Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 1535–1548
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Numerical integration in logistic-normal models

Jorge González[a,*], Francis Tuerlinckx[a], Paul De Boeck[a], Ronald Cools[b]

[a]*Department of Psychology, K.U. Leuven, Tiensestraat 102, B-3000 Leuven, Belgium*
[b]*Department of Computer Science, K.U. Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium*

Received 27 July 2005; received in revised form 5 May 2006; accepted 6 May 2006
Available online 2 June 2006

Abstract

Marginal maximum likelihood estimation is commonly used to estimate logistic-normal models. In this approach, the contribution of random effects to the likelihood is represented as an intractable integral over their distribution. Thus, numerical methods such as Gauss–Hermite quadrature (GH) are needed. However, as the dimensionality increases, the number of quadrature points becomes rapidly 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. A comparison between three integration methods for logistic-normal models: GH, QMC, and full Monte Carlo integration (MC) is presented. It turns out that, under certain conditions, the QMC and MC method perform better than the GH in terms of accuracy and computing time.
© 2006 Elsevier B.V. All rights reserved.

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 (e.g., Lesaffre and Spiessens, 2001), 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 contribution of random effects to the likelihood is represented as an integral over their distribution, 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 Casella, 2004, p. 83).

* Corresponding author. Tel.: +32 16 325909; fax: +32 16 325916.
E-mail address: jorge.gonzalez@psy.kuleuven.be (J. González).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.05.003

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. For a recent survey of QMC aimed at a statistical readership, see Hickernell et al. (2005).

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 MC EM algorithm based on QMC and applied it to a geostatistical problem in which an integral of dimensionality 16 had to be solved because the outcome covaries with 16 distinct geographical location variables.

The main aim of this paper is to compare the performance of three numerical integration methods: GH, MC and QMC, for the integrals appearing in the estimation process of 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 neither explain 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 multi-dimensional 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. Section 5 shows an application to a real data set about verbal aggression (De Boeck and Wilson, 2004), comparing the three methods of integration in the complete optimization of the likelihood function. Finally, we discuss the results and we draw some conclusions.

2. Quasi-Monte Carlo integration

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

$$\int_{\Omega} f\,(x_1,\ldots,x_d)\;\mathrm{d}x_1,\ldots,\mathrm{d}x_d \approx \frac{\mathrm{vol}(\Omega)}{N}\sum_{i=1}^{N} f\left(x_1^i,\ldots,x_d^i\right). \tag{1}$$

In the regular MC method the vector $\boldsymbol{x}^i=\left(x_1^i,\ldots,x_d^i\right)$ follows a uniform distribution in Ω. Note that when the region of integration is the unit hypercube, $\Omega=(0,1)^d$, then $\mathrm{vol}(\Omega)=1$ in (1). If an integration region not contained in 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 $(1/N)\sum_{i=1}^{N} f\left(\boldsymbol{x}^i\right)$ 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 $\boldsymbol{x}^i=\left(x_1^i,\ldots,x_d^i\right)$ is not a random sample but a set of points deterministically chosen. These points are commonly known as LDS 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 of the unit hypercube than random points, because each new point is chosen in such a way that it is located in an area 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 frequently used. Fig. 1 shows the improved uniformity, of Sobol' and Halton points in comparison with random points, for the case of two dimensions and 1000 points.

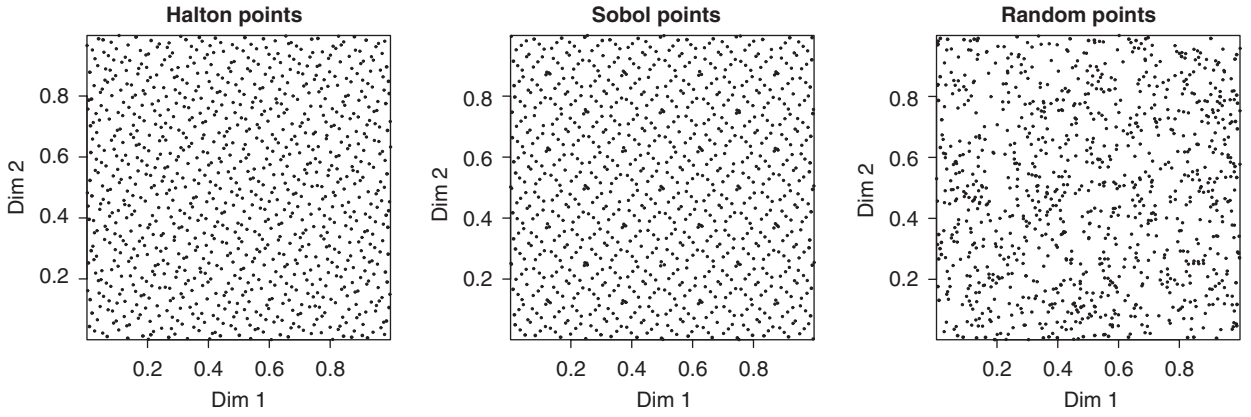

Fig. 1. Two low discrepancy sequences and random points.

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

$p=2$			$p=3$		
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

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

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

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 five 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).

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

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 3
Generation of the v_i values for Sobol' sequences

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

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 $m_i = 2b_1 m_{i-1} \oplus 2^2 b_2 m_{i-2} \oplus \cdots \oplus 2^l$ according to a primitive polynomial $P(z) = z^p + b_1 z^{p-1} + \cdots + b_{p-1} z + 1$, 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 Eq. (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 Gray 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. To obtain the v_i, the m_i must first be written in binary form and then the position of the fractional point shifted i positions to the left. Table 3 shows how to do this.

The first five Sobol' points are obtained as follows:

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

$$y_1 = y_0 \oplus v_1 = 0.0 \oplus 0.1 = 0.1 = 1 \times 2^{-1} = 0.500,$$

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

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

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

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

In the first line we added v_1 because the binary representation of $r - 1 = 0$ is 0, and 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 which is in the second position. In the third we added v_1 because 2 in binary is 10 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 a Sobol' sequence can be found in Bratley and Fox (1988).

Codes for both the Sobol' and Halton sequences have been implemented in some high-level software programs like the library fOptions (Wuertz, 2005) of R (R Development Core Team, 2005), which was used in this paper.

3. The logistic-normal model

Often the integrand of interest appearing in the likelihood function to be maximized, is the product of a function $f(t)$ 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 GH. In this section, we discuss more in detail the logistic-normal model as it is common in biostatistics and psychometrics.

Let Y_{ij} be the binary outcome variable for observation j ($j = 1, \ldots, k_i$) in cluster i ($i = 1, \ldots, c$), $\boldsymbol{\beta}$ is a p-dimensional vector of fixed effects with associated $k_i \times p$ design matrix $\boldsymbol{X}_i$, and $\boldsymbol{\theta}_i$ is a d-dimensional vector of random effects for cluster i with associated $k_i \times d$ design matrix $\boldsymbol{Z}_i$ (see Rijmen et al., 2003). The use of $f(t) = \{1 + \exp(-t)\}^{-1}$ is common in this framework. The assumption that $\boldsymbol{\theta}_i$ follows a multivariate normal distribution with $\mathbf{0}$ mean and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_{\theta_1} & \sigma_{\theta_1\theta_2} & \cdots & \sigma_{\theta_1\theta_d} \\ \sigma_{\theta_2\theta_1} & \sigma^2_{\theta_2} & \cdots & \sigma_{\theta_2\theta_d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\theta_d\theta_1} & \cdots & \cdots & \sigma^2_{\theta_d} \end{pmatrix},$$

leads to a logistic-normal model (Agresti, 2002, p. 496; Hosmer and Lemeshow, 2000, p. 310) of the form

$$\Pr\left(Y_{ij} = 1 \mid \boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\theta}_i\right) = f\left(\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right) = \frac{\exp\left(\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right)}{1 + \exp\left(\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right)}. \tag{8}$$

This models the probability of observation j in cluster i having a certain characteristic $\left(Y_{ij} = 1\right)$, with $\boldsymbol{z}_{ij}^\top$ and $\boldsymbol{x}_{ij}^\top$ the jth rows of the matrices $\boldsymbol{Z}_i$ and $\boldsymbol{X}_i$, respectively. Assuming that the observations regarding the same cluster are conditionally independent given $\boldsymbol{\theta}_i$, the binary random variable Y_{ij} follows a Bernoulli distribution with parameter $f\left(\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right)$, and the contribution of cluster i to the likelihood function is

$$\Pr\left(\boldsymbol{Y}_i = \boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}\right) = \int_{\mathbb{R}^d} \prod_{j=1}^{k_i} \frac{\exp\left\{y_{ij}\left(\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right)\right\}}{1 + \exp\left\{\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right\}} g\left(\boldsymbol{\theta}_i; \mathbf{0}, \boldsymbol{\Sigma}\right) \mathrm{d}\boldsymbol{\theta}_i, \tag{9}$$

where $g\left(\boldsymbol{\theta}_i; \mathbf{0}, \boldsymbol{\Sigma}\right)$ means that the vector $\boldsymbol{\theta}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Assuming independence between clusters, the full log-likelihood function is

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \sum_{i=1}^{c} \log \int_{\mathbb{R}^d} \prod_{j=1}^{k_i} \frac{\exp\left\{y_{ij}\left(\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right)\right\}}{1 + \exp\left\{\boldsymbol{z}_{ij}^\top \boldsymbol{\theta}_i + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right\}} g\left(\boldsymbol{\theta}_i; \mathbf{0}, \boldsymbol{\Sigma}\right) \mathrm{d}\boldsymbol{\theta}_i. \tag{10}$$

In Eq. (9), the function $f(t)$ is a product of logistic functions, because of the repeated nature of the observations ($j = 1, \ldots, k_i$) in cluster i. In order to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, the full log-likelihood must be maximized. An approximation of the integral in (10) is then necessary in each step of the maximization algorithm.

In the following we will first concentrate on $f(t)$ with a single logistic function. Then, $f(t)$ with a product of logistic functions will be investigated. Finally, we will fit a particular case of a logistic-normal model appearing in psychometrics using GH, MC, and QMC to approximate the integrals in the log-likelihood function.

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 appearing in the log-likelihood function 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 commonly used methods. As our interest in this section is in the integration process, rather than the optimization, we will do the calculations just for one cluster, hence $c = 1$. Therefore, we will not calculate the product or sum of c integrals, but only the integral over a product or sum of k terms for $k = 1$ in Section 4.1, and for $k = 3$ in Section 4.2. In the next section we will compare the methods in a real data set application with 316 clusters ($c = 316$) and 24 observations ($k = 24$).

A common problem of all three methods (GH, MC, and QMC), is that the standard sets of points need to be transformed before one can use them to approximate an arbitrary integral like in Eq. (9). For the GH method, the transformation amounts to re-centering and re-scaling the points 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, pp. 447–449).

For MC and QMC, the situation is different but 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 $\boldsymbol{P}_{N\times d}$ of N d-dimensional LDS or MC points and create a matrix $\boldsymbol{U}_{N\times d}$ with elements $u_{ij} = \Phi^{-1}\left(p_{ij}\right)$, where $p_{ij} \in (0, 1)$ are the elements of the matrix $\boldsymbol{P}$, $\Phi^{-1}(\cdot)$ is the inverse cumulative normal distribution function, and N is the total number of points to be used in the approximation of the integral.

(ii) *Adding correlation structure:* Put $\boldsymbol{R} = \boldsymbol{L}^\top \boldsymbol{U}$, where $\boldsymbol{L}$ is the Cholesky decomposition of the covariance matrix, $\boldsymbol{\Sigma} = \boldsymbol{L}^\top \boldsymbol{L}$.

(iii) *Obtaining the approximation:* From Eq. (9)

$$\Pr\left(\boldsymbol{Y}_i = \boldsymbol{y}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}\right) \approx \frac{1}{N}\sum_{n=1}^{N}\prod_{j=1}^{k_i} f\left(\boldsymbol{z}_{ij}^\top \boldsymbol{r}^{(n)} + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right),$$

where $\boldsymbol{r}^{(n)}$ is the nth row of the matrix $\boldsymbol{R}$.

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

$$\mathrm{RE} = \left|\frac{\widehat{T} - T}{T}\right|, \tag{11}$$

where $\widehat{T}$ is the approximated value of integral using one of the three methods to be compared and T is its real value. The RE is a measure of how many digits of accuracy one has.

Table 4
Values to form the $\boldsymbol{\Sigma}$ matrices used in Study 1

d	$\sigma^2_{\theta_1}, \ldots, \sigma^2_{\theta_d}$
3	(2.0, 1.5, 1.0)
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

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 $\boldsymbol{z}^\top$ is a vector of d ones denoted by $\mathbf{1}^\top$, and $x = -1$, then (9) reduces to

$$T = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp\{\theta_1 + \cdots + \theta_d - \beta\}}{1 + \exp\{\theta_1 + \cdots + \cdots \theta_d - \beta\}} g(\boldsymbol{\theta}; \mathbf{0}, \boldsymbol{\Sigma})\, \mathrm{d}\theta_1, \ldots, \mathrm{d}\theta_d. \tag{12}$$

As mentioned before, T denotes 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

$$W = \sum_{q=1}^{d} \theta_q - \beta \sim N\left(-\beta, \underbrace{\mathbf{1}^\top \boldsymbol{\Sigma} \mathbf{1}}_{\sigma_w^2}\right). \tag{13}$$

So, T can be reduced to

$$\int_{-\infty}^{\infty} \frac{\exp(w)}{1 + \exp(w)} g\left(w; -\beta, \sigma_w^2\right) \mathrm{d}w. \tag{14}$$

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.

4.1.1. Design of the study

We evaluated the cases $d = 3, 5, 7, 9$, and $\beta = 0, 1, 2$. The different $\boldsymbol{\Sigma}$ matrices were chosen using the values shown in Table 4 for each dimension, respectively. The value ρ refers to correlations implied in $\boldsymbol{\Sigma}$, to be crossed with the values of $\sigma^2_{\theta_q}$ for $d = 3, 5, 7$, and 9. Then, for instance, if $d = 3$, five $\boldsymbol{\Sigma}$ matrices are formed, all of them with diagonal entries 2.0, 1.5, and 1.0, but with different correlations.

The total number of points used to evaluate (12) 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, $5^7 = 78\,125$ points were used to evaluate the integral, and $5^9 = 1\,953\,125$ in $d = 9$.

4.1.2. Results of Study 1

Fig. 2 shows plots of the RE in $\log_{10}$ scale for each value of d, ρ, and, β, and, for each method. From Fig. 2 it can be seen that QMC was better than GH in all the cells in which the following conditions are fulfilled: $\rho \geqslant 0.3$, and $\beta \geqslant 1$, except in the cells with $d = 3$, $\rho = 0.3$, $\beta = 2$ and $\rho = 0.5$, $\beta = 1$, and $d = 5$, $\rho = 0.3$. 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 the Halton results, 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 dimensionality, high correlations and on the condition that the distributions are shifted ($\beta \geqslant 1$). In general, the RE for GH increases with the correlation

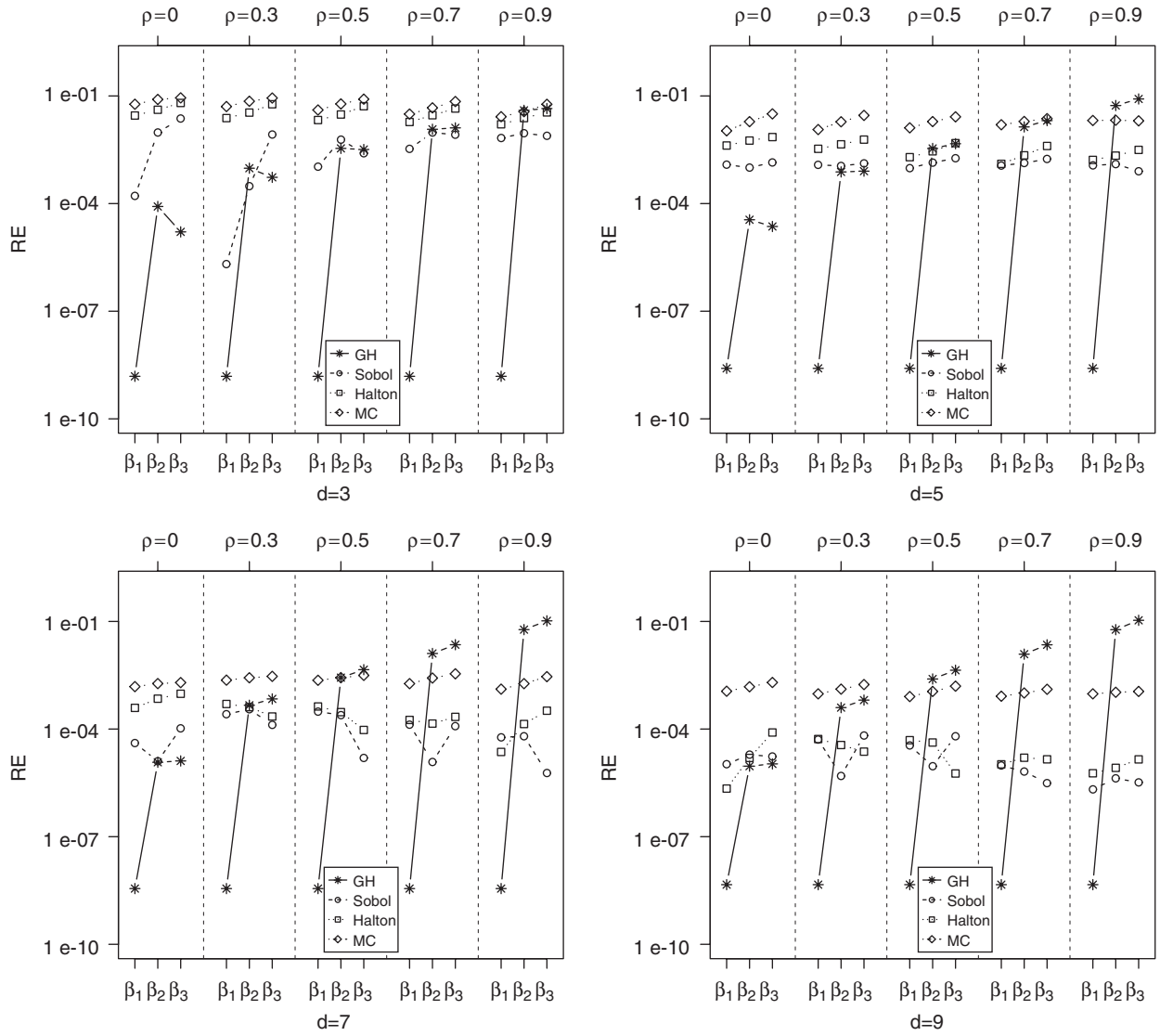

Fig. 2. Results of Study 1.

(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 1 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 the next study, we will investigate the LDS integral approximation for all points in between. Second, we did not consider the case of 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.

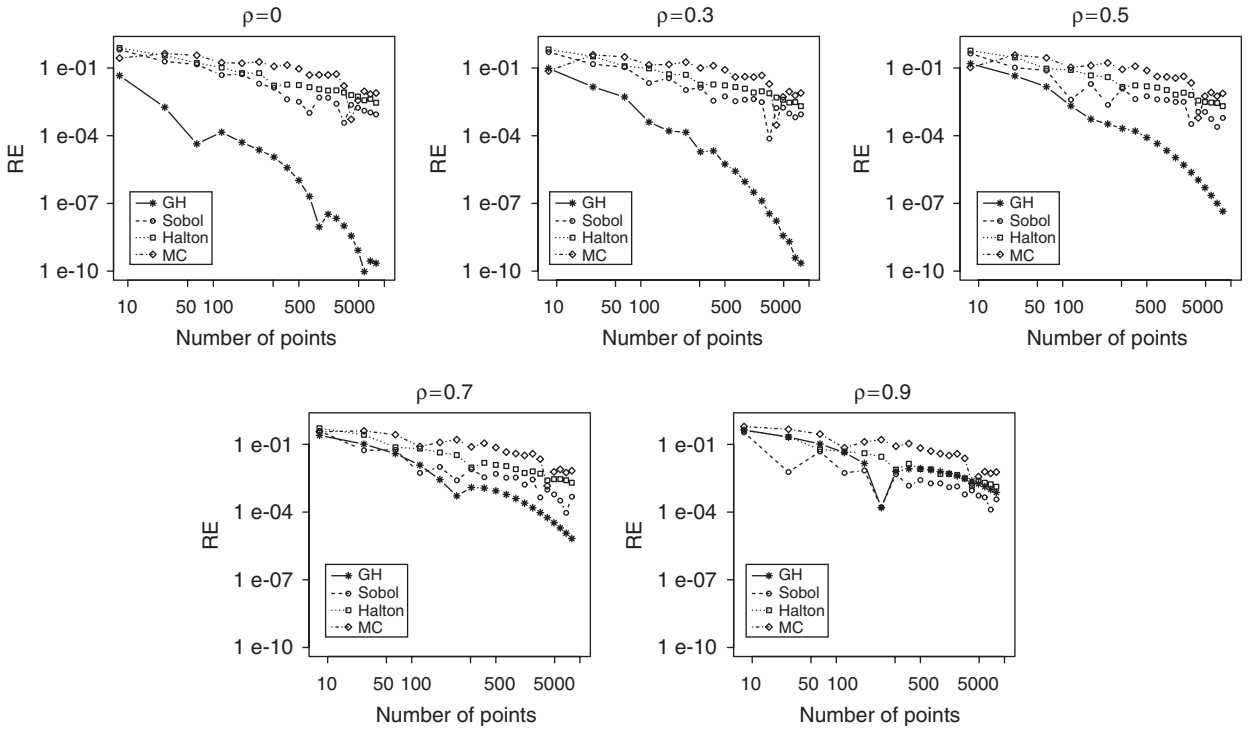

Fig. 3. Relative error versus number of points used to evaluate the test integral in Study 2, $d = 3$.

Suppose that $\boldsymbol{z}_j^\top$ is a d-dimensional vector of ones and that for each of the three observations ($k = 3$), $x_j = -1$, $j = 1, 2, 3$. Assume that the observations $\boldsymbol{y} = (1, 1, 0)$, are made, then (9) reduces to the following integral:

$$T = \int_{\mathbb{R}^d} \left(\frac{\exp\left\{\sum_{q=1}^d \theta_q - \beta_1\right\}}{1 + \exp\left\{\sum_{q=1}^d \theta_q - \beta_1\right\}} \right) \left(\frac{\exp\left\{\sum_{q=1}^d \theta_q - \beta_2\right\}}{1 + \exp\left\{\sum_{q=1}^d \theta_q - \beta_2\right\}} \right) \times \left(\frac{1}{1 + \exp\left\{\sum_{q=1}^d \theta_q - \beta_3\right\}} \right) g(\boldsymbol{\theta}; \mathbf{0}, \boldsymbol{\Sigma})\, \mathrm{d}\boldsymbol{\theta}. \tag{15}$$

This integral can accurately be approximated using (13) and (14).

4.2.1. Design of the study

We evaluated (15) for $\boldsymbol{\beta} = (0, 1, 2)$ and for the same values of d and $\boldsymbol{\Sigma}$ utilized in Study 1, see Table 4. In addition, we varied the number of points used to evaluate (15) and computed the RE (11). The number of points will not be limited to powers of quadrature points used per dimension. For the Sobol' method all the numbers in between will be used as well.

4.2.2. Results of Study 2

Fig. 3 shows the RE versus the number of points used to evaluate (15) in the case $d = 3$. Both axes in the figure are in $\log_{10}$ scale. Note that the symbols $\circ$, $\square$, $*$, $\diamond$ in the plot refer to the abscissa values $2^d, 3^d, \ldots, \ell^d$ indicating that 2 to ℓ quadrature points were used per dimension. For example, the eighth symbol refers to $9^3 = 729$ points (9 quadrature points per dimension) in total. From the figure several conclusions may be drawn. First, the GH method was always better than the other methods, except for $\rho = 0.7$ where using 3 and 5 quadrature points per dimension, the Sobol' method was better, and, more often, for $\rho = 0.9$, Sobol' was better. Generally, the QMC method seems to be better than the regular MC and Sobol' better than Halton. Second, an increasing number of points utilized to evaluate the

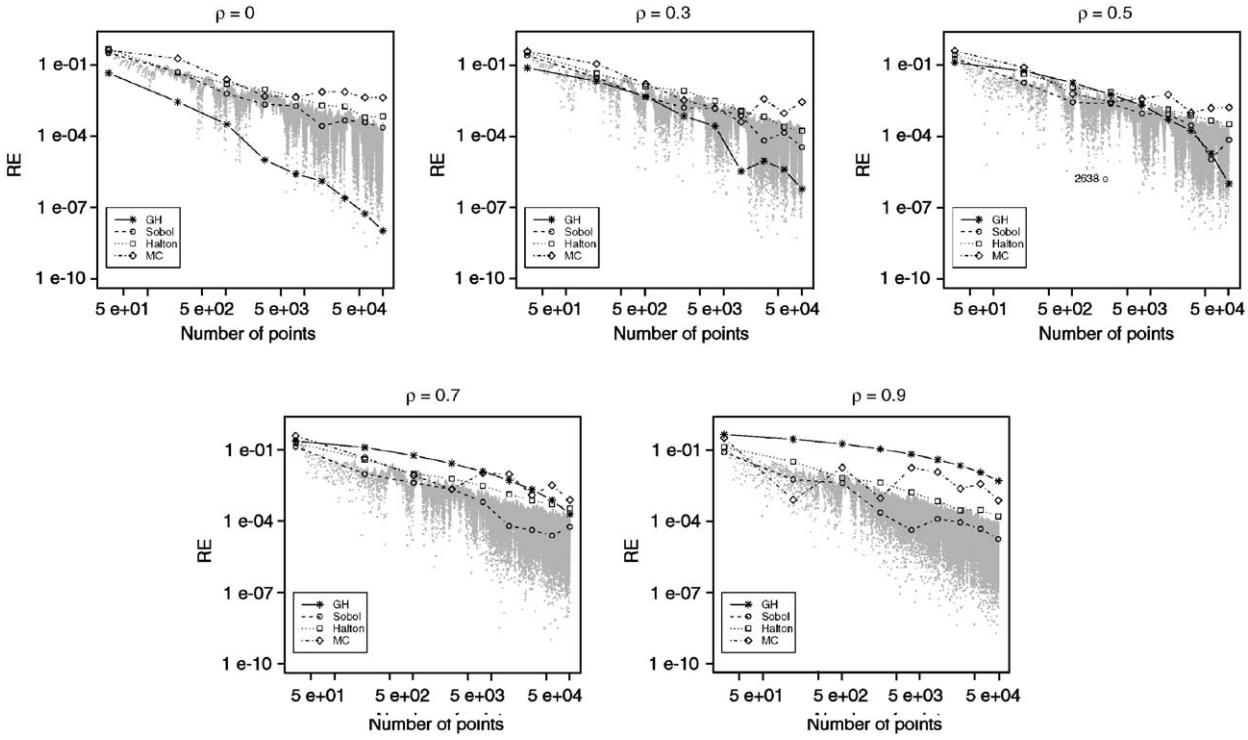

Fig. 4. Relative error versus number of points used to evaluate the test integral in Study 2, $d = 5$. The Sobol' results are shown for 2^5–10^5 points.

integrand, does not necessarily yield a more accurate result. For example, for $\rho = 0$ using 12 GH points per dimension yields a better result than using 13, 14, and 15 points. This situation is observed for the case of other ρ values as well. For the MC and QMC the behavior of the curve is very irregular and not monotonic at all. For example, when $\rho = 0$, using $11^3 = 1331$ Sobol' points seems better than using $20^3 = 8000$ Sobol' points. However, as mentioned earlier, also the GH curve is somewhat irregular and non-monotonic.

To investigate the irregularities more in detail, we computed the approximation with 2^5 up to 10^5 Sobol' points for $d = 5$, and 2^7 up to 5^7, and 2^9 up to 5^9, for $d = 7$ and $d = 9$, respectively. Sobol' points were chosen because they gave better results than the Halton points. Fig. 4 shows the case $d = 5$. 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' method is sometimes better than GH method. In the case $\rho = 0.5$, Fig. 4 shows that using 2638 or even less Sobol' points, one can obtain a result as accurate as using 9 GH quadrature points per dimension (i.e. $9^5 = 59\,049$ points). From these results, one cannot derive a general rule or not even a hint for when the Sobol' points do better. Note that when considering correlated dimensions, GH loses accuracy which is reflected in the fact that the GH curve decreases more slowly. For example, in Fig. 4 for $\rho = 0$, with 10 quadrature points per dimension we obtain approximately eight digits of precision, which is clearly more than in the case $\rho = 0.9$. Again QMC is better than MC but it is not clear whether Sobol' is better than Halton. Note that for $d = 5$, Sobol' can be much better than GH and for a much smaller number of points. From $\rho = 0.7$ on, Sobol' is always better than GH. For an even higher dimensionality, $d \geqslant 7$, and for $\rho \geqslant 0.5$, the Sobol' results are always better than any of the GH results (not shown in a figure due to space restrictions).

It may be concluded that, sometimes, the QMC method can do better than the GH, for the same number of points or with less points. For higher dimensions and highly correlated dimensions, the QMC method seems to perform better than the GH method. The fact of having a gold standard (i.e., the real value of the integral, or a very good approximation of it) to compare the methods was an advantage that in a real data application one does not have. In the next section, we will incorporate the three numerical integration methods into the complete optimization routine to obtain the parameter estimates of a two-dimensional model. The number of points will be varied as it appears that with less points, one can sometimes obtain satisfactory results.

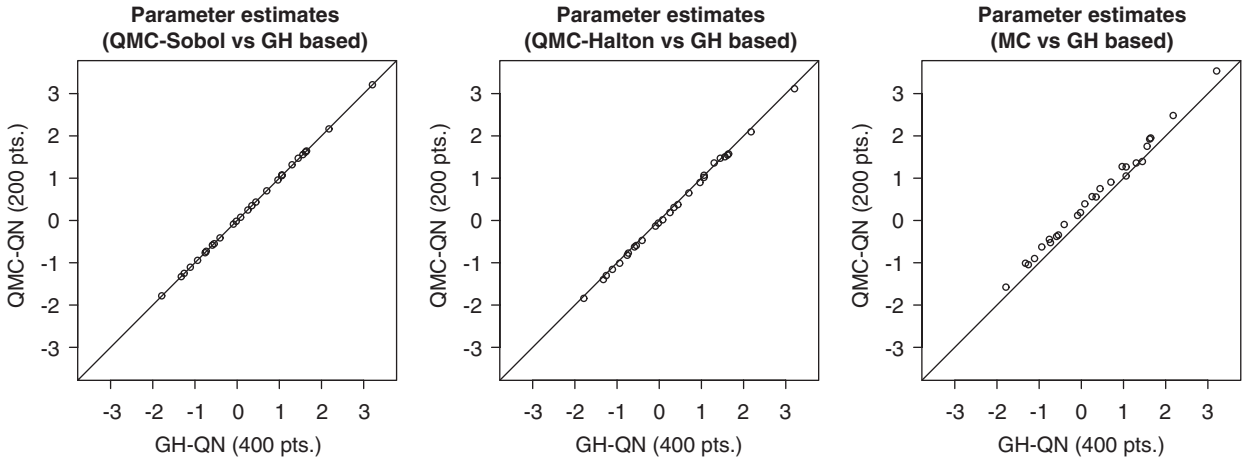

Fig. 5. Comparison of the parameter estimates obtained: GH-based versus QMC-based and MC-based for 200 points in total.

5. Application: a between-item multidimensional Rasch model

Rijmen and Briggs (2004) considered a multidimensional extension of the Rasch model (e.g., Rasch, 1960) by relaxing the assumption of just one underlying one-dimensional latent variable θ, extending it to a d-dimensional vector $\boldsymbol{\theta}_i$ of latent variables.

We will consider here a two-dimensional Rasch model, applied to a data set consisting of $k = 24$ responses (observations) of $c = 316$ persons (clusters) to questions about verbal aggression. Respondents were presented a frustrating situation and asked whether they *would want to react* in a verbally aggressive manner (want-items), or whether they *actually would react* in a verbally aggressive manner (do-items) (see more details in De Boeck and Wilson, 2004). As an application of a multidimensional model, the verbal aggression data are analyzed with a two-dimensional model assuming that want-items and do-items represent two separate dimensions.

Let Y_{ij} be the binary random variable denoting the answer to item j of individual i, taking the values 1 and 0 for a yes and no answer, respectively. Then, the model reads as

$$\Pr\left(Y_{ij} = 1 \mid \boldsymbol{\theta}_i, \beta_j\right) = \frac{\exp\left(\boldsymbol{z}_j^\top \boldsymbol{\theta}_i - \beta_j\right)}{1 + \exp\left(\boldsymbol{z}_j^\top \boldsymbol{\theta}_i - \beta_j\right)}. \tag{16}$$

It is clear that this model is a particular case of the more general model in (8) with $\boldsymbol{X}_i$ a $p \times p$ diagonal matrix with -1 as diagonal elements. Note that the index i vanishes for $\boldsymbol{Z}$ and $\boldsymbol{X}$ because these matrices are equal for all the clusters $i = 1, \ldots, c$. Each element of $\boldsymbol{z}_j^\top$ contains only one non-zero element, equal to 1, indicating the dimension the item is measuring.

The purpose of the application is to see whether with a smaller number of points in comparison with the GH method, one can obtain reasonable good estimates. Note, however, that we do not have a gold standard as in studies 1 and 2 to compare the results with.

For the estimation of the parameters $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, we use all three methods (GH, MC, and QMC) to approximate the integral appearing in the full log-likelihood function (10) to be maximized. A quasi-Newton algorithm (e.g., Kelley, 1999) is used for the maximization.

For the GH method, 20 quadrature points per dimension were used, so $20^2 = 400$ points in total to evaluate the integrals. For QMC and also for MC we consider 25, 50, 100, 200, 300 and 400 points in total. As an example, Fig. 5 shows the plot of the parameter estimates, $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\Sigma}})$, obtained with the GH-quasi-Newton (GH-QN) based optimization, versus the parameter estimates obtained with the QMC-quasi-Newton (QMC-QN) and MC-quasi-Newton (MC-QN) based optimization, respectively, for the case of 200 points.

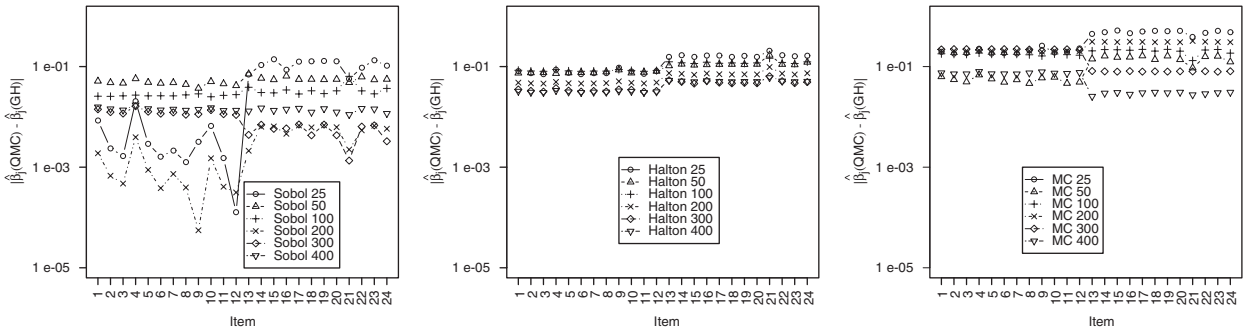

Fig. 6. Comparison of the parameter estimates obtained: GH-based versus QMC and MC-based.

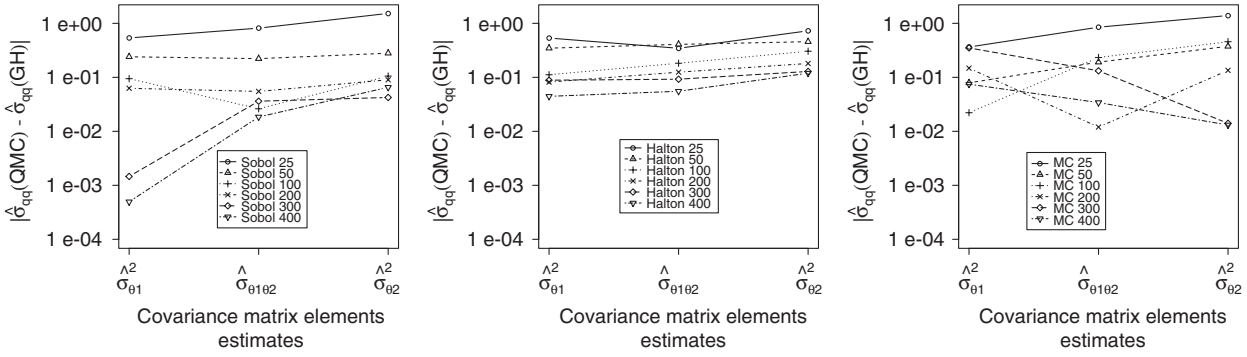

Fig. 7. Comparison of the elements of the covariance matrix estimates obtained: GH-based versus QMC and MC-based.

Figs. 6 and 7 show the absolute difference between the $\hat{\beta}$ parameter estimates, and the elements of the covariance matrix $\boldsymbol{\Sigma}$ estimates, respectively, using the three methods in the integration part. Note that the differences in the second half of the plots in Fig. 6 are sometimes bigger than in the first half. This is due to the fact that the second half of the estimates correspond to the *do* items, which were more extreme than the *want* items, and therefore they are more difficult to estimate. Something similar was found in the earlier studies, where for shifted distributions there were different results for GH and the other methods.

The estimation using 25, 50, 100, 200, 300 and 400 QMC points in total was 88.6%, 79.7%, 68.3%, 51.3%, 32.3%, and 7.0% faster, respectively, than GH using 400 points in total. It may be concluded from these results that using a QMC-QN-based estimation, one can obtain similar results to those obtained with a GH-QN-based method, or better when considering less points.

6. Discussion and conclusion

A comparison between three numerical procedures (GH, MC and QMC integration) to approximate integrals has been presented. LDS 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 sometimes do better than the GH method for the same number points or with less points. Unfortunately, when using the QMC method, there is no rule for selecting an appropriate number of points to evaluate the integrals. In this context, constructing an automated algorithm to cleverly select the number of points for a given problem is a challenge. One possibility is to consider some variant of the QMC method, called Randomized QMC, which can be used to obtain an approximation of standard errors for the QMC. Then, an automated algorithm like the one explained in Booth and Hobert (1999) could be developed.

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 ($\ell = 1, 2, 3, \ldots$) 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 seem independent of the non-monotonic relation between number of points and accuracy.

For the optimization task, the QMC method seems to be a promising method to approximate the integral, in the process of maximizing the likelihood. Although without a gold standard to compare results, Section 5 showed that, one can obtain similar results as obtained with the GH method, and with less points. In this way the computational effort and time needed for the calculations can be reduced considerably. For practical purposes, a mixed method, GH combined with QMC, may be useful, switching from GH to QMC, for the kind of problems where GH seems less accurate or needs too many points.

References

Abramowitz, M., Stegun, I. (Eds.), 1972. Handbook of Mathematical Functions. Dover Publications Inc., New York.

Agresti, A., 2002. Categorical Data Analysis. second ed. Wiley, New York.

Antonov, I., Saleev, V., 1979. An economic method of computing LP_τ-sequences. USSR Comput. Math. Math. Phys. 19, 252–256.

Booth, J.G., Hobert, J.H., 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. J. Roy. Statist. Soc. B 62, 265–285.

Bratley, P., Fox, B.L., 1988. Algorithm 659 implementing Sobol's quasirandom sequence generator. ACM Trans. Math. Software 14, 88–100.

Caflisch, R., 1998. Monte Carlo and quasi-Monte Carlo methods. Acta Numer. 7, 1–49.

Cools, R., 1997. Constructing cubature formulae: the science behind the art. Acta Numer. 6, 1–54.

Cools, R., 2002. Advances in multidimensional integration. J. Comput. Appl. Math. 149, 1–12.

Crouch, A., Spiegelman, E., 1990. The evaluations of integrals of the form $\int_{-\infty}^{\infty} f(t) \exp\left(-t^2\right) \mathrm{d}t$: application to logistic-normal models. J. Amer. Statist. Assoc. 85, 464–469.

Davis, P., Rabinowitz, P., 1984. Methods of Numerical Integration. second ed. Academic Press, Orlando, FL.

De Boeck, P., Wilson, M., 2004. Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach. Springer-Verlag, New York.

Fahrmeir, L., Tutz, G., 2001. Multivariate Statistical Modelling Based on Generalized Linear Models. second ed. Springer-Verlag, New York.

Fischer, G., Molenaar, I. (Eds.), 1995. Rasch Models: Foundations and Recent Developments. Springer-Verlag, New York.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2, 84–90.

Hickernell, F., Lemieux, C., Owen, A., 2005. Control variates for Quasi-Monte Carlo. Statist. Sci. 20, 1–31.

Hosmer, D., Lemeshow, S., 2000. Applied Logistic Regression. second ed. Wiley, New York.

Jank, W., 2005. Quasi-Monte Carlo sampling to improve the efficiency of Monte Carlo EM. Comput. Statist. Data Anal. 48, 685–701.

Judd, K., 1998. Numerical Methods in Economics. MIT Press, Boston.

Kelley, C.T., 1999. Iterative Methods for Optimization. SIAM, Philadelphia, PA.

Kocis, L., Whiten, W., 1997. Computational investigations of low-discrepancy sequences. ACM Trans. Math. Software 23, 266–294.

Lemieux, C., L'Ecuyer, P., 2001. On the use of quasi-Monte Carlo methods in computational finance. 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. Appl. Statist. 50, 325–335.

Morokoff, W., Caflisch, R., 1995. Quasi-Monte Carlo integration. J. Comput. Phys. 122, 218–230.

Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods (CBMS-NSF Regional Conference Series in Applied Mathematics, No. 63).

Paskov, S., 1995. Faster valuation of financial derivatives. J. 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, New York.

R Development Core Team, 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL ⟨http://www.R-project.org⟩.

Rasch, G., 1960. Probabilistic Models for Some Intelligence and Attainment Test. Danish Institute for Educational Research, Copenhagen, Denmark.

Rijmen, F., Briggs, D., 2004. Multiple person dimensions and latent item predictors. In: De Boeck, P., Wilson, M. (Eds.), Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach. Springer-Verlag, New York.
Rijmen, F., Tuerlinckx, F., De Boeck, P., Kuppens, P., 2003. A nonlinear mixed model framework for IRT models. Psychol. Methods 8, 185–205.
Robert, C., Casella, G., 2004. Monte Carlo Statistical Methods. second ed. Springer-Verlag, New York.
Shao, J., 2003. Mathematical Statistics. second ed. Springer-Verlag, New York.
Stroud, A., 1971. Approximate Calculation of Multiple Integrals. Prentice-Hall, Englewood Cliffs, NJ.
Wuertz, D., 2005. fOptions: Financial Software Collection—fOptions. R package version 220.10063. ⟨http://www.rmetrics.org⟩.