theta_3) \cdots p(\theta_m)d\theta_m \cdots d\theta_2. \]  

(1.1)

In this expression, \( p(\theta_{i+1}|\theta_i) \) and \( p(\theta_m) \) denote probability density functions associated with random variables \( \Theta_1, \Theta_2, \ldots, \Theta_m \). Examples of models leading to such integrals include dynamic linear models, empirical Bayes models, and hierarchical linear models.

A number of generally applicable integration techniques can be employed to approximate the marginal density appearing in (1.1). Tierney and Kadane (1986) proposed a variation of Laplace's method that often yields accurate approximations to both the posterior moments and marginal distributions resulting from such integrals, but their technique requires that the densities appearing in (1.1) be unimodal and decrease exponentially away from the conditional mode. Gaussian quadrature (Dagenais and Liem 1981; Naylor and Smith 1982) can also be employed to evaluate such integrals, but its use is predicated on the assumption that transformations of the parameters to a scale upon which the densities are approximately normal can be found. The accuracy of both methods is limited by the form of the integrand and the extent to which it satisfies the stated assumptions. Alternatively, Monte Carlo techniques (Kloek and van Dijk 1978) can be utilized, but are often too computationally expensive to be of practical use. The subject of this paper is the exposition of a numerical integration technique designed explicitly to exploit the special structure exhibited in (1.1).

This integration method, which we refer to as quantile integration, has several important properties not shared by the schemes mentioned above. Perhaps the most significant of these properties is that the approximations generated by the technique have the form of probability density functions. Asymptotically, if \( n \) denotes the number of quantile points used to approximate each integral, then subject to mild regularity conditions, the error of the approximation is \( O(\log(n)/n) \). No complications arise for parameters defined on bounded or semi-bounded intervals, and the moments of \( \Theta_1 \) can be estimated easily from the approximation obtained. In addition, suppose that the integrand in (1.1) incorporates a likelihood function for \( \theta_1 \). Then besides providing an estimate of the posterior density of \( \theta_1 \), estimates of the posterior distributions of other model parameters are also obtained as by-products of the algorithm, although the availability of individual posteriors depends on the exact form of the integral and the likelihood function. Finally, the computational effort required to evaluate (1.1) increases only linearly with \( m \), and little analytical effort is needed to implement the algorithm.
The major limitation of quantile integration is that its efficiency decreases rapidly as the effort needed to evaluate the conditional distribution functions of densities appearing in (1.1) increases. However in many statistical models, the properties of these densities are well known and numerical approximations to their distribution functions are readily available.

To illustrate the form of the approximation, consider the simplest case of (1.1) given by

$$ p(\theta_1) = \int p(\theta_1 | \theta_2) p(\theta_2) \, d\theta_2 $$

(1.2)

(Here and throughout the remainder of the paper, we assume that for each $1 \leq j \leq m - 1$, $p(\theta_j | \theta_{j+1})$ is a bounded, continuous function that has two bounded, continuous derivatives with respect to $\theta_{j+1}$. Also, the ranges of integration are over the range of the random variable indicated by the differentials, unless otherwise specified).

Equation (1.2) can be re-expressed as

$$ \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} p(\theta_1 | \theta_2) p(\theta_2) \, d\theta_2, $$

(1.3)

where $x_i$ represents the $i$th quantile of $p(\theta_2)$. Letting $w_i$ represent an arbitrary point in $[x_{i-1}, x_i]$ and expanding $p(\theta_1 | \theta_2)$ by its Taylor series expansion around $w_i$ within each interval leads to either a one term expansion given by

$$ p(\theta_1) = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \left[ p(\theta_1 | w_i) + \frac{\partial p(\theta_1 | \xi_i(\theta_2))}{\partial \theta_2} (\theta_2 - w_i) \right] p(\theta_2) \, d\theta_2 $$

(1.4)

or a two term expansion

$$ p(\theta_1) = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \left[ p(\theta_1 | w_i) + \frac{\partial p(\theta_1 | \xi_i(\theta_2))}{\partial \theta_2} (\theta_2 - w_i) + \frac{\partial^2 p(\theta_1 | \xi_i(\theta_2))}{\partial \theta_2^2} (\theta_2 - w_i)^2 \right] p(\theta_2) \, d\theta_2. $$

(1.5)

Since $\int_{x_{i-1}}^{x_i} p(\theta_2) \, d\theta_2 = 1/n$, $p(\theta_1)$ can be approximated by

$$ p(\theta) \approx \frac{1}{n} \sum_{i=1}^{n} p(\theta_1 | w_i), $$

(1.6)

with error expressible as either

$$ e_1(\theta_1) = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \frac{\partial p(\theta_1 | \xi_i(\theta_2))}{\partial \theta_2} (\theta_2 - w_i) p(\theta_2) \, d\theta_2 $$

(1.7)

or

$$ e_2(\theta_1) = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \left[ \frac{\partial p(\theta_1 | \xi_i(\theta_2))}{\partial \theta_2} (\theta_2 - w_i) + \frac{\partial^2 p(\theta_1 | \xi_i(\theta_2))}{\partial \theta_2^2} (\theta_2 - w_i)^2 \right] p(\theta_2) \, d\theta_2. $$

(1.8)
Note that the vector \( w = \{w_i\} \) is constant for all values of \( \theta_1 \).

Several heuristic interpretations of the form of the approximation are evident from (1.6). The approximation might be considered as a generalization of Simpson rule type methods, except that the product of the approximation is a continuous function rather than a scalar. Alternatively, it might be viewed as a non-random version of Monte Carlo integration in which points are chosen according to known properties of the mixing distribution, although the points drawn are used to define components of a mixture of densities.

Implications of the form of the error term on the selection of \( w \) and the asymptotic behavior of (1.6) are discussed in the Section 2. The approximation is extended to the more general case described by (1.1) in Section 3. In Section 4, we apply the technique to obtain marginal posterior distributions in two settings that typify models in which quantile integration is likely to be useful. In the first, we derive explicit expressions for the predictive and posterior densities arising in a broad class of univariate, non-Gaussian time series models, with a specific application to the robust Kalman filter model proposed by Meinhold and Singpurwalla (1989). The second class of models are the hierarchical linear models proposed by Lindley and Smith (1972). The specific case presented involves a model for multiple regression coefficients when dispersion parameters are unknown and the assumption of prior exchangeability is tenable. We conclude with a discussion of results in Section 5.

2. ONE PARAMETER INTEGRATION

The following theorems describe the asymptotic behavior of (1.6) for the one parameter integration problem described by (1.2). The first theorem applies to densities defined on arbitrary domains, while the second applies to cases in which \( p(\theta_2) \) is bounded away from zero.

Before stating the first theorem, several comments concerning the selection of \( w \) are appropriate. A number of criteria for choosing \( w \) are immediately apparent upon examination of (1.4) and (1.5), but two choices are particularly appealing. Because the quantile function for the marginal distribution of \( \Theta_2 \) is required to specify the approximation, a simple choice for \( w \) is to assign the value of the \(((2i - 1)/2n)^{th}\) quantile to \( w_i \). We denote this value of \( w_i \) as the \( i^{th} \) mid-quantile. A second but more complicated choice of \( w \) is to select \( w_i \) so that

\[
\int_{z_i-1}^{z_i} (\theta_2 - w_i)p(\theta_2) d\theta_2 = 0. \tag{2.1}
\]

This choice of \( w \) eliminates the first order error term appearing in (1.8) and is often easy to compute if the quantiles are known and \( \Theta_2 \) belongs to an exponential family. We refer to this value of \( w \) as the quantile interval mean.
Unfortunately, not all choices of $w$ lead to acceptable expansions. When the range of $\Theta_2$ is unbounded, restrictions on $w_1$ and $w_n$ are needed in order to prevent (1.7) and (1.8) from becoming arbitrarily large. For this reason, we assume that one of the following conditions is satisfied.

Condition 1. If the highest existing moment of $\Theta_2$ is $k$, then $|\omega_i - E\Theta_2| < qn^{1/k}$ for fixed $q > 0$ and $1 \leq i \leq n$.

Condition 2. Suppose that the moment generating function of $\Theta_2$ exists in some interval, say $(-s, s)$, containing the origin. Then $w$ satisfies $|\omega_i - E\Theta_2| < q \log(n)$ for fixed $q > 0$ and $1 \leq i \leq n$.

As demonstrated in the proof of Theorem 1, these conditions are easily satisfied in practice and are made strictly for technical reasons. In particular, both the choice of the mid-quantiles and quantile interval means satisfy these conditions.

**THEOREM 1.** Assume that $p(\theta_1 | \theta_2)$ is continuous and continuously differentiable with respect to $\theta_2$ and that there exists an $M$ such that $|\partial p(\theta_1 | \theta_2)/\partial \theta_2| < M$ for all $\theta_1, \theta_2$. Also, assume that the random variable $\Theta_2$ having density $p(\theta_2)$ possesses $k > 2$ moments. Then the following statements are true.

a) If Condition 1 is satisfied, then the error of the approximation given by (1.6) to $p(\theta_1)$ is $O(n^{-1+1/k})$.

b) If Condition 2 is satisfied, then the error of the approximation in (1.6) to $p(\theta_1)$ is $O(\log(n)/n)$.

**PROOF:** a) Without loss of generality, assume that $E\Theta_2 = 0$ and let $\sigma_k = E[|\Theta_2|^k]$. From Markov's inequality, $P[\Theta_2 > (n\sigma_k)^{1/k}] \leq 1/n$. Also,

$$\sigma_k = \int |\theta_2|^k p(\theta_2) d\theta_2 \geq n^{-1/k} \int_{|\theta_2| > n^{1/k}} |\theta_2|^k p(\theta_2) d\theta_2. \quad (2.2)$$

From (1.7) with $I_i = M|\theta_2 - w_i|p(\theta_2) d\theta_2$,

$$|e_1(\theta_1)| = \left| \sum_{i=1}^{n} \int_{z_{i-1}}^{z_i} \frac{\partial p(\theta_1 | \xi_i(\theta_2))}{\partial \theta_2} (\theta_2 - w_i) p(\theta_2) d\theta_2 \right| \quad (2.3)$$

$$\leq \left| \sum_{i=2}^{n-1} \int_{z_{i-1}}^{z_i} I_i + \int_{z_{n-1}}^{z_n} I_1 + \int_{z_0}^{(n\sigma_k)^{1/k}} I_n + \int_{(n\sigma_k)^{1/k}}^{(n\sigma_k)^{1/k}} I_1 \right| \quad (2.4)$$

(equate the last two integrals with 0 if the lower bound exceeds the upper bound)

$$\leq M \left| \frac{1}{n} \sum_{i=2}^{n-1} (z_i - z_{i-1}) + \frac{1}{n} (z_1 + (n\sigma_k)^{1/k}) + \frac{1}{n} ((n\sigma_k)^{1/k} - z_{n-1}) + \frac{2\sigma_k}{n^{1-1/k}} + \frac{2q}{n^{1-1/k}} \right| \quad (2.5)$$

$$= M \left| \frac{1}{n} (2(n\sigma_k)^{1/k}) + \frac{2\sigma_k}{n^{1-1/k}} + \frac{2q}{n^{1-1/k}} \right| \quad (2.6)$$

$$= O(n^{-1+1/k}).$$

4
b) If \( E[\exp(s\Theta_2)] \) exists for \( s > 0 \), then denoting \( E[\exp(s\Theta_2)] = \sigma_\infty \) gives

\[
\int \exp(s\Theta_2)p(\theta_2) \, d\theta_2 \geq \int_{\theta_2 > \log(n\sigma_\infty)/s} \exp(s\Theta_2)p(\theta_2) \, d\theta_2 \geq n\sigma_\infty \int_{\theta_2 > \log(n)/s} p(\theta_2) \, d\theta_2. \tag{2.7}
\]

Thus, \( P[\Theta_2 > \log(n\sigma_\infty)/s] \leq 1/n \), and a similar argument applied to \(-s\) gives \( P[\Theta_2 < \log(n\sigma_\infty)/s] \leq 1/n \). Also,

\[
\sigma_\infty \geq \frac{1}{\log(n)} \int_{\theta_2 > \log(n)/s} \exp(s\theta_2) \, d\theta_2 \geq \frac{s}{\log(n)} \int_{\theta_2 > \log(n)/s} |\theta_2|p(\theta_2) \, d\theta_2. \tag{2.8}
\]

Use of these inequalities in a fashion similar to that demonstrated in part (a) leads to \( |e_1(\theta_2)| = O(\log(n)/n) \).

Next, suppose that (1.2) is modified to reflect the contribution from a likelihood function, say \( l(y|\theta_1) \).

Then the posterior density of \( \theta_1 \) is given by

\[
p(\theta_1|y) = \frac{1}{K_0} \int l(y|\theta_1)p(\theta_1|\theta_2)p(\theta_2) \, d\theta_2, \tag{2.9}
\]

where

\[
K_0 = \int \int l(y|\theta_1)p(\theta_1|\theta_2)p(\theta_2) \, d\theta_2 \, d\theta_1. \tag{2.10}
\]

Hence, to apply quantile integration in this setting it is necessary to approximate not only the integral appearing in (2.9), but to also estimate the normalizing constant \( K_0 \).

If \( l(y|\theta_1) \) is bounded, then from Theorem 1

\[
\int l(y|\theta_1)p(\theta_1|\theta_2)p(\theta_2) \, d\theta_2 \tag{2.11}
\]

can be approximated by

\[
\frac{1}{n} \sum_{i=1}^{n} l(y|\theta_1)p(\theta_1|w_i), \tag{2.12}
\]

with an error having magnitude \( O(n^{-1+1/k}) \) or \( O(\log(n)/n) \), depending on the moment properties of \( \Theta_2 \).

For convenience, denote the appropriate order of magnitude by \( O(n^{-\alpha}) \).

A natural estimate for \( K_0 \) is given by \( K = \sum K_i \), where

\[
K_i = \frac{1}{n} \int l(y|\theta_1)p(\theta_1|w_i) \, d\theta_1. \tag{2.13}
\]

The error of this approximation is given by

\[
K_0 - K = \int l(y|\theta_1) \left[ \int p(\theta_1|\theta_2)p(\theta_2) \, d\theta_2 - \frac{1}{n} \sum_{i=1}^{n} p(\theta_1|w_i) \right] \, d\theta_1 \tag{2.14}
\]

\[
= O(n^{-\alpha}) \int l(y|\theta_1) \, d\theta_1. \tag{2.15}
\]
Hence, if \( l(y|\theta_1) \) is integrable with respect to \( \Theta_1 \), then

\[
\frac{1}{nK} \sum l(y|\theta_1)p(\theta_1|w_i)
\]

converges to \( p(\theta_1|y) \) at rate \( O(n^{-\alpha}) \).

An estimate of the posterior distribution \( p(\theta_2|y) \) can also be obtained from the calculations required to approximate \( p(\theta_1|y) \). Specifically, the posterior probability that \( \Theta_2 \) lies in the interval \((x_{i-1}, x_i)\) is given by

\[
\int_{x_{i-1}}^{x_i} p(\theta_2|y) d\theta_2 = \frac{1}{K_0} \int_{x_{i-1}}^{x_i} \int l(y|\theta_1)p(\theta_1|\theta_2)p(\theta_2) d\theta_1 d\theta_2
\]

(2.17)

\[
= \frac{1}{K_0} \int l(y|\theta_1) \int_{x_{i-1}}^{x_i} p(\theta_1|\theta_2)p(\theta_2) d\theta_2 d\theta_1
\]

(2.18)

\[
= \frac{1}{K_0} \int l(y|\theta_1) \left[ \frac{1}{n} p(\theta_1|w_i) + O(n^{-\alpha}) \right] d\theta_1
\]

(2.19)

\[
= \frac{K_i}{K} + O(n^{-\alpha}).
\]

(2.20)

Thus, estimates of the posterior probabilities that \( \Theta_2 \) lies within each quantile interval are obtained as by-products of the estimation algorithm. These results are summarized in the following corollary.

**Corollary 1.** Suppose that \( p(\theta_1|\theta_2) \) satisfies the conditions stated in Theorem 1 and that \( l(y|\theta_1) \) is integrable with respect to \( \theta_1 \). If \( w \) satisfies Condition 1, then

\[
p(\theta_1|y) = \frac{1}{nK} \sum_{i=1}^{n} l(y|\theta_1)p(\theta_1|w_i) = O(n^{-1+1/k}),
\]

(2.21)

and the posterior probability that \( \Theta_2 \) lies in the interval \((x_{i-1}, x_i)\) can be expressed

\[
\int_{x_{i-1}}^{x_i} p(\theta_2|y) d\theta_2 = \frac{K_i}{K} + O(n^{-1+1/k}).
\]

(2.22)

In addition, if Condition 2 is satisfied, then (2.21) and (2.22) hold if \( O(n^{-1+1/k}) \) is replaced by \( O(\log(n)/n) \).

The next theorem illustrates the efficiency of quantile integration for the special case in which \( p(\theta_2) \) is bounded away from zero, and is of interest primarily for its implications concerning the behavior of quantile integration for small and moderate \( n \).

**Theorem 2.** Assume that \( p(\theta_1|\theta_2) \) is continuous and twice differentiable with respect to \( \theta_2 \), and that there exists \( M \) such that \( \left| \partial^2 p(\theta_1|\theta_2)/\partial \theta_2^2 \right| < M \) for all \( \theta_1, \theta_2 \). Also, assume that there exists \( c > 0 \) such that \( p(\theta_2) > c \) for all \( \theta_2 \in \Theta_2 \) and that \( w \) is chosen according to (2.1). Then the rate of convergence of (1.6) to \( p(\theta_1) \) is \( O(n^{-2}) \).
PROOF: From (1.8),
\[ |e_2(\theta_1)| = \left| \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \frac{\partial^2 p(\theta_1 | \xi(x_2))}{\partial \theta_2^2} (\theta_2 - w_i)^2 p(\theta_2) \, d\theta_2 \right| \leq \sum_{i=1}^{n} M \left( \frac{1}{cn} \right)^2 \frac{1}{n} = O(n^{-2}) \cdot (2.23) \]

Although this theorem seldom applies in practice, it illustrates the performance of quantile integration for moderate \( n \). Typically, probability densities are bounded away from zero near the center of the distribution, so the error arising from the central terms in (1.6) will be almost \( O(n^{-2}) \), provided that \( w \) is near the solution to (2.1). Also, in the tails of the distribution, both \( \partial p(\theta_1 | \theta_2) / \partial \theta_2 \) and \( \int (\theta_2 - w_i)p(\theta_2) \, d\theta_2 \) are typically small when compared to \( (1/n)^2 \) for moderate values of \( n \). Hence, quantile integration can often be expected to converge at an apparent rate of \( O(n^{-2}) \) in many applications.

3. \( m \) PARAMETER INTEGRATION

Conceptually, there is little difference between applying quantile integration to one parameter integrals and to \( m \) parameter integrals. Implementation in the \( m \) parameter setting can be accomplished by iteratively applying the algorithm described in Section 2 to the individual integrals appearing in (1.1). After \( j \) iterations, this procedure leads to an approximation of the marginal distribution \( p(\theta_{m-j}) \) given by
\[ \tilde{p}(\theta_{m-j}) = \frac{1}{n} \sum_{i=1}^{n} p(\theta_{m-j} | w_{m-j-1}^{i}). \quad (3.1) \]

In this expression, \( w_{m-j+1}^{i} \in (x_{i-j-1}^{m-1}, x_{i-j+1}^{m}) \) where \( x_{i-j+1}^{m} \) is the \( i/th \) quantile from the estimated marginal density obtained in the \((j-1)th\) iteration. Hence, the marginal (posterior) density \( p(\theta_1) \) can be approximated by
\[ \tilde{p}(\theta_1) = \frac{1}{n} \sum_{i=1}^{n} p(\theta_1 | w_1^{i}). \quad (3.2) \]

Computationally, the primary difference between implementing the algorithm in the one parameter setting and the \( m \) parameter setting is that quantiles must be obtained numerically after the first iteration in the \( m \) parameter problem. However, if the distribution functions appearing in the mixture are easily approximated, this increase in computational burden is often not problematic.

The convergence rates cited in Theorem 1 for (1.2) can be extended to \( m \) parameters if the constraint given below is imposed on the conditional densities appearing in (1.1).

Condition 3. Suppose that
\[ \Psi(\theta_1, \theta_{m-j}) = \left| \frac{\partial}{\partial \theta_{m-j}} \left[ \int \cdots \int p(\theta_1 | \theta_2) \cdots \int p(\theta_{m-j-1} | \theta_{m-j}) \, d\theta_{m-j-1} \cdots d\theta_2 \right] \right|, \quad (3.3) \]

Then there exists an \( N \) such \( \Psi(\theta_1, \theta_{m-j}) < N \) for all \( \theta_1, \theta_{m-j}, j = 0, \ldots, m-2. \)
Although the behavior of $\Psi$ is unintuitive, Condition 3 is satisfied if both \( \partial p(\theta_j | \theta_{j+1}) / \partial \theta_{j+1} \), \( j = 1, \ldots, m - 1 \), and
\[
\frac{\partial}{\partial \theta_{m-k}} \mathbb{E}[\Theta_{m-k-1} | \theta_{m-k}]
\]
are bounded. Alternatively, if
\[
p(\theta_j | \theta_{j+1}) = p(\theta_j | \theta_{j+1}, \theta_{j+2}, \ldots, \theta_m)
\]
for all \( j \), then Condition 3 is satisfied if \( \partial p(\theta_k | \theta_{m-k}) / \partial \theta_{m-k} \) is bounded. Also, $\Psi$ is zero if $p(\theta_j | \theta_{j+1}) = p(\theta_j)$ for any conditional density appearing in the integrand, and the integral of $\Psi(\theta - 1, \theta_j)$ over the range of $\Theta_1$ is zero for each $j$.

**Theorem 3.** Assume that $p(\theta_j | \theta_{j+1})$ is continuous and differentiable with respect to $\theta_{j+1}$ for all $j$ and that there exists an $M$ such that $|\partial p(\theta_j | \theta_{j+1}) / \partial \theta_{j+1}| < M$ for all $\theta_j, \theta_{j+1}$. In addition, assume that the random variables $\Theta_{j+1}$ with densities $p(\theta_{j+1})$ possess $k > 2$ moments. Then if $w_j$ satisfies Condition 1 for $j = 2, \ldots, m$ and Condition 3 is satisfied, then the error of the approximation to $p(\theta_1)$ given by (3.2) is $O(n^{-1+1/k})$.

Similarly, if $w_j$ satisfies Condition 2 for $j = 2, \ldots, m$ and Condition 3 is satisfied, then (3.2) converges to $p(\theta_1)$ at rate $O(\log(n)/n)$.

**Proof:** The error of the approximation can be expressed
\[
p(\theta_1) - \tilde{p}(\theta_1) = \sum_{i=1}^{n} \int_{x_{i-1}^i}^{x_i} \frac{\partial p(\theta_1 | \xi_i^k)}{\partial \theta_2}(\theta_2 - w_i^k) \tilde{p}(\theta_2) d\theta_2
\]
\[+ \sum_{k=3}^{m} \sum_{i=1}^{n} \int_{x_{i-1}^i}^{x_i} (\theta_k - w_i^k) \tilde{p}(\theta_k)
\]
\[\times \int \cdots \int p(\theta_1 | \theta_2) \cdots p(\theta_{k-2} | \theta_{k-1}) \left[ \frac{\partial p(\theta_{k-1} | \xi_i^k)}{\partial \theta_k} \right] d\theta_{k-1} \cdots d\theta_2 d\theta_k,
\]
where $\tilde{p}(\theta_j) = \sum p(\theta_j | w_{i+1}^j) / n$. In (3.6), the functions $\xi_i^k$ represent the Cauchy form of the remainder term in the Taylor series expansion of $p(\theta_{k-1} | \theta_k)$ about $\theta_k = w_i^k$ and are functions of $\theta_k$.

Since $m$ is fixed, the result follows by applying the argument in the proof of Theorem 1 to all terms of the form
\[
\sum_{i=1}^{n} \int_{x_{i-1}^i}^{x_i} K(\theta_j - w_i^j) \tilde{p}(\theta_j) d\theta_j,
\]
where $K = \max(M, N)$.

The results obtained for estimating the marginal posterior distributions of $\theta_1$ and $\theta_2$ in one parameter integration also apply in the $m$ parameter setting. If, as in Corollary 1, the likelihood function takes the form $l(y|\theta_1)$, then the posterior density of $\theta_1$ and the posterior distribution of $\theta_2$ can be estimated using (2.21).
and (2.22). When (1.1) or the likelihood function takes a more general form, the posterior distributions of other model parameters may also be obtained, but the subset of available posteriors depends on the final form of the approximation. An illustration of this phenomenon is provided in Example 3 of Section 4.

4. APPLICATIONS

EXAMPLE 1. Here we demonstrate the performance of quantile integration in a familiar setting; namely, determining a marginal density from a bivariate normal distribution. Let \( \Theta_1, \Theta_2 \) be \( N(0,1) \) random variables with \( \text{Cov}(\Theta_1, \Theta_2) = 0.5 \), and suppose for purposes of illustration that we are interested in determining the marginal distribution of \( \Theta_1 \) given the marginal density of \( \Theta_2 \) and the conditional density of \( \Theta_1 \) given \( \Theta_2 \). The relevant marginal and conditional densities are then

\[
p(\theta_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \theta_2^2 \right) \quad \text{and} \quad p(\theta_1 | \theta_2) = \frac{1}{\sqrt{1.5\pi}} \exp\left\{ -\frac{1}{2} (\theta_1 - 0.5\theta_2)^2 \right\}.
\]

(4.1)

Initially, we select \( w \) as the vector of mid-quantiles and take \( n = 3 \). Numerical routines to evaluate the normal quantile function are commonly available and lead to \( w = (\Phi^{-1}(1/6), \Phi^{-1}(3/6), \Phi^{-1}(5/6)) = (-0.967, 0, 0.967) \). Hence, for this choice of \( n \) and \( w \), the approximation to the marginal distribution of \( \theta_1 \) is

\[
\tilde{p}(\theta_1) \doteq \frac{1}{3} \left[ n(\theta_1; -0.484, 0.75) + n(\theta_1; 0, 0.75) + n(\theta_1; 0.484, 0.75) \right],
\]

(4.2)

where \( n(\theta_1; \mu, \sigma^2) \) denotes a normal density with mean \( \mu \) and variance \( \sigma^2 \).

From this expression, both the moments and distribution function of \( \Theta_1 \) can be estimated. For example, estimates of the first three moments are \( E\Theta_1 \doteq 0 \), \( E\Theta^2_1 \doteq 0.906 \), and \( E\Theta^3_1 \doteq 0 \). Estimated densities using the mid-quantile points for \( n = 3, 5, 10 \) are depicted in Fig. 1a.

Next, suppose quantile interval means are used. For Gaussian models, as with most exponential family models, quantile interval means are available analytically if the quantiles are known, and in the case \( n = 3 \) are given by \(( -1.091, 0, 1.091 ) \). The approximation is then similar in form to (4.2) with \( \pm 0.546 \) replacing \( \pm 0.484 \). The corresponding estimates of the first three moments are \( 0.0, 0.948, \) and \( 0.0 \). The density estimates for \( n = 3, 5, 10 \) using quantile interval means are displayed in Fig. 1b. Figure 2 provides the relative errors of the estimates displayed in Figures 1a and 1b, e.g. \( (p - \tilde{p}) / p \).

figures 1 and 2 here

Finally, under the same model suppose that a random variable having a Gaussian distribution with conditional mean \( \theta_1 \) and variance \( 1/25 \) is observed to take the value 1.0. Then (2.21) and (2.22) can be used
to estimate the marginal posterior distributions of $\theta_1$ and $\theta_2$. Taking $n = 50$ and $w$ the quantile interval mean produces the estimated marginal posterior distributions depicted in Figure 3. For comparison, the exact densities are also shown. Note that the true and estimated posterior densities of $\theta_1$ are not distinguishable.

EXAMPLE 2. The previous example provides a benchmark for comparing quantile integration to other standard techniques. In this example, we demonstrate how quantile integration can be used to efficiently evaluate predictive and posterior densities arising in univariate, non-Gaussian dynamic linear models.

Suppose a sequence of observations $Y_1, Y_2, \ldots, Y_t$ are generated according to an observation equation

$$Y_t = \Theta_t + \nu_t,$$

(4.3)

where the underlying system parameters evolve according to the system equation

$$\Theta_t = \Theta_{t-1} + \nu_t.$$

(4.4)

The errors $\nu_t$ and $\nu_t$ are assumed to be serially and pairwise independent with densities $lik(\cdot)$ and $pr_1(\cdot)$, and for convenience we assume that the innovations $\nu_t$ are identically distributed and belong to a location-scale family. Also, take $\Theta_0 = 0$ and let $D_t$ denote the history of the series to time $t$, i.e. $y_1, y_2, \ldots, y_t$.

After the first observation, the posterior density for $\Theta_1$ can be written

$$p(\theta_1|D_1) = \frac{1}{K_1} lik(y_1 - \theta_1) pr_1(\theta_1),$$

(4.5)

where

$$K_1 = \int lik(y_1 - \theta_1) pr_1(\theta_1) d\theta_1.$$

(4.6)

The normalizing constant $K_1$ can be approximated using a degenerative version of quantile integration that amounts to little more than the rectangular rule applied on a transformed scale. Letting $w$ represent a vector of $n$ quantile points from the density $lik(\cdot)$, the normalizing constant $K_1$ can be estimated as

$$K_1 \approx \frac{1}{n} \sum_{i=1}^{n} pr_1(y_1 - w_i).$$

(4.7)

Next, consider estimation of the predictive density for $\Theta_2$ given $D_1$. Noting that $\Theta_2|\Theta_1 \sim pr_2(\theta_2 - \theta_1)$, the predictive density for $\Theta_2$ given $D_1$ is thus proportional to

$$\int pr_2(\theta_2 - \theta_1) lik(y - \theta_1) pr_1(\theta_1) d\theta_1.$$

(4.8)

Again expanding around the likelihood function and using the vector $w$ previously determined, the estimated predictive density for $\Theta_2$ is proportional to

$$p(\theta_2|D_1) \propto \frac{1}{n} \sum_{i} pr_2(\theta_2 - y_i + w_i) pr_1(y_1 - w_i).$$

(4.9)
The normalization constant is simply \( \sum pr_1(y_1 - w_i)/n \). Denote \( pr_1(y_1 - w_i)/n \) by \( c_1^i \)

Similarly, the posterior density for \( \Theta_2 \) given \( D_2 \) can be written

\[
p(\theta_2|D_2) = \frac{1}{K_2} \int lik(y_2 - \theta_2) pr_2(\theta_2 - \theta_1) lik(y_1 - \theta_1) pr_1(\theta_1) d\theta_1
\]

\[
\overset{\text{def}}{=} \frac{1}{K_2} \sum_{i=1}^{n} lik(y_2 - \theta_2) pr(\theta_2 - y_i + w_i) c_1^i
\]

(4.10)

where

\[
K_2 \overset{\text{def}}{=} \int \sum_{i=1}^{n} lik(y_2 - \theta_2) pr(\theta_2 - y_i + w_i) c_1^i \ d\theta_2
\]

\[
\overset{\text{def}}{=} \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} pr_2(y_j - y_i) c_1^i
\]

\[
\equiv \sum_{j} c_2^j.
\]

(4.11)

The generalization of this iteration to obtain estimates of future predictive and posterior densities is apparent. Defining

\[
c_j^{m} = \frac{1}{n} \sum_{i=1}^{n} pr_m(y_m - w_j - y_{m-1} + w_i) c_{i}^{m-1},
\]

(4.12)

the estimated predictive density for \( \Theta_{m+1} \) given \( D_m \) is given by

\[
\frac{\sum_j pr_{m+1}(\theta_{m+1} - y_m + w_j) c_j^m}{\sum_j c_j^m},
\]

(4.13)

and the estimated posterior density given \( D_{m+1} \) is

\[
\frac{1}{K_{m+1}} lik(y_{m+1} - \theta_{m+1}) \sum_j pr_{m+1}(\theta_{m+1} - y_m + w_j) c_j^m
\]

\[
K_{m+1} = \sum_j c_j^{(m+1)}.
\]

(4.14)

A convenient property of this updating scheme for location family models is that the vector of quantile points, \( w \), need be computed only once. Additionally, because computation of predictive and posterior densities for the \( m^{th} \) system parameter depends only on the weight vectors \( c_j^m \) and \( c_j^{m-1} \), future observations can be processed by retaining only the two previous weights. The method can also be extended to include more realistic situations in which the scale parameters of the error distribution are not known a priori. The details of this extension will be discussed in a future article.

As a numerical illustration, we consider a robustification of the Kalman filter model proposed by Meinhold and Singpurwalla (1989) in which the innovations \( v_r \) and \( u_s \) are assumed to have Student-\( t \) distributions.
Advantages of such models over more standard Gaussian models lie in the treatment of extreme observations. If the Student-\(t\) distribution of the system innovations are modeled with more degrees of freedom than the likelihood, then extreme outliers are essentially ignored while moderate outliers result in bimodal posteriors. Also, the precision of the posterior can decrease if the new observation indicates a drastic change in the level of the system equation.

Meinhold and Singpurwalla applied their model to the data depicted in Table 1. The observation innovations were assumed to follow a Student-\(t\) density on 2 degrees of freedom with scale parameter 2, and the system innovations were assigned a Student-\(t\) distribution on 3 degrees of freedom with scale 1. The estimated posterior densities for \(t = 1, 3, \ldots, 9\) are depicted in Fig. 4. For comparison, the estimated posteriors obtained using the first order approximation proposed by Meinhold and Singpurwalls are also shown in dashed lines. The estimates depicted are based on \(n = 200\) quantile points and are not discernible from those based on \(n = 1000\). Determining the nine weight vectors \(c^m\) required six seconds on a DECstation 3100.

Table 1. Simulated Data

<table>
<thead>
<tr>
<th>(t)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_t)</td>
<td>1.638</td>
<td>-224</td>
<td>9.540</td>
<td>-2.079</td>
<td>0.181</td>
<td>1.902</td>
<td>6.721</td>
<td>8.134</td>
<td>5.882</td>
</tr>
</tbody>
</table>

EXEMPLARY 3. As a final example we illustrate the use of quantile integration in estimating marginal posterior densities arising in hierarchical linear models (HLM) with unknown dispersion parameters. We follow the development provided by Lindley and Smith (1972), who proposed several formulations of HLMs based on differing assumptions of exchangeability. Here, we consider their model for exchangeability within multiple regression equations.

Specifically, let \(y\) denote a vector of \(n\) observations, \(X\) a known \(n \times p\) design matrix, \(\beta\) an unknown vector of \(p\) parameters, and \(\sigma\) an unknown scale parameter. Then given \(\beta\) and \(\sigma\)

\[
y|\beta, \sigma \sim N(X\beta, \sigma^2I_n) .
\]

The second stage of the hierarchical model, representing our prior information on the (possibly rescaled) parameter \(\beta\), can be specified by

\[
\beta_j|\xi, \sigma^2_\beta \sim N(\xi, \sigma^2_\beta) .
\]
where $\xi$ and $\sigma_\beta$ are unknown hyperparameters.

To complete the specification of the model, we assign vague reference priors to $\xi$, $\sigma_\beta$, and $\sigma$. In particular, we assume that $\xi$ is uniformly distributed on the real line, and that the prior densities for $\sigma_\beta$ and $\sigma$ are given by $p(\sigma_\beta) \propto \sigma_\beta^{-1}$ and $p(\sigma) \propto \sigma^{-1}$.

The estimation of the marginal posterior densities in HLMs is greatly simplified by the fact that analytic results are available when the dispersion parameters are known a priori. In the present case, integration with respect to $\xi$ to obtain the marginal joint posterior distribution of $(\beta, \sigma, \sigma_\beta)$ can be accomplished directly and leads to

$$
p(\beta, \sigma, \sigma_\beta | y) \propto \sigma^{-n} \exp\left[-\frac{(y - X\beta)^T (y - X\beta)}{2\sigma^2}\right] \times \sigma_\beta^{-p} \exp\left[-\frac{1}{\sigma_\beta} \sum_{j=1}^{p} (\beta_j - \bar{\beta})^2 / 2\sigma_\beta^2\right] \times \frac{1}{\sigma} \times \frac{1}{\sigma_\beta},
$$

(4.16)

where $\bar{\beta}$ is the arithmetic mean of the components of $\beta$.

Because the prior distributions for the dispersion parameters were specified to be uniform on the logarithmic scale, the quantile interval means can be obtained by exponentiating evenly spaced points on the real line. Clearly, an infinite number of quantile points are not wanted, as would be formally required with an improper prior, so we choose an arbitrary number of such quantile intervals around a crudely obtained estimate for each of the scale parameters (of course, a check should be made to verify that the extreme intervals are assigned negligible posterior probability using (2.22)).

Letting $n_\sigma$ and $n_{\sigma_\beta}$ represent the number of quantile interval means selected from $p(\sigma)$ and $p(\sigma_\beta)$, quantile integration results in an estimate of the marginal posterior distribution of $\beta$ given by

$$
p(\beta | y) \propto \sum_{i=1}^{n_\sigma} \sum_{j=1}^{n_{\sigma_\beta}} \sigma_i^{-n} \exp\left[-\frac{1}{\sigma_i} \sum_{j=1}^{p} (\beta_j - \bar{\beta})^2 / 2\sigma_\beta^2\right] \times \sigma_\beta^{-p} \exp\left[-\frac{1}{\sigma_\beta} \sum_{j=1}^{p} (\beta_j - \bar{\beta})^2 / 2\sigma_\beta^2\right].
$$

(4.17)

Hence, to determine the posterior distribution of $\beta$ we need only complete the square in each term and retain the normalizing constants. So doing leads to

$$
p(\beta | y) \doteq \frac{1}{K} \sum_{i=1}^{n_\sigma} \sum_{j=1}^{n_{\sigma_\beta}} w_{ij} N(B_{ij}X^T y / \sigma_i^2, B_{ij}),
$$

(4.18)

where

$$B_{ij} = \left[ X^T X / \sigma_i^2 + (I_p - J_p/p) / \sigma_\beta^2 \right]^{-1},$$

$$w_{ij} = |B_{ij}|^{1/2} \exp\left\{ \frac{1}{2} \left[ \frac{y^T X}{\sigma_i^2} - B_{ij} \frac{X^T y}{\sigma_i^2} \right] \right\},$$

$$K = \sum_{ij} w_{ij},
$$

(4.19)

and $J_p$ represents a $p$ square matrix of 1's.
This form of the approximation is particularly useful since the marginal distributions of each component or subset of components of $\beta$ are directly available. This is because their joint posterior distribution is estimated as a mixture of multivariate normal densities. Additionally, histogram estimates of the posterior distributions of $\sigma$ and $\sigma_\beta$ can be obtained using (2.22) directly from the weights $w_{ij}$.

Lindley and Smith (1972) applied this model to a 10 factor, non-orthogonal multiple regression model previously analyzed by Gorman and Toman (1966) and Hoerl and Kennard (1970). Due to the unavailability of the marginal posterior distributions in this example, Lindley and Smith proposed an iterative algorithm to obtain approximations to the mode of each component of $\beta$. For brevity, we provide only a plot of the marginal density for $\beta_1$ (Fig. 5) and the histogram estimates of the posterior distributions of $\sigma$ and $\sigma_\beta$ (Fig. 6) obtained using quantile integration. Eleven quantile points were used for each of the scale parameters, where each quantile point was taken to be the midpoint on the logarithmic scale of equally spaced intervals around the least squares estimates of $\sigma$ and $\sigma_\beta$. Also shown in Fig. 5 is the estimated mode (*) determined by Lindley and Smith. The computations required to obtain all results took two seconds on a DECstation 3100.

5. DISCUSSION

Quantile integration can be applied to a variety of integrals related to (1.1). The primary requirement for its use is that the integral defining the marginal (posterior) density be factorable into a product of conditional densities, preferably so that the corresponding conditional distributions can be evaluated easily.

Equation (1.1) depicts the most parsimonious form for such a factorization. At the opposite end of the spectrum lie integrals having the form

$$p(\theta_1) = \int \cdots \int p(\theta_1 | \theta_2, \theta_3, \ldots, \theta_m)p(\theta_2 | \theta_3, \ldots, \theta_m) \cdots p(\theta_m) \, d\theta_m \cdots d\theta_2. \quad (5.1)$$

Theoretically, quantile integration applies equally well in both cases, but its application in the latter results in an approximation expressible as

$$\frac{1}{n_2n_3 \cdots n_m} \sum_{i_2} \sum_{i_3} \cdots \sum_{i_m} p(\theta_1 | w_{i_2}^2, w_{i_3}^3, \ldots, w_{i_m}^m). \quad (5.2)$$

This equation demonstrates an obvious limitation of quantile integration; for integrals of the type (5.1), the computational effort needed to implement the algorithm grows exponentially with $m$. Thus, the method may not provide a practical means for evaluating high dimensional integrals having this form (in contrast, recall that the computational requirements increase only linearly with $m$ for integrals of the type (1.1)).
Another variation on (1.1) occurs when the parameters appearing in the integrand represent vectors rather than scalars. Quantile integration can be extended to such integrals in a rather obvious fashion, the primary difference being that quantile intervals are arbitrarily redefined to be quantile regions representing partitions of the parameter space into partition elements of known measure. Under regularity conditions similar to those cited in the scalar case, the convergence of the approximations obtained for vector-valued parameters is $O((\log(n))^p/n)$, where $p$ is the dimension of the parameter and $n$ is the number of partitioning elements from which quantile points are selected. Unfortunately, the constant associated with this order of approximation grows exponentially with $p$, although its rate of growth appears to be considerably lessened as the dependence between parameters increases. Additionally, issues of identifiability often are important in higher dimensions, and both issues are the topic of current investigation.

Another outstanding issue involves the manner in which outliers are handled. In Example 2, we integrated with respect to a Student-$t$ likelihood function and tacitly assumed that the quantile interval means covered the center of the predictive densities. However in the presence of extreme outliers, this procedure would not prove adequate without either drastically increasing the number of quantile points or assigning unequal weights to the extreme intervals in such a way as to restrict the maximum width of each interval. Preliminary investigation of this problem suggests that the latter provides an adequate solution, and a more detailed discussion of this point will appear in a forthcoming article.
REFERENCES


Figure legends

Figure 1. Integration results for bivariate normal distribution. a) Estimates of the marginal density of $\Theta_1$ obtained using the mid-quantile points corresponding to $n=3$, $5$, and $10$. Also shown is the exact marginal, in this case a standard normal density. At the mode, the estimates decrease monotonically with $n$; the largest value corresponds to $n=3$, followed by $n=5$, $n=10$, and the exact density. b) Estimates of the marginal density of $\Theta_1$ obtained using the quantile interval means corresponding to $n=3$, $5$, and $10$. Note the improved performance of the estimates using this choice of $w$. Again, at the mode the estimates decrease monotonically with $n$.

Figure 2. Relative errors from the bivariate normal integration. a) Relative errors of estimates obtained using mid-quantiles with $n=3$, $5$, $10$. b) Relative errors of estimates obtained using quantile interval means for $n=3$, $5$, $10$.

Figure 3. Estimates of marginal posterior distributions. a) Estimated posterior density of $\Theta_1$ based on 50 quantile interval means. The true posterior density is also depicted, but cannot be distinguished from the estimated posterior. b) Histogram estimate of the posterior distribution of $\Theta_2$. The plots have been scaled so that the areas in each quantile interval represent the probabilities estimated by Equation 2.22. The true posterior density is also shown.

Figure 4. Estimated posterior densities for time intervals 2-9 for the data in Example 2. The first order approximations proposed by Meinhold and Singpurwalla are shown in dashed lines for comparison.

Figure 5. Posterior distribution of $\beta_1$. The posterior mode estimated by Lindley and Smith is indicated by the (•).

Figure 6. Histogram estimates of the posterior distribution of $\sigma$ and $\sigma_\beta$ obtained using reference priors and (2.22). a) Histogram estimate for $\sigma$. b) Histogram estimate for $\sigma_\beta$. 