Available online at http://scik.org
J. Math. Comput. Sci. 3 (2013), No. 4, 929-944
ISSN: 1927-5307

# APPROXIMATE RISK ANALYSIS USING NUMERICAL INTEGRATION ON SPARSE GRIDS

S. CHEN*, X. WANG

Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, M3J 1P3, Canada

**Abstract.** In risk analysis with prior information, one often needs to evaluate multi-dimensional integrals in order to obtain various characteristics of posterior density functions. Monte Carlo Markov Chain method has been widely used. However, the MCMC method could be computationally intensive. The traditional method for numerical integration requires a full grid evaluation which is computationally intensive when the dimensions are not low. We introduce a novel approach to approximate Bayesian computation by numerical integrations on sparse grids. The number of required grid points for numerical integrations by using sparse grids does not rise exponentially with the dimensions. The proposed method is computationally efficient compared with the traditional numerical integration approach. The posterior density including the normalizing factor can be computed numerically. The posterior mean, median and confidence intervals can then be approximated directly. Both simulated and real data sets are used to evaluate the performance of the proposed method. Numerical experiments suggest that the proposed method could provide fast and efficient approximations with relatively high level of accuracy.

**Keywords**: Bayesian computation, multivariate quadrature, numerical integration, sparse grids.

**2000 AMS Subject Classification:** 65D30; 65C60; 97K80 47H09

## 1. Introduction

In risk analysis, Bayesian inference plays an important role since it integrates prior information or expert opinion with actual measurements or observation. In Bayesian data analysis,

*Corresponding author
Email address: chensy@mathstat.yorku.ca (S. Chen)
Received May 24, 2013

one often needs to evaluate multi-dimensional integrals in order to obtain various characteristics of posterior or marginal densities such as mean, median and confidence intervals. This requires finding the normalizing factor which ensures that the posterior density is indeed a probability density. The normalizing factor is defined by integrating a function that is proportional to the joint posterior density. If the problem under consideration does not assume a conjugate structure for the likelihood and prior distributions, the multi-dimensional integrals often do not have close forms.

Monte Carlo simulation (also called *Markov chain Monte Carlo*, or MCMC) is a general method based drawing values randomly from approximate distributions. The random numbers are then used to approximate the target posterior distribution. There are many excellent reviews of the MCMC method, such as Gelman *et. al* (2004). The MCMC method is very powerful and has many applications. However, Gelman *et. al* (2004) state that the Gibbs sampler and Metropolis algorithms have inherent inefficiency due to their random walk behaviors. Although reparametrerization and jumping rules can improve the situation, the problem remains for complicated models in high dimensional distributions such as Bayesian models for environmental space-time processes described in Le and Zidek (2006). Multimodal posterior distribution could also pose serious problems for MCMC techniques. It is quite easy for the MCMC simulations to stay in one single mode for a long period of time. Excellent review and detailed discussions can be found in Gelman *et. al* (2004).

For problems with only a few parameters, numerical integration using adaptive quadratures for numerical integration work well in low dimensions. Detailed descriptions of various quadrature rules can be found in Kennedy and Gentle (1980) and Ralston and Rabinowitz (1978). In general, numerical integration in multiple dimensions poses a serious numerical challenge in the past. This is due to the well known fact that it is inefficient to directly extend the univariate quadrature to multiple dimensions by applying univariate quadrature rule to each dimension. Such an extension will cause computational cost to rise exponentially in multiple dimensions. This is also known as the *curse of dimensionality*, a term coined in Bellman (1961).

Recently, numerical integration in high dimensions has attracted research interests in numerical mathematics. The approach of numerical integration on sparse grids for high dimensions has emerged as an effective and efficient method in numerical analysis in recent years. The original idea of sparse grid can be traced back to Smolyak (1963). The high-dimensional basis is derived from one-dimensional multi-scale basis by a tenor product construction. The sparse grid method based on Smolyak's rules is exact for polynomial function. It can also be used to approximate functions that are not polynomial. The sparse grid method aims to be exact in the class of complete polynomial instead of tensor products of univariate polynomial. Therefore, the required grid points for numerical integration does not rise exponentially with the dimensionality. Bungartz and Griebel (2004) give an excellent review of the sparse grid method.

Using the sparse grids integration, we directly compute the normalizing factor of the posterior density. Hence, posterior density functions are approximated numerically. Consequently, the posterior mean and median are obtained numerically as well. By solving appropriate nonlinear integration equations iteratively, the Bayesian posterior confidence intervals are constructed as well. Unlike approximations proposed in the literature based on normal approximation or Laplace's method, the proposed approach of numerical integration on sparse grids does not require the posterior density to be approximately normal or to satisfy some regularity conditions.

We demonstrate our method by using simulation studies on the single parameter model. We also evaluate the performance of our proposed method with hierarchical models on two well-studied data sets. The first one is the tumor incidence data set from Tarone (1982). We apply hierarchical model of binomial distributions with Beta prior. The hyperprior is assumed to be non-informative. It only takes about 0.55 seconds to derive an accurate approximation of the marginal distributions of the hyperparameters. The second one is the coagulation time data set from Box, Hunter and Hunter (1978). Normal sampling model with uniform priors are used to analyze this data set. We also provide computing time for approximating the marginal posterior density functions. It takes only 3.31 seconds to derive the marginal confidence intervals for all seven parameters.

In Section 3, we introduce the problem of approximating the posterior density function numerically. We then discusses the sparse grid method in detail and presents some theoretical results. Section 3 demonstrates the proposed method through simulation studies and data analyses. Discussions are provided in Section 4.

## 2. Approximate Bayesian Inference by Numerical Integration

In Bayesian analysis, the observed data $\mathbf{y} = (y_1, y_2, \cdots, y_n)$, given a vector of unknown parameters $\boldsymbol{\theta}$, follow a probability distribution $f(\boldsymbol{y}|\boldsymbol{\theta})$ with a prior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\eta})$, where $\boldsymbol{\eta}$ is a vector of hyperparameters sampled from a hyperprior density $h(\boldsymbol{\eta})$. One key element in Bayesian computation is to evaluate various integrals associated with the posterior densities such as the normalizing factor.

For hierarchical models, we assume that $\theta_1, \theta_2, \ldots, \theta_m$ are *i.i.d.* random variables from a prior distribution $\pi_{\boldsymbol{\eta}}$ where the hyeperparameter $\boldsymbol{\eta}$ follows a hyperprior distribution $h(\boldsymbol{\eta})$. Therefore the joint posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ is given by

$$p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) \ \ \pi(\boldsymbol{\theta}|\boldsymbol{\eta}) \ h(\boldsymbol{\eta}). \tag{1}$$

The normalizing factor is then defined as

$$C(\boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{\theta}) \ \ \pi(\boldsymbol{\theta}|\boldsymbol{\eta}) \ h(\boldsymbol{\eta}) \ d\boldsymbol{\eta}. \tag{2}$$

0.1. **Numerical Integration on Sparse Grid Algorithm.** In the following introduction of the sparse grid algorithm, we let $g(u) = f(y|u)\pi(u|\boldsymbol{\eta})$ and $C(y, \boldsymbol{\eta}) = \int g(u)du$. For simplicity, we assume that the hyperparameters take fixed values although the numerical integration can certainly handle integration involving hyperprior distributions. We will first describe the numerical method to approximate the normalizing factor $C(y, \boldsymbol{\eta})$ when $u$ is a scalar, which will be utilized by the high dimensional method in the following context.

0.1.1. *Univariate quadrature.* A quadrature approximation of an integral requires only a finite number of function evaluations, and uses the weighted sum as the approximation:

$$\int g(u)du \approx \sum_{i=1}^{K} w_i g(u_i). \tag{3}$$

Davis and Rabinowitz (1975) and Neumaier (2001) have detailed treatments of Gauss quadrature and other quadrature rules as well, such as Clenshaw-Curtis rule, Newton-Cotes rules (midpoint, rectangle, trapezoidal), and Gauss quadratures for special purposes ( Chebyshev, Laguerre, Hermite, Jacobi, Kronrod, Patterson).

More generally, we define an operator $Q_j$:

$$Q_j[g] = \sum_{u \in U_j} w(u)g(u), \tag{4}$$

where $U_j$ specifies the set of evaluation points; and $w : U_j \to \mathbb{R}$ provides the corresponding weights. Note that for different quadrature rules the cardinality of $U_j$ are different; however they all achieve improving polynomial exactness with increasing $j$.

0.1.2. *Multivariate quadrature.* For a multivariate function $g$, the multivariate quadrature rules seek the optimal $U_K$ and $W_K$ to achieve highest possible polynomial exactness with $K$ function evaluations. We first define the order of differentiability of a multivariate polynomial to be its *total order*. Let a $d$-variate polynomial

$$g(u_1, \cdots, u_d) = \sum_{l=1}^{L} c_l u_1^{e_{l1}} \cdots u_d^{e_{ld}},$$

then its total order is the maximal $e_{l1} + \cdots + e_{ld}$ for all $l = 1, \cdots, L$.

We define the multivariate quadrature rule through the tensor product and allow different polynomial exactness among different dimensions:

$$(Q_{j_1} \otimes \cdots \otimes Q_{j_d})[g] = \sum_{u_1 \in U_{j_1}} \cdots \sum_{u_d \in U_{j_d}} w_{j_1}(x_1) \cdots w_{j_d}(x_d) g(x_1, \cdots, x_d), \tag{5}$$

where $U_{j_1}, \cdots, U_{j_d}$ and weights $w_{j_1}, \cdots, w_{j_d}$ are exactly the same as in the univariate quadrature rules. The approach is also widely known as the *full grid* method. However it suffers from the curse of dimensionality. To see this, let's fix $j_1 = j_2 = \cdots = j$, and suppose there are $K$

points in $U_j$, then the number of points used in the tensor product is exactly $K^d$. For a given accuracy, the exponential growth of the number of function evaluations is a serious challenge for a high dimensional problem.

In contrast to the full grid approach, the sparse grid approach to be introduced in the next section does not experience an exponential growth of the number of function evaluations. Although the sparse grid approach also uses the tensor product of the underlying univariate quadrature rules, it only selects evaluation points of the highest marginal benefit. We note that, for a certain type of particular problem, one might be able to directly solve the high dimensional problem without resorting to the tensor approach, see Stroud (1971) and Cools (2003). However such a solution usually does not extend to arbitrary dimensions. The tensor product approach of the sparse grid is general and easy to implement as shown in the next section.

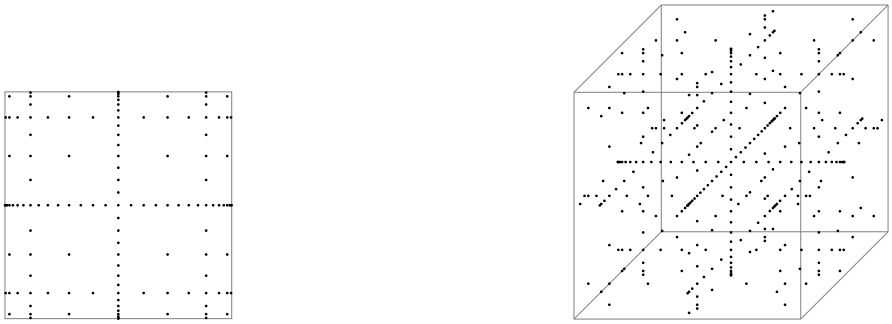
0.1.3. *Sparse Grid Algorithm.* The sparse grid idea dates back to Smolyak (1960), which can utilize any univariate quadrature rules $\{Q_j : j \in N\}$ to approximate a $d$-dimensional problem. Let $\|j\|_1 = j_1 + \cdots + j_d$, then the set of evaluation points in the sparse grid method is simply

$$\mathcal{U}(q,d) = \bigcup_{\|j\|_1 \le q+d-1} (U_{j_1} \otimes \cdots \otimes U_{j_d}),\ \forall q \in N. \tag{6}$$

The cardinality of $\mathcal{U}(q,d)$ increases monotonically with $q$, which measures our computation effort. We illustrate the difference between the sparse grid and the full grid using a bivariate example. Let $d = 2, q = 2$, then the requirement $\|j\|_1 \le 3$ allows the following multi-index $j = [j_1, j_2]$: $[1,1], [1,2], [2,1]$. For Gauss-Kronrod-Patterson univariate quadrature rule, $U_1 = \{0.5\}, U_2 = \{0.1127, 0.5, 0.8873\}$, let $a = 0.1127, b = 0.5, c = 0.8873$, then $U_1 \otimes U_1 = \{(b,b)\}$; $U_1 \otimes U_2 = \{(b,a),(b,b),(b,c)\}$; $U_2 \otimes U_1 = \{(a,b),(b,b),(c,b)\}$; and the sparse grid $\mathcal{U}(2,2) = \{(a,a),(b,a),(b,c),(a,b),(c,b)\}$. In contrast, the full grid approach will include four additional points $\{(a,a),(c,c),(a,c),(c,a)\}$, which almost doubles the number of function evaluations. The Fig 1 shows the sparse grid for $d = 2,\ q = 5$ and $d = 3,\ q = 5$ using Gauss-Kronrod-Patterson univariate rule.

Let $Q^0 = 0$, and define the operator

$$\Delta_j = Q_j - Q_{j-1}, j \in N, \tag{7}$$



(a) $d = 2$ (b) $d = 3$

FIGURE 1. Sparse grid on unit square and unit cube for $q = 5$ with underlying Gauss-Kronrod-Patterson univariate quadrature. There are $129$ gird points in the unit square and $351$ grid points in the unit cube.

then the Symolk's definition of sparse grid is

$$A(q,d) = \sum_{\|j\|_1 \leq q+d-1} \Delta_{j_1} \otimes \cdots \otimes \Delta_{j_d}, q \in N. \tag{8}$$

Wasilkowski and Wozniakowski (1995) showed a combination technique to calculate the (8) in terms of the original univariate quadrature:

$$A(q,d) = \sum_{q \leq \|j\|_1 \leq q+d-1} (-1)^{q+d-1-\|j\|_1} \cdot \binom{d-1}{\|j\|_1 - q} \cdot (U^{j_1} \otimes \cdots \otimes U^{j_d}). \tag{9}$$

We illustrate the equivalence between (8) and (9) through a two dimensional example. In this case,

$$\begin{aligned} A(q,2) &= \sum_{\|j\|_1 \leq q+1} \Delta_{j_1} \otimes \Delta_{j_2} \\ &= \Delta_1 \otimes (\Delta_1 \oplus \cdots \oplus \Delta_{q-1} \oplus \Delta_q) + \\ &\quad \Delta_2 \otimes (\Delta_1 \oplus \cdots \oplus \Delta_{q-1}) + \\ &\quad \vdots \\ &\quad \Delta_q \otimes \Delta_1 \\ &= \Delta_1 \otimes U_q + \Delta_2 \otimes U_{q-1} + \cdots + \Delta_q \otimes U_1 \\ &= \sum_{\|j\|_1 = q+1} U_{j_1} \otimes U_{j_2} - \sum_{\|j\|_1 = q} U_{j_1} \otimes U_{j_2}, \end{aligned} \tag{10}$$

which is equivalent to (9).

The implementation of (9) is best described by its explicit formula:

$$A(q,d) = \sum_{\|j\|_1=q}^{q+d-1} (-1)^{q+d-1-\|j\|_1} \binom{d-1}{\|j\|_1 - 1} \sum_{u_1 \in U_{j_1}} \cdots \sum_{u_d \in U_{j_d}} w_{j_1}(u_1) \ldots w_{j_d}(u_d) g(u_1, \ldots, u_d), \tag{11}$$

where $w_{j_k}(u_k)$ is the weighting function of the univariate quadrature rule. It is ready to see that the sparse grid approximation is in fact a weighted sum of the function values at the points in $\mathcal{U}(q,d)$. However a naive expansion of the formula (11) will incur redundant function evaluations at certain nested quadrature points. A more efficient way is to use the following procedure: (i)generate $\mathcal{U}(q,d)$; (ii) evaluate $g$ on $\mathcal{U}(q,d)$, (iii) use equation (11); (iii) compute the combined weight $\tilde{w}$ for each point in $\mathcal{U}(q,d)$; (iv) finally compute

$$A(q,d) = \sum_{u \in \mathcal{U}(q,d)} \tilde{w}(u) g(u). \tag{12}$$

Bungartz and Griebel (2004) provide a comprehensive survey of the sparse grid method with discussions.

Given a non-normalized posterior density function, the algorithm to approximate Bayesian inference is then quite straightforward:

1. Determine a desirable level of sparse grid.

2. Obtain the normalizing factor using sparse grid integration.

3. Use the computed normalizing factor compute the posterior mean.

4. Use a root-finding algorithm to obtain the median, lower and upper bounds of posterior confidence interval by solving nonlinear integral equations.

We remark that there is no known theoretical results to select the level of sparse grid. This is a value specified by the user. High level of sparse grids will deliver better accuracy with increasing computational costs. A common practice, which is also followed in our numerical experiments, is to increase the level of sparse grid sequentially, and then find an acceptable trade-off between the computation time and required accuracy.

## 3. Numerical Experiments

In this section, we demonstrate our proposed method on both single parameter models and hierarchical models with real world data sets.

Many statistical applications involve multiple parameters or data sets that are considered to be related. We now consider a well studied data set concerning tumor rates in historical control groups and current group of rats from Tarone (1982). Suppose that one is concerned about the probability of tumor in a population of certain female rats assigned to a control group. The data show that 4 out of 14 rats in this control group developed a certain kind of tumor called *endometrial stromal polyps*. Suppose that 70 historical data sets are available. For the purpose of demonstration, we assume exchangeability in order to apply the hierarchical model described in Section 2. Detailed discussions and analysis of this data set can be found in Gelman *et al. (2004)*.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/19 | 0/19 | 0/19 |
| 0/19 | 0/18 | 0/18 | 0/17 | 1/20 | 1/20 | 1/20 | 1/20 | 1/20 | 1/20 |
| 1/18 | 1/18 | 2/25 | 2/24 | 2/23 | 2/20 | 2/20 | 2/20 | 2/20 | 2/20 |
| 2/20 | 1/10 | 5/49 | 2/19 | 5/46 | 3/27 | 2/17 | 7/49 | 7/47 | 3/20 |
| 3/20 | 2/13 | 9/48 | 10/50 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 |
| 4/20 | 10/48, | 4/19 | 4/19 | 4/19 | 5/22 | 11/46 | 12/49 | 5/20 | 5/20 |
| 6/23 | 5/19 | 6/22 | 6/20 | 6/20 | 6/20 | 16/52 | 15/47 | 15/46 | 9/24 |

Current experiment: 4/14.

TABLE 1. Rat Tumor incidence in historical control groups and current groups of rats, from Tarone (1982). The table displays the values of $y_j/n_j$ (number of rats with tumors) / (total number of rats).

| Sparse Grid Level | Q=5 (145 points) | Q=6 (321 points) | Q=7 (705 points) | Q=8 (1537 points) |
|---|---|---|---|---|
| Est. Max. Abs. Error. | $2.4 \times 10^{-2}$ | $1.1 \times 10^{-2}$ | $4.2 \times 10^{-4}$ | $4.9 \times 10^{-6}$ |
| Est. Max. Rel. Error | $26\%$ | $3.5\%$ | $0.1\%$ | $0\%$ |
| Run Time | 0.26 sec. | 0.55 sec. | 1.18 sec. | 2.52 sec. |

TABLE 2. Comparison of accuracy by using different levels of sparse grids. The computation times were recorded using MATLAB on Leveno T400S laptop.

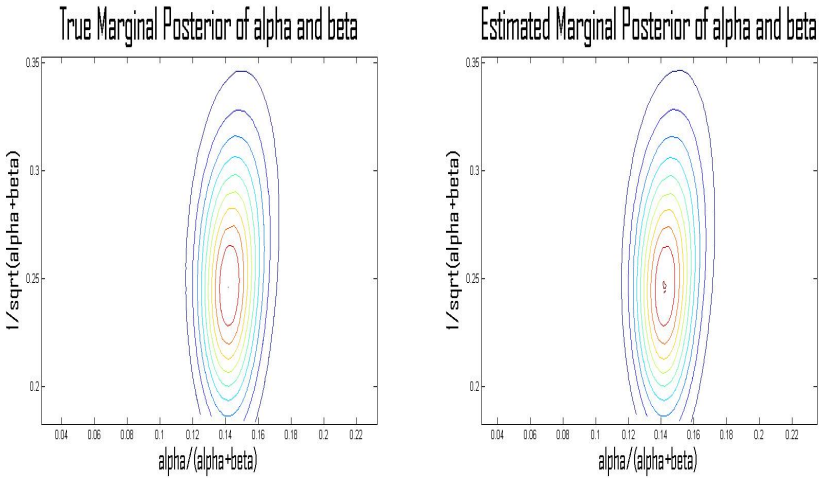



FIGURE 2. Contour plots of true Posterior density (left) and estimated Posterior density (right)
of $(\alpha/(\alpha+\beta), 1/\sqrt{\alpha+\beta})$ using 145 sparse grid points. To avoid computational overflow, we subtract the maximum value from the log density from each point on the grid and exponentiate.

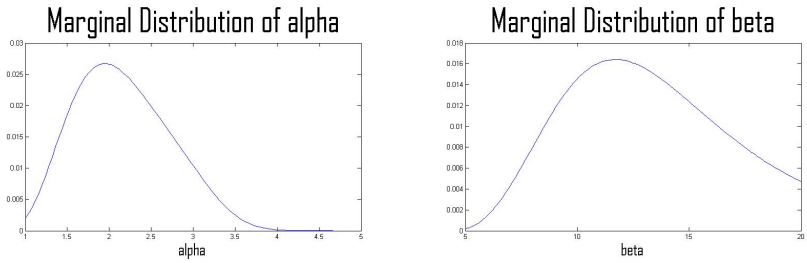



FIGURE 3. Marginal Densities for $(\alpha, \beta)$, $\alpha$ and $\beta$ respectively.

Consider the Binomial sampling model with $Beta$ prior and $h(\alpha, \beta)$ as the hyperprior distribution. The posterior density is given by equation (5), we restate the result here for convenience:

$$p(\alpha, \beta|\boldsymbol{y}) \propto h(\alpha, \beta) \prod_{j=1}^{m} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma(\alpha+y_j)\,\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}. \tag{13}$$

Following the discussions in Gelman *et al.* (2004), we employ a noninformative uniform prior on $(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2})$ which implies that

$$p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}. \tag{14}$$

Based solely on the sample mean and standard deviation of the 70 historical data sets without any prior distribution, a crude estimate for $(\alpha, \beta)$ is $(1.4, 8.6)$. Therefore, we choose the domain

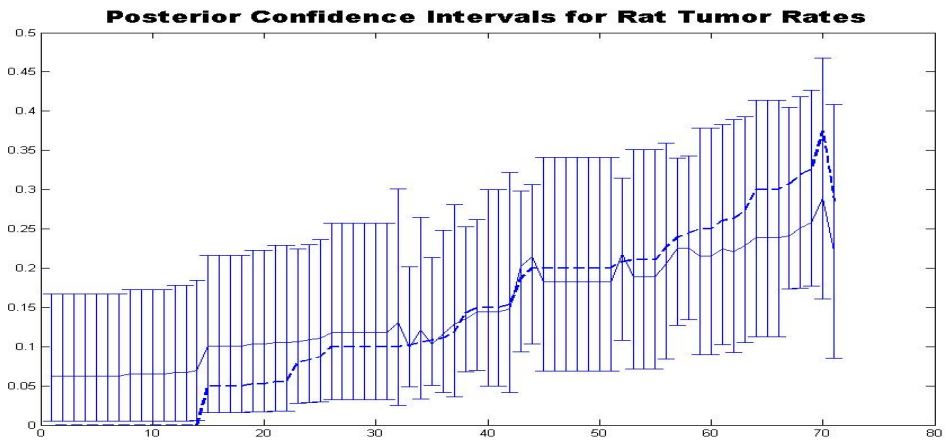



FIGURE 4. Plot of marginal posterior confidence intervals for $\theta_j,\ j = 1, 2, \cdots, 71$. The solid line corresponds the posterior mean. The broken line corresponds to the observed rate $y(i)/n(i)$.

of our prior distribution to be $[1, 5] \times [5, 20]$ in order to cover a reasonable range for these two parameters.

We compute the true posterior density function of $(\alpha, \beta)$ and approximated posterior density by using 145 grid points. The contours plots are presented in Figure 3. The results of using different number of grid points to estimate the posterior density are also provided in Table 5. It can be seen that the level 6 sparse grid with 321 points already achieves high level of accuracy for approximating the posterior density. The estimated relative error is close to 3.5%. We emphasize that this is achieved in 0.55 second by a computer program in MATLAB on a Lenevo T400s laptop.

By using our method, we obtain the posterior mean for $(\alpha, \beta)$ $(2.11, 13.02)$. By using the posterior mean, we can then compute the point estimate for $\alpha/(\alpha + \beta)$ which equals to 0.1395. The mean of all 71 experiments is 0.1381. This is not surprising since we have chosen a non-informative uniform prior for $\alpha/(\alpha+\beta)$ to represents our ignorance about the mean of the prior distribution. We can also obtain the confidence intervals for $\alpha, \beta$. We also compute the marginal distributions for $\alpha$ and $\beta$ respectively as shown in Figure 4.

The 95% posterior confidence interval of $\alpha$ is $[1.21, 4.82]$ with posterior median of 2.06. The 95% posterior confidence interval of $\beta$ is $[7.06, 16.11]$ with posterior median of 12.45. The computational time is 8.24 seconds to find the confidence intervals for both $\alpha$ and $\beta$ by using

level 6 (321 points) sparse grid. The confidence interval for the last experiment is of particular interest since it is the current experiment. The non-Bayesian estimate is $4/14 = 0.2857$. We obtain that the confidence interval is $[0.09, 0.41]$ with posterior median of $0.2191$ and posterior mean of $0.2221$. The posterior means of $\theta_i$ are pulled towards 0.1395. Figure 5 shows the approximated posterior confidence interval of $\theta_i$ and posterior means.

In order to find the confidence intervals for all the $\theta_i$, $i = 1, 2, \cdots, 71$, one needs to solve the nonlinear equations described in Section 2. For this step, an efficient search algorithm is essential. We tried two search algorithms in MATLAB with their default settings. Our experiences indicate that a search algorithm not using derivatives provide better numerical results. The search algorithm using derivatives might be able to deliver same or better results if the appropriate setting is chosen. We also notice that the starting points for any search algorithm could be very important. The starting points for the bisection search algorithm are set to be either 0, 1 or 0.5. Since our focus is not on the search algorithms, we do not investigate this issue further. The computational time including using the derivative-free search algorithm is 135 seconds using level 5 sparse grid.

We also demonstrate our method on hierarchical model with 7 dimensions. This data set has been considered as an example in the statistical literature. Table 3 shows the data set. Consider $J$ independent experiments with parameter $\theta_j$ for each experiment with $n_j$ independent normally distributed observations:

$$y_{ij}|\theta_j \ \sim \ N(\theta_j, \sigma^2), \ \ i = 1, 2, \cdots n_j; \ j = 1, \cdots, J. \tag{15}$$

We assume that the parameter $\theta_j$ are drawn from a normal distribution with hyperparameters $(\mu, \tau)$:

$$p(\theta_1, \cdots, \theta_J|\ \mu, \tau^2) \ \ = \ \ \prod_{j=1}^{J} N(\theta_j|\mu, \sigma^2). \tag{16}$$

We use noninformative uniform priors for all the hyperparameters.

The total run time for the using sparse grid method is recorded as $3.31$ seconds on a Lenevo T400S laptop computer using a MATLAB program. The sparse grid method only uses 113

| Diet | Measurements |
|---|---|
| A | 63, 60, 63, 59 |
| B | 63, 67, 71, 64, 65, 66 |
| C | 68, 66, 71, 67, 68, 68, |
| D | 56, 62, ,60, 61, 63, 64, 63, 59 |

TABLE 3. Cogulation time in seconds for blood drawn from 24 animals randomly allocated to four different diets. Originally from Box, Hunter, and Hunter (1978).

points for integrations on 7 dimensions which is much less than the required grid points for the full grid approach.

| Parameter | $2.5\%$ | $Median$ | $97.5\%$ |
|---|---|---|---|
| $\theta_1$ | 57.70 | 64.50 | 71.29 |
| $\theta_2$ | 60.62 | 66.00 | 71.37 |
| $\theta_3$ | 64.82 | 68.00 | 71.17 |
| $\theta_4$ | 56.02 | 61.00 | 65.97 |
| $\mu$ | 59.56 | 64.00 | 68.43 |
| $\sigma$ | 1.33 | 3.64 | 5.95 |
| $\tau$ | 1.48 | 6.68 | 11.88 |

TABLE 4. Summary of posterior quantiles at 25% and 97.5% and median using sparse grid integration.

| Spars Grid | Full Grid | | |
|---|---|---|---|
| Points | Points(q=5) | Points (q=7) | Points (q=10) |
| 113 | $78,000$ | $820,000$ | $1,000,000$ |

TABLE 5. Comparison of Grid Points for Full Grid and Spars Grid.

We assume non-informative uniform hyperprior distribution to represent our ignorance on all the hyperparameters. The range of integration is chosen by using the frequent's 95% confidence intervals or proportional to the point estimates. The 95% posterior confidence interval for all seven parameters are listed in Table 4. The shrinkage effect is most significant for the first experiment which has the smallest sample size among the four groups. The medians of other groups do not seem to be affected. We also compare the total number sparse grid points used with possible full grid points in Table 5. The computational efficiency of using the sparse grid method is quite clear.

We now consider a classical example for Bayesian model selection in Carlin and Chib (1995). They study a model selection example in which there are two competing explanatory

variables to explain a single response variable. For 42 specimens of radiata pine, the maximum compressive strength parallel to the grain as $y_i$ were observed together with its density $x_i$ and density adjusted for resin content $z_i$.

It is desired to compare the following two models:

$$M = 1: \quad y_i = \alpha + \beta x_i + \epsilon_i, \qquad \epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2), i = 1, 2, \cdots, n, \tag{17}$$

$$M = 2: \quad y_i = \gamma + \delta x_i + \epsilon_i, \qquad \epsilon_i \overset{i.i.d.}{\sim} N(0, \tau^2), i = 1, 2, \cdots, n. \tag{18}$$

where $M = 1, or 2$, specifies the model choice. Thus, we have $\theta_1 = (\alpha, \beta, \sigma)$ and $\theta_2 = (\gamma, \delta, \tau)$. It then follows that

$$P(\mathbf{y}|M = j) = \int f(\mathbf{y}|\theta_j, M = j)\ p(\theta_\mathbf{j}|M = j) d\theta_j. \tag{19}$$

We adopt the same setting as that in Carlin and Chib (1995). Namely, we place $N(3000, 185)^t$, $diag(10^6, 10^4))$ priors on $(\alpha, \beta)^t$ and $(\gamma, \delta)^t$, and inverse gamma priors on $\sigma^2$ and $\tau^2$, both mean and variance equal to $300^2$. To be consistent, the model prior probabilities are also set to be $\pi_1 = 0.9995$ and $\pi_2 = 0.0005$.

Carlin and Chib (1995) reported that the Bayes factor is 4420 with posterior probabilities $(0.3114, 0.6686)$ for the two competing models. Our estimated Bayes factor is 4309 and posterior probabilities $(0.3169, 0.6831)$ while using the domain of integration to be $[0, \quad 9000] \times [-215, \quad 485] \times [1, \quad 2 * 90000]$ which roughly correspond to two or three standard deviations around the prior mean. Therefore, we are able to find a very reasonable approximation by using the proposed numerical approach.

The run time is 0.4836 second on a 64-bit desktop compute with Intel i7CPU with 6GB memory. It can provide fast and efficient approximation when other priors are employed.

Although the proposed method could provide fast and relatively accurate approximation, significant numerical challenges still remain. Since the conditional likelihood by definition could be very small based on the data, it could be numerically zero for very large data set or some non-informative prior. To alleviate this problem, we multiply the conditional likelihood by $10^6$ or any other constant that will bring the integrand to a reasonable numerical level. We also observe that the result is dependent on the domain of integration. For example, a domain of

$[0,\ \ 9000] \times [-215,\ \ 585] \times [1,\ \ 2 * 90000]$ produces an estimate Bayes factor of 5020. Carlin and Chib (1995) reported an average of 4420 with the 95% interval of $(4353, 4487)$. This is clearly not comparable with their estimates. We believe that this dependence on the domain of integration might be related to the aforementioned numerical challenge.

## 3. Concluding Remarks

The Monte Carlo Markov Chain methods have proven to be very powerful and effective. However, they could be computationally intensive especially for high dimensions. We propose to apply the sparse grid method for numerical integration to conduct approximate Bayesian inference especially for exploratory purposes. The main advantage of using sparse grid integration is that it does not require exponentially increasing computational cost as the traditional full grid method does. Furthermore, it requires far fewer points than the Monte Carlo methods. Results from our numerical experiments suggest that this method is very fast for our selected case studies and could be promising for providing efficient approximations for Bayesian models in high dimensions. Work is in progress to investigate an effective and adaptive strategy to stabilize the numerical integration and choose the domain of integration judiciously.

**ACKNOWLEDGEMENTS:** This research was supported in part by the Natural Sciences and Engineering Research Council of Canada, the New Opportunity Fund by the Canadian Foundation of Innovation and Ontario Innovation of Trust.

### REFERENCES

[1] Bellman, R. (1961).Adaptive Control Process: A Guided Tour. Princeton University Press.

[2] Brent, R. (1973). Algorithms for Minimization Without Derivatives, Prentice-Hall, 1973.

[3] Bungartz, H.J. and Griebel, M. (2004). Sparse grids. *Acta numerica*, 147-269.

[4] Burkardt, J. (2010). Slow exponential growth for Gauss Patterson sparse grids. Technical Report, ICAM & ITD, Virginia Tech.

[5] Carlin, B.J. and Chib, S. (1995). Bayesian model via Monte Carlo Markov Chain methods. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 57, 473-484.

[6] Cools, R. (2003). An encyclopedia of cubature formulas, Journal of Complexity, Vol. 19, 445-453.

[7] Davis, P.J. and Rabinowitz, P. (1975). Methods of Numerical Integration. Academic Press.

[8] Delvos, F.J. (1982). D-variate Boolean Interpolation. Journal of Approximation Theory, vol. 34, 99-114.

[9] Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D. B. (2003). Bayesian Data Analysis, 2nd edition, Chapman and Hall.CRC.

[10] Kennedy, W.J. and Gentle. J. E. (1980). Statistical Computing. Marcel Dekker, Inc, New York.

[11] Neumaier, A. (2001). Introduction to Numerical Analysis. Cambridge University Press.

[12] Novak, E. and Ritter, K. (1996). High dimensional integration of smooth functions over cubes. Numerische Mathematik, vol. 75, 79?7.

[13] Ralston, A. and Rabinowitz, P. (1978). A First Course in Numerical Analysis. Dover Publications, Inc, New York.

[14] Smolyak, S.A. (1960). Interpolation And Quadrature Formula For The Class $W_s^a$ And $E_s^a$, *Dokl. Akad. Nauk SSSR*, 131, 1028-1031, (in Russian, English Translation: Soviet Math. Dokl. 4, 240-243 (1963)).

[15] Stroud, A. H. (1971). Approximate Calculation of Multiple Intergrals. Prentice-Hall, Englewood Cliffs, New Jersy.

[16] Tarone, R.E. (1982). The use of historical control information in testing for a trend in propotions. *Biometrics*, 38, 215-220.

[17] Wasilkowsi, W.G. and Wozniakowski, H. (1995). Explicit cost bounds of algorithms for multivariate tensor product problems. Journal of Complexity, Vol 11, 1?6.

[18] Le, N.D. and Zidek, J.V. (2006). Statistical Analysis of Environmental Space-Time Processes, Springer Series in Statistics.