T E C H N I C A L

R E P O R T

0543

NUMERICAL INTEGRATION IN LOGISTIC-NORMAL MODELS

GONZALEZ J., TUERLINCKX F., DE BOECK P., and R. COOLS

I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Numerical integration in logistic-normal models

J. González, F. Tuerlinckx, P. De Boeck and R. Cools*

Department of Psychology, K. U. Leuven

Tiensestraat 102, B-3000 Leuven, Belgium

**Department of Computer Science, K. U. Leuven*

Celestijnenlaan 200A, B-3001 Heverlee, Belgium

July 29, 2005

Abstract

When estimating logistic-normal models, the integral appearing in the marginal likelihood is analytically intractable, so that numerical methods such as Gauss-Hermite quadrature (GH) are needed. When the dimensionality increases, the number of quadrature points becomes too high. A possible solution can be found among the Quasi-Monte Carlo (QMC) methods, because these techniques yield quite good approximations for high dimensional integrals with a much lower number of points, chosen for their optimal location. In this paper a comparison will be made between three integration methods: GH, QMC, and full Monte Carlo integration (MC).

Keywords: (Quasi-)Monte Carlo, low discrepancy sequences, Gaussian quadrature, multidimensional integrals.

1 Introduction

In many different fields such as physics (e.g., Morokoff and Caflisch, 1995), geostatistics (e.g., Jank, 2005), psychometrics (e.g., Fischer and Molenaar, 1995; De Boeck and Wilson, 2004), biostatistics (see e.g., Lesaffre and Spiessens, 2001 for an application in a longitudinal clinical trial), and finance (e.g., Paskov, 1995) among others, the estimation process of the parameters of a model involves the calculation of an integral. In general, random parameters are included in a model to account for heterogeneity and related dependence between outcome variables. Although in principle these random parameters can follow any probability distribution, the normal and multivariate normal distribution are typically assumed depending of whether one or more random effects are included in the model. Unfortunately, the resulting (multidimensional) integral usually does not have an analytic solution, and therefore it is necessary to use numerical methods to approximate it. Well-known numerical methods are Gauss-Hermite quadrature (GH) and Monte Carlo (MC) integration. Both methods utilize the same kind of *cubature formula* in which the integral is approximated by a weighted sum of the integrand evaluated in a set of points. The GH method makes use of a set of fixed (known) points and weights (available from standard tables; Abramowitz and Stegun, 1972, p. 924). On the contrary, MC is based on a uniformly random distributed set of points (see e.g., Robert and Cassella, 2004, p. 83).

A major disadvantage of the GH method is that the number of quadrature points increases as an exponential function of the number of dimensions, so that the method rapidly becomes unfeasible. As an alternative, the MC method may be considered more suited for problems with high dimensionality. However, locating points at random does not guarantee an optimal distribution of the points (i.e., because of the sampling error, the points may not be distributed exactly uniformly.

An alternative to both the GH and MC methods is the so-called Quasi-Monte Carlo (QMC) method (e.g. Niederreiter, 1992; Judd, 1998; Caflisch, 1998). QMC works like the regular MC but instead of using an uniformly and randomly distributed set of points, ‘uniformly distributed’ deterministic sequences, called low discrepancy sequences (LDS) (Niederreiter, 1992), are utilized. The strength of the QMC method is that the distribution of points is optimal indeed, following some criterion to be explained.

Applications of QMC in finance (Lemieux and L’Ecuyer, 2001) have shown that the method works remarkably well in problems with integrals of very high dimensionality (see e.g., Paskov, 1995 where a 360-dimensional integral is evaluated). Jank (2005) implemented the Monte Carlo EM algorithm based on QMC and applied it to a geostatistical problem in which an integral of dimensionality 16 has to be solved because the outcome covaries with 16 distinct geographical location variables.

The aim of this paper is to compare the performance of three numerical integration methods: GH, MC and QMC, for relatively high-dimensional integrals appearing in one step of the estimation process in logistic-normal models. Because QMC has been successfully used in other fields, mainly in finance, we investigate its use for mixed logistic regression models because of its promising performance for high dimensional integration.

In this text, we will not explain neither GH and related cubature (i.e. multivariate quadrature) methods, nor MC methods in detail, as there are many good references (e.g., Stroud, 1971; Davis and Rabinowitz, 1984; Cools, 1997, 2002; Robert and Casella, 2004; Caflisch, 1998).

The paper is organized as follows. Section 2 explains how QMC methods work and briefly introduces two LDS, the Halton and Sobol sequences. Section 3 presents the logistic-normal model and motivates the problem of multidimensional (intractable) integrals. In Section 4 we compare the three methods of integration: GH, QMC and MC for the simple case of one integral over a logistic function, and next we report on a small case study with an integral over a product of logistic functions. Finally, we give some conclusions.

2 Quasi-Monte Carlo integration

Both MC and QMC methods approximate the integral of interest by means of a sample average in the following way

$$\int_{\Omega} f(x_1, \ldots, x_d) dx_1, \ldots, dx_d \approx \frac{vol(\Omega)}{N} \sum_{i=1}^{N} f(x_1^i, \ldots, x_d^i) \tag{1}$$

In the regular MC method the vector $\boldsymbol{x}^i = (x_1^i, \ldots, x_d^i)$ follows a uniform distribution in Ω. Note that when the region of integration is the unit hypercube, $\Omega = (0,1)^d$, then $vol(\Omega) = 1$ in (1). If an integration region different from the unit hypercube is considered, a suitable transformation of variables has to be applied. By the Law of large numbers (see e.g., Shao, 2003) the average $\frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}^i)$ converges almost surely to the expected value $E(f(\boldsymbol{x}))$, which is in this case the integral we are interested in.

Instead of using randomly distributed points, QMC uses deterministic points. So, in (1) the vector $\boldsymbol{x}^i = (x_1^i, \ldots, x_d^i)$ is not a random sample but a set of points deterministically chosen. These points are commonly known as Low Discrepancy Sequences because they are a better approximation of uniformity (discrepancy is a measure of deviation from uniformity, for details see Niederreiter, 1992). LDS provide better coverage than random points, because each new point is chosen in such a way that it covers the unit hypercube in the areas not covered by previous points.

Although other LDS exist, in this paper we will use the Halton and Sobol sequences (e.g., Kocis and Whiten, 1997), because they are the most common. Figure 1 shows the improved uniformity of Sobol and Halton points in comparison with random points, for the case of two dimensions and 1000 points.

INSERT FIGURE 1 ABOUT HERE

In the next two subsections, the generation of the Halton and Sobol sequences is briefly explained. Both sequences belong to the family of p-adic expansion of integers (see later).

In the case of Halton sequences, prime numbers and some mathematical definitions are used to create sequences of high uniformity whereas in generating Sobol sequences, the coefficients of primitive polynomials are used with a recursive formula in order to generate the points.

2.1 Halton sequences

Let p be a fixed prime number. Then any positive integer r can be uniquely written as its p-adic expansion in the form

$$r = \sum_{i=0}^{m} a_i p^i \qquad a_i \in \{0, \ldots, p-1\} \quad i = 0, \ldots, m \ . \tag{2}$$

The rth number of the one-dimensional Halton sequence is defined by

$$y_r = \sum_{i=0}^{m} \frac{a_i}{p^{i+1}} \ . \tag{3}$$

The d-dimensional Halton sequence is generated taking d different prime numbers (usually the first d) and putting together the resulting d one-dimensional sequences. Table 1 shows how to obtain the first 5 points using the values $p = 2, 3$. Note that by construction, all the resulting Halton points y_r lie in the interval $(0, 1)$. More details can be found in Halton (1960).

INSERT TABLE 1 ABOUT HERE

2.2 Sobol sequences

Let $v_i = m_i 2^{-i}, i = 1, 2, \ldots,$ where m_i are odd positive integers chosen using the recursion $m_i = 2c_1 m_{i-1} \oplus 2^2 c_2 m_{i-2} \oplus \cdots \oplus 2^{p-1} c_{p-1} m_{i-p+1} \oplus 2^p m_{i-p} \oplus m_{i-p}$ according to a primitive polynomial $P(z) = z^p + c_1 z^{p-1} + \cdots + c_{p-1} z + 1$, and $\oplus$ is the addition using binary arithmetic. The rth number of the one-dimensional Sobol sequence is defined by

$$y_r = a_1 v_1 \oplus a_2 v_2 + \cdots \tag{4}$$

where $a_1, a_2, \ldots$ is the binary representation of r (see equation (2)).

Antonov and Saleev (1979) improved Sobol's original algorithm and proposed to use the following scheme

$$y_r = g_1 v_1 \oplus g_2 v_2 \oplus \cdots \tag{5}$$

where $\ldots g_3 g_2 g_1$ is the Grade code representation of r defined by

$$G(r) = r \oplus \lfloor r/2 \rfloor \tag{6}$$

$\lfloor x \rfloor$ being the largest integer smaller than or equal to x. Combining (5) and (6), the rth term of the Sobol sequence can be obtained as

$$y_r = y_{r-1} \oplus v_c \tag{7}$$

where v_c is the v_i number associated with the rightmost zero in the binary representation of $r - 1$. If no zeroes appear, a leading zero must be added.

As an example consider the primitive polynomial $P(z) = z^3 + z + 1$ and initial values $m_1 = 1, m_2 = 3$, and $m_3 = 7$. The corresponding recurrence is $m_i = 4m_{i-2} \oplus 8m_{i-3} \oplus m_{i-3}$. Table 2 shows how to obtain the values m_4 and m_5.

INSERT TABLE 2 ABOUT HERE

To obtain the v_i, the m_i must be first written in binary form and shifted i positions to the left of the fractional point. Table 3 shows how to do this.

INSERT TABLE 3 ABOUT HERE

The first 5 Sobol points are obtained as follows:

Let $y_0 = 0$ (initial value), using (7) we have

$$y_1 = y_0 \oplus v_1 = 0.1 = 1 \times 2^{-1} = 0.500$$

$$y_2 = y_1 \oplus v_2 = 0.1 \oplus 0.11 = 0.01 = 0 \times 2^{-1} + 1 \times 2^{-2} = 0.250$$

$$y_3 = y_2 \oplus v_1 = 0.01 \oplus 0.10 = 0.11 = 1 \times 2^{-1} + 1 \times 2^{-2} = 0.750$$

$$y_4 = y_3 \oplus v_3 = 0.11 \oplus 0.111 = 0.001 = 0 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} = 0.125$$

$$y_5 = y_4 \oplus v_2 = 0.001 \oplus 0.11 = 0.111 = 1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3} = 0.875$$

In the first line we added v_1 because in the binary representation of $y_0 = 0$ the rightmost 0 is in the first position. In the second, we added v_2 because 1 is 1 in binary representation, so we have to add a leading 0 and it is in the second position. In the third we added v_1 because 2 in binary is 01 and then the rightmost 0 is in the first position, and so on.

The d-dimensional Sobol sequence is obtained considering d primitive polynomials and putting together the corresponding one-dimensional sequences generated with polynomial P_i, $i = 1, \ldots, d$. Note that like in the Halton sequence, all the values y_r lie in the unit interval. More details and an implementation of the Sobol sequence can be found in Bratley et al. (1988).

3 The logistic-normal model

In many of the applications mentioned earlier, the integrand of interest is the product of a function $f(\cdot)$ times the normal distribution. This kind of integral has been studied before for instance by Crouch and Spiegelman (1990) who compare their proposed method with the GH. In this section, we discuss more in detail the logistic-normal model as it is common in biostatistics and psychometrics.

Let Y_{ij} and $\boldsymbol{x}'_{ij} = (1, x_{1ij}, \ldots, x_{pij})$ be the binary outcome variable and covariate vector for observation j ($j = 1, \ldots, k$) in cluster i ($i = 1, \ldots, n$), respectively. A commonly used $f(\cdot)$ function in this framework is $f(t) = \{1 + \exp(-t)\}^{-1}$ leading to a logistic-normal model (Agresti, 2002, p.496; Hosmer and Lemeshow, 2000 p. 310) of the form

$$Pr(Y_{ij} = 1 \mid \boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\theta}_i) = f(\boldsymbol{z}'_{ij}\boldsymbol{\theta}_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta}) \tag{8}$$

which models the probability of observation j in cluster i having a certain characteristic ($Y_{ij} = 1$). $\boldsymbol{\beta}$ is a k-dimensional vector of fixed effects with associated covariate vector $\boldsymbol{x}_{ij}$, and $\boldsymbol{\theta}_i$ is a d-dimensional vector of random effects for cluster i with associated covariate vector $\boldsymbol{z}_{ij}$ (see Rijmen et al., 2003). To formulate a likelihood function for model (8), it is assumed that $\boldsymbol{\theta}_i$ follows a multivariate normal distribution with $\mathbf{0}$ mean and covariance matrix $\boldsymbol{\Sigma}$ of the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{\theta_1} & Cov(\theta_1, \theta_2) & \cdots & Cov(\theta_1, \theta_d) \\ Cov(\theta_2, \theta_1) & \sigma^2_{\theta_2} & \cdots & Cov(\theta_2, \theta_d) \\ \vdots & \vdots & \cdots & \vdots \\ Cov(\theta_d, \theta_1) & \cdots & \cdots & \sigma^2_{\theta_d} \end{pmatrix}.$$

The contribution of cluster i to the likelihood function is

$$Pr(\boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{k} \frac{\exp\{y_{ij}(\boldsymbol{z}'_{ij}\boldsymbol{\theta}_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta})\}}{1 + \exp\{\boldsymbol{z}'_{ij}\boldsymbol{\theta}_i + \boldsymbol{x}'_{ij}\boldsymbol{\beta}\}} N(\boldsymbol{\theta}_i; \mathbf{0}, \boldsymbol{\Sigma}) d\theta_{i1} \cdots d\theta_{id} \tag{9}$$

where $N(\boldsymbol{\theta}_i; \mathbf{0}, \boldsymbol{\Sigma})$ means that the vector $\boldsymbol{\theta}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. The full likelihood is then the product of (9) over the n clusters.

In equation (9), the function $f(\cdot)$ is a product of logistic functions, because of the repeated nature of the observations ($j = 1, \ldots, k$) in cluster i. In order to estimate $\boldsymbol{\beta}$ and

$\boldsymbol{\Sigma}$, the full likelihood must be maximized. An approximation of the integral in (9) is then necessary in each step of the maximization algorithm.

In the following we will first concentrate on $f(\cdot)$ with a single logistic function, and afterwards also $f(\cdot)$ with a product of logistic functions will be investigated.

4 Evaluation of the methods

Three methods of integration will be compared: GH, MC, and QMC using Sobol and Halton sequences, for a general type of logistic-normal integral for which a very close approximate value can be derived. The comparison focuses on the number of points used to evaluate (9) for different values of d, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, and the precision obtained. In this way, it can be evaluated whether for higher dimensionality the QMC method can do better with the same number of points, or equally well with less points than GH and MC, the two more common methods. As our interest is in the integration process, rather than the optimization, we will do the calculations just for one cluster, hence $n = 1$. Therefore, we will not calculate the product of n integrals, but only the integral over a product of k terms.

Let T denote the true but unknown value of the integral we are interested in, for a particular combination of d, β and $\boldsymbol{\Sigma}$. To evaluate our approaches, we need to know the 'exact' value of the integral T. This value can be approximated very accurately by using the fact that the sum of random variables with a normal distribution is normally distributed as well. Indeed, note that if $\boldsymbol{\theta} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$ then

$$W = \sum_{q=1}^{d} \theta_q \sim N\left(0, \underbrace{\sum_{q=1}^{d} \sigma_{\theta_q}^2 + 2\sum_{q<q'} Cov(\theta_q, \theta_{q'})}_{\sigma_z^2}\right) . \tag{10}$$

So, T can be reduced to

$$\int_{-\infty}^{\infty} \frac{\exp(w)}{1+\exp(w)} N(w;0,\sigma_w^2)dw \ , \tag{11}$$

Although there exists no analytical solution for this one-dimensional integral, standard routines (e.g., Piessens et al. 1983) that yield highly accurate results are available.

A common problem with both the GH method and the LDS is that the standard sets of nodes and points, respectively, need to be transformed before one can use them to approximate an arbitrary integral like in equation (9). For the GH method, the transformation amounts to recentering and rescaling the nodes such that they reflect the mean vector and covariance matrix of the normal distribution. This transformation is standard and has been discussed for example by Fahrmeir and Tutz (2001, p. 447-449).

For the LDS (but also for MC), the situation is straightforward as well. LDS are defined in the unit hypercube, and in order to evaluate (9) using QMC and MC, we first need to transform the integration domain $(0,1)^d$ into $\mathbb{R}^d$. We chose the inverse normal transformation and obtained the points used to evaluate the integrals and the corresponding approximations in the following way:

i) *Change of integration region:* Draw a matrix $P_{N\times d}$ of LDS or MC points and create a matrix $X_{N\times d} = \Phi^{-1}(p_{ij})$, where N is the total number of points to be used.

ii) *Adding correlation structure:* Put $Y = L'X'$, where L is the Cholesky decomposition of the covariance matrix, $\Sigma = L'L$.

iii) *Obtaining the approximation:* $\widehat{T} = \frac{1}{N}\sum_{i=1}^{N} f(y_1^i, \ldots, y_d^i)$.

Here $p_{ij} \in (0,1)$ are the elements of the matrix P, and $\Phi(\cdot)$ is the cumulative normal distribution function.

To be able to compare how well the methods perform, we use as a dependent variable, the relative error (RE),

$$RE = \left| \frac{\widehat{T} - T}{T} \right| , \tag{12}$$

where $\widehat{T}$ is the approximated value of integral using one of the three methods to be compared

4.1 Study 1: An integral with a single logistic function

We will consider first the simplest case with one observation per cluster, $k = 1$. Suppose that $y = 1$, so that $\boldsymbol{z}'$ is a vector of d ones, and $x = -1$, then (9) reduces to

$$T = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp\{\theta_1 + \cdots + \theta_d - \beta\}}{1 + \exp\{\theta_1 + \cdots + \cdots \theta_d - \beta\}} N(\boldsymbol{\theta}; \mathbf{0}, \Sigma) d\theta_1 \cdots, d\theta_d \tag{13}$$

4.1.1 Design of the study

We evaluated the cases $d = 1, 5, 7, 9$, and $\beta = 0, 1, 2$. When $d = 1$, $\Sigma = \sigma^2_{\theta_1}$ was fixed to 1. For $d > 1$, the different Σ matrices were chosen using the values showed in Table 4 for each dimension respectively. The value ρ refers to correlations implied in Σ, to be crossed with the values of $\sigma^2_{\theta_q}$ for $d = 5, 7, 9$.

INSERT TABLE 4 ABOUT HERE

The total number of points used to evaluate (13) in each problem with dimension d was $N = 5^d$. As mentioned before, the number of quadrature points for the GH method increases exponentially with the number of dimensions, so we chose this number of points to make the calculations feasible, mainly for the GH method. Thus, for instance in dimension 7, $5^7 = 78125$ points were used to evaluate the integral, and $5^9 = 1953125$ in $d = 9$. For the MC method we used as an estimate the average of $S = 10$ integrals obtained from samples of size 5^d. An estimate of the standard error (SE) can be calculated as

$$SE(\widehat{I}) = \sqrt{\frac{1}{S^2} \sum_{s=1}^{S} Var(\tilde{I}_s)}, \qquad \widehat{I} = \frac{\sum_{s=1}^{S} \tilde{I}_s}{S} \tag{14}$$

where $\tilde{I}_s$ is the integral for sample s evaluated using MC and

$$\sqrt{Var(\tilde{I}_s)} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum_{i=1}^{n}(h(x_i) - \tilde{I}_s)^2}{n-1}} \tag{15}$$

is the regular MC standard error, with $h(\cdot)$ the integrand of $\tilde{I}_s$ (see Tanner, 1996).

The choice of $S = 10$ is rather arbitrary. We chose to do more than one MC sample in order to calculate a SE. However, in comparison with the other methods, the MC method is favored in this way.

4.1.2 Results of study 1

Table 5 shows in the first row for each value of d and for each method the estimated value of the integral, and in the second row the associated relative error. In the case of MC the standard error (SE) is reported as well. Figure 2 shows the results reported in Table 5 using plots of the RE in $\log_{10}$ scale. The log-plot in connection with the relative error is a measure of how many digits of accuracy one has.

INSERT TABLE 5 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

From Table 5, and from Figure 2 it can be seen that QMC was better than GH in all the cells in which the following conditions are fulfilled: $d > 1$, $\rho \geq 0.3$, and $\beta \geq 1$, except in the cells with $d = 5$ and $\rho = 0.3$. On the other hand MC was better than GH in all the above mentioned cells except in the cells with $d = 5, \rho = 0.5, \beta = 1, 2$; and with $d = 7, \rho = 0.3, \beta = 1$. Note that the Sobol results were often better than Halton, and that in general QMC was better than MC. From these results it is clear that there is room for improvement on GH for higher dimensionalities, high correlations and on the condition that the distributions are shifted ($\beta \geq 1$). In general, the RE for GH increases with the correlation (ρ) and with β, and the effect of the latter is very large. Given these results, there is some room left for both the MC and QMC methods to outperform the GH method, especially when the latter is not centered on the distribution.

Study one is limited in several ways. First of all, note that the comparisons are based on fixed numbers of points corresponding to the values of 5^d, whereas in fact for the MC and QMC methods, any number of points less than or equal to 5^d can be used. In part of the next study, we will investigate the LDS integral approximation for all points in between. Second, in study one $k = 1$, whereas the case we are interested in is one with repeated observations. Therefore a larger value of k will be chosen in the next study.

4.2 Study 2: An integral with a product of logistic functions

The more general case for the kind of logistic model we are interested in is one with $k > 1$, meaning that repeated observations are made for the clusters. We will consider here a case with $k = 3$, but again only one cluster i will be considered.

Suppose that $\boldsymbol{z}'$ is a d-dimensional vector of ones and that for each of the three observations ($k = 3$) we have an indicator variable, $x_{ij} = -1$ for $i = j$ and $x_{ij} = 0$ for $i \neq j$, and no intercept (fixed intercept is zero). Assume that the observations $\boldsymbol{y}_k = (1, 1, 0)$, are made, then, the estimation process involves the following integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\exp\{\sum_{q=1}^{d} \theta_q - \beta_1\}}{1 + \exp\{\sum_{q=1}^{d} \theta_q - \beta_1\}} \right) \left(\frac{\exp\{\sum_{q=1}^{d} \theta_q - \beta_2\}}{1 + \exp\{\sum_{q=1}^{d} \theta_q - \beta_2\}} \right) \times \left(\frac{1}{1 + \exp\{\sum_{q=1}^{d} \theta_q - \beta_3\}} \right) N(\boldsymbol{\theta}; \mathbf{0}, \boldsymbol{\Sigma}) d\theta_1 \cdots d\theta_d \ . \tag{16}$$

4.2.1 Design of the study

We evaluated (16) for $\boldsymbol{\beta} = (0, 1, 2)$ and different values of d and Σ, see Table 6. In addition we varied the number of points used to evaluate (16) and computed the relative error as in (12). The number of points for the multidimensional case will not be limited to powers of quadrature points used per dimension, but for the best performing QMC method all the numbers in between will be used as well.

INSERT TABLE 6 ABOUT HERE

4.2.2 Results study 2

Figure 3 shows the RE versus the number of points used to evaluate (16) in one dimension, $d = 1$, and an increasing number of points from 1 to 20. From the figure several conclusions may be drawn. First, the GH method was always better than the other methods, and, generally, the QMC method seems to be slightly better than the regular MC. Second, an increasing number of points utilized to evaluate the integrand, does not necessarily yield a more accurate result. For example, using 7 GH points per dimension yields a better result than using 8, 9, and 10 points. For the MC and QMC the behavior of the curve is very irregular and not monotonic at all. For example using 3 Halton points seems better than using 4, 5, and 6 Halton points. However, also the GH curve is somewhat irregular and non-monotonic.

INSERT FIGURE 3 ABOUT HERE

Let us now consider more than one dimension. For zero correlation, and $d = 5$, the results are very similar to those for $d = 1$. Figure 4 shows first that again GH is better than the other methods, and that QMC is better than MC, and second that the shape of the curves is again irregular. When $d > 1$, note that the symbols $\circ, \Box, *, \diamond$ in the plots refer to the abscissa values $1^d, 2^d, \ldots, \ell^d$ indicating that 1 to ℓ quadrature points were used per dimension. For instance if $d = 5$ the symbols refer to $1^5, 2^5, \ldots, 10^5$, as number of points used. For example, the eighth symbol refers to $8^5 = 32768$ points (8 quadrature points per dimension) in total.

INSERT FIGURE 4 ABOUT HERE

To investigate the irregularities more in detail, we computed the approximation with 1 up to 10^5 Sobol points, because the Sobol points gave slightly better results than the Halton points. The resulting plot is shown in Figure 5. From the figure one can see that the result is highly variable depending on the number of points. For *the same number of points* the Sobol is sometimes better than GH, but most often GH is better. The points indicated with arrows in Figure 5 show that using 27021 or even 4486 Sobol points, one can obtain a result as accurate as using 8 GH quadrature points per dimension (i.e. $8^5 = 32768$ points). From these results, one cannot derive a general rule or not even a hint for when the Sobol points do better.

INSERT FIGURE 5 ABOUT HERE

Now consider a distribution with five dimensions and $\rho = 0.5$. The results are shown in Figure 6. First note that the GH method is still the better of the three. However when considering correlated dimensions, GH loses accuracy which is reflected in the fact that now the GH curve decreases more slowly. For example, with 10 quadrature points per dimension we obtain approximately 4 digits of precision, which is clearly less than in the uncorrelated case. Again QMC is better than MC but it is not clear whether Sobol is better than Halton. Note also that again Sobol can be much better than GH and for a much smaller number of points. For instance, an arrow in the figure points to the Sobol result for 1247 points which is much better than any of the GH results.

INSERT FIGURE 6 ABOUT HERE

Finally, we evaluated a 10-dimensional integral for both uncorrelated ($\rho = 0$) and correlated ($\rho = 0.5$) dimensions, using 1 to 3 points per dimension. Figures 7 and 8, using $\log_{10}$ scale in both x and y axes, show the results. Note that once more GH looses precision when correlated dimensions are considered. There are again a few excellent results obtained with Sobol points, for example, 527 Sobol points do extremely well, and much better than much larger numbers of GH points. In the correlated case, the MC method is better than all other methods. The perhaps surprisingly good performance of the MC method may be attributed to the fact that ten samples were used each time. In order to check the difference this can make, we also calculated the RE for each of the ten samples (RE10s) and its corresponding average (AvRE), when using $1^{10}, 2^{10}$, and 3^{10} points, see Figure 8. It can be seen that if one would like to compare the methods with the same number of points (i.e., just one MC sample), the results of the MC method are inferior to those of the other methods from 2^{10} points on.

INSERT FIGURE 7 ABOUT HERE

INSERT FIGURE 8 ABOUT HERE

5 Discussion and conclusion

A comparison between three numerical procedures (Gauss-Hermite quadrature, Monte Carlo and Quasi-Monte Carlo integration) to approximate analytically intractable integrals has been presented. Low discrepancy sequences and its use in QMC were discussed for the evaluation of logistic-normal integrals.

We have found that in most of the examples shown, the Halton and Sobol methods can do better than the GH method with less points, on the condition that the appropriate number of points are selected. Unfortunately, when using the QMC method, there is no clear rule for selecting an appropriate number of points to evaluate the integrals. In this context, constructing an automated algorithm to cleverly select the points for a given problem is a challenge.

Because of the non-monotonic relation between number of points and accuracy, it follows that a larger number of quadrature points to approximate the integral, does not necessarily give more accurate results for the Sobol and Halton sequences. Moreover and perhaps surprisingly, the same phenomenon also applies to the GH method when used for evaluating logistic-normal integrals.

An advantage of QMC or MC compared to GH in higher dimensions is that one can use any numbers of points while the GH method is restricted to the use of ℓ^d ($\ell = 1, 2, 3, \dots$) number of points. For instance using GH in dimension 5 with 3 quadrature points per dimension ($3^5 = 243$ points in total) one cannot use 242, 241, 240, etc. points. For GH, there is no way of reducing the number of points by units.

Apart from this a priori advantage, it appears that the QMC and MC beat the GH method in higher dimensions when the distributions are shifted, and when also the dimensions are correlated. Not only do they have more precision, but the same precision as that of the GH can be reached with less (and often much less) quadrature points. These results seems independent from the non-monotonic relation between number of points and accuracy.

In both studies we used the inverse normal transformation to change the region of integration from the unit hypercube in $\mathbb{R}^d$. However, other transformation could be used as well. As a matter of fact, we have replicated the analysis reported in this paper using a logit transformation, which gave rather poor results in comparison with the inverse normal transformation (the results are available from the authors upon request). Robert and Casella (2004, p. 77) point out that one must be careful in the choice of the transformation as it could be crucial for the efficiency of the method. This suggestion is supported by our results.

Finally, we focused on the approximation of intractable integrals, rather than on the optimization problem yielding the estimation of the parameters of a model. For the optimization task, the QMC seems to be a promising method to approximate the integral, in the process of maximizing the likelihood, especially for the case of correlated dimensions and shifted distributions. One may consider to use a mixed method, GH combined with QMC for the kind of problems where GH seems less accurate or needs too many points.

References

- Abramovitz, M., Stegun, I., (eds) 1972. *Handbook of Mathematical Functions*. Dover Publications Inc, New York.
- Agresti, A., 2002. Categorical data analysis (2nd ed.). N.Y.: Wiley.
- Antonov, I., Saleev, V., 1979. An economic method of computing LP_τ-sequences. *USSR computational mathematics and mathematical physics*. 19, 252-256.
- Bratley, P., Bennett L. Fox, 1988. Algorithm 659 Implementing Sobol's Quasirandom sequence generator. *ACM Transactions on Mathematical Software*, 14, 88-100.
- Caflisch, R., 1998. Monte Carlo and quasi-Monte Carlo methods. Acta Numerica, 7, 1-49.
- Cools, R. 1997. Constructing cubature formulae: the science behind the art. Acta Numerica 6, 1-54. Cambridge University Press.
- Cools, R., 2002. Advances in multidimensional integration. *Journal of Computational and Applied Mathematics*, 149, 1-12.
- Crouch, A., Spiegelman, E., 1990. The evaluations of integrals of the form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)dt$: application to logistic-normal models. *Journal of the American Statistical Association*, 85, 464-469.
- De Boeck, P., Wilson, M., 2004. Explanatory item response models: A generalized linear and nonlinear approach. N.Y: Springer-Verlag.
- Davis, P., Rabinowitz, P., 1984. Methods of numerical integration. Academic Press, Orlando, Florida, second edition.
- Fahrmeir, L., Tutz, G., 2001. Multivariate statistical modelling based on generalized linear models (2nd ed.). N.Y: Springer-Verlag.

- Fischer, G., Molenaar, I. (Eds.), 1995. Rasch models: Foundations and recent developments. N.Y: Springer-Verlag.
- Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik, 2, 84-90.
- Hosmer, D., Lemeshow, S., 2000. Applied logistic regression (2nd ed.), N.Y.: Wiley.
- Jank, W., 2005. Quasi-Monte Carlo sampling to improve the efficiency of Monte Carlo EM. *Computational Statistics and Data Analysis*, 48, 685-701.
- Judd, K., 1998. Numerical methods in economics. Boston: MIT Press.
- Kocis, L., Withen, W., 1997. Computational investigations of Low-Discrepancy sequences. *ACM Transactions on Mathematical Software*, 23, 266-294.
- Lemieux, C., L'Ecuyer, P., 2001. On the use of quasi-Monte Carlo methods in computational finance. Computational Science – ICCS 2001, Lecture Notes in Computer Science, Springer-Verlag, 607-616. Available in

  `http://www.iro.umontreal.ca/~lecuyer/myftp/papers/iccs01.pdf`
- Lesaffre, E., Spiessens, B., 2001. On the effect of the number of quadrature points in a logistic random-effects model: An example. *Applied Statistics*, 50, 325-335.
- Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods (CBMS-NSF Regional Conference Series in Applied Mathematics, No 63).
- Morokoff, W. and Caflisch, R., 1995. Quasi-Monte Carlo integration. *Journal of Computational Physics*, 122, 218-230.
- Paskov, S., 1995. Faster valuation of financial derivatives. *Journal of Portfolio Management*, 22, 113-120.
- Piessens, R., de Doncker-Kapenga, E., Uberhuber, C., Kahaner, D., 1983. QUADPACK: A subroutine package for automatic integration. Springer Verlag.

- Robert, C., Casella, G., 2004. Monte Carlo statistical methods (2nd ed.). N.Y.: Springer-Verlag.
- Rijmen, F., Tuerlinckx, F., De Boeck, P., Kuppens, P. (2003). A nonlinear mixed model framework for IRT models. *Psychological Methods*, 8, 185-205.
- Shao, J., 2003. Mathematical Statistics (2nd. ed.). N.Y.: Springer-Verlag.
- Stroud, A., 1971. Approximate calculation of multiple integrals. Englewood Cliffs, NJ. Prentice Hall.
- Tanner, M., 1996. Tools for statistical inference (3rd ed.). N.Y.: Springer-Verlag.

$p=2$			$p=3$		
r	$\sum_{i=0}^{m} a_i p^i$	y_r	r	$\sum_{i=0}^{m} a_i p^i$	y_r
1	$1 \times 2^0 + 0 \times 2^1$	0.500	1	1×3^0	0.333
2	$0 \times 2^0 + 1 \times 2^1$	0.250	2	2×3^0	0.667
3	$1 \times 2^0 + 1 \times 2^1$	0.750	3	$0 \times 3^0 + 1 \times 3^1$	0.111
4	$0 \times 2^0 + 0 \times 2^1 + 1 \times 2^2$	0.125	4	$1 \times 3^0 + 1 \times 3^1$	0.444
5	$1 \times 2^0 + 0 \times 2^1 + 1 \times 2^2$	0.625	5	$2 \times 3^0 + 1 \times 3^1$	0.778

Table 1: First 5 Halton points using the primes $p = 2, 3$

i	m_i	$m_i = 4m_{i-2} \oplus 8m_{i-3} \oplus m_{i-3}$	Binary sum	Decimal representation
1	1			
2	3			
3	7			
4	5	$12 \oplus 8 \oplus 1$	$1100 \oplus 1000 \oplus 0001 = 0101$	5
5	7	$28 \oplus 24 \oplus 3$	$11100 \oplus 11000 \oplus 00011 = 00111$	7

Table 2: Generation of the m_i values for Sobol sequences (The first three values are given)

i	m_i	Binary representation	v_i
1	1	1	0.1
2	3	11	0.11
3	7	111	0.111
4	5	101	0.0101
5	7	111	0.00111

Table 3: Generation of the v_i values for Sobol sequences

d	$\sigma^2_{\theta_1}, \ldots, \sigma^2_{\theta_d}$
5	(2.0, 1.5, 1.0, 0.5, 0.3)
7	(2.0, 1.5, 1.0, 0.5, 0.3, 0.6, 0.7)
9	(2.0, 1.5, 1.0, 0.5, 0.3, 0.6, 0.7, 0.1, 0.25)
ρ	0.0, 0.3, 0.5, 0.7, 0.9

Table 4: Values to form the Σ matrices used in Study 1

Dimension $d=1$															
							$\rho=0$								
							$\beta=0$	$\beta=1$	$\beta=2$						
GH							5.00E-01	3.03E-01	1.56E-01						
RE							3.05E-10	1.56E-04	2.59E-04						
Sobol							5.36E-01	3.19E-01	1.57E-01						
RE							7.22E-02	5.17E-02	6.93E-03						
Halton							4.64E-01	2.57E-01	1.18E-01						
RE							7.22E-02	1.52E-01	2.39E-01						
MC							5.14E-01	2.85E-01	1.48E-01						
RE							2.74E-02	6.08E-02	4.52E-02						
(se)							2.95E-02	2.31E-02	1.88E-02						
Dim 5															
	$\rho=0$			$\rho=0.3$			$\rho=0.5$			$\rho=0.7$			$\rho=0.9$		
	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$
GH	5.00E-01	3.64E-01	2.44E-01	5.00E-01	3.94E-01	2.96E-01	5.00E-01	4.05E-01	3.16E-01	5.00E-01	4.09E-01	3.26E-01	5.00E-01	3.98E-01	3.17E-01
RE	2.56E-09	3.54E-05	2.30E-05	2.55E-09	7.63E-04	8.02E-04	2.56E-09	3.44E-03	4.66E-03	2.54E-09	1.37E-02	2.06E-02	2.55E-09	5.42E-02	8.25E-02
Sobol	4.99E-01	3.64E-01	2.44E-01	4.99E-01	3.94E-01	2.96E-01	5.00E-01	4.06E-01	3.17E-01	4.99E-01	4.14E-01	3.33E-01	4.99E-01	4.21E-01	3.45E-01
RE	1.22E-03	1.00E-03	1.41E-03	1.19E-03	1.13E-03	1.31E-03	9.75E-04	1.39E-03	1.83E-03	1.15E-03	1.35E-03	1.75E-03	1.16E-03	1.26E-03	7.96E-04
Halton	4.98E-01	3.62E-01	2.42E-01	4.98E-01	3.93E-01	2.95E-01	4.99E-01	4.05E-01	3.16E-01	4.99E-01	4.14E-01	3.32E-01	4.99E-01	4.20E-01	3.44E-01
RE	4.16E-03	5.71E-03	7.14E-03	3.37E-03	4.49E-03	6.08E-03	1.97E-03	2.87E-03	4.84E-03	1.28E-03	2.22E-03	4.04E-03	1.65E-03	2.17E-03	3.14E-03
MC	5.00E-01	3.65E-01	2.44E-01	4.99E-01	3.97E-01	2.94E-01	5.01E-01	4.12E-01	3.21E-01	5.02E-01	4.14E-01	3.33E-01	5.03E-01	4.20E-01	3.46E-01
RE	1.41E-04	8.59E-04	9.28E-04	1.90E-03	5.40E-03	9.22E-03	1.12E-03	1.32E-02	9.33E-03	4.36E-03	2.28E-03	3.91E-04	5.31E-03	2.59E-03	3.09E-03
(se)	1.88E-03	1.80E-03	1.56E-03	2.14E-03	2.10E-03	1.93E-03	2.23E-03	2.18E-03	2.06E-03	2.29E-03	2.25E-03	2.14E-03	2.34E-03	2.30E-03	2.21E-03
Dim 7															
	$\rho=0$			$\rho=0.3$			$\rho=0.5$			$\rho=0.7$			$\rho=0.9$		
	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$
GH	5.00E-01	3.74E-01	2.60E-01	5.00E-01	4.12E-01	3.29E-01	5.00E-01	4.23E-01	3.50E-01	5.00E-01	4.27E-01	3.59E-01	5.00E-01	4.13E-01	3.39E-01
RE	3.94E-09	1.19E-05	1.29E-05	3.58E-09	4.61E-04	6.92E-04	3.77E-09	2.71E-03	4.52E-03	2.58E-09	1.28E-02	2.25E-02	4.95E-09	5.91E-02	1.04E-01
Sobol	5.00E-01	3.74E-01	2.60E-01	5.00E-01	4.12E-01	3.29E-01	5.00E-01	4.24E-01	3.52E-01	5.00E-01	4.33E-01	3.67E-01	5.00E-01	4.39E-01	3.79E-01
RE	4.05E-05	1.27E-05	1.04E-04	2.59E-04	3.56E-04	1.31E-04	3.08E-04	2.44E-04	1.57E-05	1.33E-04	1.21E-05	1.20E-04	5.86E-05	6.32E-05	5.95E-06
Halton	5.00E-01	3.73E-01	2.60E-01	5.00E-01	4.12E-01	3.29E-01	5.00E-01	4.24E-01	3.52E-01	5.00E-01	4.33E-01	3.67E-01	5.00E-01	4.39E-01	3.79E-01
RE	3.90E-04	7.02E-04	9.63E-04	5.00E-04	4.08E-04	2.24E-04	4.26E-04	2.97E-04	9.45E-05	1.80E-04	1.42E-04	2.19E-04	2.31E-05	1.38E-04	3.25E-04
MC	5.00E-01	3.73E-01	2.60E-01	5.00E-01	4.13E-01	3.29E-01	5.00E-01	4.25E-01	3.51E-01	5.00E-01	4.33E-01	3.67E-01	4.99E-01	4.39E-01	3.79E-01
RE	6.39E-04	4.63E-04	6.27E-04	6.57E-04	1.41E-03	4.24E-04	4.70E-04	9.01E-04	1.51E-03	5.46E-04	3.40E-04	1.38E-04	1.83E-03	1.10E-03	4.74E-05
(se)	3.94E-04	3.79E-04	3.36E-04	4.55E-04	4.47E-04	4.24E-04	4.72E-04	4.66E-04	4.49E-04	4.83E-04	4.79E-04	4.64E-04	4.91E-04	4.87E-04	4.76E-04
Dim 9															
	$\rho=0$			$\rho=0.3$			$\rho=0.5$			$\rho=0.7$			$\rho=0.9$		
	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$	$\beta=0$	$\beta=1$	$\beta=2$
GH	5.00E-01	3.76E-01	2.64E-01	5.00E-01	4.19E-01	3.42E-01	5.00E-01	4.30E-01	3.63E-01	5.00E-01	4.34E-01	3.71E-01	5.00E-01	4.19E-01	3.49E-01
RE	4.42E-09	9.24E-06	1.07E-05	4.55E-09	3.96E-04	6.42E-04	4.08E-09	2.49E-03	4.36E-03	7.00E-09	1.22E-02	2.23E-02	4.60E-09	5.90E-02	1.08E-01
Sobol	5.00E-01	3.76E-01	2.64E-01	5.00E-01	4.19E-01	3.42E-01	5.00E-01	4.31E-01	3.65E-01	5.00E-01	4.39E-01	3.80E-01	5.00E-01	4.45E-01	3.91E-01
RE	1.05E-05	1.94E-05	1.71E-05	5.10E-05	4.93E-06	6.59E-05	3.47E-05	9.26E-06	6.32E-05	9.63E-06	6.57E-06	3.13E-06	2.07E-06	4.21E-06	3.25E-06
Halton	5.00E-01	3.76E-01	2.64E-01	5.00E-01	4.19E-01	3.42E-01	5.00E-01	4.31E-01	3.65E-01	5.00E-01	4.39E-01	3.80E-01	5.00E-01	4.45E-01	3.91E-01
RE	2.33E-06	1.59E-05	8.05E-05	5.28E-05	3.61E-05	2.33E-05	4.91E-05	4.23E-05	5.94E-06	1.05E-05	1.60E-05	1.44E-05	5.91E-06	8.36E-06	1.43E-05
MC	5.00E-01	3.76E-01	2.64E-01	5.00E-01	4.19E-01	3.42E-01	5.00E-01	4.31E-01	3.65E-01	5.00E-01	4.39E-01	3.80E-01	5.00E-01	4.45E-01	3.91E-01
RE	2.33E-04	6.86E-05	5.11E-04	3.59E-04	2.21E-04	1.90E-04	1.06E-04	2.06E-04	3.82E-05	2.87E-04	4.68E-04	3.37E-05	2.09E-04	2.07E-04	1.76E-04
(se)	7.96E-05	7.66E-05	6.83E-05	9.30E-05	9.16E-05	8.77E-05	9.63E-05	9.53E-05	9.24E-05	9.84E-05	9.76E-05	9.53E-05	9.99E-05	9.92E-05	9.73E-05

Table 5: Integral value and relative error for the four methods $d = 1, 5, 7, 9$

d	$\sigma^2_{\theta_1}, \dots, \sigma^2_{\theta_d}$
1	1
5	(1, 1, 1, 1, 1)
10	(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
ρ	0.0, 0.5

Table 6: Values to form the Σ matrices used in Study 2

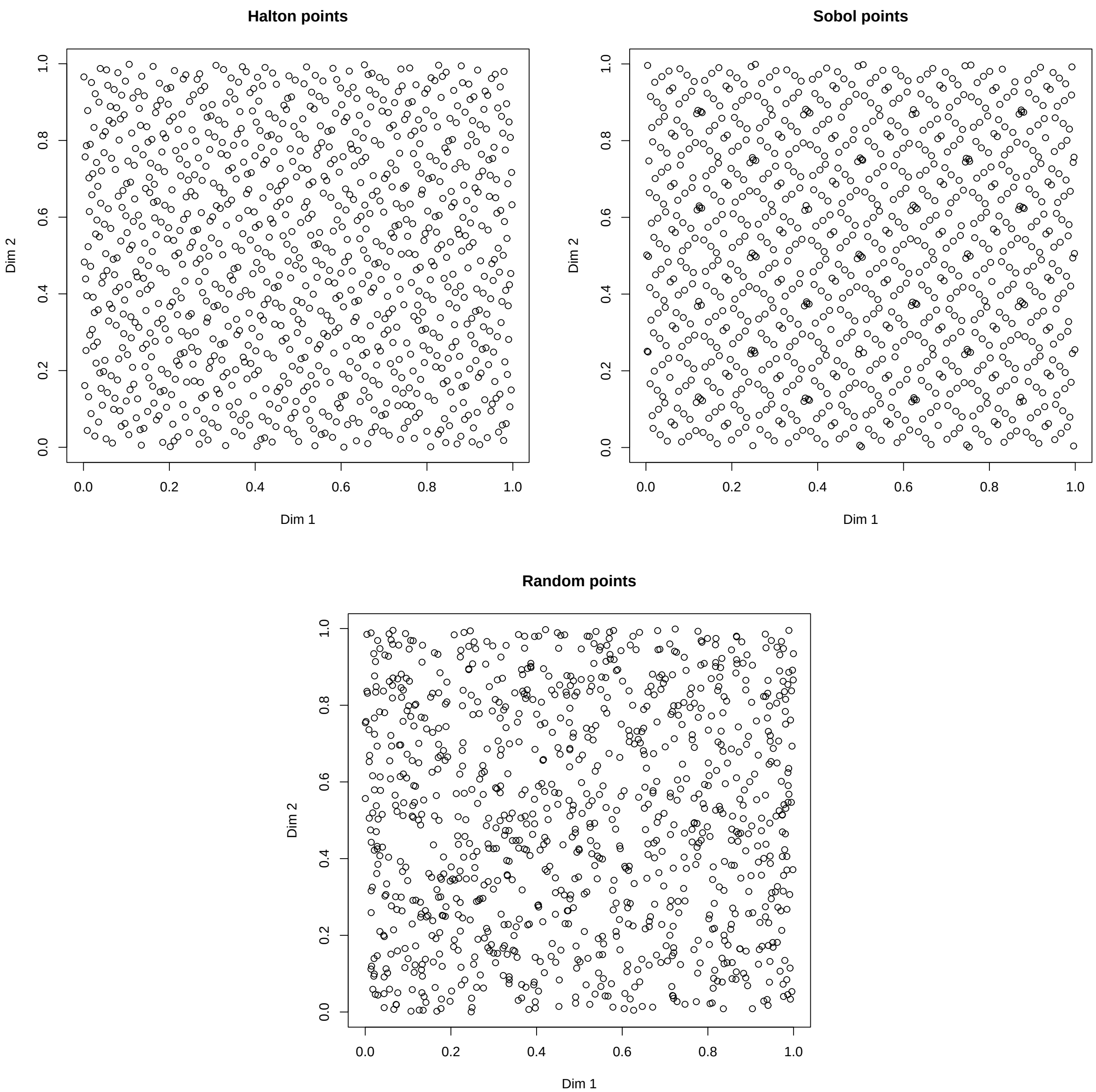

Figure 1: Two Low Discrepancy Sequences and random points

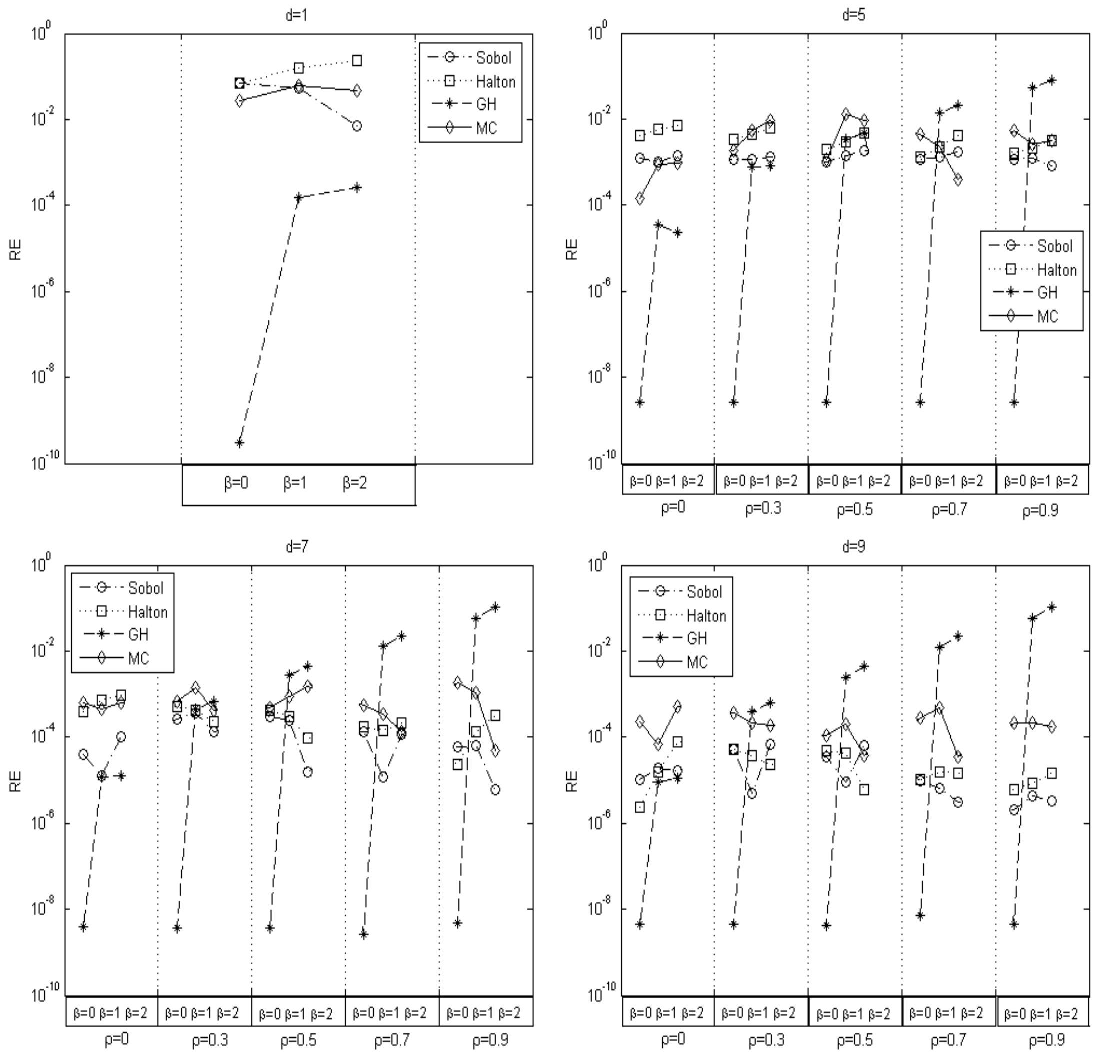

Figure 2: Results of study 1

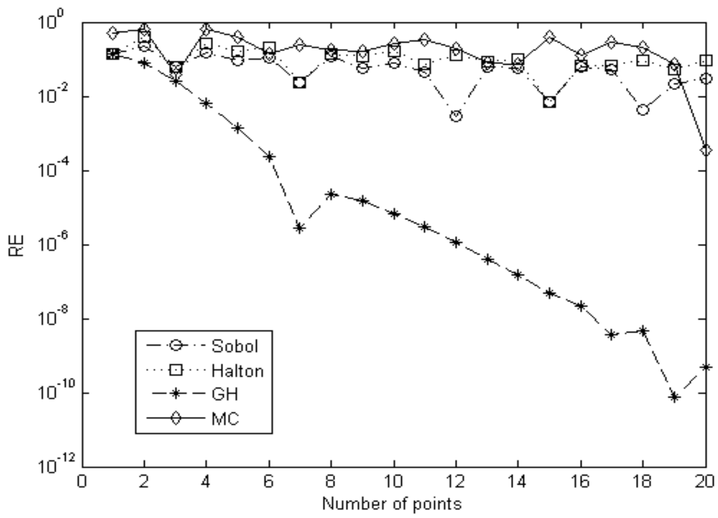

Figure 3: Relative error versus number of points used to evaluate the test integral in Study 2, $d = 1$

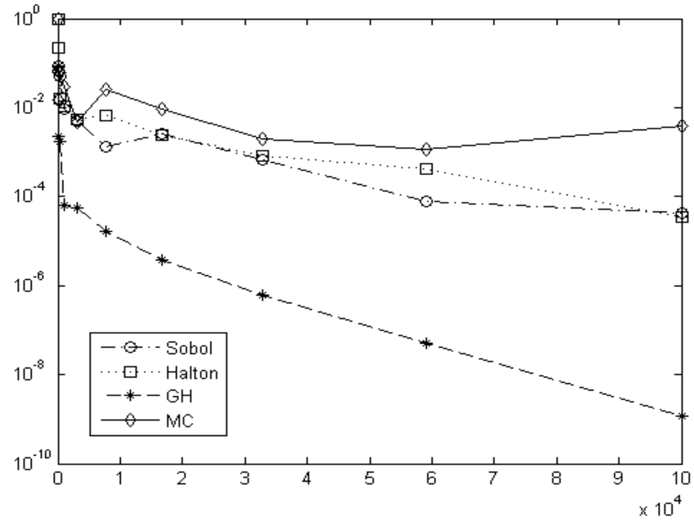

Figure 4: Relative error versus number of points used to evaluate the test integral, $d = 5$, and $\rho = 0$

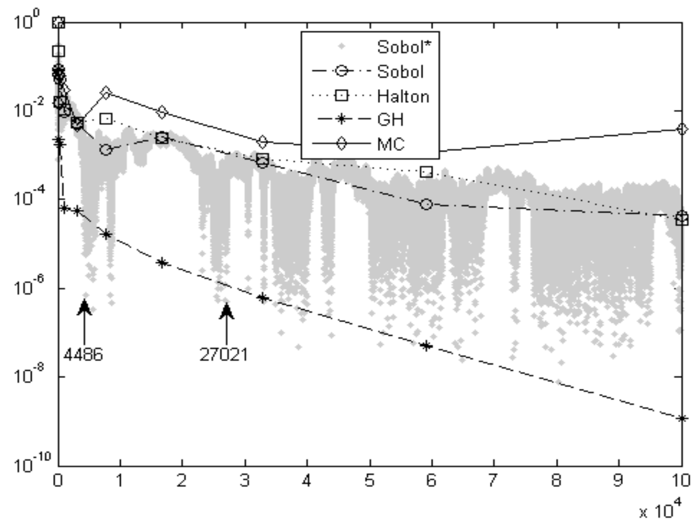

Figure 5: Relative error versus number of points used to evaluate the test integral, $d = 5$, and $\rho = 0$. *The Sobol results are shown for 1 to 10^5 points.

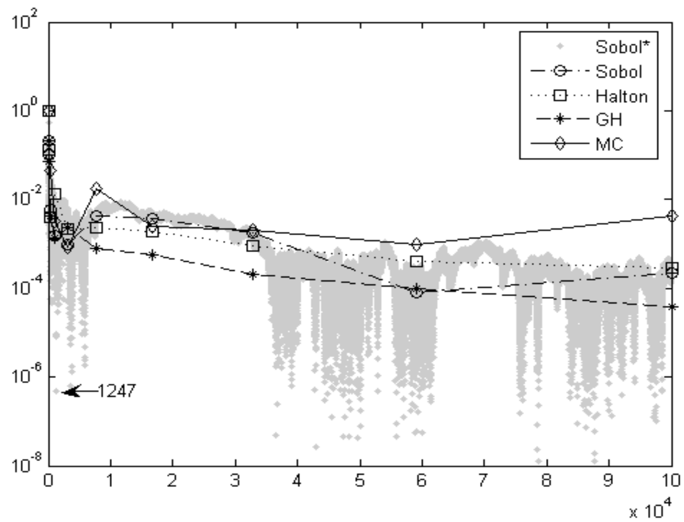

Figure 6: Relative error versus number of points used to evaluate the test integral, $d = 5$, and $\rho = 0.5$. *The Sobol results are shown for 1 to 10^5 points.

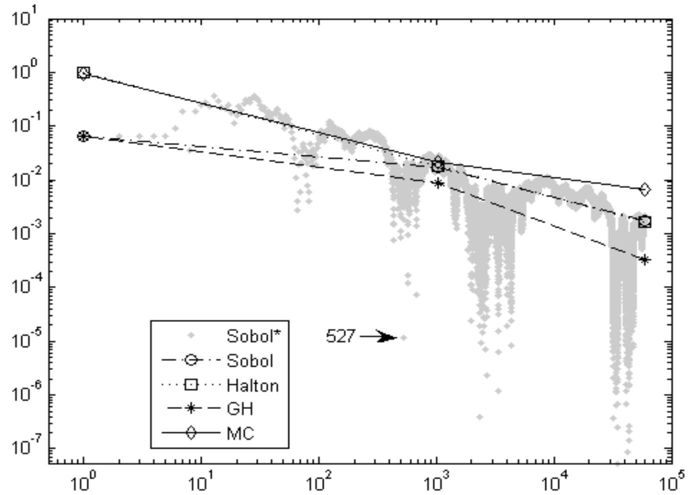

Figure 7: Relative error versus number of points used to evaluate the test integral, $d = 10$, and $\rho = 0$. *The Sobol results are shown for 1 to 3^{10} points.

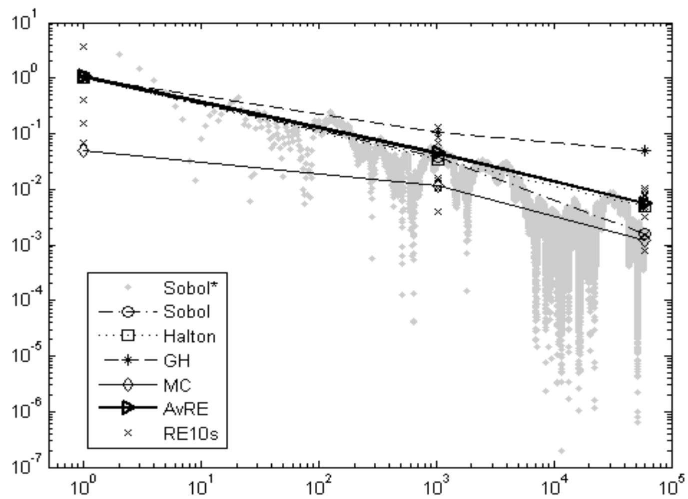

Figure 8: Relative error versus number of points used to evaluate the test integral, $d = 10$, $\rho = 0.5$. *The Sobol results are shown for 1 to 3^{10} points.