CIRRELT

Centre interuniversitaire de recherche
sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre
on Enterprise Networks, Logistics and Transportation

Estimation of the Mixed Logit Likelihood Function by Randomized Quasi-Monte Carlo

David Munger
Pierre L'Écuyer
Fabian Bastin
Cinzia Cirillo
Bruno Tuffin

November 2010

CIRRELT-2010-47

Bureaux de Montréal :
Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec)
Canada H3C 3J7
Téléphone : 514 343-7575
Télécopie : 514 343-7121

Bureaux de Québec :
Université Laval
2325, de la Terrasse, bureau 2642
Québec (Québec)
Canada G1V 0A6
Téléphone : 418 656-2073
Télécopie : 418 656-2624

www.cirrelt.ca

Estimation of the Mixed Logit Likelihood Function by Randomized Quasi-Monte Carlo

David Munger[1,2], Pierre L'Écuyer[1,2], Fabian Bastin[1,2,*], Cinzia Cirillo[1,3], Bruno Tuffin[4]

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Computer Science and Operations Research, Université de Montréal, P.O. Box 6128, Station Centre-ville, Montréal, Canada H3C 3J7

[3] Department of Civil & Environmental Engineering, University of Maryland, College Park, MD 20742, USA

[4] INRIA Rennes Bretagne-Atlantique, Campus de Beaulieu, 35042 Rennes cedex, France

Abstract. Advanced discrete choice models, such as parametric/non-parametric mixed logit and hybrid choice models, are heavily used in travel behavior research. The complexity of model formulation nevertheless remains limited by the associated estimation difficulties, even if important progress has been made these last years. In this piece of work, we examine the effectiveness of randomized quasi-Monte Carlo (RQMC) techniques to estimate the integrals that express the discrete choice probabilities in a mixed logit model, for which no closed form formula is available. We review some popular RQMC constructions, discuss the choice of their parameters as a function of the considered class of integrands, and compare their effectiveness to reduce both variance and bias, in comparison with standard Monte Carlo (MC), when estimating the log-likelihood function at a given parameter value. In our numerical experiments, randomized rank-1 lattice rules (with carefully selected parameters) and digital nets in base 2 outperform randomized Halton sequences and standard MC. Interestingly, they also reduce the bias as much as the variance.

Keywords. Discrete choice modeling, mixed logit model, Monte Carlo, quasi-Monte Carlo, variance reduction, bias reduction, lattice rules, weighted discrepancy, ANOVA.

Acknowledgements. This work was partly supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), Discovery Grants to P. L'Ecuyer and F. Bastin, a Canada Research Chair to P. L'Ecuyer, a GERAD scholarship to D. Munger, the EuroNF Network of Excellence to B. Tuffin, and INRIA's associated team MOCQUASIN to all authors.

---

* Corresponding author: Fabian.Bastin@cirrelt.ca

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 2010

1 Introduction

Travel behavior analysis makes heavy use of discrete choice models. Recent modeling frameworks account for a large number of (random) effects: heterogeneity in preferences (Hess et al., 2005; Cirillo and Axhausen, 2006) and/or in scale factor (Hess et al., 2009), variability in willingness to pay (Bastin et al., 2010), correlation across alternatives (Brownstone et al., 2000), in space (Bhat and Sener, 2009), etc. Advanced discrete choice models are often associated with choice probabilities that can be written as multivariate integrals, but do not admit a closed-form formula. In mixed logit models, for example, we have an integral with respect to the mixing density, which depends on unknown parameters that we want to estimate. These integrals are typically estimated by Monte Carlo (MC) or quasi-Monte Carlo (QMC) methods (McFadden and Train, 2000; Train, 2003), In particular, Halton sequences have found widespread application for mixed logit model estimation in transportation (Train, 2000; Bhat, 2001, 2003a). Certain digital nets and sequences, including Sobol' nets, have also been experimented, e.g., by Sándor and Train (2004). Sivakumar et al. (2005) have compared Faure sequences against Halton sequences and found that they perform better empirically in the evaluation of the multidimensional integrals that they considered.

In other areas of applications, for example in finance (L'Ecuyer, 2009), the best empirical results are typically obtained by Sobol' nets with certain types of scrambles and by randomly-shifted lattice rules with a baker's transformation, so we thought it would be interesting to try these types of RQMC methods for estimating the integrals involved in mixed logit models. For the lattice rules, we are also interested in selecting the parameters to minimize a measure of discrepancy adapted to the problem, and see if it can make a difference. We focus solely on the estimation of the log-likelihood function at a given parameter value, for a given data set. This is the basic building block and the main source of error when we estimate the parameter value that maximizes the log-likelihood function (and use it as an estimator of the unknown parameter value in the model). Just estimating the log-likelihood function is simpler than estimating its maximizer, and avoids additional error sources from the maximization. It also permits us to study both the bias and the variance of the estimator. The lattice rules and Sobol nets examined in this paper perform better than the Halton sequences. They reduce both the variance and the simulation bias.

The remainder is organized as follows. In Section 2, we collect the basic definitions and properties of the mixed logit model considered in this paper, and we analyze the bias and variance of an MC estimator of the log-likelihood. In Section 3, we provide a short state-of-the-art review of RQMC methods, with a particular focus on randomly-shifted lattice rules and the choice of their parameters. Numerical experiments that examine and compare the bias and variance of various RQMC estimators of the log-likelihood, first for synthetic data and then for a real-life data

set, are reported in Section 4. A brief experiment on parameter estimation (maximizing the log-likelihood), with the real-life data, is also reported. A conclusion and ideas for future research are given in Section 5.

2 The mixed logit model and log-likelihood estimation

2.1 The model

In a popular form of *multinomial mixed logit* model (McFadden and Train, 2000; Train, 2003; Sivakumar et al., 2005; Bastin et al., 2006b), the utility of alternative j for individual q has the form

$$U_{q,j} = \boldsymbol{\beta}_q^{\mathsf{t}} \mathbf{x}_{q,j} + \epsilon_{q,j} = \sum_{\ell=1}^{s} \beta_{q,\ell} x_{q,j,\ell} + \epsilon_{q,j}$$

where $\boldsymbol{\beta}_q = (\beta_{q,1}, \ldots, \beta_{q,s})^{\mathsf{t}}$ is an unobserved random vector of *taste parameters* (or coefficients) for each individual q, $\mathbf{x}_{q,j} = (x_{q,j,1}, \ldots, x_{q,j,s})^{\mathsf{t}}$ gives the *observed attributes* for choice j and individual q, and the $\epsilon_{q,j}$ are random variables that represent unobserved *random noise*. The $\epsilon_{q,j}$ are assumed to be independent and to have a Gumbel distribution of mean 0 and scale factor of 1 (the choice of mean has no importance, because only utility differences matter).

It is assumed that if q is an individual selected randomly from the entire population, then its associated (random) vector $\boldsymbol{\beta}_q$ of taste coefficients has a multivariate density $f_{\boldsymbol{\theta}}$ that depends on some parameter (vector) $\boldsymbol{\theta}$. This accounts for the fact that different individuals may have different tastes. Individual q always selects the alternative j having the largest utility $U_{q,j}$.

Conditional on a given (fixed) $\boldsymbol{\beta}_q$, we have an ordinary *logit* model, and one can show in that case that the individual selects alternative j with probability

$$L_q(j, \boldsymbol{\beta}_q) = \frac{\exp[\boldsymbol{\beta}_q^{\mathsf{t}} \mathbf{x}_{q,j}]}{\sum_{a \in \mathcal{A}(q)} \exp[\boldsymbol{\beta}_q^{\mathsf{t}} \mathbf{x}_{q,a}]}, \tag{1}$$

independently of other individuals, where $\mathcal{A}(q)$ is the set of his alternatives. The unconditional probability that a random individual selects alternative j is then

$$p_q(j, \boldsymbol{\theta}) = \mathbb{E}[L_q(j, \boldsymbol{\beta}_q)] = \int_{\mathbb{R}^s} L_q(j, \boldsymbol{\beta}) f_{\boldsymbol{\theta}}(\boldsymbol{\beta}) \, \mathrm{d}\boldsymbol{\beta}. \tag{2}$$

We assume that a random vector $\boldsymbol{\beta}$ with density $f_{\boldsymbol{\theta}}$ can be written as $\boldsymbol{\beta} = h(\boldsymbol{\theta}, \mathbf{U})$ for some explicit function h, where $\mathbf{U}$ is a vector of independent uniform random variables over $(0, 1)$. This

assumption is standard and is required to simulate realizations of $\boldsymbol{\beta}$ from independent uniform random numbers. In the examples considered in this paper, $\mathbf{U}$ has dimension s. Then we have

$$p_q(j,\boldsymbol{\theta}) = \mathbb{E}\left[L_q(j,h(\boldsymbol{\theta},\mathbf{U}))\right] = \int_{(0,1)^s} L_q(j,h(\boldsymbol{\theta},\mathbf{u}))\,\mathrm{d}\mathbf{u}. \tag{3}$$

In general, no explicit formula is available for $p_q(j,\boldsymbol{\theta})$, and MC is often the most practical way of estimating the integral in (2) or (3) for each q. To do that, for a given $\boldsymbol{\theta}$ and q, we generate n_q independent random points $\mathbf{U}_q^{(1)},\ldots,\mathbf{U}_q^{(n_q)}$, put $\boldsymbol{\beta}_q^{(i)}(\boldsymbol{\theta}) = h(\boldsymbol{\theta},\mathbf{U}_q^{(i)})$ for $i = 1,\ldots,n_q$, and we estimate $p_q(j,\boldsymbol{\theta})$ by

$$\hat{p}_q^{n_q}(j,\boldsymbol{\theta}) = \frac{1}{n_q}\sum_{i=1}^{n_q} L_q(j,\boldsymbol{\beta}_q^{(i)}(\boldsymbol{\theta})) = \frac{1}{n_q}\sum_{i=1}^{n_q} L_q(j,h(\boldsymbol{\theta},\mathbf{U}_q^{(i)})), \tag{4}$$

where each $L_q(j,\boldsymbol{\beta}_q^{(i)}(\boldsymbol{\theta}))$ can be computed via (1).

When the same individual delivers several observations rather than only one, these repeated choices are usually correlated, and we must consider the probability of the individual's choice sequence, instead of the particular observations. Typically, the tastes of a given decision-maker are assumed to remain constant across choice situations for each particular respondent, such that tastes vary across individuals, but not across observations for the same individual. If decision-maker q makes a sequence of T_q independent selections (conditionally on $\boldsymbol{\beta}_q$) and chooses alternative j_t for his tth decision, the probability of this choice sequence is

$$L_q^{T_q}(j_1,\ldots,j_{T_q},\boldsymbol{\beta}_q) = \prod_{t=1}^{T_q} L_q(j_t,\boldsymbol{\beta}_q),$$

where each $L_q(j_t,\boldsymbol{\beta}_q)$ is computed as in (1), and the unconditional probability for the choices sequence is

$$p_q(j_1,\ldots,j_{T_q},\boldsymbol{\theta}) = \mathbb{E}\left[L_q^{T_q}(j_1,\ldots,j_{T_q},h(\boldsymbol{\theta},\mathbf{U}))\right].$$

2.2 Estimating the log-likelihood of a sample

Suppose we have a data set of one observation per individual for m individuals (the case of several observations per individual is similar), in which we observe that individual q was given the vector of attributes $\mathbf{x}_{q,j}$ for each alternative j and made the choice y_q, for $q = 1,\ldots,m$. The goal is to estimate $\boldsymbol{\theta}$ from this data. The *maximum likelihood method* estimates the true $\boldsymbol{\theta}$ by the value that maximizes the joint probability $L(\boldsymbol{\theta})$ (the likelihood) of the observed population sample,

as a function of $\boldsymbol{\theta}$, or the *logarithm* of this probability, which is mathematically equivalent and computationally simpler. For convenience, we also divide by m, so we have an average over individuals and the expression does not blow up when $m \to \infty$. That is, we want to maximize

$$\frac{\ln L(\boldsymbol{\theta})}{m} = \frac{1}{m} \ln \prod_{q=1}^{m} p_q(y_q, \boldsymbol{\theta}) = \frac{1}{m} \sum_{q=1}^{m} \ln p_q(y_q, \boldsymbol{\theta}). \tag{5}$$

Here the $p_q(y_q, \boldsymbol{\theta})$ are unknown but we can replace them by their estimators $\hat{p}_q^{n_q}(y_q, \boldsymbol{\theta})$ defined in (4). This gives the following estimator of $\ln L(\boldsymbol{\theta})/m$:

$$\frac{\ln(\hat{L}(\boldsymbol{\theta}))}{m} = \frac{1}{m} \sum_{q=1}^{m} \ln\left(\hat{p}_q^{n_q}(y_q, \boldsymbol{\theta})\right) = \frac{1}{m} \sum_{q=1}^{m} \ln\left(\frac{1}{n_q} \sum_{i=1}^{n_q} L_q(y_q, h(\boldsymbol{\theta}, \mathbf{U}_q^{(i)}))\right). \tag{6}$$

The log-likelihood estimator is biased, because $\ln$ is not a linear function, and the bias is negative because $\ln$ is concave. The dominant term of the bias can be found as follows. Let

$$R_q = \frac{\hat{p}_q^{n_q}(y_q, \boldsymbol{\theta}) - p_q(y_q, \boldsymbol{\theta})}{p_q(y_q, \boldsymbol{\theta})},$$

the relative estimation error in $p_q(y_q, \boldsymbol{\theta})$. A Taylor expansion of $\ln(\hat{p}_q^{n_q}(y_q, \boldsymbol{\theta}))$ around $\ln(p_q(y_q, \boldsymbol{\theta}))$ gives

$$\ln(\hat{p}_q^{n_q}(y_q, \boldsymbol{\theta})) - \ln(p_q(y_q, \boldsymbol{\theta})) \;=\; R_q - R_q^2/2 + \mathcal{O}(|R_q|^3). \tag{7}$$

The total bias in (6) is

$$\begin{aligned} \mathbb{E}\left[\frac{\ln(\hat{L}(\boldsymbol{\theta})) - \ln(L(\boldsymbol{\theta}))}{m}\right] &= \frac{1}{m} \sum_{q=1}^{m} \mathbb{E}\left[-\frac{R_q^2}{2} + \mathcal{O}(|R_q|^3)\right] \\ &\approx -\frac{1}{2m} \sum_{q=1}^{m} \mathrm{Var}\left[\frac{\hat{p}_q^{n_q}(y_q, \boldsymbol{\theta})}{p_q(y_q, \boldsymbol{\theta})}\right]. \end{aligned} \tag{8}$$

If $n_q = n$ for all q, then this bias is $\mathcal{O}(1/n)$ and the variance of $\ln(\hat{L}(\boldsymbol{\theta}))/m$ is $\mathcal{O}(1/(mn))$. For fixed m, the contribution of the square bias to the mean square error (MSE) becomes negligible compared to that of the variance when n is large enough: $\mathcal{O}(n^{-2})$ compared with $\mathcal{O}((mn)^{-1})$. But in practice, n is not always very large and the bias can be important, and it is not reduced when we increase m, in contrast to the variance. To reduce the bias, one can subtract an estimate of (8) to the estimator $\ln(\hat{L}(\boldsymbol{\theta}))/m$. Bastin and Cirillo (2010) examined this strategy in the context

of MC estimation. They observed empirically that the correction eliminated most of the bias without significantly increasing the variance. In the remainder, we focus on the bias and variance of estimators of $\ln L(\boldsymbol{\theta})/m$.

2.3 Multinormal density for β

For some examples in this paper, we assume that $f_{\boldsymbol{\theta}}$ is the multinormal density with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_s)^{\mathrm{t}}$ and covariance matrix $\boldsymbol{\Sigma}$ with elements $\sigma_{i,\ell}$, and that $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This multinormal assumption is frequent in practice, and $\boldsymbol{\Sigma}$ is also often assumed to be diagonal. Here we also consider non-diagonal matrices, i.e., nonzero correlations between the coordinates of $\boldsymbol{\beta}$.

To generate a realization of $\boldsymbol{\beta}$, we first decompose $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{t}}$, and then use inversion:

$$\boldsymbol{\beta} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Z} \tag{9}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_s)^{\mathrm{t}}$, $\mathbf{U} = (U_1, \ldots, U_s)^{\mathrm{t}}$ is a uniform random vector over $(0,1)^s$, $Z_\ell = \Phi^{-1}(U_\ell)$, and Φ is the standard normal distribution function (pdf). A standard way of decomposing $\boldsymbol{\Sigma}$ as $\mathbf{A}\mathbf{A}^{\mathrm{t}}$ is the Cholesky decomposition, but there are many other ways (an infinite number, in fact). The choice of decomposition makes no difference when the U_ℓ's are generated independently by standard MC, but it can have a large impact on the variance when we use RQMC (L'Ecuyer, 2009). A choice that often performs much better with RQMC is the eigen-decomposition used in principal component analysis (PCA) (L'Ecuyer, 2009); it gives $\mathbf{A} = \mathbf{P}\mathbf{D}^{1/2}$ where $\mathbf{D}$ is a diagonal matrix that contains the eigenvalues of $\boldsymbol{\Sigma}$ in decreasing order and the columns of $\mathbf{P}$ are the corresponding unit-length eigenvectors. With this decomposition, the randomness in $\boldsymbol{\beta}$ depends as much as possible on the first few coordinates of $\mathbf{U}$, that is, on the first few coordinates of the RQMC points. We will use this PCA decomposition for all our reported experiments with RQMC. We also tried the Cholesky decomposition and the RQMC variances were typically slightly larger, by a factor between 1 and 2.

3 RQMC Methods

3.1 RQMC and variance bounds

RQMC methods are designed to estimate integrals over the unit hypercube, as in (3), with smaller variance than ordinary MC (Owen, 1998; L'Ecuyer and Lemieux, 2000; L'Ecuyer, 2009; Lemieux, 2009). An RQMC method estimates the integral of a function f over the s-dimensional unit cube

$(0,1)^s$ by evaluating f over a set of n points $P_n = \{\mathbf{U}_0, \ldots, \mathbf{U}_{n-1}\}$ and taking the average,

$$\hat{\mu}_{n,\mathrm{rqmc}} = \frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{U}_i). \tag{10}$$

These points must form an RQMC point set, which means that

(a) P_n covers $(0,1)^s$ very uniformly when taken as a set and
(b) each point $\mathbf{U}_i$ has the uniform distribution over $(0,1)^s$ when taken individually.

Condition (b) ensures that the average $\hat{\mu}_{n,\mathrm{rqmc}}$ is an unbiased estimator of the integral $\mu = \int_{(0,1)^s} f(\mathbf{u})\,\mathrm{d}\mathbf{u}$. We recognize that for (a), we need a precise definition of what "very uniformly" really means. There are in fact many different ways of measuring the uniformity of a point set, and different measures are used for different types of constructions and classes of integrands to obtain error and variance bounds (Niederreiter, 1992; L'Ecuyer, 2009). For this, one usually specifies a class $\mathcal{H}$ of functions f, often a reproducing kernel Hilbert space, and one derives a worst-case bound on the integration error of the form

$$|\hat{\mu}_{n,\mathrm{rqmc}} - \mu| \le D(P_n)V(f) \tag{11}$$

for all $f \in \mathcal{H}$ and any point set $P_n \subset (0,1)^s$, where $D(P_n)$ measures the discrepancy of P_n from the uniform distribution and $V(f) = \|f - \mu\|_{\mathcal{H}}$ measures the variability of f in $\mathcal{H}$ (Hickernell, 1998, 2000; L'Ecuyer, 2009). The definitions of $D(P_n)$ and $V(f)$ must depend on each other, and a definition that makes $V(f)$ smaller will generally make $D(P_n)$ larger, and vice-versa. One special case of (11) is the classical Koksma-Hlawka inequality often associated with QMC methods, where P_n is deterministic (Niederreiter, 1992). When P_n is randomized, we obtain the variance bound

$$\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] = \mathbb{E}[(\hat{\mu}_{n,\mathrm{rqmc}} - \mu)^2] \le \mathbb{E}[D^2(P_n)]\, V^2(f). \tag{12}$$

Then, if $V(f) < \infty$, the variance converges at the same rate as $\mathbb{E}[D^2(P_n)]$ as a function of n. We will return to this bound later.

3.2 Functional ANOVA decomposition

We know that covering the s-dimensional unit hypercube very uniformly requires a number of points n that increases exponentially with s, so accurate high-dimensional integration by RQMC looks hopeless at first sight. Yet empirically, the method works well even in hundreds of dimensions in some cases. The usual explanation is that in those cases, the integrand f can be well

approximated by a sum of low-dimensional functions that are accurately integrated by RQMC, and the residual has small variance (Owen, 1998; L'Ecuyer, 2009). This can be formalized via the *functional ANOVA decomposition* of f, defined as follows. If

$$\sigma^2 = \mathrm{Var}[f(\mathbf{U})] = \int_{(0,1)^s} f(\mathbf{u})\,\mathrm{d}\mathbf{u} - \mu^2 < \infty$$

for $\mathbf{U}$ uniformly distributed over $(0,1)^s$, one can write

$$f(\mathbf{u}) = \mu + \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,\ldots,s\}} f_{\mathfrak{u}}(\mathbf{u}) \tag{13}$$

where $\mathfrak{u}$ denotes an arbitrary subset of coordinates (this is standard notation, not to be confused with $\mathbf{u}$), each $f_{\mathfrak{u}} : (0,1)^s \to \mathbb{R}$ depends only on $\{u_i,\, i \in \mathfrak{u}\}$, the $f_{\mathfrak{u}}$'s integrate to zero and are orthogonal, and the variance admits the corresponding decomposition $\sigma^2 = \sum_{\mathfrak{u} \subseteq \{1,\ldots,s\}} \sigma_{\mathfrak{u}}^2$ where $\sigma_{\mathfrak{u}}^2 = \mathrm{Var}[f_{\mathfrak{u}}(\mathbf{U})]$.

If $\sum_{\mathfrak{u} \in \mathcal{U}} \sigma_{\mathfrak{u}}^2/\sigma^2$ is very close to 1 for a relatively small set $\mathcal{U}$ of the subsets of $\{1,\ldots,s\}$, this means that the approximation of f by $\sum_{\mathfrak{u} \in \mathcal{U}} f_{\mathfrak{u}}$ accounts for most of the MC variance. Then, we can construct the RQMC point set P_n by focusing on the uniformity of its projections over the subsets of coordinates $\mathfrak{u} \in \mathcal{U}$, by giving them an importance in relation with $\sigma_{\mathfrak{u}}^2$, and neglect the other projections. For a special case, define

$$\sigma_r^2 = \sum_{|\mathfrak{u}|=r} \sigma_{\mathfrak{u}}^2,$$

the total variance for the projections of dimension (or order) r, for $r = 1,\ldots,s$. If $\sum_{r=1}^{s'} \sigma_r^2/\sigma^2$ is very close to 1 for some small s', then we can construct P_n by giving a global weight to the uniformity of all projections of order r, for each $r \leq s'$, in relation to the size of σ_r^2.

3.3 Randomly-shifted lattice rules

A *rank-1 lattice rule* with n points in s dimensions is defined as follows (Niederreiter, 1992; Sloan and Joe, 1994). Select a vector $\mathbf{a}_1 = (a_1,\ldots,a_s)$ whose coordinates belong to $\mathbb{Z}_n = \{0,\ldots,n-1\}$, let $\mathbf{v}_1 = \mathbf{a}_1/n$, and define $P_n^0 = \{\mathbf{v} = i\mathbf{v}_1 \bmod 1,\ i = 0,1,\ldots,n-1\}$, where the division and the modulo operation are coordinate-wise. This point set is the intersection of a lattice with the unit hypercube in s dimensions. The a_j are usually taken relatively prime to n, so that the projection of P_n^0 over any of its coordinates contains n distinct points, namely $\{0, 1/n, \ldots, (n-1)/n\}$.

Thus, there is no need to measure the uniformity of the one-dimensional projections. Here we randomize P_n^0 by applying a random shift modulo 1, which consists in generating a single point $\mathbf{U}$ uniformly over $(0,1)^s$ and adding it to each point of P_n^0, modulo 1, coordinate-wise (Cranley and Patterson, 1976; L'Ecuyer and Lemieux, 2000), to obtain P_n. We follow this by a baker's transformation, which replaces each coordinate u of each point by $2u$ if $u < 1/2$ and by $2-2u$ otherwise (Hickernell, 2002; L'Ecuyer, 2009).

The vector $\mathbf{v}_1$ is selected to try to minimize a given discrepancy measure of P_n^0. In this paper, we use the weighted $\mathcal{P}_{2\alpha}$ criterion

$$\mathcal{P}_{2\alpha}(P_n^0) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,\ldots,s\}} \frac{1}{n} \sum_{i=0}^{n-1} \gamma_{\mathfrak{u}} \left[\frac{(-4\pi^2)^\alpha}{(2\alpha)!}\right]^{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} B_{2\alpha}(u_{i,j}), \tag{14}$$

where $\mathbf{u}_i = (u_{i,1}, \ldots, u_{i,s}) = i\mathbf{v}_1 \bmod 1$ is the ith lattice point before the shift, α is a positive integer, the *projection-dependent weights* $\gamma_{\mathfrak{u}}$ are arbitrary positive real numbers, $|\mathfrak{u}|$ is the cardinality of $\mathfrak{u}$, and $B_{2\alpha}$ is the Bernoulli polynomial of order 2α (for example, for $\alpha = 1$ we have $B_2(u) = x^2 - x + 1/6$). This criterion can be motivated as follows. Consider the class $\mathcal{F}_\alpha$ of functions f for which for each subset $\mathfrak{u}$ of coordinates, the partial derivative of order α with respect to these coordinates is square integrable, and the partial derivatives of orders 0 to $\alpha - 2$ of the periodic continuation of f over $\mathbb{R}^s$ are continuous. For $\alpha = 1$, this continuity condition just disappears, but for $\alpha = 2$ the periodic continuation of f must be continuous. It turns out that the square variation of $f \in \mathcal{F}_\alpha$ defined as

$$V^2(f) = \sum_{\mathfrak{u} \subseteq \{1,\ldots,s\}} V^2(f_{\mathfrak{u}}) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,\ldots,s\}} \frac{1}{\gamma_{\mathfrak{u}}(4\pi^2)^{\alpha|\mathfrak{u}|}} \int_{[0,1]^{|\mathfrak{u}|}} \left| \frac{\partial^{\alpha|\mathfrak{u}|}}{\partial \mathbf{u}^\alpha} f_{\mathfrak{u}}(\mathbf{u}) \right|^2 d\mathbf{u}, \tag{15}$$

corresponds to a square discrepancy for which $\mathbb{E}[D^2(P_n)] = \mathcal{P}_{2\alpha}(P_n^0)$, and the variance bound (12) holds for this pair (Dick et al., 2004; L'Ecuyer, 2009). More interestingly, it is known that for any $\alpha > 1$, any $\delta > 0$, and any choices of weights $\gamma_{\mathfrak{u}}$, there exists a sequence of rank-1 lattices for which $\mathcal{P}_{2\alpha}(P_n^0) = \mathcal{O}(n^{-2\alpha+\delta})$, and the corresponding vectors $\mathbf{v}_1$ can be constructed explicitly one coordinate at a time, by a so-called *component-by-component* (CBC) construction method (Dick et al., 2004). This means that for any $f \in \mathcal{F}_\alpha$, it is possible to construct lattice rules for which $\mathrm{Var}[\hat{\mu}_{n,\mathrm{rqmc}}] = \mathcal{O}(n^{-2\alpha+\delta})$. The role of the baker's transformation mentioned earlier is to make the periodic condition of f continuous, so we can have $\alpha = 2$ instead of $\alpha = 1$ when f is smooth enough; see Hickernell (2002); L'Ecuyer (2009) for detailed explanations.

3.4 Selecting the weights

We want to construct lattice rules that minimize (14), but how do we choose the weights $\gamma_{\mathfrak{u}}$? It can be shown that $\mathcal{P}_{2\alpha}(P_n^0)$ is equal to the RQMC variance for a worst-case function f_α^* defined by

$$f_\alpha^*(\mathbf{u}) = \sum_{\mathfrak{u}\subseteq\{1,\ldots,s\}} \sqrt{\gamma_{\mathfrak{u}}} \prod_{j\in\mathfrak{u}} \frac{(2\pi)^\alpha}{\alpha!} B_\alpha(u_j),$$

whose ANOVA variance components are

$$\sigma_{\mathfrak{u}}^2 = \gamma_{\mathfrak{u}} \left[\mathrm{Var}[B_\alpha(U)]\frac{(4\pi^2)^\alpha}{(\alpha!)^2}\right]^{|\mathfrak{u}|} = \gamma_{\mathfrak{u}} \left[|B_{2\alpha}(0)|\frac{(4\pi^2)^\alpha}{(2\alpha)!}\right]^{|\mathfrak{u}|} \stackrel{\text{def}}{=} \gamma_{\mathfrak{u}}(\kappa(\alpha))^{-|\mathfrak{u}|} \tag{16}$$

where $\kappa(\alpha)$ is a constant that depends on α. In particular, we have $\kappa(1) = 3/\pi^2 \approx 0.30396$ and $\kappa(2) = 45/\pi^4 \approx 0.46197$. Acting as if we were integrating this f_α^*, a natural strategy is to adopt weights $\gamma_{\mathfrak{u}}$ given by these formulas, in which the $\sigma_{\mathfrak{u}}^2$ are replaced by estimates. In our experiments, we did estimate those $\sigma_{\mathfrak{u}}^2$ using the algorithm of Sobol' and Myshetskaya (2007) and we used $\alpha = 1$ to compute the weights $\gamma_{\mathfrak{u}}$. We took $\alpha = 1$ instead of $\alpha = 2$ because we observed empirically that in the range of values of n that we used, the variance of the estimator does not decrease faster than $\mathcal{O}(n^{-1.7})$, even with the baker's transformation. In our results, we will denote the lattices constructed based on $\mathcal{P}_{2\alpha}(P_n^0)$ with those weights by *lattice-*$\gamma_{\mathfrak{u}}$.

Estimating all those weights becomes impractical when s increases, because there are $2^s - 1$ variance components to estimate. Note that the weights of one-dimensional projections (for which $|\mathfrak{u}| = 1$) are irrelevant for selecting $\mathbf{v}_1$, because all one-dimensional projections are the same under our assumptions, so there is no need to specify them. This gives s fewer parameters to estimate. Also, multiplying all weights by a given constant has no impact on the selection of $\mathbf{v}_1$, since it does not change the relative importance of the projections. To reduce the number of parameters to estimate (and at the same time reduce the likelihood of overfitting), we may bundle the projections $\mathfrak{u}$ in subgroups, and attach one weight to each subgroup. For example, we can have *order-dependent weights*, where all projections of cardinality (or order) r are given the weight γ_r, for $r = 2, \ldots, s$. For this, we can estimate $\sigma_r^2 = \sum_{\{\mathfrak{u}:|\mathfrak{u}|=r\}} \sigma_{\mathfrak{u}}^2$, and plug it in the formula $\gamma_r = C(\kappa(\alpha))^r \sigma_r^2 / \binom{s}{r}$, where C is an arbitrary positive constant and $\binom{s}{r}$ is the number of projections of order r. This gives $s - 1$ parameters to estimate. In our results, we will use *lattice-order* to refer to these rules. To reduce this number even further, we can simply assume that $\gamma_r = C\gamma^r$ for all r, for some constant γ, and estimate a γ that best fits this model, for instance by

fitting a linear regression model of the form

$$r \ln \kappa(\alpha) + 2 \ln \sigma_r = \ln C + r \ln \gamma + \epsilon_r$$

by finding C and γ that minimize $\sum_{r=2}^{\infty} \epsilon_r^2$. We call the resulting weights *geometric order-dependent weights*. In our experiments, we will end up taking $\gamma = 0.1$, $\gamma = 0.25$, $\gamma = 0.5$, and we refer to the corresponding rules by *lattice-0.1*, *lattice-0.25*, and *lattice-0.5*. Note that multiplying γ by some factor here is equivalent to multiplying $\kappa(\alpha)$ in (16) by the same factor.

The integral (5) that we estimate by RQMC is in fact an average of m integrals, one per individual. For this, it is most convenient to use the same lattice rule, with independent randomizations, across individuals. Another possibility is to use a different lattice for each individual, in which case the lattice parameters could conceivably be selected based on an ANOVA analysis and specific set of weights γ_u for each individual. That is, instead of adopting a one-lattice-fits-all solution, each individual could have his own specially tailored lattice. This is unlikely to be practical, but we have nevertheless explored the potential MSE improvement that could be achieved by doing this, by some empirical experiments with one example (based on real data), in Section 4.3. As we will see, the gain was very small. Note that for classical deterministic QMC methods, it is it is customary to use a different point sets for each individual (otherwise, the "positive dependence" across individuals typically increases the error on the average). With RQMC, we can also do it, but we do not have to, because the independent randomizations remove the dependence (L'Ecuyer and Lemieux, 2000).

We also considered for comparison the Korobov lattices tabulated in L'Ecuyer and Lemieux (2000), whose parameters were selected based on the $M_{32,24,16,12}$ criterion defined in that paper, and which accounts for projections in up to 32 dimensions over successive coordinates, and a selected set of projections of order 2, 3, and 4 only. Also worthy of note is that only the "worst" projection of a lattice contributes to this criterion, whereas (14) is a weighted average over all projections. These point sets were constructed with no particular application in mind. We refer to them by *lattice-*$M32$.

3.5 Sobol' nets

A Sobol' net with $n = 2^k$ points in s dimensions, for some positive integer k, contains the first n points of a Sobol' sequence in s dimensions (Sobol', 1967; Lemieux, 2009). These points are defined by $w \times k$ binary matrices $\mathbf{C}_1, \ldots, \mathbf{C}_s$ called the *generator matrices*, where w is the number of bits retained in the binary expansion of the points. To define the ith point $\mathbf{u}_i$, for $i = 0, \ldots, 2^k - 1$,

we write the digital expansion of i in base 2 and multiply the vector of its digits by $\mathbf{C}_j$, modulo 2, to obtain the digits of the binary expansion of the jth coordinate of $\mathbf{u}_i$. The matrices $\mathbf{C}_j$ are not unique; they depend on the choice of *direction numbers*, which determine their columns. For our experiments here, we have used the default direction numbers of SSJ (L'Ecuyer, 2008), taken from (Lemieux et al., 2004). We randomize our Sobol' nets by a left matrix scramble followed by a digital random shift. The left matrix scramble multiplies each matrix $\mathbf{C}_j$ (modulo 2) by a lower triangular binary matrix $\mathbf{M}_j$ with 1's on the diagonal and random bits below the diagonal (Owen, 2003). The digital random shift generates a single point $\mathbf{U}$ uniformly over $(0,1)^s$ and adds each digit of the binary expansion of each of its coordinate to the corresponding digit of each point, modulo 2 (L'Ecuyer and Lemieux, 1999; Owen, 2003).

3.6 Halton points

We will also try P_n defined as the first n points of randomized versions of the Halton sequence in s dimensions, motivated by the fact that these Halton points are popular in discrete choice modeling and analysis. We randomize them by random shifts modulo 1, independent across individuals, and we refer to these points as shifted *Halton points*. In some cases, we also compare with the more traditional practice of defining P_n for individual q as the points $(q-1)n-1$ through qn of the Halton sequence, i.e., individual 1 has the first n points, individual 2 has the next n points, and so on (Train, 2000; Bhat, 2001, 2003b). This strategy is quite standard and appropriate when using deterministic QMC sequences, for the reasons mentioned earlier for lattice rules, but here we also apply a random shift modulo 1 to the entire sequence to obtain an unbiased estimator. We will refer to this as a shifted *Halton sequence* to distinguish from the case where we re-use the same Halton points as described above, even though formally speaking both are Halton points taken from the Halton sequence. Faure and Lemieux (2009) provide a recent extensive study of various scrambles and randomizations for the Halton sequence, but we did not implement these methods.

4 Numerical experiments

4.1 General experimental setting

We report experimental results for two examples. The first one uses artificially generated data and has two variants. The second one is based on real-life data. For each example, we selected a parameter value $\boldsymbol{\theta}$ not far from the optimizer of the log-likelihood. Because we want to use the same lattices for all individuals, we constructed lattices adapted to an "average" individual.

To do that, for each $\mathfrak{u}$ of cardinality $|\mathfrak{u}| \leq 6$, we estimated $\sigma^2_{\mathfrak{u}}$, defined as the average over the m individuals q of the variance components $\sigma^2_{q,\mathfrak{u}}$ in the ANOVA decompositions of the functions $f_q(\mathbf{u}) = L_q(y_q,\, h(\boldsymbol{\theta}, \mathbf{u}))$, which are the integrands that correspond to the estimators (4) (or their extension to multiple choices per individual). This heuristic procedure is equivalent to constructing lattices adapted to the integration of the average over q of the f_q's, that is, for simulating an estimator defined as in (6), but without the logarithm on the right-hand side. We have to use such a heuristic because the estimator (6) is not expressed in the same form as (10) with the $\mathbf{U}_i$ uniformly distributed over $(0,1)^s$, so the ANOVA decomposition does not apply to it in the form described in (13). We estimated the variance components using the algorithm of Sobol' and Myshetskaya (2007) with 5000 independent RQMC runs (or more in some cases until sufficient precision was reached) with 16381 lattice points, for each $\mathfrak{u}$. Computing them for all $\mathfrak{u}$ would have been too expensive for large s. Moreover, typical values of $\sigma^2_{\mathfrak{u}}$ decrease quickly when $|\mathfrak{u}|$ increases, and they become much more difficult to estimate accurately, because the relative error of their estimators becomes too large. We selected our weights $\gamma_{\mathfrak{u}}$, γ_r and γ^r based on those estimates of the $\sigma^2_{\mathfrak{u}}$, as explained earlier. We also estimated the variance components $\sigma^2_{q,\mathfrak{u}}$ for selected individuals, in order to construct specialized lattices based on specific weights for each of those individuals.

For MC and each RQMC method, and each n considered, we computed 5000 independent realizations of the estimator (6) of (5), then computed the empirical mean and variance of these realizations. We estimated the bias using the approximation in (8), based on the 5000 independent realizations of (4) for each q. For MC and for the lattice rules, the values of n considered were $n = 31, 67, 127, 257, 521, 1021, 2039, 4093, 8191$ and 16381 (each one is a prime number close to a power of two). For the Sobol' nets and Halton points, we considered for n all powers of two from $n = 2^5 = 32$ to $n = 2^{14} = 16384$, and matched each one with its nearby prime n for lattice rules.

We expect the variance and the bias of (6) with RQMC to behave approximately as $\mathrm{Var}[\ln(\hat{L}(\boldsymbol{\theta}))/m] \approx V_0 m^{-1} n^{-\nu_1}$ and $\mathrm{Bias}[\ln(\hat{L}(\boldsymbol{\theta}))/m] \approx B_0 n^{-\nu_2}$, where the constants V_0, B_0, ν_1 and ν_2 depend on the RQMC method. Our numerical results will confirm this. For MC, we know that $\nu_1 = \nu_2 = 1$, at least for large n. We wanted to fit this model in the range of values of n of practical interest, say from 2^8 to 2^{14}, instead of in the limit when $n \to \infty$ (where the parameters might differ). We did this by applying linear regression to the logarithm of the observations for $n \geq 2^8 = 256$. We discarded the smaller values of n because the exponents ν_1 and ν_2 were sometimes changing in that range and this was distorting the results. These models for the variance and bias give the MSE expression:

$$\mathrm{MSE}[\ln(\hat{L}(\boldsymbol{\theta}))/m] \approx V_0 m^{-1} n^{-\nu_1} + B_0^2 n^{-2\nu_2}. \tag{17}$$

Assuming that $1 \leq \nu_1 < 2\nu_2$ (which is typical), we see from this expression that for small enough n and large enough m, the contribution from the square bias dominates, and that this contribution becomes negligible compared with that of the variance when n becomes large enough. This means that the MSE decreases roughly as $\mathcal{O}(n^{-2\nu_2})$ for small n and $\mathcal{O}(n^{-\nu_1})$ for large n. We define the *MSE reduction factor* of an RQMC estimator with n points, with respect to an MC estimator based on n independent simulation runs, as the MSE of the MC estimator divided by that of the RQMC estimator. We estimate this factor using the fitted versions of (17). When n is large, it should increase approximately as $\mathcal{O}(n^{\nu_1 - 1})$. But for small or moderate n, it is not always increasing in n, because of the effect of the bias in (17).

4.2 Examples with synthetic data

Our first set of numerical experiments are with an artificial data set generated from a known model, as in Sivakumar et al. (2005). We consider three values of s, namely $s = 5$, $s = 10$, and $s = 15$. For each s, we take $|\mathcal{A}(q)| = 4$ for each q, for $m = 2000$ individuals. The coordinates $x_{q,j,\ell}$ of the attribute vectors are independent $N(1,1)$ random variables (normal with mean 1 and variance 1) for alternatives $j = 1, 2$ and $N(0.5, 1)$ for alternatives $j = 3, 4$. The s coordinates of $\boldsymbol{\beta}_q = (\beta_{q,1}, \ldots, \beta_{q,s})$ are also independent $N(1,1)$ random variables. Then we repeated the experiments for $\boldsymbol{\beta}_q$ multinormal with the same $N(1,1)$ marginals, but with correlations of 0.3 across its s components. That is, the covariance matrix has 1's on the diagonal and 0.3 everywhere else. We refer to these two distributions as the *independent* and *correlated* cases. After generating the artificial data from the model in a first stage, in the second stage we estimate the log-likelihood function for this data set, at the value of $\boldsymbol{\theta}$ used to generate the data, by simulation. Here we give a representative subset of the results and a qualitative summary. More detailed results can be found in the online appendix.

The distribution of the ANOVA variances among the different projections vary significantly across individuals, as illustrated in Figures 1 and 2 for selected individuals, for the independent case with $s = 5$ and $T_q = 1$. For example, the proportion of variance contributed by the projections of order 3 and higher is less than 2% for individual $q = 1$ and more than 9% for individual $q = 4$. We thus expect RQMC to be more effective for the first individual than for the fourth one. Figure 3 confirms this; it shows the variance of the likelihood estimator for these two individuals, with a specialized lattice for each individual (lattice-γ_u) and for other RQMC point sets. As it turns out, all the lattice rules (except lattice-$M32$) as well as the Sobol' nets have comparable performances. They perform much better than the Halton points, and they provide a large MSE reduction over MC. The lattice-$M32$ rules have an erratic behavior; they come close to the good

r	fraction of total variance	γ_r
1	0.83	1
2	0.12	0.023
3	0.035	0.0019
4	0.0079	0.00027
5	0.0011	0.000055

Table 1: Fraction of ANOVA variance per projection order (r), with associated order-dependent weights, for the independent case with $s = 5$ and $T_q = 1$.

ones for some values of n and they do much worse for other values. The explanation might be that for certain values of n, we were more lucky with the uniformity of the most important projections not considered in the $M_{32,24,16,12}$ criterion. For the other rank-1 lattices, no choice of weights seems to offer a solid advantage over the other choices and their differences in MSE reductions fluctuate somewhat randomly (but not too much) across values of n. For example, the ratio of variances for lattice-0.1 and lattice-γ_{u} is larger than 1 on average but ranges from 0.5 to 3.5 for the different values of n, and is not at all monotone in n. These ratios are similar for the other choices of weights (compared with lattice-γ_{u}). Thus, specializing the lattices to the different individuals brings a modest improvement that can hardly be qualified as significant here.

Figure 1 (bottom) shows the average over all individuals of the ANOVA variances, for the independent case with $s = 5$ and $T_q = 1$. We see that all projections of the same order contribute almost the same variance. This symmetry is due to the homogeneity of the synthetic population. It suggests that order-dependent weights are perfectly appropriate for this situation. The ANOVA variances regrouped by projection order, with the associated order-dependent weights, are shown in Table 1. The variance, bias, and MSE on the log-likelihood function, obtained with the lattices constructed based on those weights and with other point sets, are given in Figure 5. In particular, it can be seen from the bottom right plot that the share of MSE contributed by the square bias decreases faster with RQMC than with MC when n increases. The reason for this is that the ratio of the square bias to the variance decreases as n^{-1} with MC whereas it decreases as $n^{-(2\nu_2-\nu_1)}$ with RQMC, and, at least in our examples, we have $2\nu_2 - \nu_1 > 1$. For small n (say less than about 200), the MSE reduction is mostly due to the reduction of the bias.

Figure 4 shows the ANOVA variances σ_r^2, averaged over individuals and regrouped by projection order r, for $s = 5$ and $s = 10$. We find that there is not much difference between the independent and correlated cases, except that the lower-order projections have a slightly larger share of the variance for the correlated case, so we expect RQMC to work slightly better in that case. Their share also decreases (so we expect RQMC to be less effective) when s or T_q increases. By fitting the parameter γ for the exponential order-dependent weights, in the independent case,

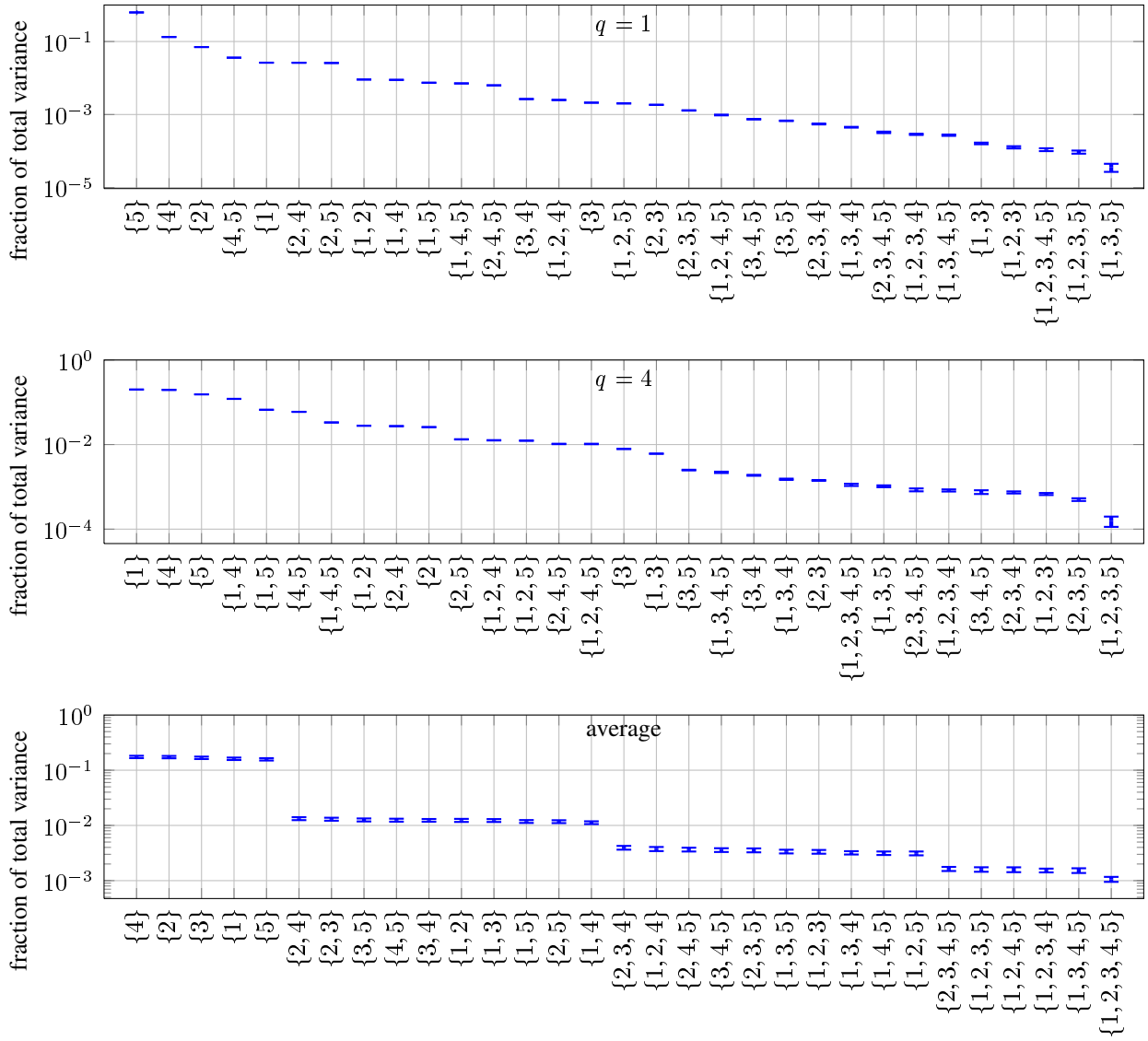

Figure 1: Fractional ANOVA variances of the conditional likelihood function for a single individual $q = 1$ (top) and $q = 4$ (middle) and for the average over all individuals (bottom), sorted by decreasing order, for the independent case with $s = 5$ and $T_q = 1$. The projections are listed on the horizontal axis, and their fraction of total variance is plotted along the vertical axis. The limits of each vertical bar correspond to a normal confidence interval at 95 % on the variance estimate, the standard deviation of which was estimated across random draws.

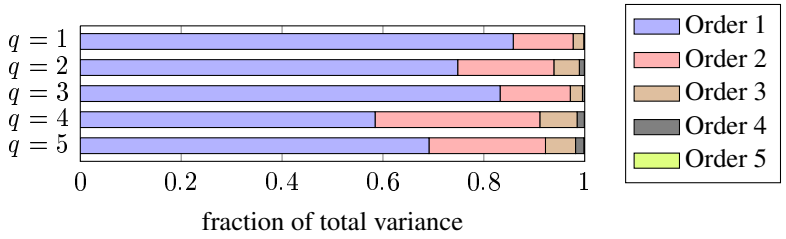

Figure 2: Fractional ANOVA variance per order for the conditional likelihood function of selected individuals, for the independent case with $s = 5$ and $T_q = 1$.

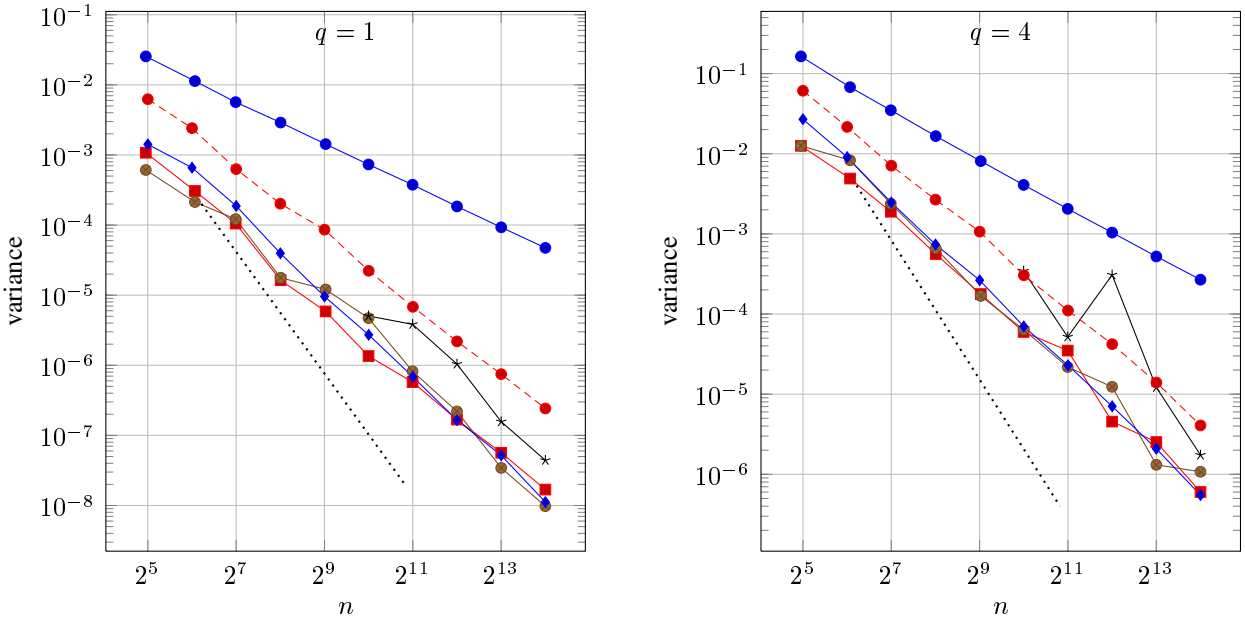

Figure 3: Estimated variance of the MC and RQMC estimators of the log-likelihood of a single individual for the independent case with $s = 5$ and $T_q = 1$, for individuals $q = 1$ (left) and 4 (right), using MC (—●—), lattice-γ_u (—■—), lattice-0.1 (—●—), lattice-$M32$ (—⋆—), Sobol' nets (—♦—), and Halton points (- -●- -). For lattice-order and lattice-0.5, the variances are very similar to those of lattice-γ_u and lattice-0.1, and we do not show them to reduce the number of curve superpositions. The results for the Halton sequence (not shown here) are practically undistinguishable from those for the Halton points. The dotted line indicates the n^{-2} slope, for reference.

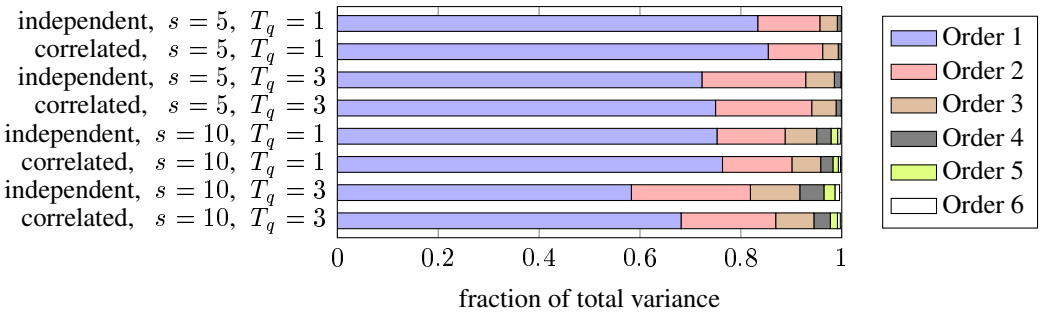

Figure 4: Average over all individuals of the fractional ANOVA variance per projection order r (up to $r = 6$), for $s = 5$ and 10, with $T_q = 1$ and 3, for the independent and correlated cases.

we obtained $\gamma = 0.13$ for $s = 5$ and $\gamma \approx 0.09$ for $s = 10$. This provides support for the lattice-0.1 rules.

Table 2 summarizes the estimated exponents ν_1 and ν_2 for various RQMC point sets. We see that ν_1 and ν_2 are generally close to each other, and for the range of values of n considered here, they are smaller when s or T_q are larger, as expected. These exponents are only part of the story; the variance and MSE reduction factors also depend on the constants V_0 and B_0. These constants are much larger for the Halton points than for the other RQMC point sets, as indicated by the vertical distance between the corresponding lines in Figures 3 and 5. Table 3 shows the MSE reduction factors approximated by (17), evaluated at 16381. Sobol' nets generally give the best MSE reduction factors in low dimensions and when $T_q = 1$, but rank-1 lattices win in higher dimension and with $T_q > 1$. In general, the MSE reduction factors are much higher in situations where the low-order projections have a large share of the variance (smaller s, smaller T_q, and higher correlation), as expected. In fact, the Sobol' nets generally do a bit better in situations where the projections of order two have a larger share of the variance (smaller s and higher correlation). This is consistent with the fact that the parameters we use for those nets were selected mainly on the basis of the uniformity of the two-dimensional projections.

As we said earlier, the MSE reduction factor does not always increase in n, even if $\nu_1 > 1$ and $\nu_2 > 1$. This is illustrated in Figure 6 (top), which shows curves of the fitted MSE and MSE reduction factors with MC and RQMC, for the independent case with $s = 10$. Here, when n is in the range from 200 to 1000 (approximately), the square bias dominates the MSE for MC, and it decreases faster in that range than the MSE for the RQMC method. This is an example where, in the range of values of n used in practice, the bias is neither dominant nor negligible with respect to the variance. Figure 6 (bottom) illustrates the fact that RQMC can remain efficient when the population size m grows. We see in the figure that the MSE reduction factor increases with m.

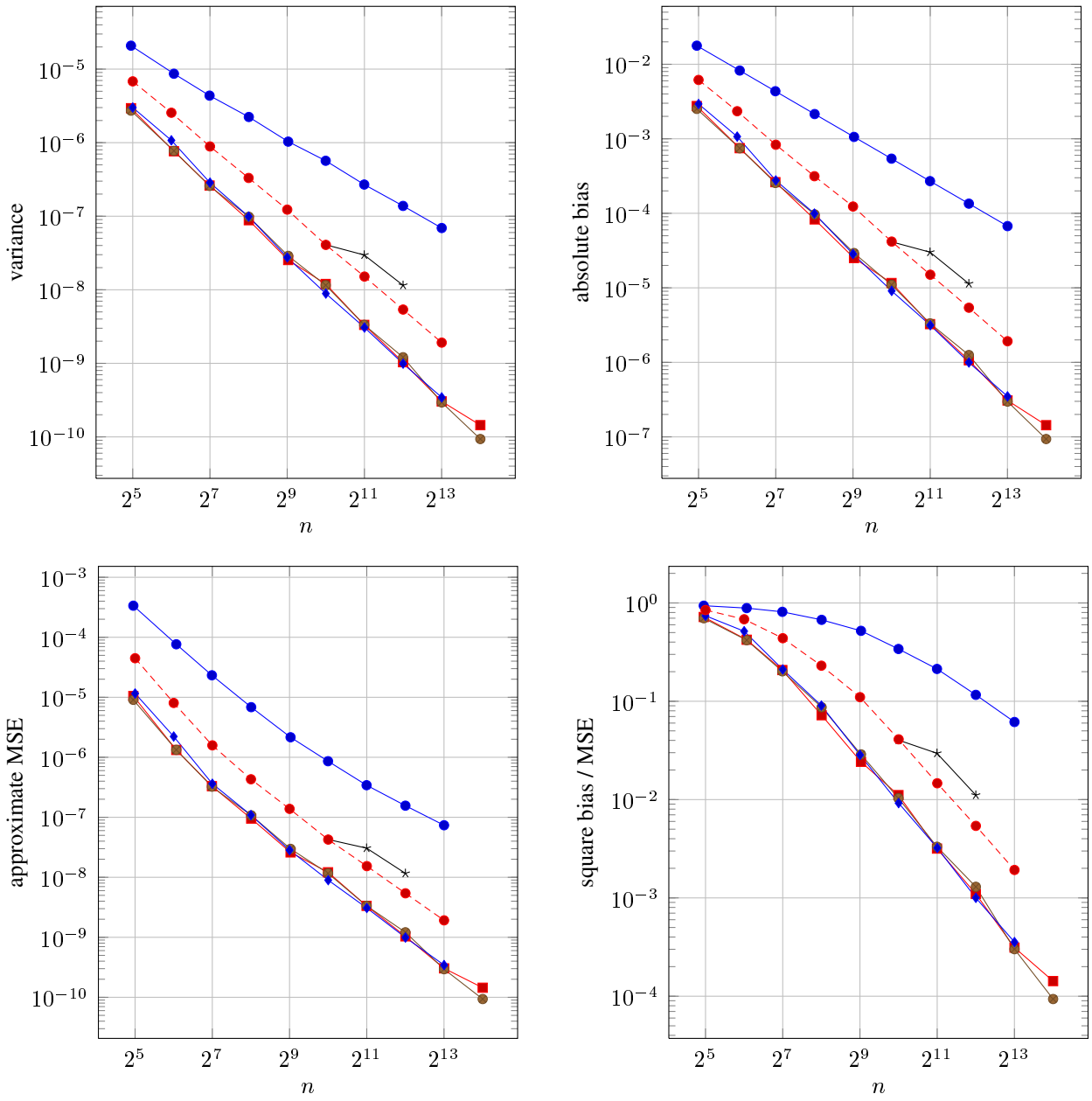

Figure 5: Estimated variance (top left), bias (top right), MSE (bottom left), and fraction of the MSE contributed by the square bias (bottom right) of the MC and RQMC estimators of the log-likelihood function for the independent case with $s = 5$ and $T_q = 1$, using MC (—●—), lattice-γ_u (—■—), lattice-0.1 (—●—), lattice-$M32$ (—⋆—), Sobol' nets (—♦—) and Halton points (--●--). The dotted line indicates the n^{-2} slope, for reference. The results for lattice-order and lattice-0.5 (not shown here) are very similar to those of lattice-γ_u and lattice-0.1, while those for the Halton sequence are almost indentical to those for the Halton points. The dotted line indicates the n^{-2} slope, for reference.

Independent case

Point set	$s = 5$		$s = 10$		$s = 15$	
	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$
Monte Carlo	1.00 (1 %)	1.01 (0.4 %)	1.02 (0.6 %)	0.97 (0.6 %)	1.02 (0.9 %)	0.98 (0.9 %)
	1.00 (0.03 %)	1.00 (0.03 %)	1.00 (0.04 %)	1.00 (0.2 %)	1.00 (0.07 %)	1.00 (0.1 %)
lattice-0.1	**1.67 (2 %)**	1.50 (5 %)	**1.30 (2 %)**	**1.11 (2 %)**	1.24 (2 %)	**1.08 (0.7 %)**
	1.66 (2 %)	1.51 (4 %)	**1.29 (0.9 %)**	**1.17 (0.7 %)**	1.20 (0.8 %)	**1.13 (0.2 %)**
Sobol' nets	1.62 (2 %)	**1.53 (1 %)**	1.29 (2 %)	1.10 (1 %)	1.19 (3 %)	1.05 (3 %)
	1.62 (2 %)	**1.54 (1 %)**	1.27 (2 %)	1.15 (1 %)	1.16 (3 %)	1.08 (2 %)
Halton points	1.49 (0.7 %)	1.35 (1 %)	1.28 (0.8 %)	1.06 (1 %)	**1.26 (2 %)**	1.04 (3 %)
	1.48 (0.9 %)	1.36 (0.7 %)	1.26 (0.8 %)	1.10 (1 %)	**1.22 (2 %)**	1.07 (2 %)

Correlated case

Point set	$s = 5$		$s = 10$		$s = 15$	
	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$
Monte Carlo	1.02 (0.5 %)	1.02 (0.8 %)	1.01 (0.5 %)	1.01 (0.5 %)	1.01 (0.6 %)	1.03 (0.4 %)
	1.00 (0.02 %)	1.00 (0.02 %)	1.00 (0.02 %)	1.00 (0.03 %)	1.00 (0.03 %)	1.00 (0.04 %)
lattice-0.1	**1.73 (3 %)**	**1.62 (3 %)**	**1.36 (0.4 %)**	**1.26 (1 %)**	1.27 (1 %)	**1.18 (0.8 %)**
	1.73 (3 %)	**1.62 (3 %)**	**1.36 (2 %)**	**1.25 (1 %)**	1.26 (1 %)	**1.17 (0.5 %)**
Sobol' nets	1.70 (1 %)	**1.62 (1 %)**	1.31 (2 %)	1.23 (1 %)	1.21 (1 %)	1.15 (1 %)
	1.70 (1 %)	**1.62 (1 %)**	1.31 (2 %)	1.23 (1 %)	1.21 (1 %)	1.13 (1 %)
Halton points	1.52 (1 %)	1.41 (0.9 %)	1.32 (0.6 %)	1.17 (1 %)	**1.35 (3 %)**	1.15 (4 %)
	1.52 (1 %)	1.41 (0.7 %)	1.31 (0.3 %)	1.17 (1 %)	**1.32 (4 %)**	1.11 (3 %)

Table 2: Estimates of ν_1 and ν_2 for the independent case (top table) and the correlated case (bottom table). Each entry contains ν_1 above ν_2, with the relative half-width of a 95% confidence interval in parentheses.

Independent case

Point set	$s = 5$		$s = 10$		$s = 15$	
	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$
Halton points	52	14	8.3	2.6	3.7	1.6
lattice-0.1	**352**	43	**22**	**4.8**	**9.5**	**3.2**
Sobol' net	335	**53**	16	3.8	5.7	2.2

Correlated case

Point set	$s = 5$		$s = 10$		$s = 15$	
	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$	$T_q = 1$	$T_q = 3$
Halton points	72	20	14	4	9.9	3.1
lattice-0.1	583	76	**39**	**7.7**	**16**	**4.4**
Sobol' net	**592**	**89**	28	5.9	12	3.3

Table 3: MSE reduction factors with respect to MC, approximated by (17), evaluated at $n = 16381$, for the independent case (top table) and the correlated case (bottom table).

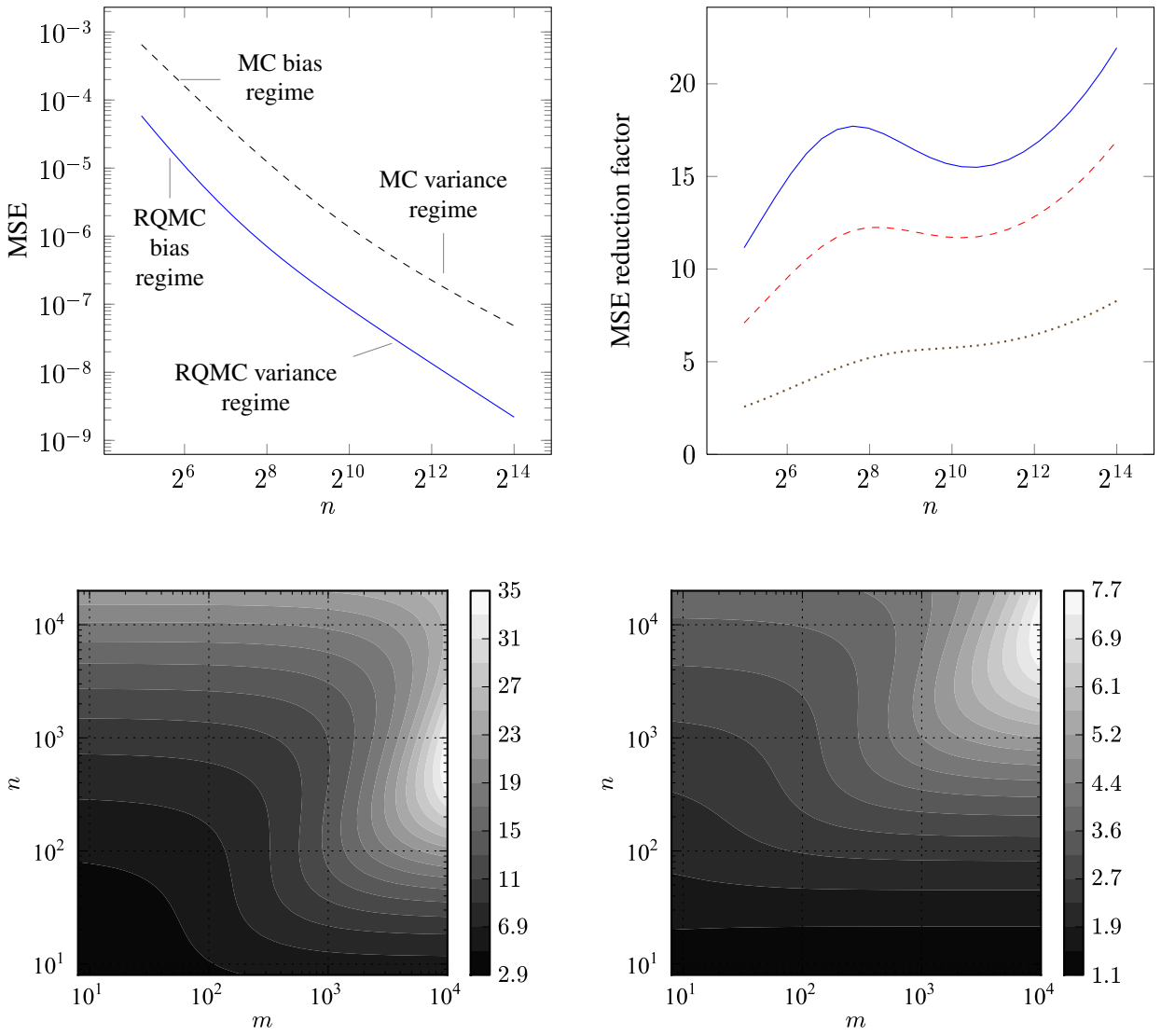

Figure 6: Top left: fitted MSE reduction factor showing the bias and variance regimes, for the independent case with $s = 10$, $T_q = 1$, using MC and RQMC (lattice-0.1 rules). Top right: MSE reduction based on the fitted MSE, for the same case, using lattice-0.1 (——), Sobol' nets (- - -) and Halton points (·····). Bottom: fitted MSE reduction factor as a function of m and n, for the independent case with $s = 10$, with $T_q = 1$ (left) and $T_q = 3$ (right), using lattice-0.1 rules.

4.3 An example with real-life data

For our second example, we consider behavioral data collected in April 2008 at the Baltimore/Washington International (BWI) airport, on airport ground access with automated vehicle technology (called cybercars) (Cirillo and Xu, 2010). The respondents were intercepted in the waiting area of the airport and the responses recorded during a face-to-face interview. The final sample contains information from 274 respondents. Both Revealed Preference (RP) data and Stated Preferences (SP) information were collected. The SP experiment includes 2 parts: a between-mode experiment (SP game 1) and a within mode experiment (SP game 2). SP game 1 is mainly about ground access mode choice and includes the hypothetical cybercar service as long as three other existing modes (car, transit and taxi). SP game 2 proposes two different cybercar services over which the respondents are called to express their preferences. In each game the respondent was presented with 9 scenarios; attribute level of variations were based upon the respondents real trip to the airport as reported in the RP questionnaire. We therefore have a total of 2466 observations. In this paper we make use of just SP game 2; the variables that describe the service in this game are described in Table 4. The following variables enter the SP game 2 model: (a) cybercar waiting time (ranging from 10, 15 and 20 minutes) (b) cybercar travel cost (USD) (c) passenger being dropped at the terminal, (d) fully automated driven cybercar, (e) human driven cybercar, and (f) guideway cybercar.

variables	possible levels
1. dropping area	terminal building, parking lot
2. maneuvering system	fully automated, human driver with ITS, human driver
3. waiting time	5, 10, 15, 20 (in minutes)
4. travel Cost	70% of taxi, 85% of taxi, same as taxi
5. track structure	guideway, grade with rubber tire

Table 4: The variables and their admissible levels, for SP game 2

A number of parametric models for the distributions were estimated and compared, and the retained model assumes that the waiting time distribution parameters are fixed across individuals, that the cost is lognormal, and that the remaining service-level variables are normally distributed. These distributions of the components of $\boldsymbol{\beta}_q$ are given in Table 5, where $N(\mu, \sigma^2)$ and $\ln N(\mu, \sigma^2)$ refer the the normal and lognormal distributions, respectively, with parameters μ and σ^2. The first three components of $\boldsymbol{\beta}_q$ have constant values, so we simulate only the last five, to which we assign the indexes 1 through 5. Thus, $s = 5$.

Like for the example of Subsection 4.2, the distribution of the ANOVA variances among the

coordinate index	distribution
	constant $= -0.6141158$
	constant $= -1.0036583$
	constant $= -1.7356732$
1	$\ln N(-1.9953313, 1.8167162)$
2	$N(1.7088261, 1.5988351)$
3	$N(-0.23137212, 1.1745619)$
4	$N(0.13015642, 0.71851975)$
5	$N(-0.10063324, 1.042506)$

Table 5: Distributions of the components of $\boldsymbol{\beta}_q$ used in the simulation experiments with the real data.

different projections varies across individuals, as shown in Figures 7 and 8 for selected individuals (the coordinate indices correspond to those in Table 5). For individual $q = 116$, less than 10 % of the total variance goes in projections of order $r \geq 3$, whereas this percentage is more than 45% for individual $q = 79$. Figure 9 shows the variance of the likelihood estimator for constructed lattices and other point sets for these two individuals. As expected, RQMC is more effective for individual $q = 116$ than for $q = 79$. The respective performances of the different point sets follow the same pattern as in the example with synthetic data, and here too, no choice of weights for lattices clearly stands out for all values of n. For example, the ratio of variances for lattice-0.1 to those for lattice-γ_u is larger than 1 on average but ranges from 0.4 to 2.7 for the different values of n, and is not at all monotone in n.

The average over all individuals of the ANOVA variances (estimated here with $60,000$ independent replications to obtain sufficient precision), are given in Figure 7 (bottom). Unlike in the example with synthetic data, not only the variances are not uniform across all projections of the same order, but neither do they consistently decrease with projection order. This suggests that order-dependent or geometric weights are probably not ideal in this case, but we will see that they nevertheless perform well empirically. Table 6 shows the ANOVA variances regrouped per projection order r, and the corresponding order-dependent weights. If we insist on exponentially-decreasing weights, the best fit for the parameter γ is $\gamma = 0.077$, which is again not far from 0.1. This gives support to the lattice-0.1 rules. The variance, bias, and MSE on the log-likelihood function, obtained with the lattices constructed based on those weights and with other point sets, are plotted in Figure 10. Despite the fact that order-dependent weights and geometric weights appear less suited for this example than in the previous example with synthetic data, the lattices constructed with these weights offer comparable efficiency as the lattice-γ_u rules. The ratio of the MSE obtained with lattice-γ_u to that obtained with lattice-order, lattice-0.25 or lattice-0.5 rules

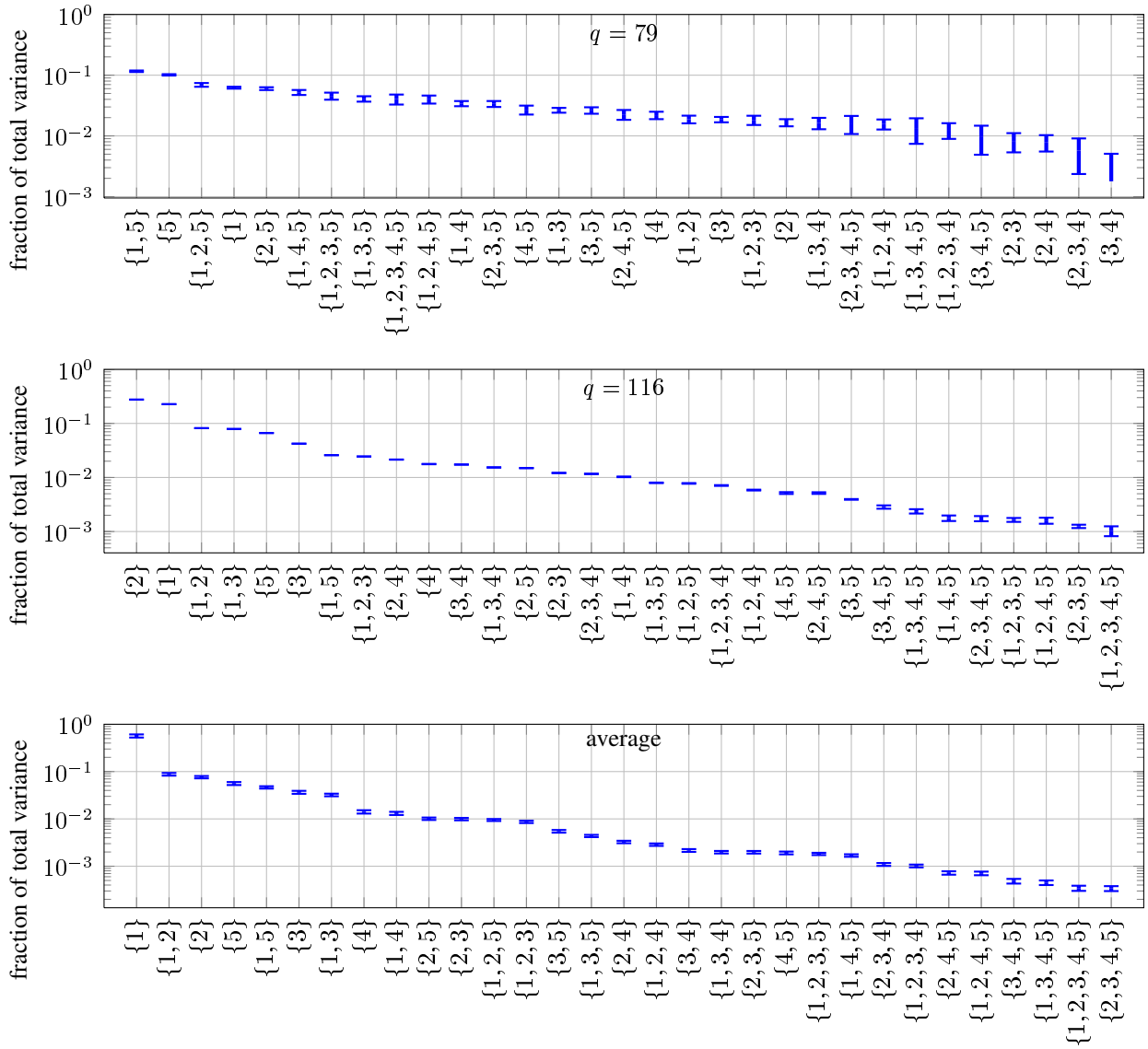

Figure 7: Similar to Figure 1, but for the real-life data, with individuals $q = 79$ (top) and $q = 116$ (middle), and for the average over all individuals (bottom).

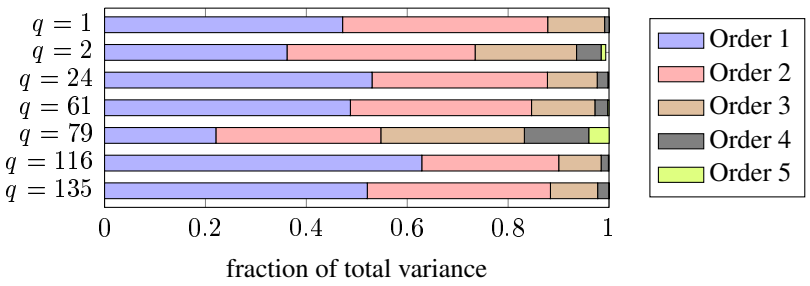

Figure 8: Fractional ANOVA variance per order for the conditional likelihood function of selected individuals, for the example with real-life data.

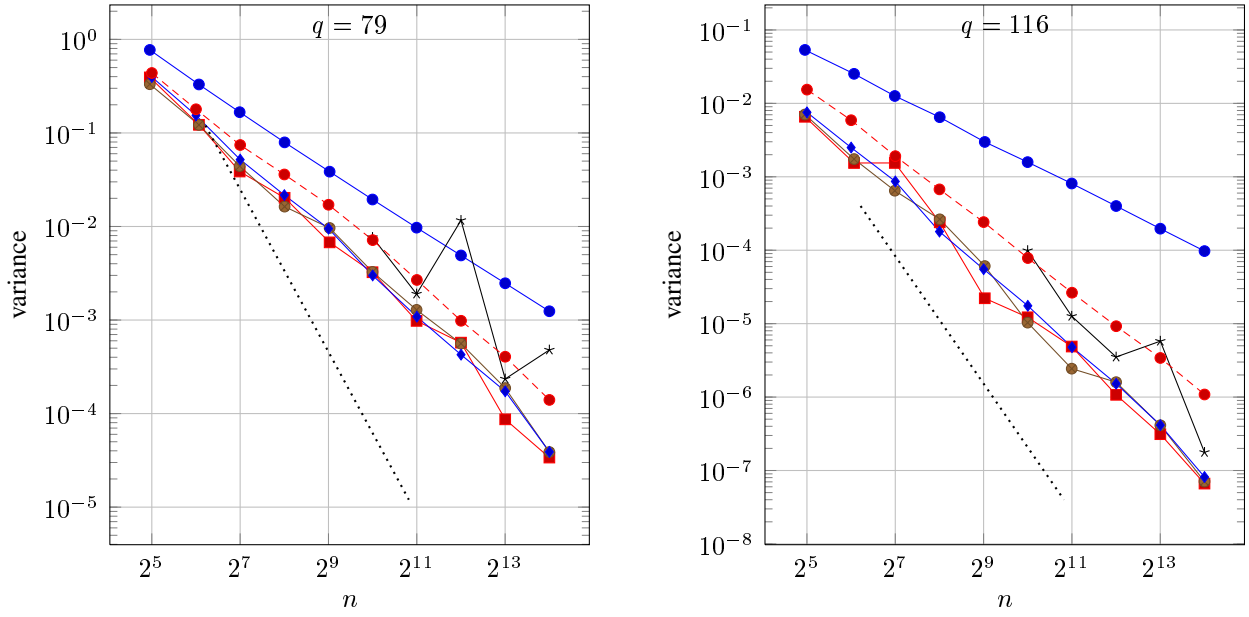

Figure 9: Estimated variance of the MC and RQMC estimators of the log-likelihood of a single individual for the example with real-life data, for individuals $q = 79$ (left) and 116 (right), using MC (—●—), lattice-γ_u (—■—), lattice-0.1 (—●—), lattice-$M32$ (—⋆—), Sobol' nets (—◆—), and Halton points (- -●- -). The dotted line indicates the n^{-2} slope for reference.

r	fraction of total variance	γ_r
1	0.75	1
2	0.21	0.043
3	0.033	0.0021
4	0.0043	0.00016
5	0.0034	0.00002

Table 6: Fractional ANOVA variance per projection order (r) associated order-dependent weights computed with (16).

point set	$n = 4093$ (4096)	$n = 8191$ (8192)	$n = 16381$ (16384)
lattice-γ_u	23	32	30
lattice-order	**27**	30	54
lattice-0.1	19	31	48
lattice-0.25	17	32	46
lattice-0.5	**27**	**35**	**59**
Sobol' net	**27**	28	49
Halton points	8	9	11
Halton sequence	7	9	11

Table 7: Observed MSE reduction factors at two values of $n = 4093, 8191$ and 16381 for lattice rules, and $n = 4096, 8192$ and 16384 for Sobol' nets and Halton points.

range from 0.7 to 1.5, and from 0.9 to 1.1 with lattice-0.1 rules. Figure 11 shows plots of the fitted MSE reduction factors with respect to MC, as a function of n, for several RQMC point sets. The actual observed MSE reduction factors at a few values of n are given in Table 7. These results point to the robustness of lattice rules constructed with criterion (14) with weights of a simple form, such as lattice-0.5 rules.

4.4 Optimization with RQMC for the real-life data

While our main target in this paper was to develop a better understanding of RQMC methods for the evaluation of choice probabilities, the ultimate goal should be to improve parameters estimation. As an empirical test of the improvement provided by our randomized lattice rules for this estimation, we generated 50 independent realizations of the log-likelihood function estimator (6), using $n = 1021$ independent draws per individuals (standard MC) for each realization. Then we maximized each of those functions with respect to $\boldsymbol{\theta}$, using a modified version of AMLET (Bastin et al., 2006a), and we computed the sample mean and variance of these 50 optimizers. The vector $\boldsymbol{\theta}$ contains 13 parameters: three constants for the waiting times, and the two parameters μ and σ for each of the five normal or lognormal distributions. We repeated the same experiment using the lattice-0.5 rule with $n = 1021$ points for each individual, with independent random shifts across individuals and across the 50 runs. Table 8 reports the empirical means of the 13 parameter estimates

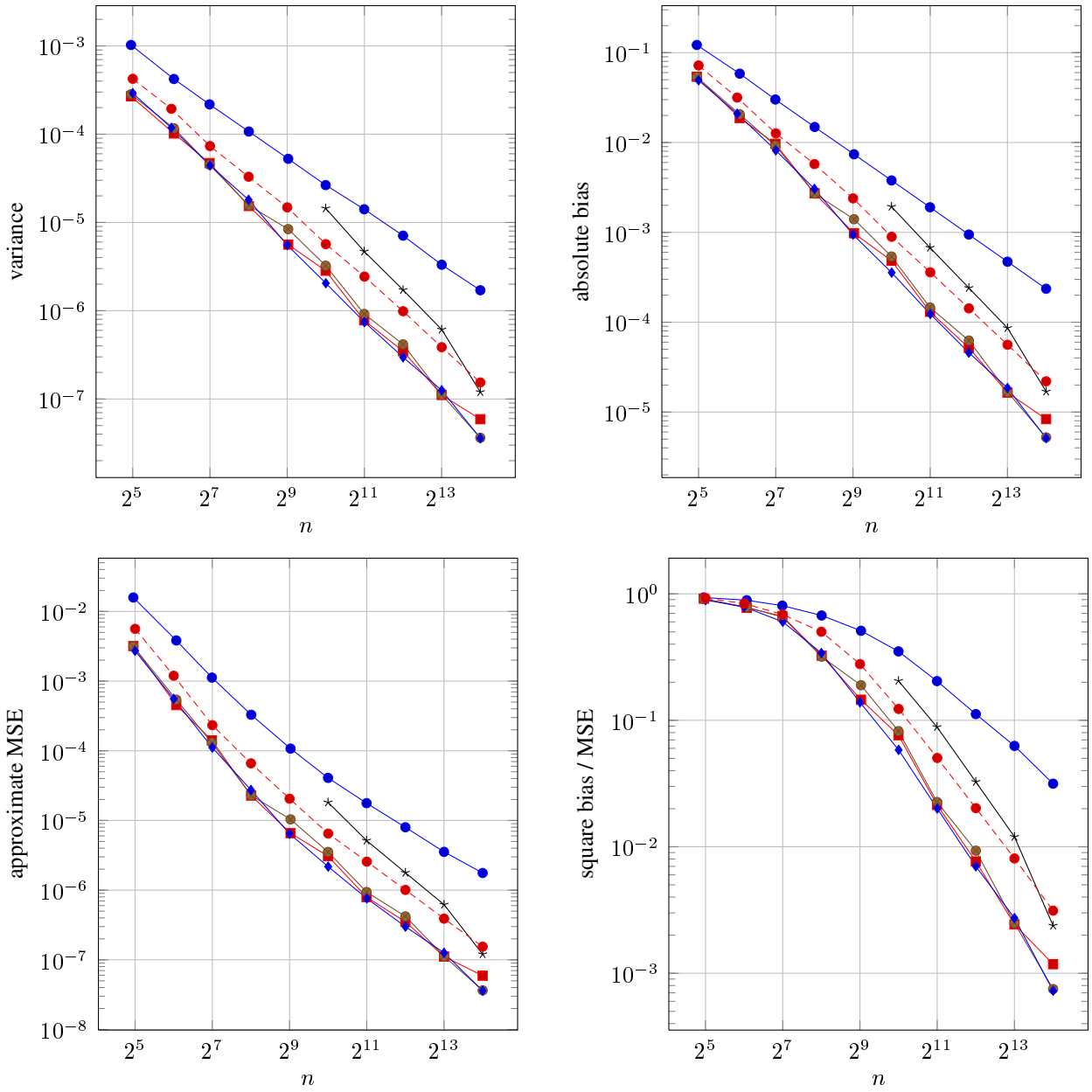

Figure 10: Estimated variance (top left), bias (top right), MSE (bottom left), and fraction of the MSE contributed by the square bias (bottom right) of the MC and RQMC estimators of the log-likelihood function for the example with real-life data, using MC (—●—), lattice-γ_u (—■—), lattice-0.1 (—●—), lattice-$M32$ (—⋆—), Sobol' nets (—♦—) and Halton points (-●-). The dotted line indicates the n^{-2} slope for reference. The results for the lattice-order, lattice-0.25, and lattice-0.5 rules, are very similar to those of lattice-γ_u and lattice-0.1 rules and Sobol' nets. The results for the Halton sequence are again essentially the same as those of Halton points.

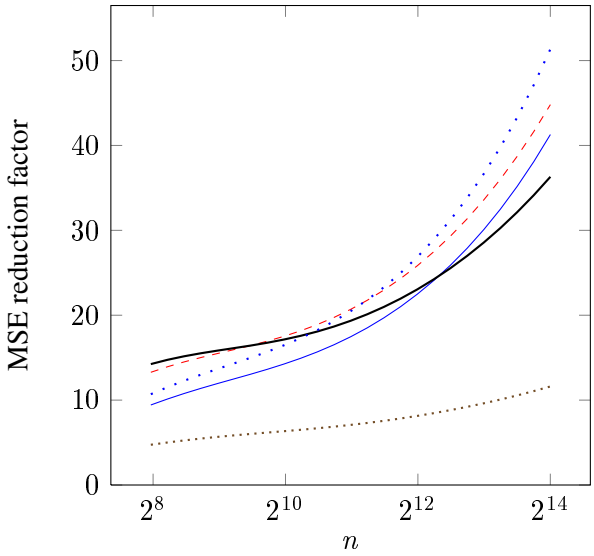

Figure 11: MSE reduction factors based on the fitted MSE, for the example with real-life data, using lattice-γ_u (——), lattice-0.1 (——), lattice-0.5 (· · · ·), Sobol' nets (- - -) and Halton points (·····).

for MC and for lattice-0.5 (they agree quite well), and the variance reduction factor (VRF) defined as the MC variance divided by the RQMC variance. It also reports the means and the VRF for the 50 estimates of the log-likelihood maximum value returned by the algorithm, for comparison. This VRF is not the same as in the results of Figure 10, where $\boldsymbol{\theta}$ is fixed instead of being optimized. We see that the VRF for the parameter estimates is more modest than for the log-likelihood estimate, but it is still significant. We did not examine the bias on the parameter estimates, because various sources of bias (from the simulation and from the optimization) can interfere, as observed by Bastin and Cirillo (2010), and we have no good way of estimating the overall bias.

5 Conclusion

We have studied the application of randomly-shifted lattice rules to estimate the mixed-logit log-likelihood function with better accuracy than standard MC for an equivalent computing effort, and compared the performance with that of other RQMC methods. The lattice parameters were selected based on a weighted $\mathcal{P}_{2\alpha}$ discrepancy criterion, with weights on the projections of the RQMC point set. We made an attempt at "optimizing" the projection weights (and select different specialized lattice parameters) for each problem instance, and even for each individual in the population. But it turned out that much simpler choices of weights (geometrically decreasing with the cardinality of the projection) and using the same lattice for all individuals gave comparable MSE

Parameter	MC mean	Lattice-0.5 mean	VRF
Waiting time 10 minutes	-0.614	-0.617	4.2
Waiting time 15 minutes	-1.003	-1.009	3.9
Waiting time 20 minutes	-1.736	-1.743	3.9
Cost (μ)	-2.013	-2.009	5.4
Cost (σ)	1.815	1.808	4.7
Passenger dropped (μ)	1.704	1.704	4.8
Passenger dropped (σ)	1.594	1.602	5.5
Automated cybercar (μ)	-0.230	-0.232	2.0
Automated cybercar (σ)	1.168	1.178	4.2
Human driven cybercar (μ)	0.134	0.135	2.5
Human driven cybercar (σ)	0.715	0.732	2.6
Guided way cybercar (μ)	-0.098	-0.101	2.9
Guided way cybercar (σ)	1.042	1.051	2.4
Loglikelihood	-4.413	-4.410	15.8

Table 8: Parameter estimates with MC and with lattice-0.5, and variance reduction factor (VRF) for lattice-0.5 compared with MC, for the example with real data

improvements in our experiments. This is very encouraging, because it indicates that the method is relatively robust to the choice of weights and there is no need to spend a huge effort to estimate the appropriate weights. The improvement was comparable to that obtained with Sobol' nets with a random digital shift, better than with Korobov lattices based on the $M_{32,24,16,12}$ criterion, and much better than randomly-shifted Halton points. Interestingly, RQMC improves the convergence rate (as a function of the number n of draws per individual) of the bias as well as that of the variance, compared with MC, so it reduces the MSE in two ways. In fact, the reduction of square bias dominates the variance reduction when n is small, and we saw this happening for n as large as 1000. This bias reduction means that a smaller increase of n is required to compensate a growth of the population size m (because increasing m reduces only the variance, not the bias). As expected, we observed that the efficiency improvement of RQMC compared to MC is reduced when the dimension of the integration problem increases (either from an increase of the dimension s of the parameter vector or because we have panel data), presumably because a larger proportion of the variance falls into higher-order projections. Yet, in all cases, RQMC provided a net advantage over MC. Finally, we saw in one example that this RQMC improvement translates into a lower variance for the parameter estimates obtained by maximizing the sample log-likelihood. The improvement is less spectacular than for estimating the log-likelihood at a single point, but still significant.

One issue that would require further examination and better understanding is the specific contributions to the MSE from individuals drawn from a heterogeneous population, and in particular

from individuals who make low probability choices. Moreover, here we focused on point estimation of the log-likelihood near its maximizer, but much remains to be said about the impact of RQMC on the behavior of the maximum-likelihood estimator itself and on the whole log-likelihood optimization process.

Acknowledgement

This work was supported by NSERC-Canada Discovery Grants to P. L'Ecuyer and F. Bastin, a Canada Research Chair to P. L'Ecuyer, a GERAD scholarship to D. Munger, the EuroNF Network of Excellence to B. Tuffin, and INRIA's associated team MOCQUASIN to all authors. This paper is an outgrow from a preliminary (and partial) version entitled "Estimation strategies for mixed logit models" (Paper 555), presented at the *12th Conference of the The International Association for Travel Behaviour Research*, in Jaipur, India, 2009.

References

Bastin, F., Cirillo, C., 2010. Reducing simulation bias in mixed logit model estimation. Journal of Choice Modelling 3 (2), 71–88.

Bastin, F., Cirillo, C., Toint, Ph. L., 2006a. An adaptive Monte Carlo algorithm for computing mixed logit estimators. Computational Management Science 3 (1), 55–79.

Bastin, F., Cirillo, C., Toint, Ph. L., 2006b. Convergence theory for nonconvex stochastic programming with an application to mixed logit. Mathematical Programming, Series B 108 (2–3), 207–234.

Bastin, F., Cirillo, C., Toint, Ph. L., 2010. Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations. Transportation Science, to appear.

Bhat, C. R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research 35B (7), 677–693.

Bhat, C. R., 2003a. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research B 37 (3), 837–855.

Bhat, C. R., 2003b. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B: Methodological 37 (9), 837–855.

Bhat, C. R., Sener, I. N., 2009. A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. Journal of Geographical Systems 11 (3), 243–272.

Brownstone, D., Bunch, D. S., Train, K., 2000. Joint mixed logit models of stated and revealed references for alternative-fuel vehicles. Transportation Research Part B: Methodological 34 (5), 315–338.

Cirillo, C., Axhausen, K. W., 2006. Evidence on the distribution of values of travel time savings from a six-week diary. Transportation Research A 40, 444–457.

Cirillo, C., Xu, R., 2010. Forecasting cybercar use for airport ground access: A case study at BWI (Baltimore Washington International Airport). Journal of Urban Planning and Development 136 (3), 186–194.

Cranley, R., Patterson, T. N. L., 1976. Randomization of number theoretic methods for multiple integration. SIAM Journal on Numerical Analysis 13 (6), 904–914.

Dick, J., Sloan, I. H., Wang, X., Wozniakowski, H., 2004. Liberating the weights. Journal of Complexity 20 (5), 593–623.

Faure, H., Lemieux, C., 2009. Generalized Halton sequences in 2008: A comparative study. ACM Transactions on Modeling and Computer Simulation 19 (4), Article 15.

Hess, S., Bierlaire, M., Polak, J. W., 2005. Estimation of value of travel-time savings using mixed logit models. Tranportation Research Part A 39 (3), 221–236.

Hess, S., Rose, J., Bain, S., 2009. Random scale heterogeneity in discrete choice models. In: Proceedings of the European Transport Conference. Leeuwenhorst, The Netherlands.

Hickernell, F. J., 1998. A generalized discrepancy and quadrature error bound. Mathematics of Computation 67, 299–322.

Hickernell, F. J., 2000. What affects the accuracy of quasi-Monte Carlo quadrature? In: Niederreiter, H., Spanier, J. (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 1998. Springer-Verlag, Berlin, pp. 16–55.

Hickernell, F. J., 2002. Obtaining $O(N^{-2+\epsilon})$ convergence for lattice quadrature rules. In: Fang, K.-T., Hickernell, F. J., Niederreiter, H. (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2000. Springer-Verlag, Berlin, pp. 274–289.

L'Ecuyer, P., 2008. SSJ: A Java Library for Stochastic Simulation. Software user's guide, available at http://www.iro.umontreal.ca/~lecuyer.

L'Ecuyer, P., 2009. Quasi-Monte Carlo methods with applications in finance. Finance and Stochastics 13 (3), 307–349.

L'Ecuyer, P., Lemieux, C., 1999. Quasi-Monte Carlo via linear shift-register sequences. In: Proceedings of the 1999 Winter Simulation Conference. IEEE Press, pp. 632–639.

L'Ecuyer, P., Lemieux, C., 2000. Variance reduction via lattice rules. Management Science 46 (9), 1214–1235.

Lemieux, C., 2009. Monte Carlo and Quasi-Monte Carlo Sampling. Springer-Verlag, New York, NY.

Lemieux, C., Cieslak, M., Luttmer, K., 2004. RandQMC User's Guide: A Package for Randomized Quasi-Monte Carlo Methods in C. Software user's guide, available at http://www.math.uwaterloo.ca/~clemieux/randqmc.html.

McFadden, D. L., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15 (5), 447–270.

Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods. Vol. 63 of SIAM CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, PA.

Owen, A. B., 1998. Latin supercube sampling for very high-dimensional simulations. ACM Transactions on Modeling and Computer Simulation 8 (1), 71–102.

Owen, A. B., 2003. Variance with alternative scramblings of digital nets. ACM Transactions on Modeling and Computer Simulation 13 (4), 363–378.

Sándor, Z., Train, K., 2004. Quasi-random simulation of discrete choice models. Transportation Research Part B 38 (4), 313–327.

Sivakumar, A., Bhat, C. R., Ökten, G., 2005. Simulation estimation of mixed discrete choice models with the use of randomized quasi-monte carlo sequences: A comparative study. Transportation Research Record 1921, 112–122.

Sloan, I. H., Joe, S., 1994. Lattice Methods for Multiple Integration. Clarendon Press, Oxford.

Sobol', I. M., 1967. The distribution of points in a cube and the approximate evaluation of integrals. U.S.S.R. Comput. Math. and Math. Phys. 7, 86–112.

Sobol', I. M., Myshetskaya, E. E., 2007. Monte Carlo estimators for small sensitivity indices. Monte Carlo Methods and Applications 13 (5–6), 455–465.

Train, K., 2000. Halton sequences for mixed logit. Working paper No. E00-278, Department of Economics, University of California, Berkeley, USA.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, New York, USA.

In these appendices, we present additional results, for the examples with synthetic data (A) as well as with real-life data (B). In particular, we show the ANOVA variances for more cases, together with the variance, bias and MSE on the log-likelihood estimators for all examples. We also detail the analysis for a few more single individuals. Finally, lattice parameters for geometric weights are given in C.

A Additional results for the examples with synthetic data

Figures 1 and 2 show the distribution of the ANOVA variances among all projections for the conditional likelihood of selected individuals, for the independent case with $s = 5$ and $T_q = 1$. The estimated variances of the individual likelihood estimator with constructed lattices and other point sets are reported in Figures 3 and 4. The empirical distributions of the values of the estimator, obtained with 5000 independent replications, are shown in Figure 5. Figure 6 presents the average over all individuals of these ANOVA variances, for the independent and correlated cases with $s = 5$. The estimated variance, bias, MSE and fraction of the MSE due to the square bias on the likelihood function estimator with constructed lattices and other point sets are plotted in Figures 7 through 18 for all independent and correlated cases with $s = 5$, 10 and 15 with $T_q = 1$ and 3. We have constructed the lattice-γ_{u} and lattice-order rules only for $s = 5$. In higher dimension, especially with $s = 15$, the lattice-0.1 rules are consistently more efficient than the Sobol' nets. We also observe that in presence of panel data ($T_q > 1$), the variance is larger than for cross-sectional data ($T_q = 1$), and the square bias dominates in the MSE until higher values of n. We estimated the log-likelihood for each individual; we plot the distribution of these values in Figure 19. The empirical distribution of the 5000 independent replications of the simulated log-likelihood function is shown in Figure 20. It is in particular interesting to note that the distribution of the individual contributions to the average log-likelihood has a tail on the left, meaning that some individuals have a lower choice probability and penalize the overall log-likelihood.

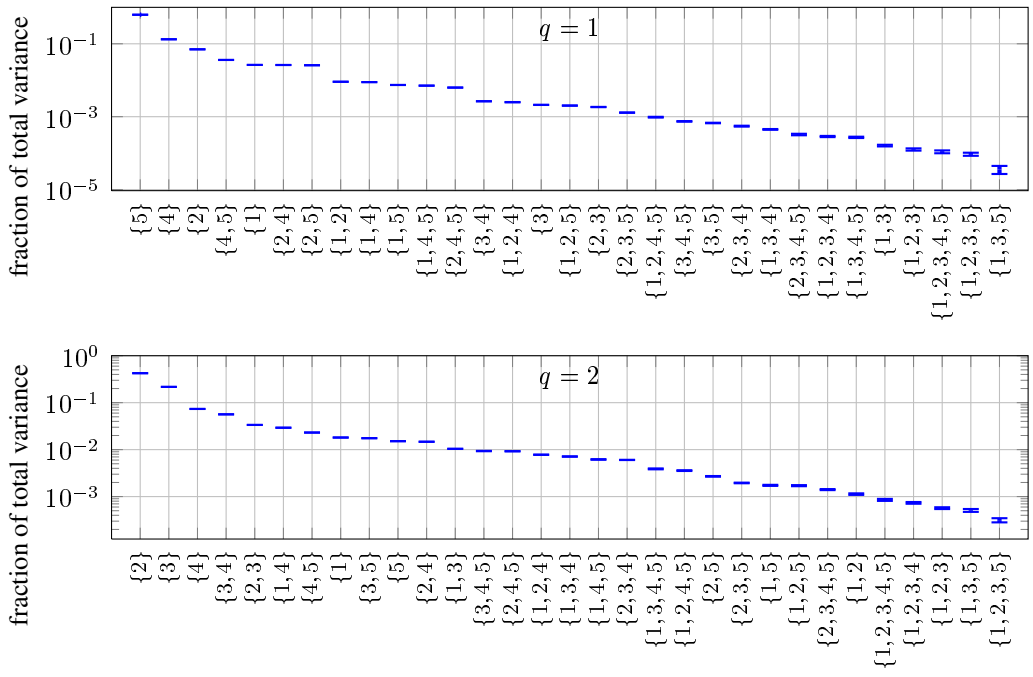

Figure 1: Fractional ANOVA variances of the conditional likelihood function for a single individual $q = 1$ (top) and 2 (bottom), sorted by decreasing order, for the independent case with $s = 5$ and $T_q = 1$. The projections are listed on the horizontal axis, and their fraction of total variance is plotted along the vertical axis. The limits of the vertical bars correspond to a normal confidence interval at 95 % on the variance estimate, the standard deviation of which was estimated across random draws.

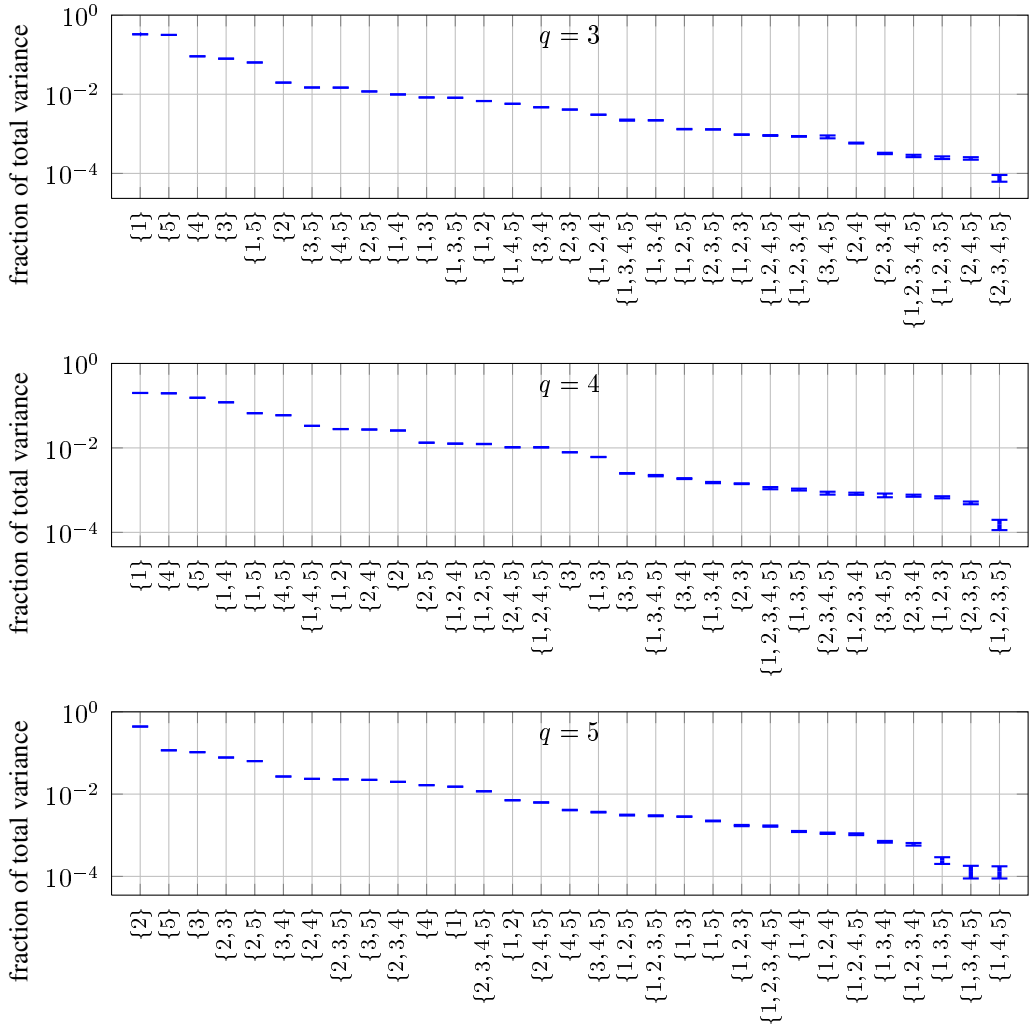

Figure 2: Fractional ANOVA variances of the conditional likelihood function for a single individual $q = 3$, 4, and 5 (from top to bottom). See Figure 1 for further details.

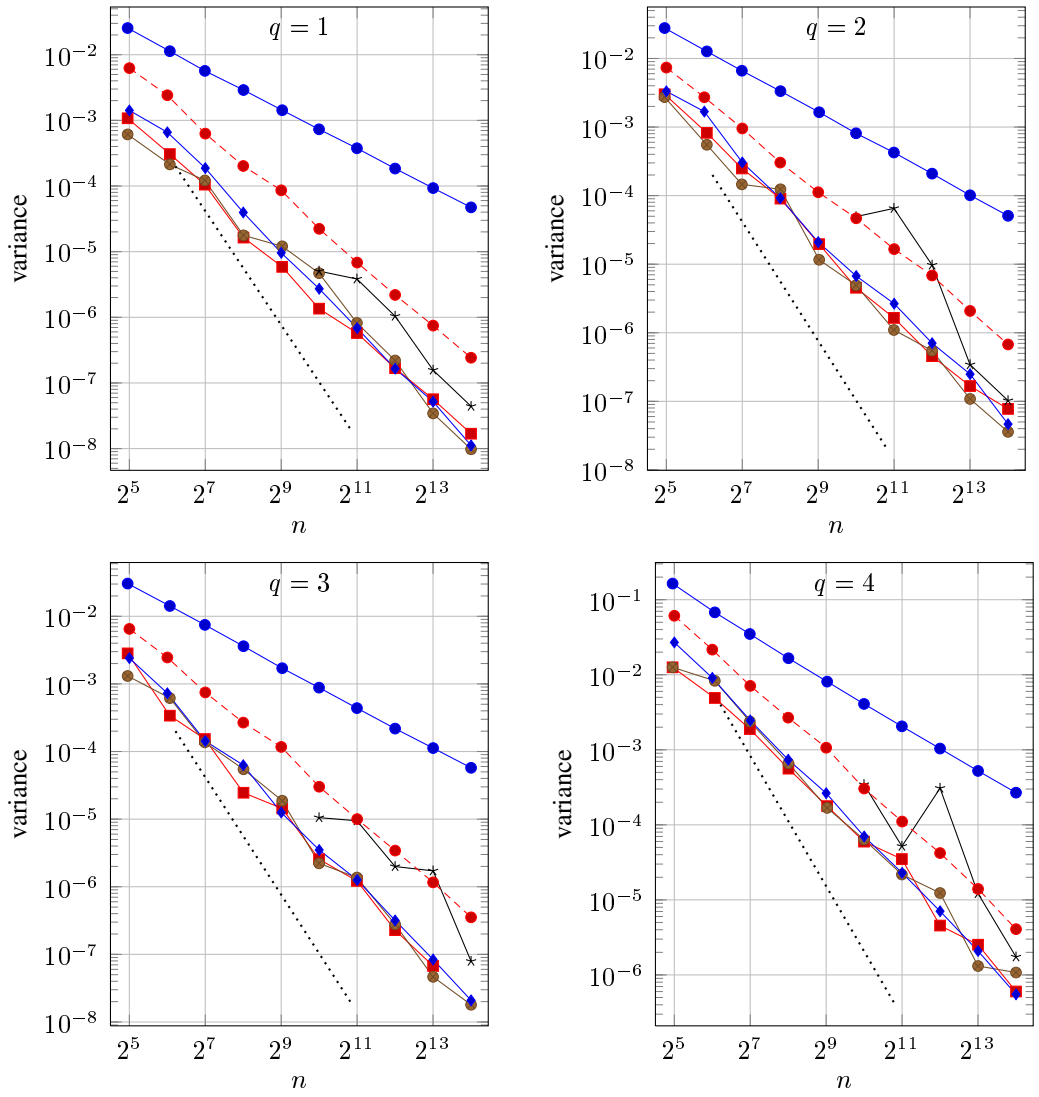

Figure 3: Estimated variance of the MC and RQMC estimators of the log-likelihood of a single individual for the independent case with $s = 5$ and $T_q = 1$, for individuals $q = 1$ (top left), 2 (top right), 3 (bottom left) and 4 (bottom right), using MC (—●—), lattice-γ_u (—■—), lattice-0.1 (—●—), lattice-$M32$ (—⋆—), Sobol' nets (—♦—), and Halton points (- -●- -). For lattice-order and lattice-0.5, the variances are very similar to those of lattice-γ_u and lattice-0.1, and we do not show them to reduce the number of curve superpositions. The results for the Halton sequence (not shown here) are practically undistinguishable from those for the Halton points. The dotted line indicates the n^{-2} slope, for reference.

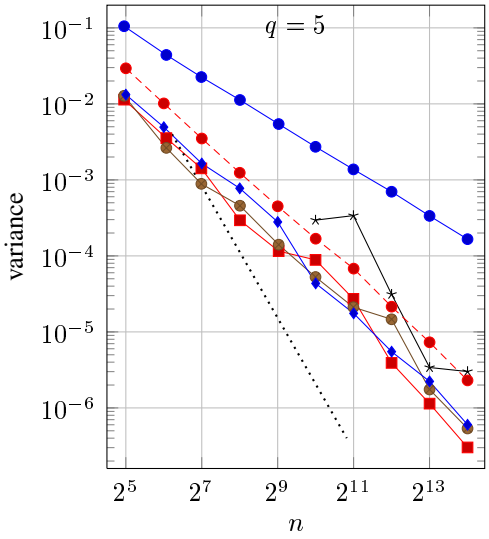

Figure 4: Same as Figure 3, but for individual $q = 5$.

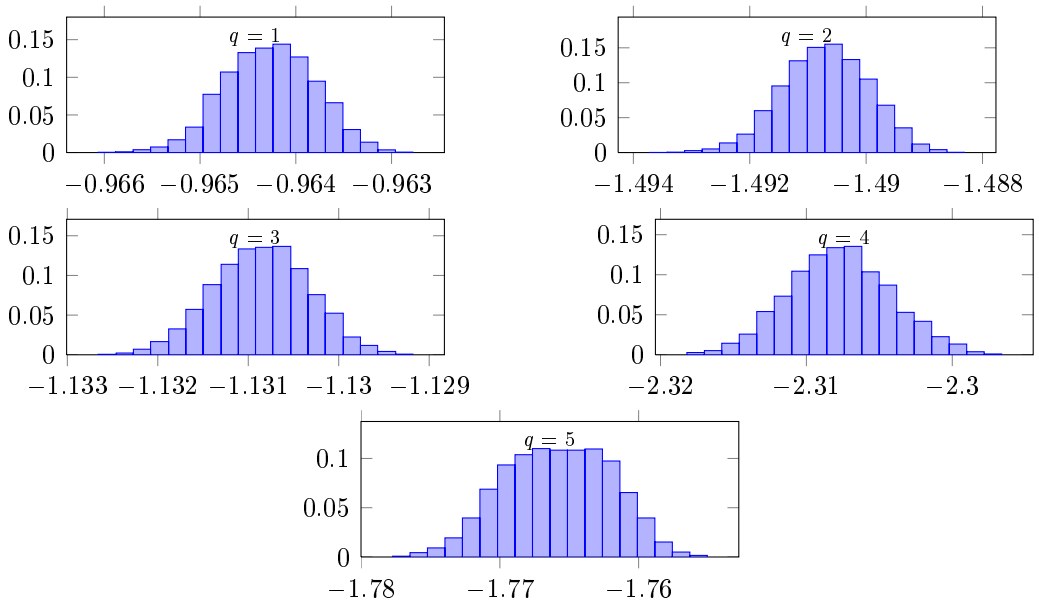

Figure 5: Empirical distribution of the simulated log-likelihood $\ln \hat{p}_q^n$ of a single individual for the independent case with $s = 5$ and $T_q = 1$, for individuals $q = 1$, 2, 3, 4 and 5 (row-wise), using lattice-0.1 with $n = 4093$.

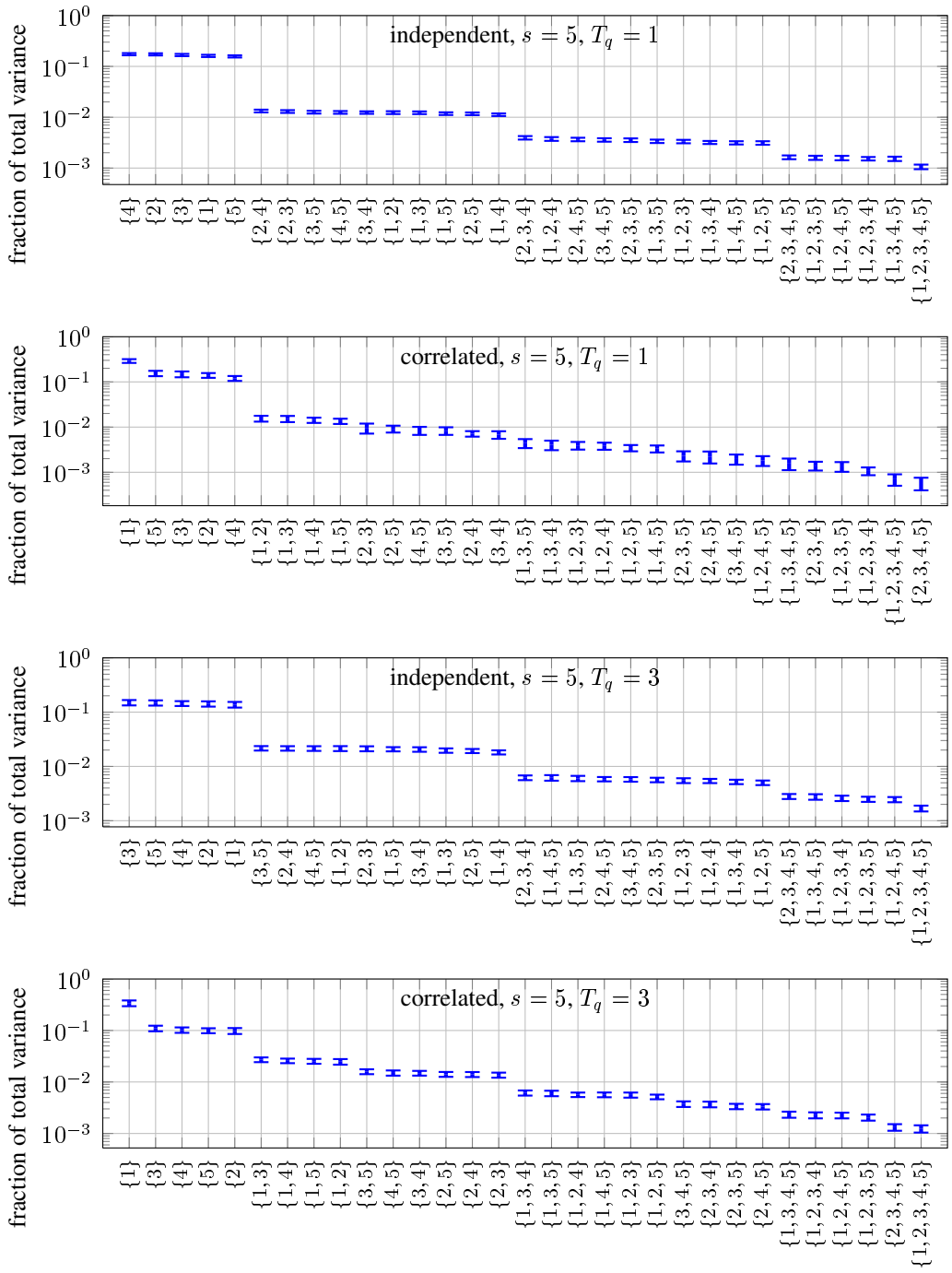

Figure 6: Average over all individuals of the ANOVA variances of the conditional likelihood functions, sorted by and displayed as fraction of the total variance, for the independent and correlated cases with $s = 5$ and $T_q = 1$ and 3. See 1 for further details.

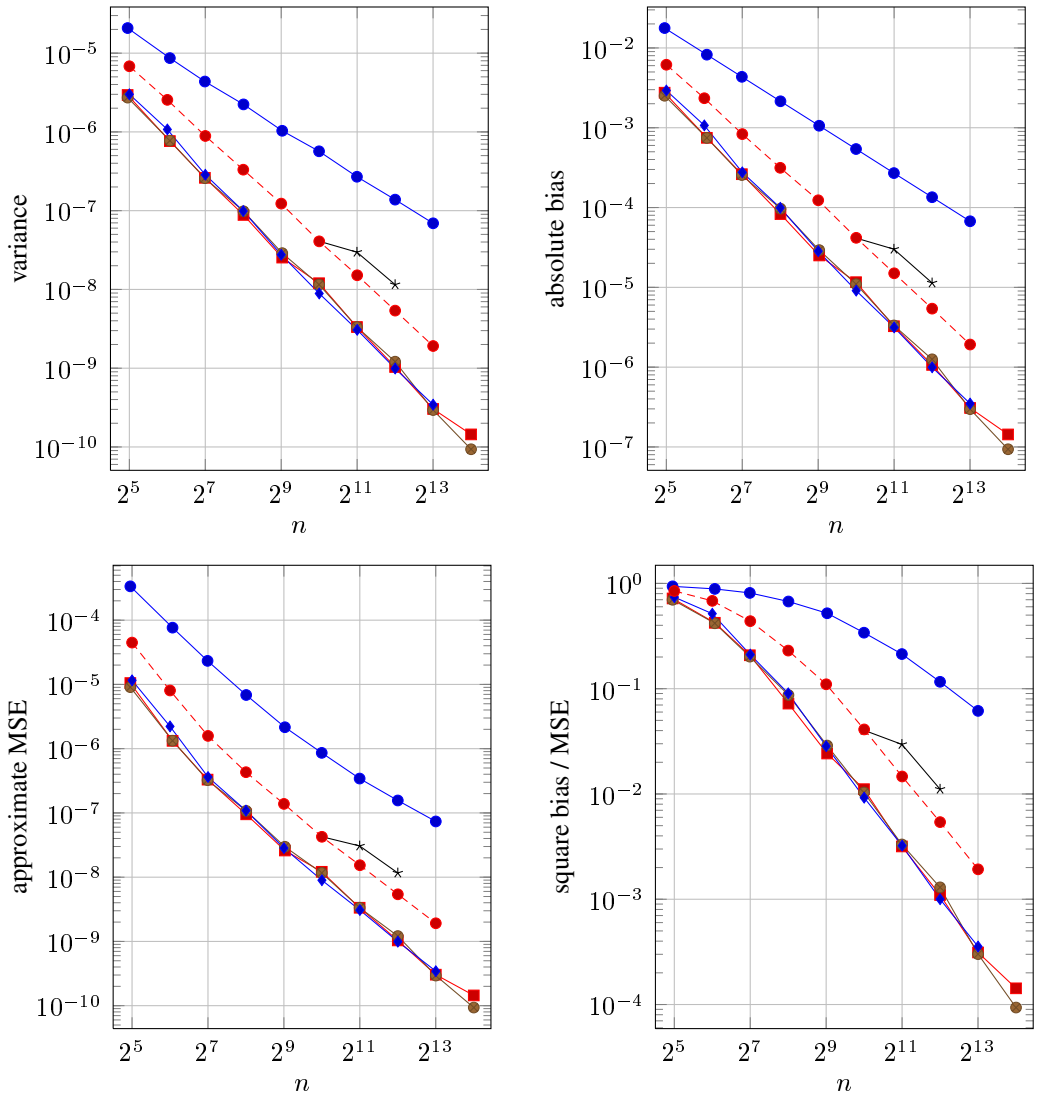

Figure 7: Estimated variance (top left), bias (top right), MSE (bottom left), and fraction of the MSE contributed by the square bias (bottom right) of the MC and RQMC estimators of the log-likelihood function for the independent case with $s = 5$ and $T_q = 1$, using MC (—●—), lattice-γ_u (—■—), lattice-0.1 (—●—), lattice-$M32$ (—⋆—), Sobol' nets (—♦—) and Halton points (- -●- -). For lattice-order and lattice-0.5, the variances are very similar to those of lattice-γ_u and lattice-0.1, and we do not show them to reduce the number of curve superpositions. The results for the Halton sequence (not shown here) are practically undistinguishable from those for the Halton points. The dotted line indicates the n^{-2} slope, for reference.

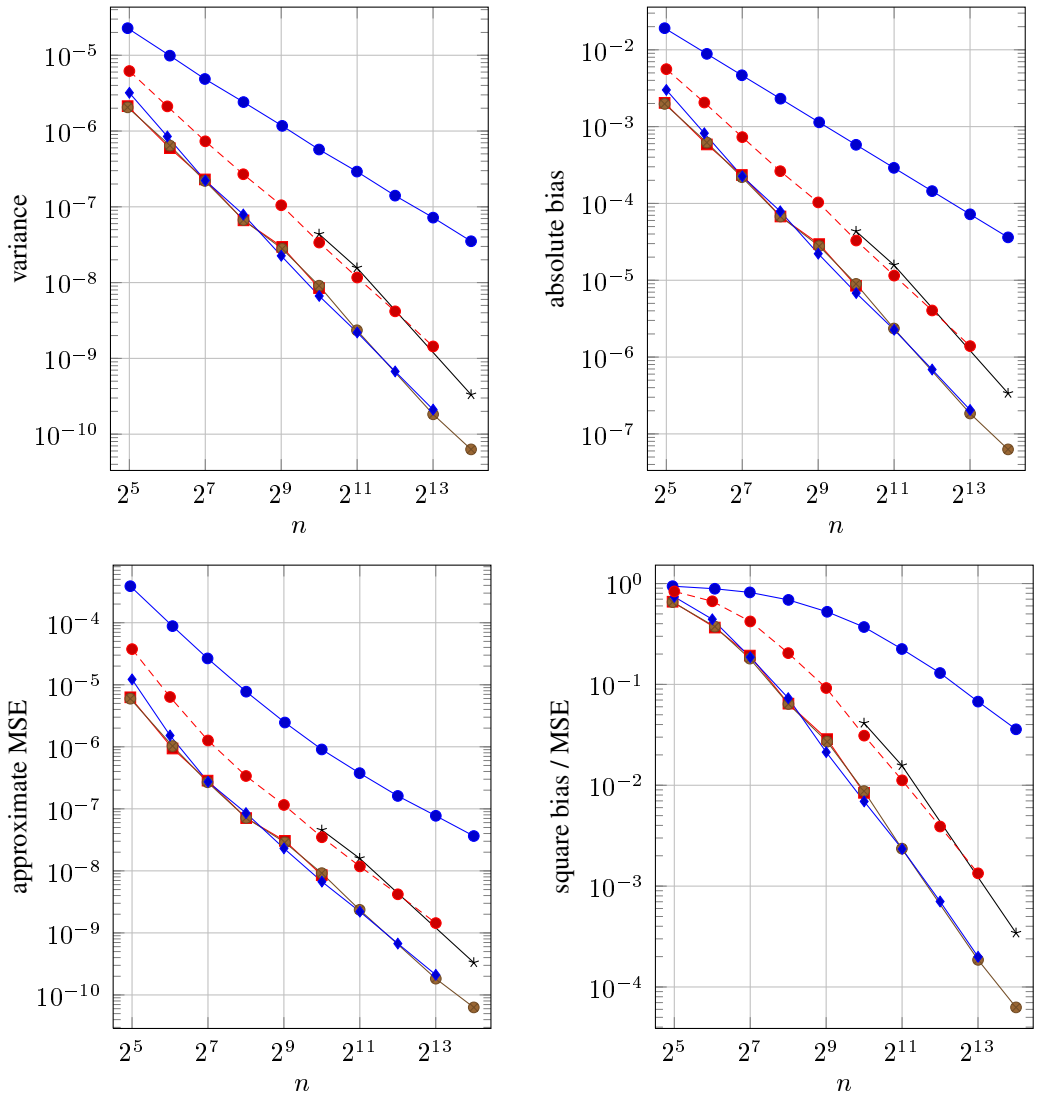

Figure 8: Same as Figure 7, but for the correlated case with $s = 5$ and $T_q = 1$.

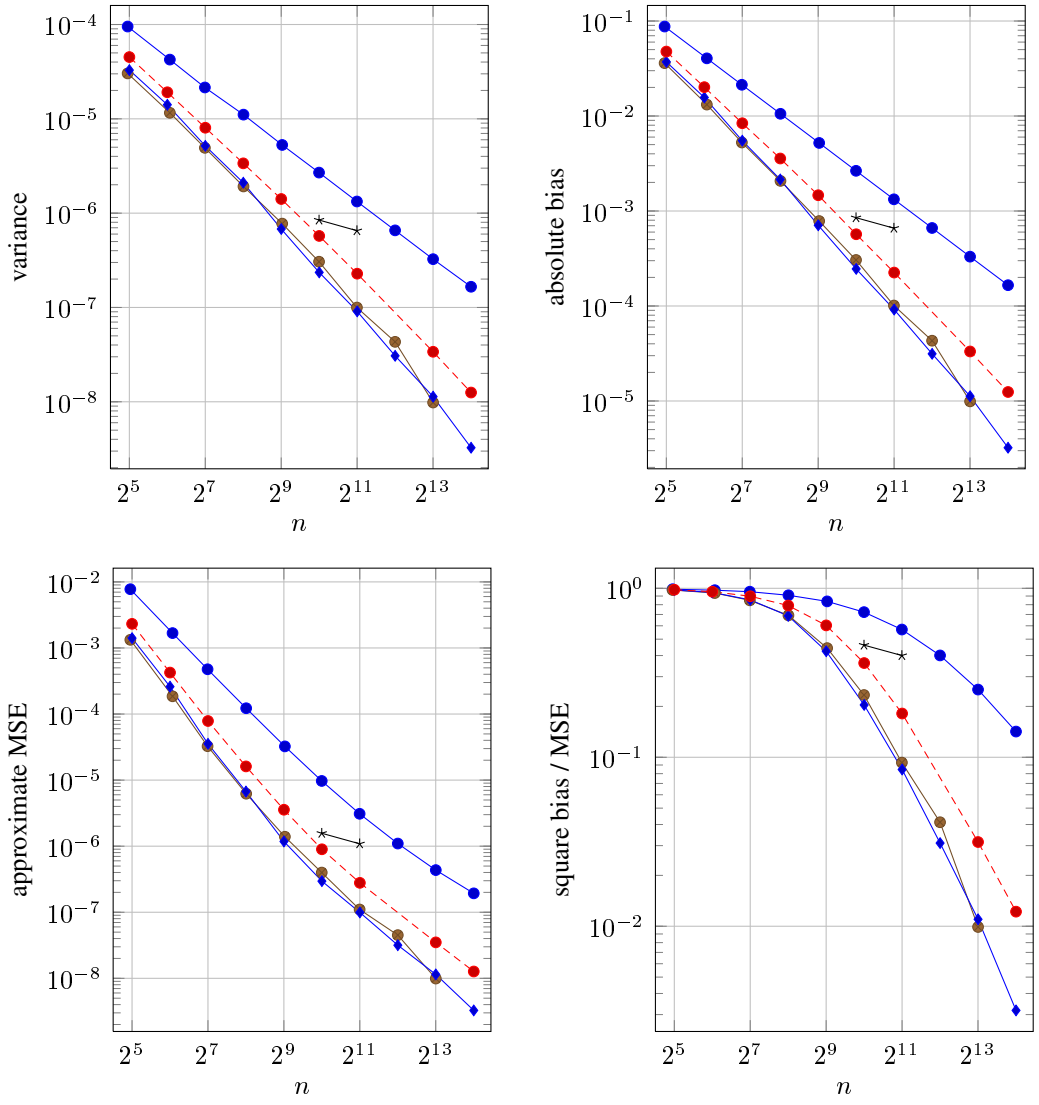

Figure 9: Same as Figure 7, but for the independent case with $s = 5$ and $T_q = 3$.

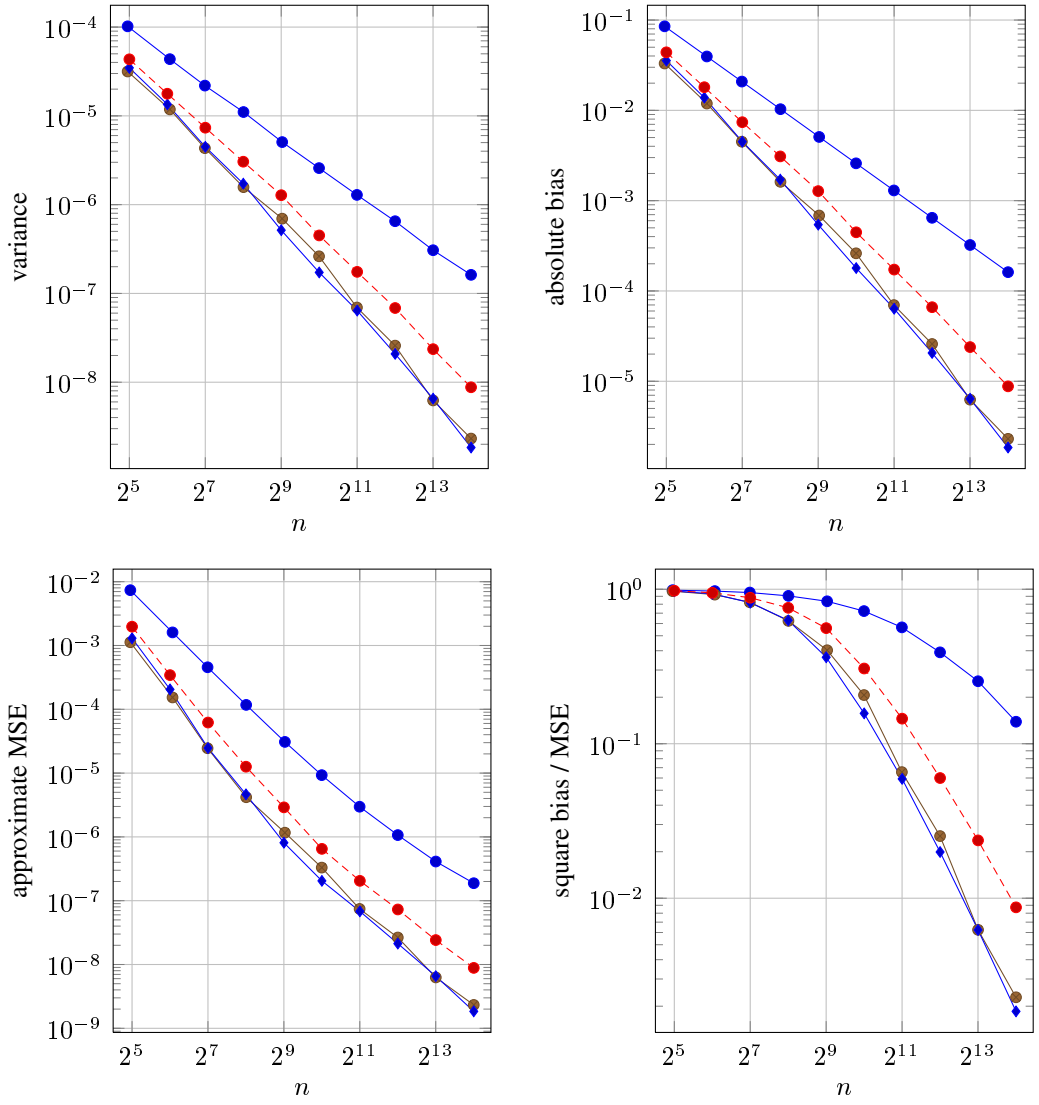

Figure 10: Same as Figure 7, but for the correlated case with $s = 5$ and $T_q = 3$.

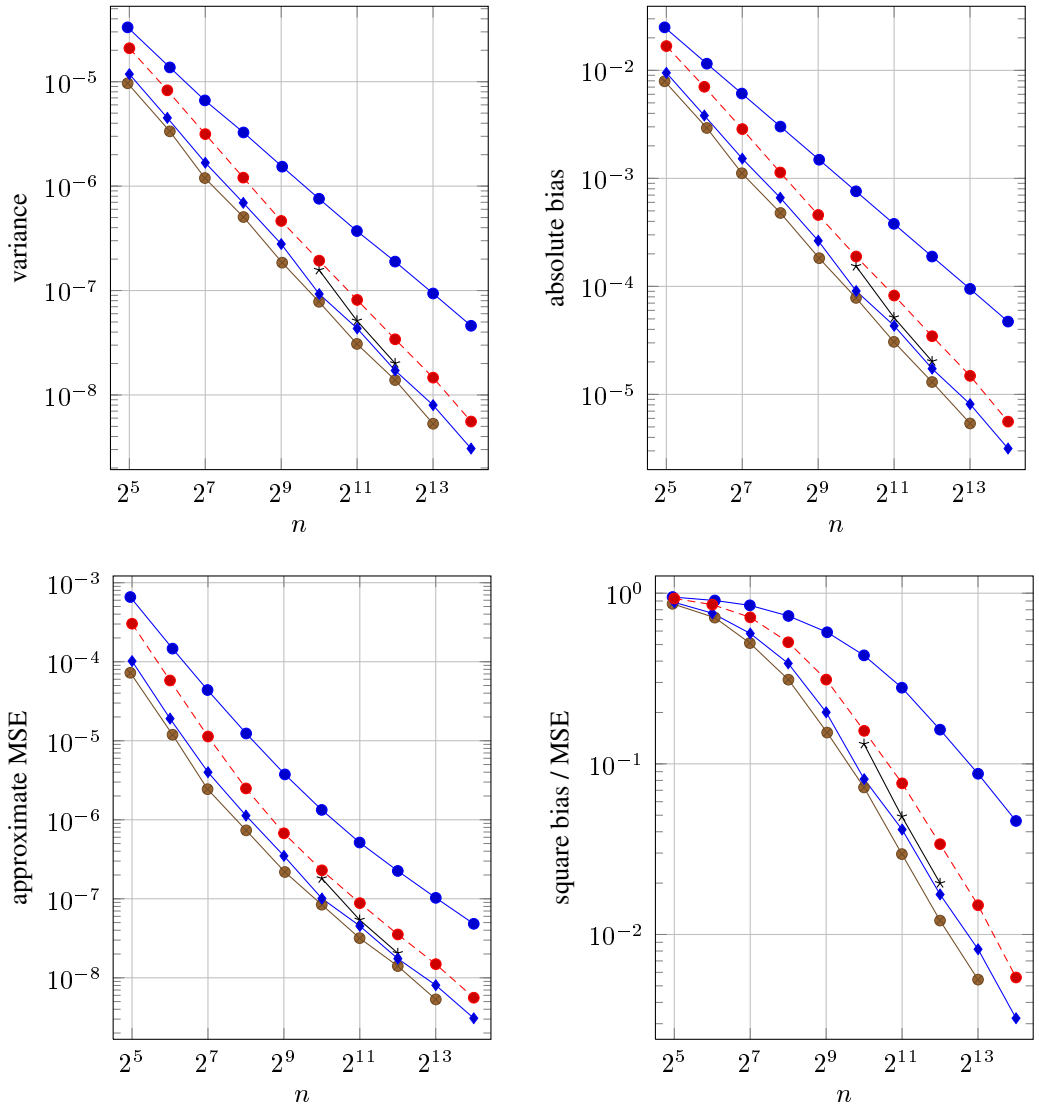

Figure 11: Same as Figure 7, but for the independent case with $s = 10$ and $T_q = 1$.

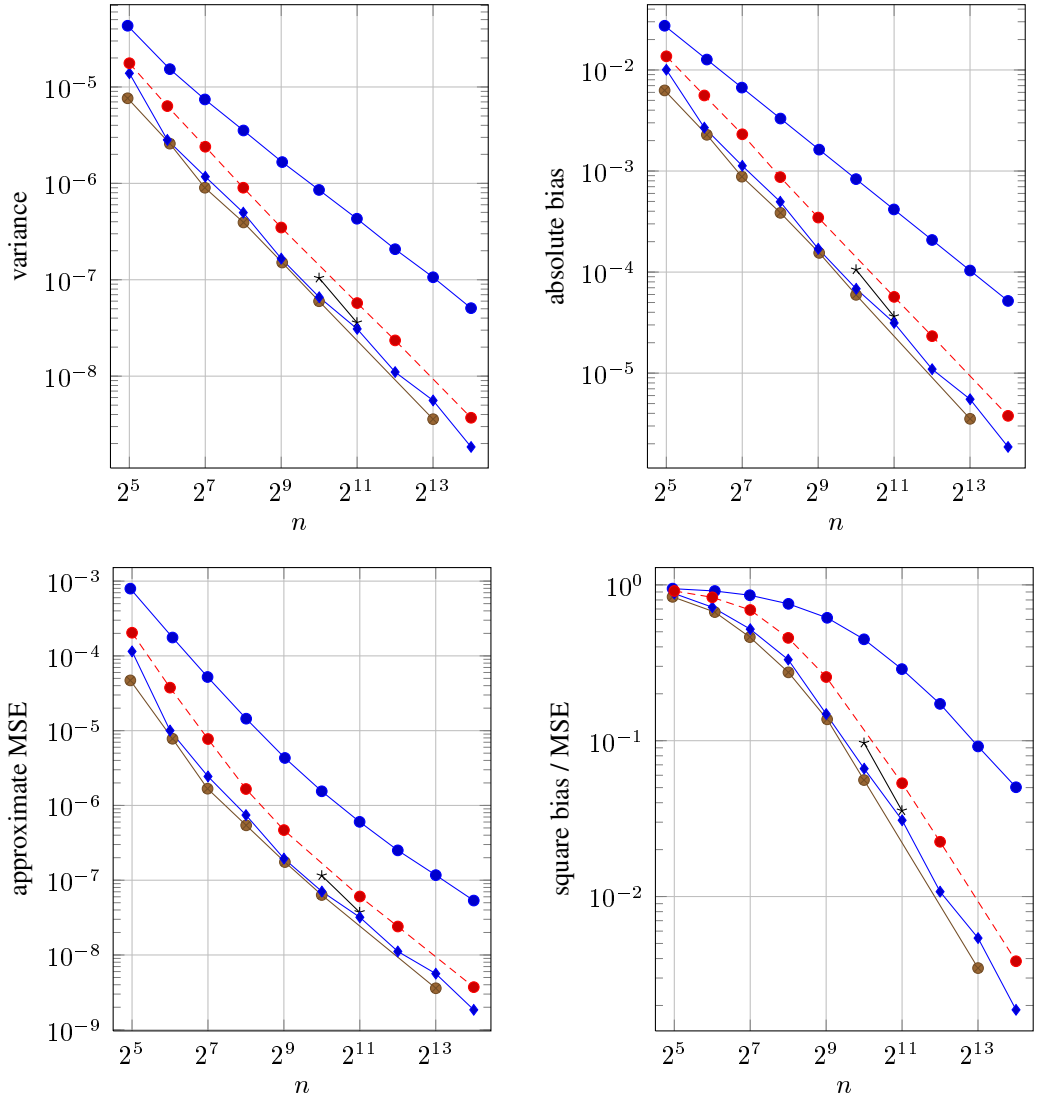

Figure 12: Same as Figure 7, but for the correlated case with $s = 10$ and $T_q = 1$.

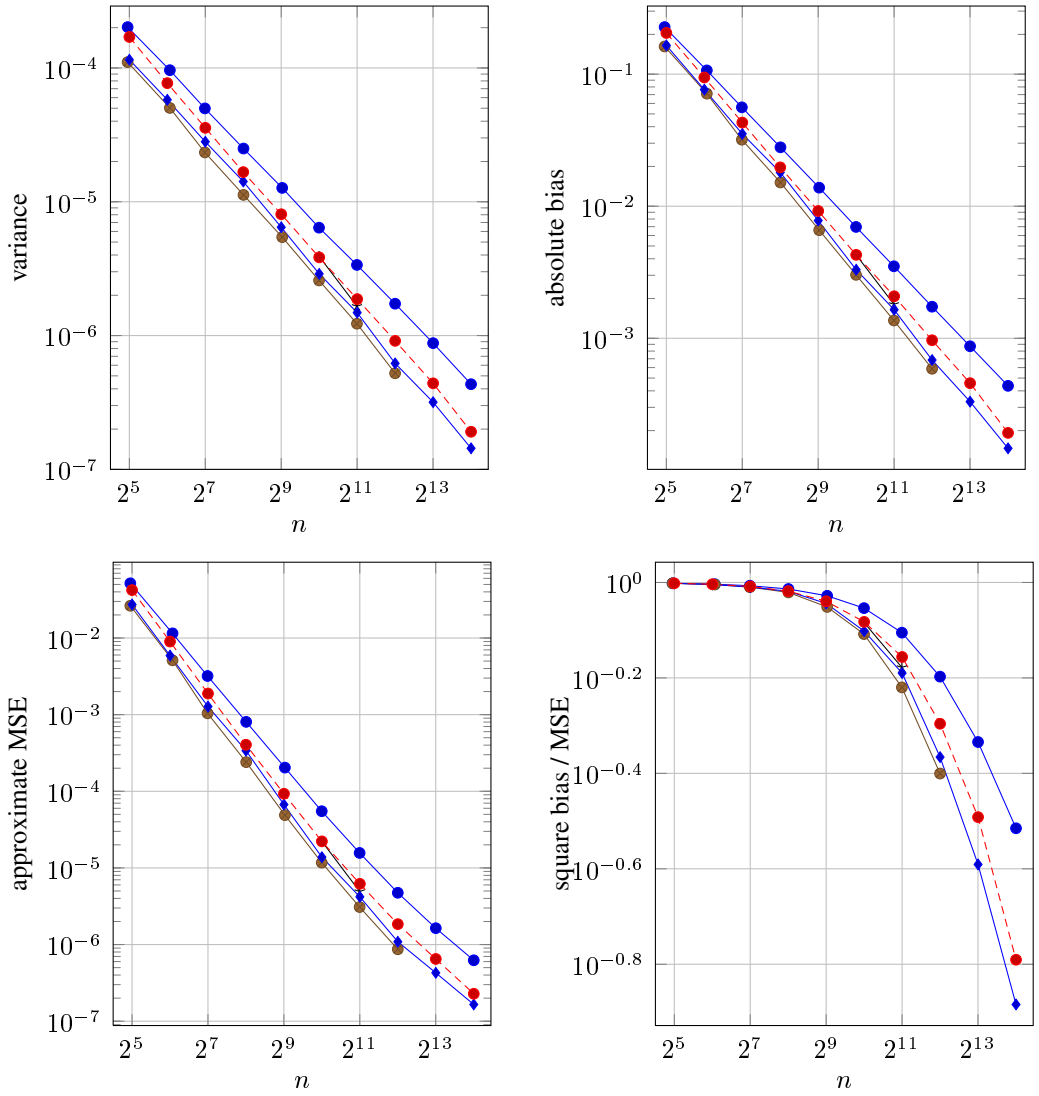

Figure 13: Same as Figure 7, but for the independent case with $s = 10$ and $T_q = 3$.

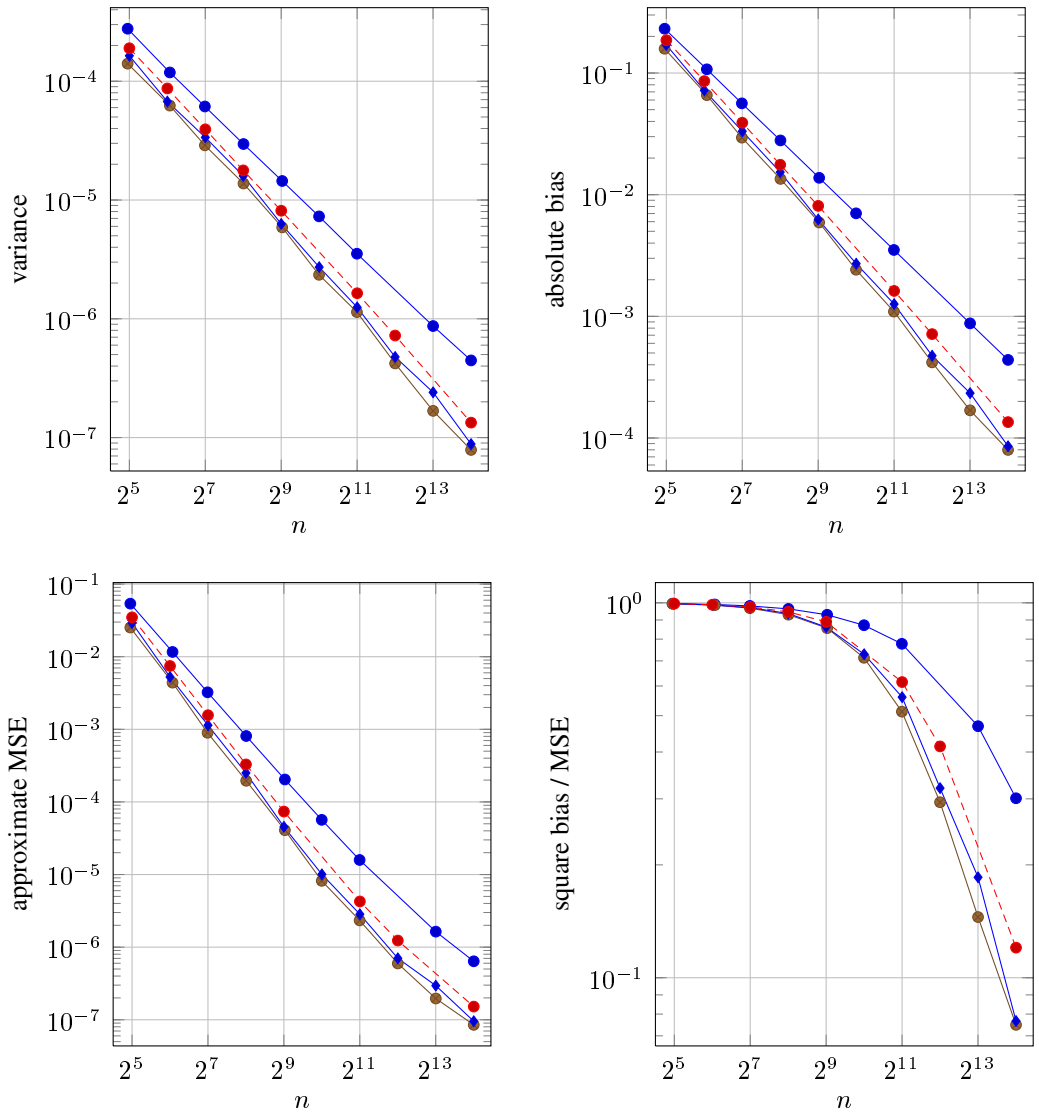

Figure 14: Same as Figure 7, but for the correlated case with $s = 10$ and $T_q = 3$.

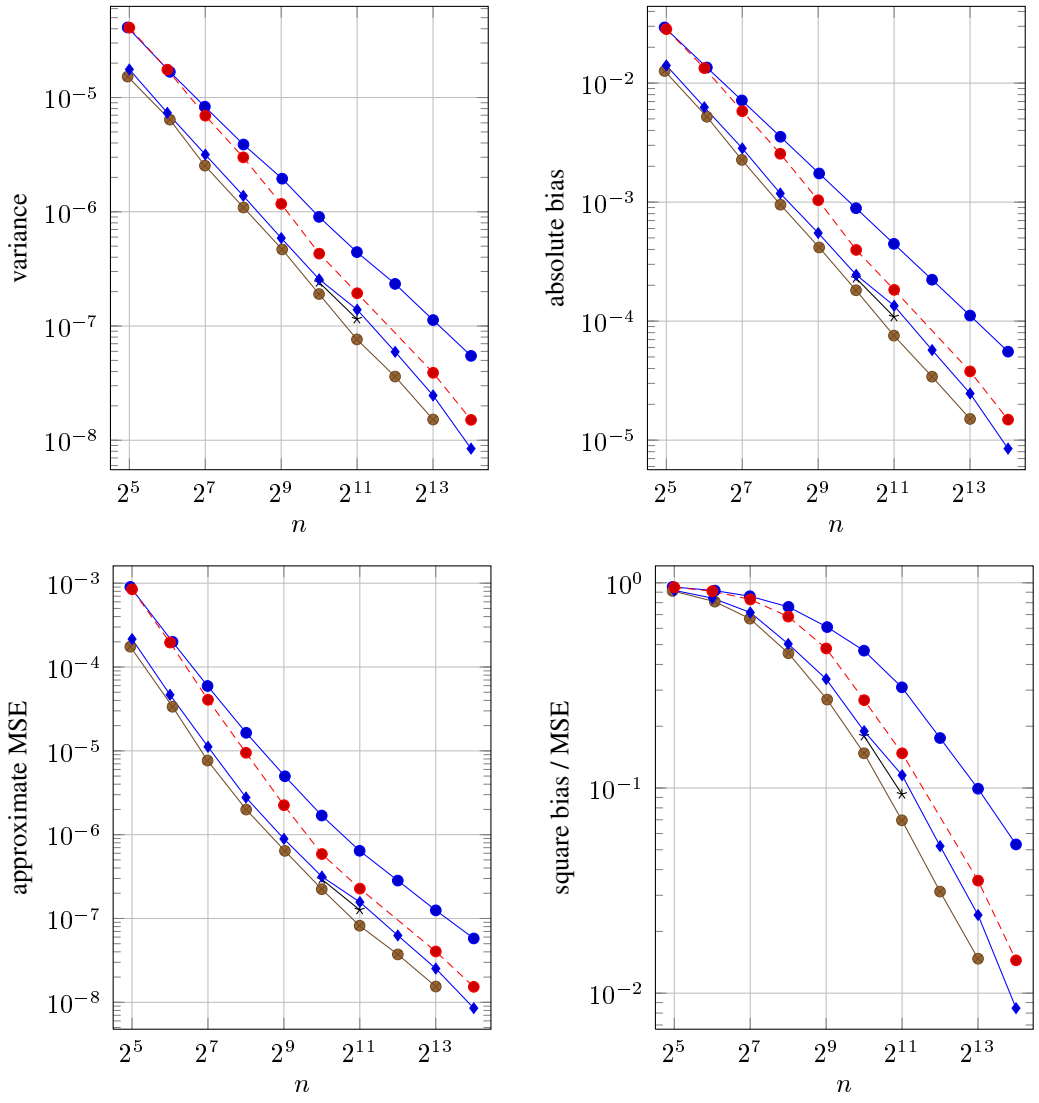

Figure 15: Same as Figure 7, but for the independent case with $s = 15$ and $T_q = 1$.

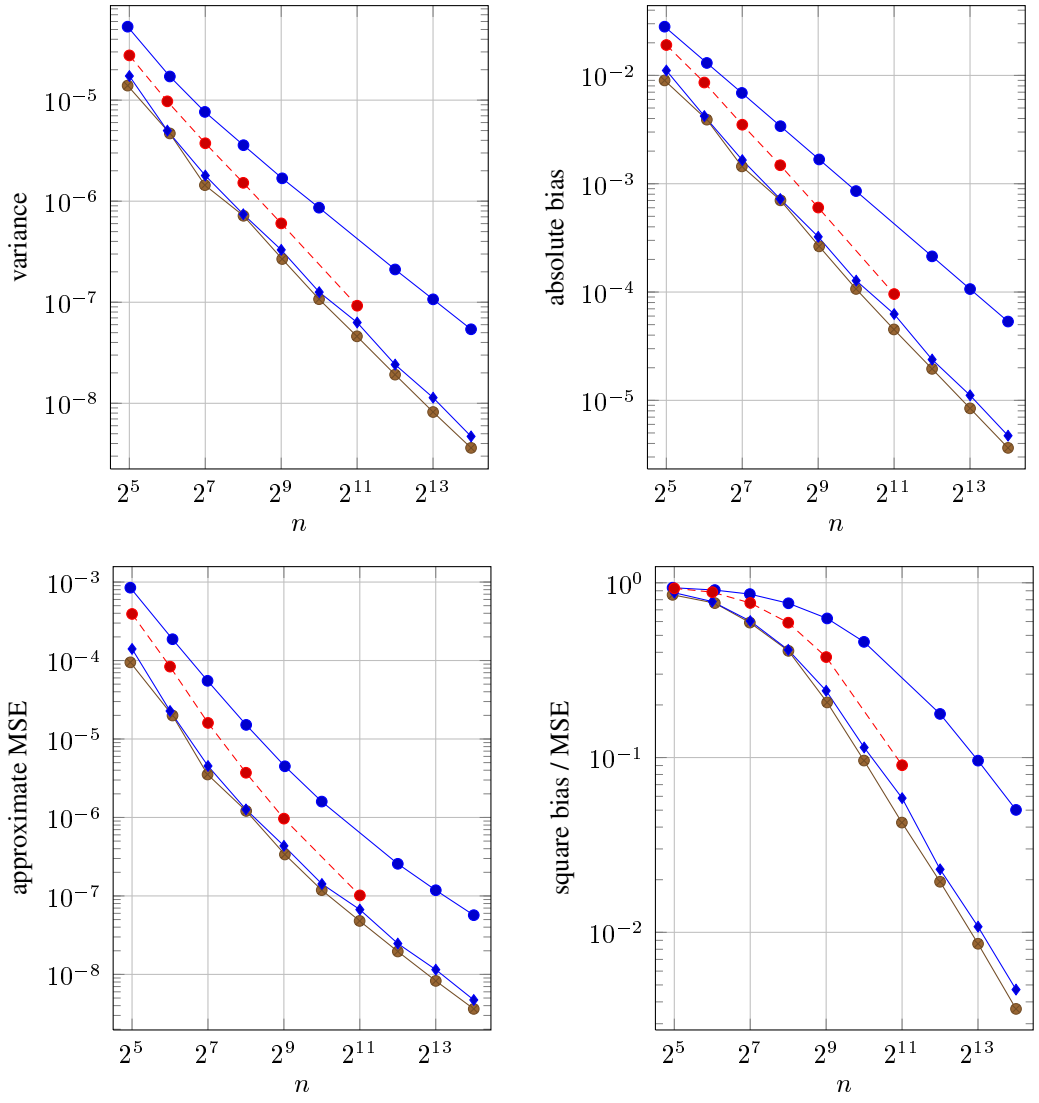

Figure 16: Same as Figure 7, but for the correlated case with $s = 15$ and $T_q = 1$.

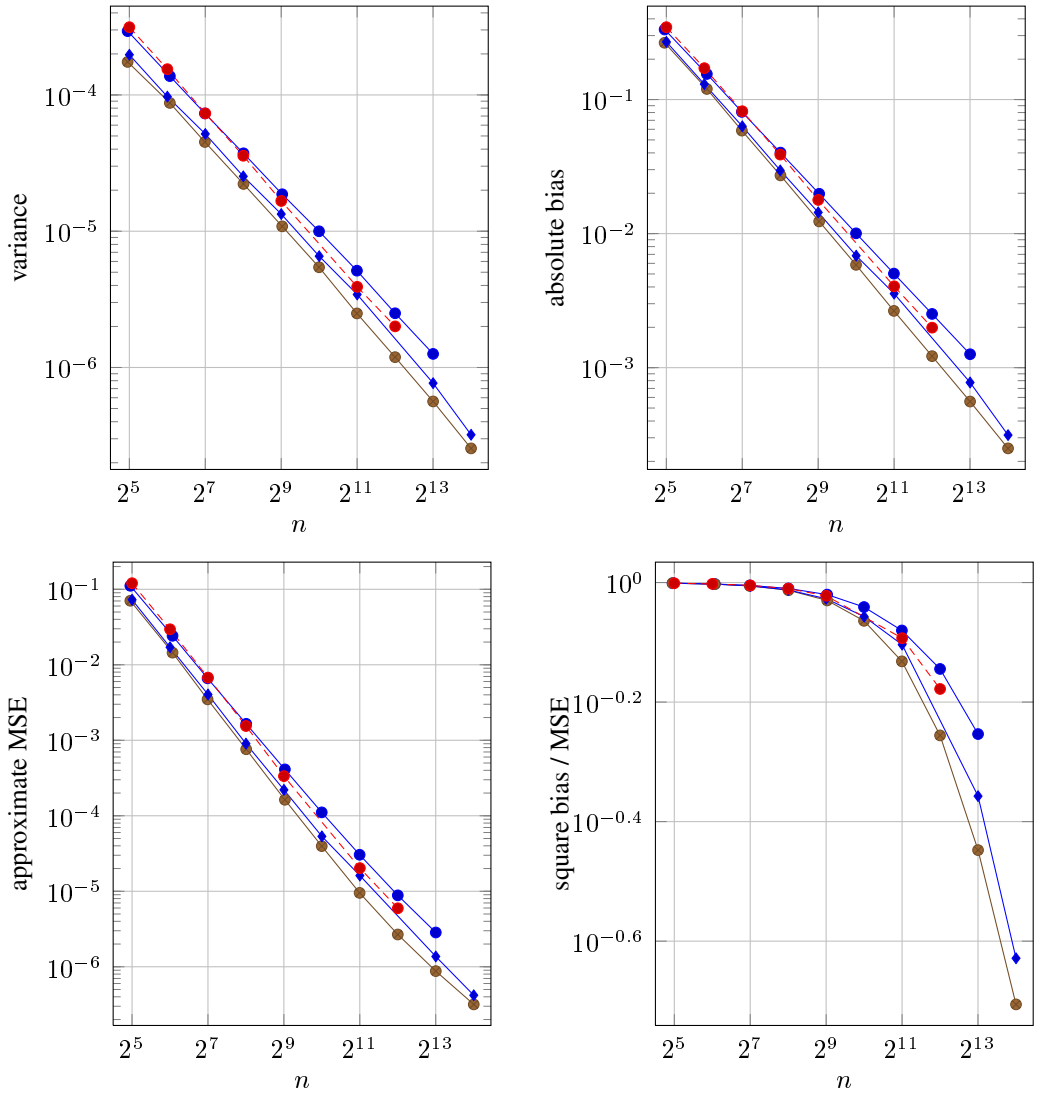

Figure 17: Same as Figure 7, but for the independent case with $s = 15$ and $T_q = 3$.

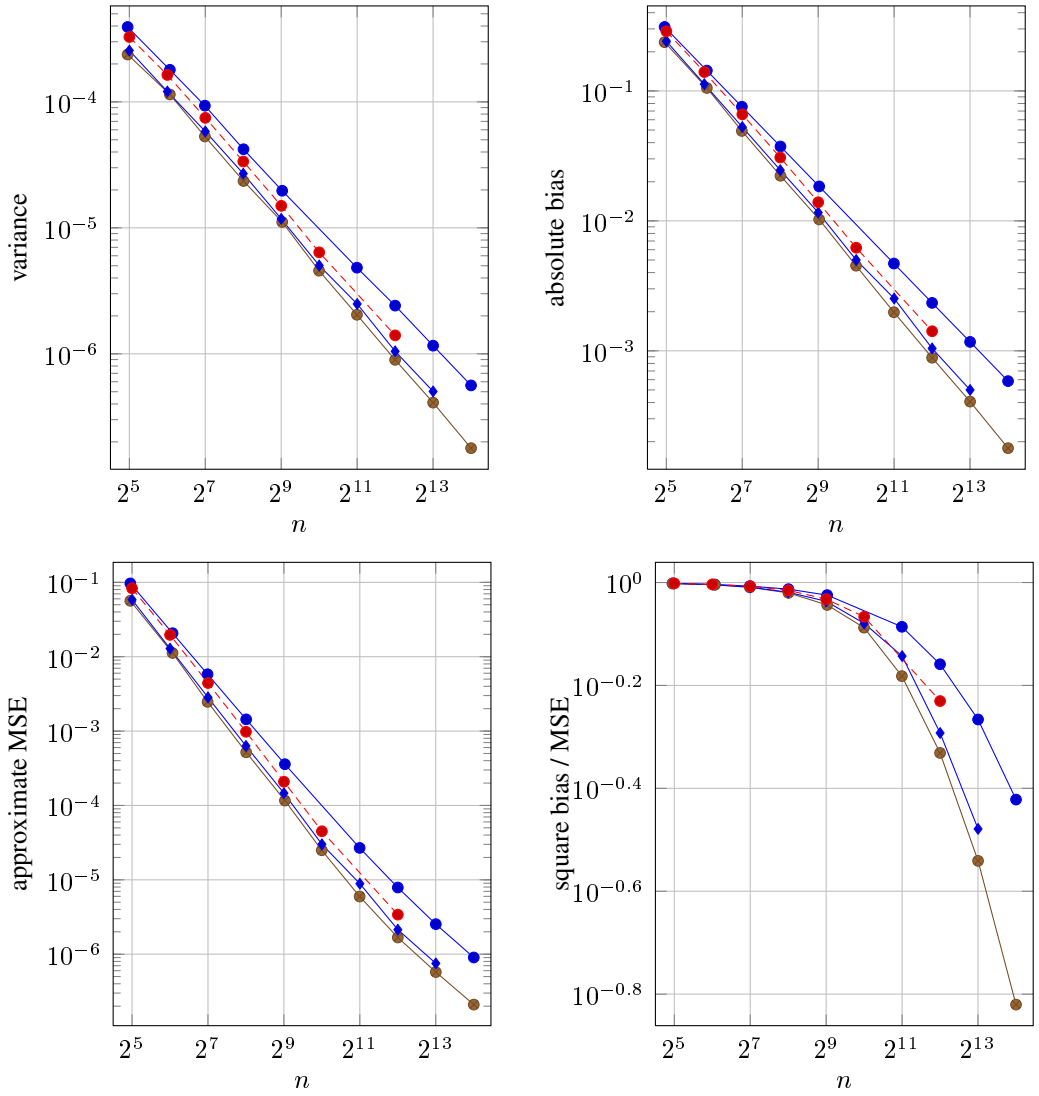

Figure 18: Same as Figure 7, but for the correlated case with $s = 15$ and $T_q = 3$.

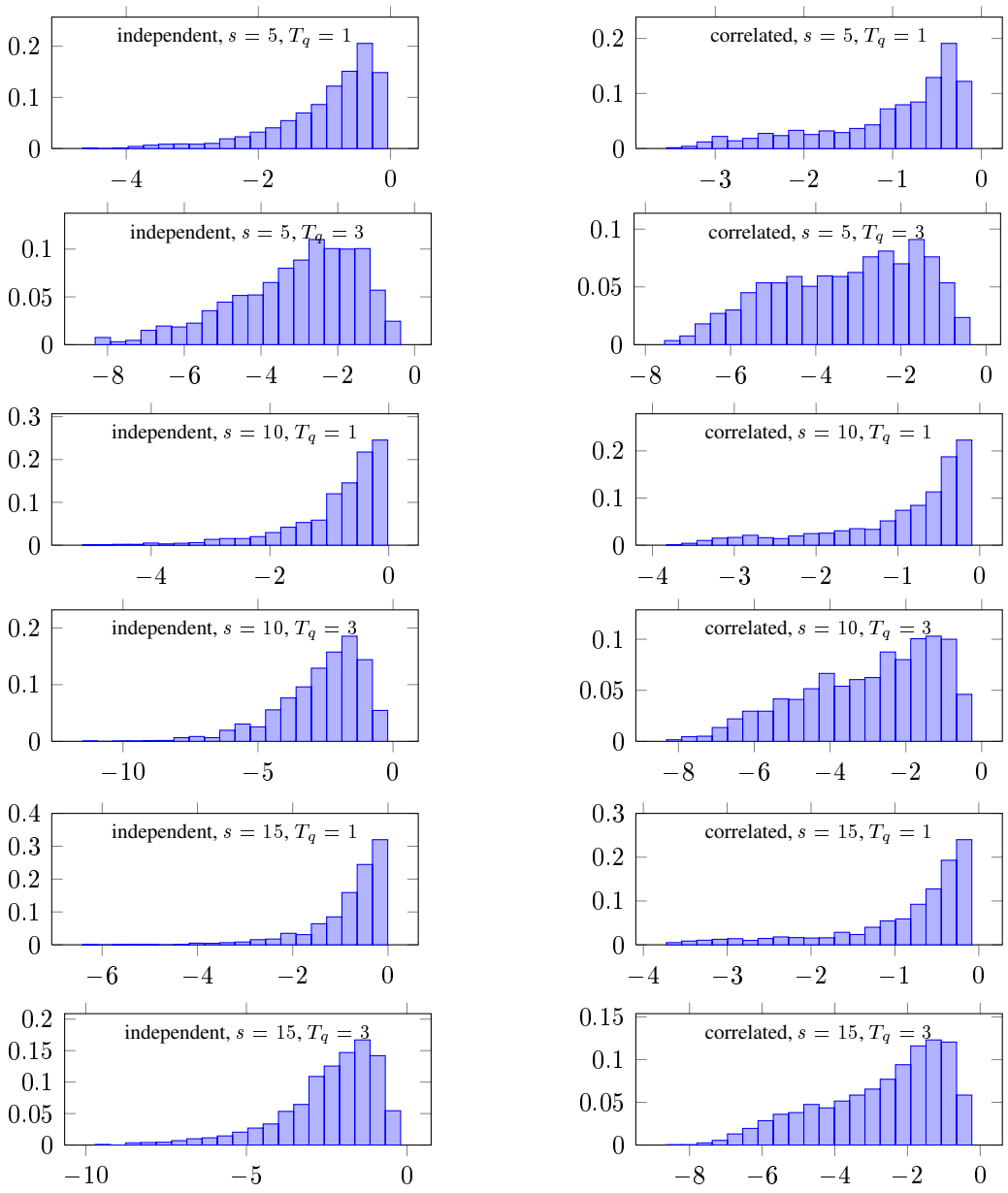

Figure 19: Empirical distribution of the $m = 2000$ estimated individual log-likelihoods across individuals, for all independent and correlated cases with $s = 5$, 10 and 15 and $T_q = 1$ and 3, using lattice-0.1 with $n = 4093$.

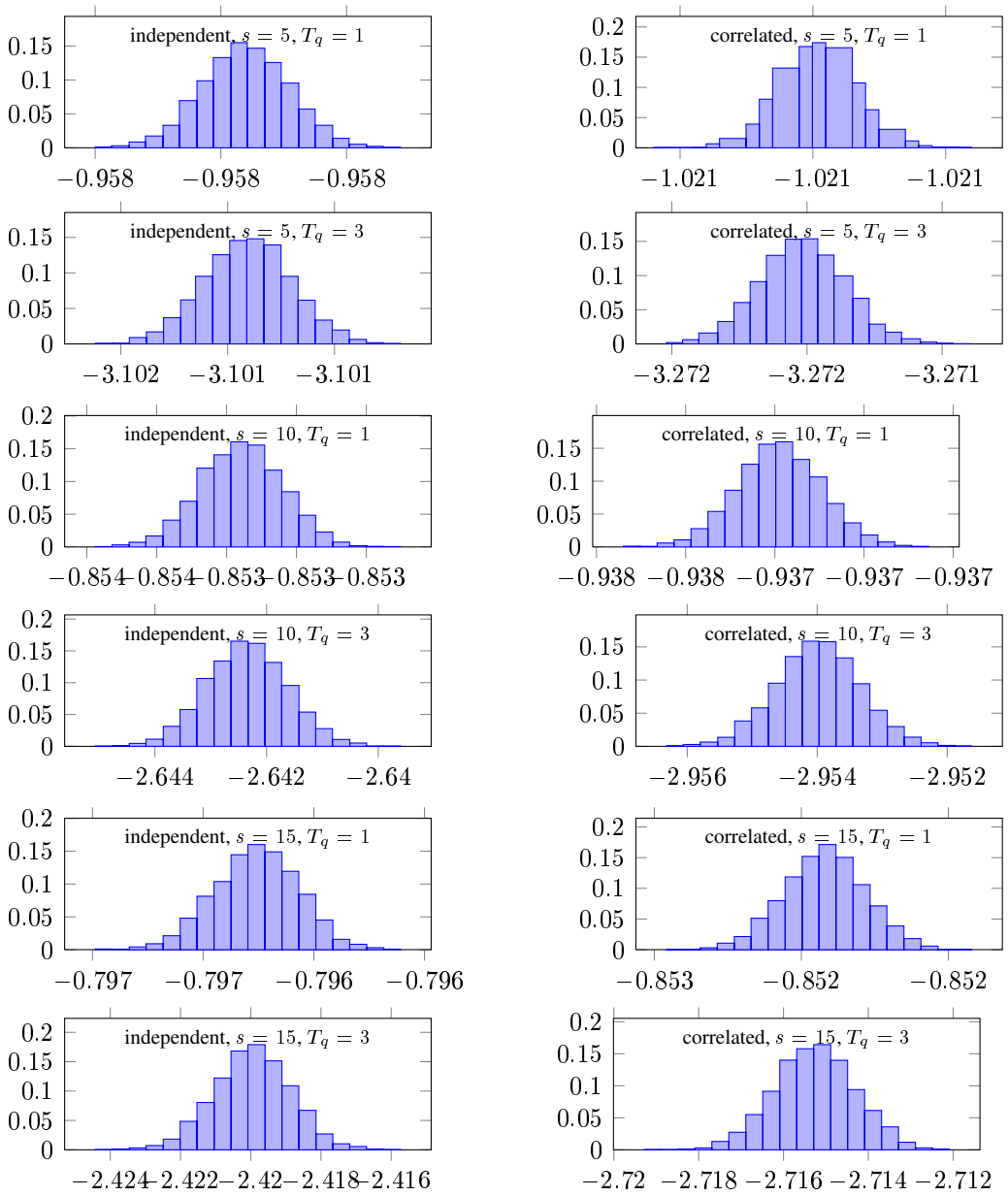

Figure 20: Empirical distribution of the simulated values of the log-likelihood function across independent replications, for all independent and correlated cases with $s = 5$, 10 and 15 and $T_q = 1$ and 3, using lattice-0.1 with $n = 4093$.

B Additional results for the example with real-life data

We also performed additional analysis of the real-life data. Figures 21 and 22 show the distribution of the ANOVA variances among the different projections for the conditional likelihood of selected individuals. For individuals $q = 1$, 61, 116 and 135, whose log-likelihoods range from -4.33 to -2.81, the ANOVA variances were estimated with 5000 replications. For individuals $q = 24$ and 79 with log-likelihoods -5.70 and -6.44, respectively, it can be seen that the ANOVA variances are less accurate, despite the fact that we increased the number of replications to 20,000 for these individuals. The relative error on the ANOVA variances thus seems larger for individuals with a low choice probability. The estimated variances of the individual likelihood estimator with constructed lattices and other point sets are plotted in Figures 23 and 24. The empirical distributions of the values of the estimator are presented in Figure 25. Finally, histograms showing the empirical distribution of the individual log-likelihoods across individuals, and of the simulated values of the average log-likelihood function across independent replications, are given in Figure 26.

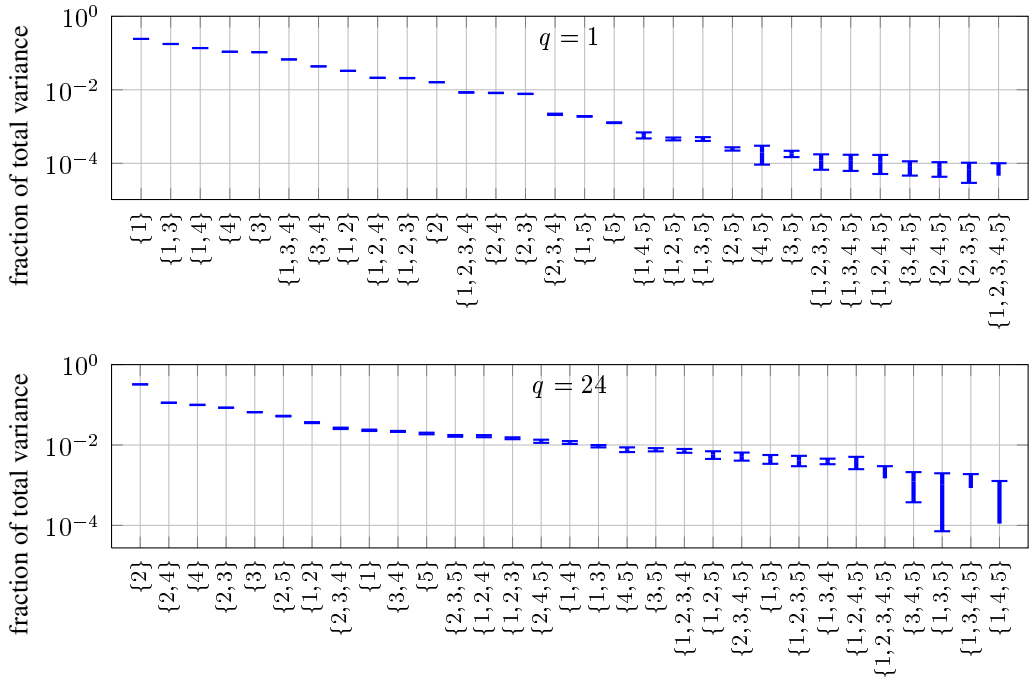

Figure 21: Fractional ANOVA variances of the conditional likelihood function for a single individual $q = 1$ (top) and 24 (bottom), for the example with real-life data. See 1 for further details.

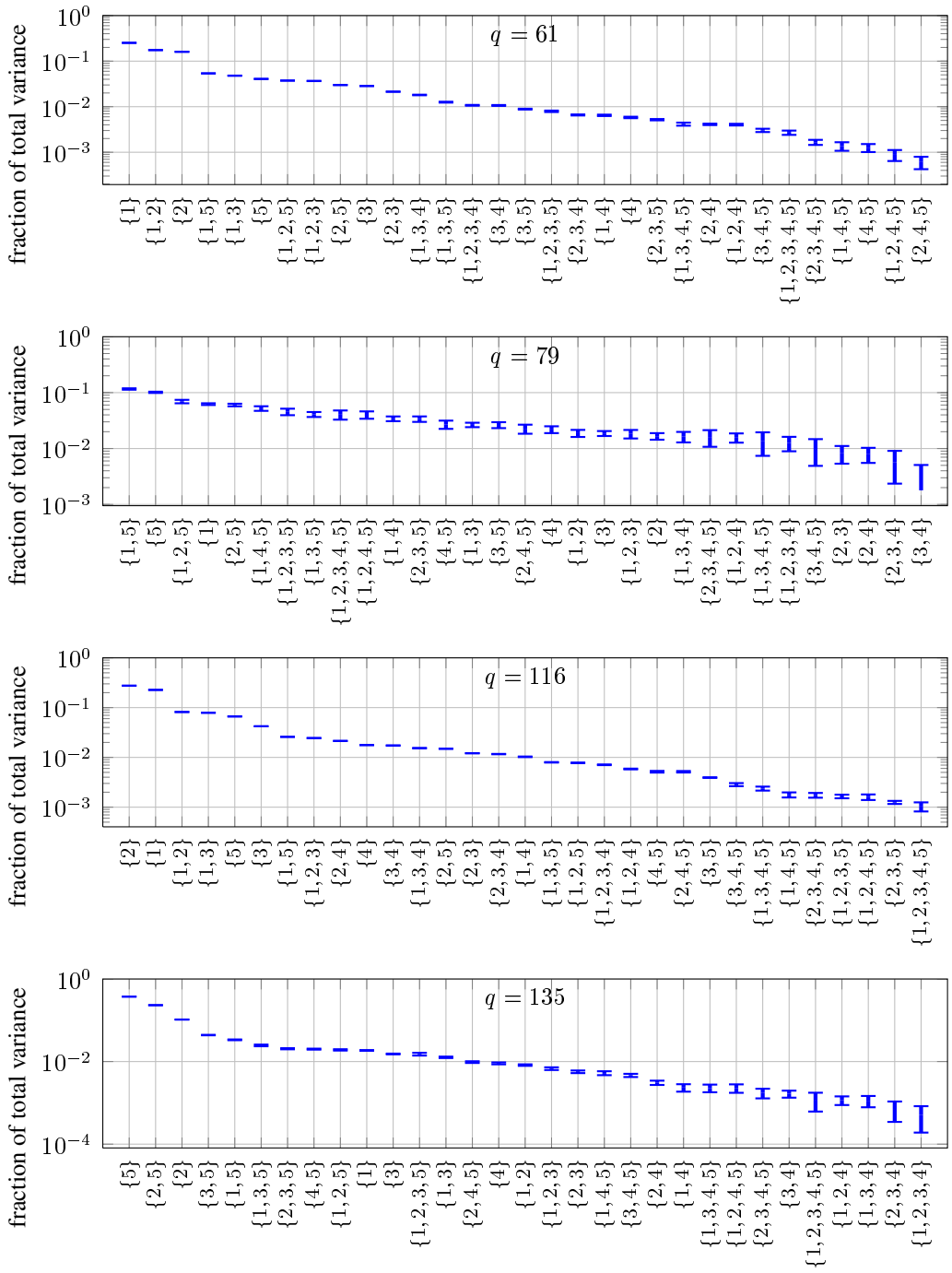

Figure 22: Fractional ANOVA variances of the conditional likelihood function for a single individual $q = 61$, 79, 116 and 135 (from top to bottom), for the example with real-life data. See 1 for further details.

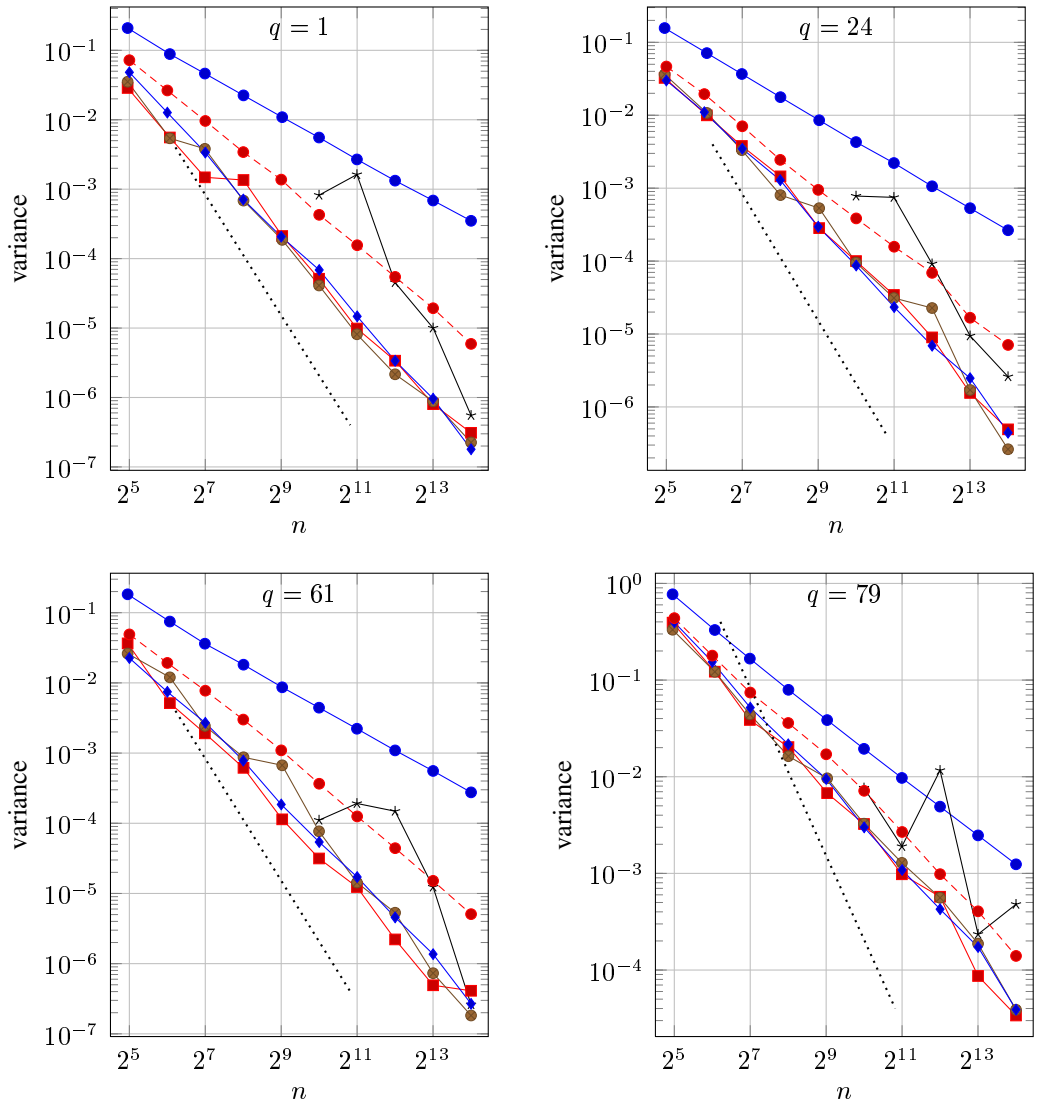

Figure 23: Estimated variance of the MC and RQMC estimators of the log-likelihood of a single individual for the examples with real-life data, for individuals $q = 1$ (top left), 24 (top right), 61 (bottom left) and 79 (bottom right), using MC (—●—), lattice-γ_u (—■—), lattice-0.1 (—●—), lattice-$M32$ (—⋆—), Sobol' nets (—♦—), and Halton points (-●-). For lattice-order and lattice-0.5, the variances are very similar to those of lattice-γ_u and lattice-0.1, and we do not show them to reduce the number of curve superpositions. The results for the Halton sequence (not shown here) are practically undistinguishable from those for the Halton points. The dotted line indicates the n^{-2} slope, for reference.

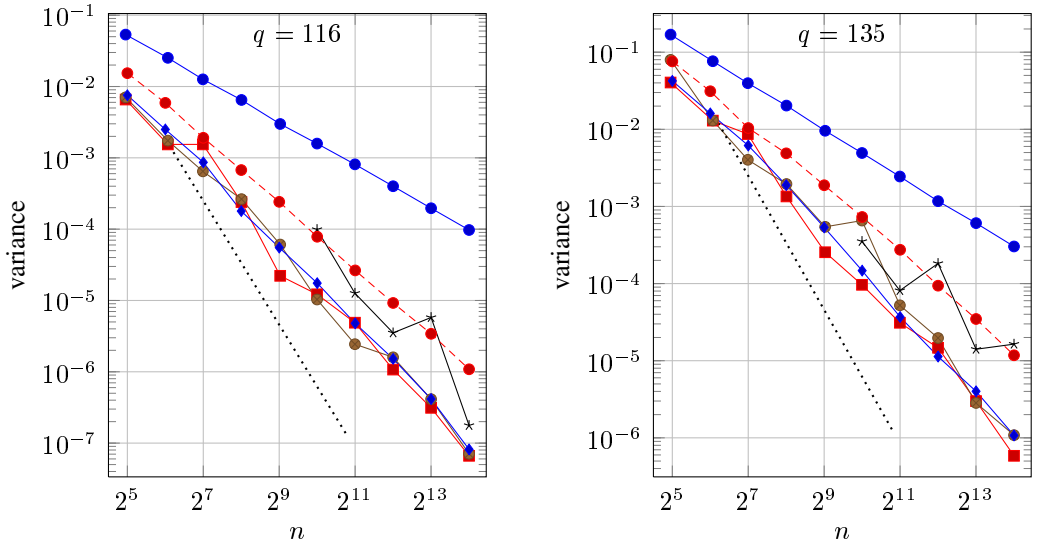

Figure 24: Same as Figure 23, but for individuals $q = 116$ (left) and 135 (right).

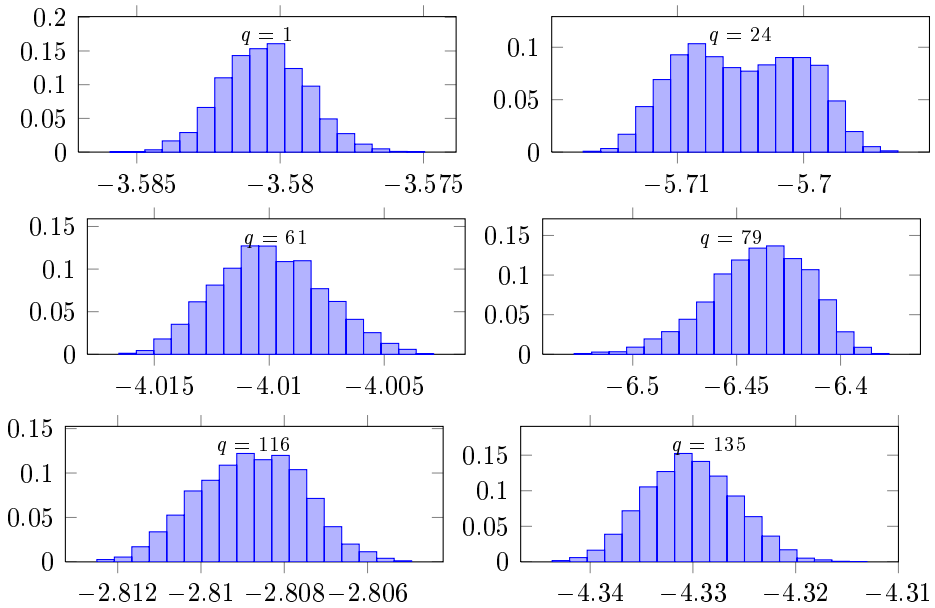

Figure 25: Empirical distribution of the simulated log-likelihood $\ln \hat{p}_q^n$ of a single individual for the example with real-life data, for individuals $q = 1$, 24, 61, 79, 116 and 135 (row-wise), using lattice-0.1 with $n = 4093$.

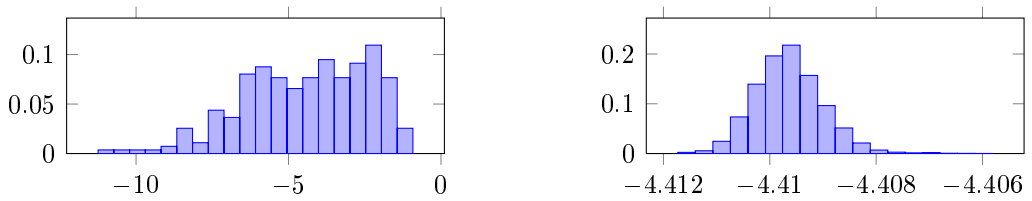

Figure 26: Empirical distribution of the $m = 274$ estimated individual log-likelihoods across individuals (left) and of the simulated values of the log-likelihood function across independent replications (right), for the example with real-life data, using lattice-0.1 with $n = 4093$.

C Lattice Parameters

The parameters for lattice-0.1, lattice-0.25 and lattice-0.5 are given in Tables 1, 2, and 3, respectively. To generate a lattice in s dimensions, use $\mathbf{v}_1 = (a_1, a_2, \ldots, a_s)/n$.

n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
31	1	12	9	17	4	6	10	16	13	7	2	5	11	3	8
32	1	9	13	7	15	5	3	11	5	3	11	13	1	15	7
64	1	27	15	23	25	29	19	17	3	11	7	9	31	13	21
67	1	18	14	8	20	23	12	17	5	26	11	19	32	2	30
127	1	29	24	56	38	35	10	43	16	50	52	31	18	44	7
128	1	49	37	23	29	47	39	53	63	9	5	57	45	51	33
256	1	99	67	37	107	47	117	53	19	13	31	83	127	29	61
257	1	76	113	54	44	97	231	83	12	211	124	33	117	5	60
274	1	115	127	85	35	59	133	31	69	25	263	117	123	65	43
512	1	149	115	87	55	123	45	153	193	139	37	109	181	79	191
521	1	199	226	53	127	135	109	190	230	409	511	22	17	337	79
1021	1	647	154	420	214	456	473	295	96	63	891	104	354	426	401
1024	1	275	421	231	71	453	83	483	105	325	27	411	19	371	345
2039	1	462	705	520	775	1640	348	182	882	1788	570	236	675	32	962
2048	1	791	549	207	493	659	535	225	87	277	541	477	131	595	631
4093	1	1210	984	1577	1785	612	439	1110	1467	1244	2023	1486	1092	947	1288
4096	1	1557	1237	1119	481	175	295	2025	429	747	1197	201	863	1271	1393
8191	1	2431	3799	1570	1690	992	806	2083	2924	2714	1337	3462	3669	1878	220
8192	1	2431	3739	1689	3185	2609	3849	1525	71	2109	2585	679	3083	3657	433
16381	1	6789	1848	3501	6232	5261	2010	13207	2720	2974	3100	3669	3747	3551	986
16384	1	6229	2691	1399	7751	2865	3221	379	2211	1593	4075	2911	3051	7907	2063

Table 1: Lattice parameters for geometric weights with $\gamma = 0.1$.

n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
31	1	12	9	14	4	6	20	7	15	2	10	13	3	5	8
32	1	9	15	7	5	11	3	13	1	7	9	15	3	13	5
64	1	19	29	11	3	17	21	5	27	25	15	13	31	7	9
67	1	26	6	23	10	14	19	8	32	28	17	30	21	12	5
127	1	98	54	61	46	13	9	31	43	51	6	39	56	24	50
128	1	49	37	23	29	5	63	45	11	13	19	51	43	35	9
256	1	75	47	111	125	87	15	27	65	123	71	89	39	23	105
257	1	71	56	21	120	75	12	26	114	53	10	95	103	100	7
512	1	149	115	193	225	155	27	245	207	145	131	105	151	215	139
521	1	144	249	79	163	420	231	134	53	176	476	184	220	181	107
1021	1	374	420	154	130	37	104	214	16	402	980	237	322	496	302
1024	1	275	167	403	195	481	253	131	321	371	365	101	111	499	215
2039	1	462	711	140	398	505	956	745	165	1522	642	1868	1001	593	18
2048	1	791	549	207	287	659	641	271	611	385	445	759	95	989	361
4093	1	2378	1422	499	1559	92	1136	1939	1314	2257	1388	1579	830	856	681
4096	1	1557	1741	1449	1873	1009	371	47	1673	787	127	215	365	1289	265
8191	1	2431	3799	1570	6501	992	806	1072	3662	1914	4798	356	127	328	5674
8192	1	3457	2879	3047	1631	975	2383	3665	1751	3175	1343	261	887	1325	1953
16381	1	3711	5711	3321	8615	9236	6041	5832	670	11056	5153	1779	323	6091	3623
16384	1	6915	3959	7525	1123	7817	3185	6091	6655	5519	7241	2535	4815	931	635

Table 2: Lattice parameters for geometric weights with $\gamma = 0.25$.

n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
31	1	12	5	3	10	8	14	6	15	4	7	13	2	9	11
32	1	7	15	5	3	9	11	13	1	7	9	13	5	15	3
64	1	27	15	31	25	29	9	21	11	7	23	13	5	17	3
67	1	41	6	28	9	23	21	10	14	25	8	12	24	5	7
127	1	29	73	66	46	50	59	35	41	3	10	24	48	8	31
128	1	47	19	11	53	15	59	45	21	31	55	3	41	23	5
256	1	75	47	111	125	87	53	113	7	95	99	109	33	43	117
257	1	71	20	104	57	169	59	5	106	120	9	36	123	81	55
512	1	149	113	193	51	187	167	109	179	93	41	215	249	217	91
521	1	144	272	79	163	37	94	255	152	211	81	90	34	190	51
1021	1	374	154	420	352	61	322	302	89	231	247	289	271	496	245
1024	1	275	167	403	195	61	145	283	35	349	267	165	251	125	359
2039	1	462	711	140	398	26	241	670	96	777	326	968	553	459	522
2048	1	791	213	957	761	37	697	375	775	471	891	255	69	825	477
4093	1	1210	2551	1785	842	1113	910	418	2775	822	460	1003	1714	897	1031
4096	1	1557	1741	1873	1449	1061	1213	735	709	437	169	1541	1023	1735	1577
8191	1	2431	3799	1141	520	2865	3896	3528	3514	971	788	851	3562	717	1842
8192	1	2433	3867	1159	2847	3779	3191	1447	1615	2183	671	97	3221	45	1869
16381	1	9592	1848	6013	7065	13117	4236	5320	1907	413	6127	8168	7284	6739	2486
16384	1	6229	2691	3349	5893	3723	1143	4779	6569	6173	2619	2029	2195	4415	2383

Table 3: Lattice parameters for geometric weights with $\gamma = 0.5$.