Computational Statistics and Data Analysis 151 (2020) 107010

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Faster Monte Carlo estimation of joint models for time-to-event and multivariate longitudinal data

Pete Philipson [a,*], Graeme L. Hickey [b], Michael J. Crowther [c], Ruwanthi Kolamunnage-Dona [b]

[a] School of Mathematics, Statistics & Physics, Newcastle University, United Kingdom of Great Britain and Northern Ireland
[b] Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, United Kingdom of Great Britain and Northern Ireland
[c] Biostatistics Research Group, Department of Health Sciences, University of Leicester, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Article history:
Received 3 December 2018
Received in revised form 5 May 2020
Accepted 6 May 2020
Available online 27 May 2020

Keywords:
Quasi Monte Carlo
Joint modelling
Multivariate longitudinal
Time-to-event
EM algorithms

ABSTRACT

Quasi-Monte Carlo (QMC) methods using quasi-random sequences, as opposed to pseudo-random samples, are proposed for use in the joint modelling of time-to-event and multivariate longitudinal data. The QMC integration framework extends the Monte Carlo Expectation Maximisation approaches that are commonly adopted, namely using ordinary and antithetic variates. The motivation of QMC integration is to increase the convergence speed by using nodes that are scattered more uniformly. Through simulation, estimates and computational times are compared and this is followed with an application to a clinical dataset. There is a distinct speed advantage in using QMC methods for small sample sizes and QMC is comparable to the antithetic MC method for moderate sample sizes. The new method is available in an updated version of the R package `joineRML`.

1. Introduction

Longitudinal studies in clinical research involve subjects who are followed-up repeatedly and on whom response data are collected such as, for example, one or more biomarkers (Gould et al., 2015). The time to an event is also usually of interest, for example death. The longitudinal data may be censored by this time-to-event outcome. Modelling these two outcome processes separately is generally inefficient, and can lead to biased effect size estimates if the two outcome processes are correlated (Ibrahim et al., 2010). Consequently, during the past two decades, there has been a rapid and substantial development in research on joint modelling of longitudinal and time-to-event data (Wulfsohn and Tsiatis, 1997; Henderson et al., 2000; Ibrahim et al., 2010; Rizopoulos, 2010; Asar et al., 2015). Motivation has stemmed from three broad scientific objectives (Henderson et al., 2000): drawing inference about a repeated measurement outcome subject to an informative dropout mechanism; drawing inference for a time-to-event outcome, whilst taking account of an intermittently and possibly error-prone measured endogenous time-dependent variable; and studying the joint evolution of the correlated processes.

* Correspondence to: School of Mathematics, Statistics & Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom of Great Britain and Northern Ireland.
E-mail address: peter.philipson1@newcastle.ac.uk (P. Philipson).

https://doi.org/10.1016/j.csda.2020.107010
0167-9473/

One avenue of current research is on the incorporation of multiple longitudinal outcomes, and despite several software packages being developed to fit joint models (Rizopoulos, 2010; Crowther et al., 2013; Kim, 2016; Philipson et al., 2017; Xu et al., 2020), they have mostly been limited to the setting of univariate longitudinal outcomes. This is despite a plethora of research into joint models of multivariate longitudinal data and time-to-event outcomes (Hickey et al., 2016). From a theoretical perspective, the extension of the classical univariate joint model to that of the multivariate case is straightforward. However, from a practical viewpoint, the estimation algorithms for fitting these extended models are computationally expensive. This is due to the need to integrate out subject-specific random effects from the generally intractable likelihood, of which the number increases with each additional longitudinal outcome. Hence, the search for efficient integration methods in multivariate joint models is highly motivated with the first work in this area beginning to emerge (Crowther, 2018; Martin et al., 2020).

There have been several proposals on how to fit joint models in recent years. Aside from the fitting paradigm, for example Bayesian (Xu and Zeger, 2001), frequentist (Henderson et al., 2000; Crowther et al., 2013), generalised estimating equations (Song et al., 2002), or latent class model estimation (Proust-Lima et al., 2012), a common theme is the evaluation of complex integrals. Common approaches to evaluating these integrals are Gaussian quadrature (GQ), including adaptive (Wulfsohn and Tsiatis, 1997) and pseudo-adaptive (Rizopoulos, 2012) variants ; Laplace approximations (Rizopoulos et al., 2009) and Monte Carlo (MC) estimation (Lin et al., 2002).

These approaches have several pros and cons. GQ can be very accurate, however the number of nodes increases at an exponential rate (Cools, 2002). Therefore, it is only useful for low dimensional random effects. Adaptive and pseudo-adaptive GQ approaches can reduce some of this burden, but still remain constrained by the exponential growth in nodes. Laplace approximations are particularly amenable to high dimensional data, and have been shown to reduce computational burden; however, the approximation may be inaccurate when some subjects contribute very few observations (Rizopoulos et al., 2009). MC methods benefit from a convergence rate that is independent of the dimensionality of the problem (Lemieux, 2009). However, estimates are subject to MC error, which can only be reduced by increasing the number of MC draws.

In this paper, we describe the estimation of a joint model with a solitary time-to-event outcome and multivariate longitudinal data, with a linear random effects structure. We consider an alternative method, which has, to-date, not been routinely adopted into statistical estimation: Quasi-Monte Carlo (QMC) (Caflisch, 1998). QMC methods are comprised of a set of heuristic algorithms that generate low discrepancy sequences. Such methods have been shown to be highly suitable to high-dimensional generalised linear mixed models (Pan and Thompson, 2007). Our primary objective is therefore to compare three different variations of MC: (1) ordinary MC (OMC); (2) antithetic variables MC (AMC); and (3) QMC. In particular, we contrast the estimates and the computational requirements using simulation studies with two and three longitudinal biomarkers across three sample sizes, and analogous applications to a clinical dataset.

The latest version of the R package `joineRML` (Hickey et al., 2018) has been extended to fit the aforementioned model using a Monte Carlo Expectation Maximisation algorithm (MCEM) (Wei and Tanner, 1990; Lin et al., 2002) under QMC. Previously, the package was predicated on exploiting a variance reduction technique for the MC E-step — antithetic variables simulation.

2. Joint models for multivariate longitudinal data and time-to-event data

2.1. Model

For each subject $i = 1, \ldots, n$, $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^\top, \ldots, \boldsymbol{y}_{iK}^\top)$ is the K-variate continuous outcome vector, where each $\boldsymbol{y}_{ik}$ denotes an $(n_{ik} \times 1)$-vector of observed longitudinal measurements for the kth outcome type: $\boldsymbol{y}_{ik} = (y_{i1k}, \ldots, y_{in_{ik}k})^\top$. Each outcome is measured at observed (possibly pre-specified) times t_{ijk} for $j = 1, \ldots, n_{ik}$, which can differ between subjects and outcomes. Additionally, for each subject there is an event time T_i^*, which is subject to right censoring. Therefore, we observe $T_i = \min(T_i^*, C_i)$, where C_i corresponds to a potential censoring time, and the failure indicator δ_i, which is equal to 1 if the failure is observed ($T_i^* \leq C_i$) and 0 otherwise. We assume that censoring is independent and, along with the measurement times, non-informative.

The model we describe is the natural extension of the model proposed by Henderson et al. (2000) to the case of multivariate longitudinal data. The kth longitudinal data submodel is given by

$$y_{ik}(t) = \mu_{ik}(t) + W_{1i}^{(k)}(t) + \varepsilon_{ik}(t), \tag{1}$$

where $\mu_{ik}(t)$ is the mean response, $W_{1i}^{(k)}(t)$ is a latent process and $\varepsilon_{ik}(t)$ is the model error term, which we assume to be independent and identically distributed normal, with mean zero and variance σ_k^2. The mean response is specified as a linear model

$$\mu_{ik}(t) = \boldsymbol{x}_{ik}^\top(t)\boldsymbol{\beta}_k, \tag{2}$$

where $\boldsymbol{x}_{ik}(t)$ is a p_k-vector of (possibly) time-varying covariates with corresponding fixed effect terms $\boldsymbol{\beta}_k$. In the models considered here $W_{1i}^{(k)}(t)$ is specified as a linear combination of random effects, namely

$$W_{1i}^{(k)}(t) = \boldsymbol{z}_{ik}^\top(t)\boldsymbol{b}_{ik}, \tag{3}$$

where $\boldsymbol{z}_{ik}(t)$ is an r_k-vector of (possibly) time-varying covariates with corresponding subject-and-outcome random effect terms $\boldsymbol{b}_{ik}$, which follow a zero-mean multivariate normal distribution with $(r_k \times r_k)$-variance–covariance matrix $\boldsymbol{D}_{kk}$. To account for dependence between the different longitudinal outcome outcomes, we let $\text{cov}(\boldsymbol{b}_{ik}, \boldsymbol{b}_{il}) = \boldsymbol{D}_{kl}$ for $k \neq l$. This latent process subsequently links the separate submodels via association parameters. Furthermore, we assume $\varepsilon_{ik}(t)$ and $\boldsymbol{b}_{ik}$ are uncorrelated, and that the censoring times are independent of the random effects.

The submodel for the time-to-event outcome is given by the hazard model

$$\lambda_i(t) = \lambda_0(t) \exp\left\{\boldsymbol{v}_i^\top \boldsymbol{\gamma}_v + W_{2i}(t)\right\},$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard, and $\boldsymbol{v}_i$ is a q-vector of baseline measured covariates with corresponding fixed effect terms $\boldsymbol{\gamma}_v$. Conditional on $W_i(t)$ and the observed covariate data, the longitudinal and time-to-event data generating processes are independent. To establish a latent association, we specify $W_{2i}(t)$ as a linear combination of $\left\{W_{1i}^{(1)}(t), \ldots, W_{1i}^{(K)}(t)\right\}$:

$$W_{2i}(t) = \sum_{k=1}^{K} \gamma_{yk} W_{1i}^{(k)}(t),$$

where $\boldsymbol{\gamma}_y = (\gamma_{y1}, \ldots, \gamma_{yK})$ are the corresponding association parameters. To emphasise the dependence of $W_{2i}(t)$ on the random effects, we explicitly write it as $W_{2i}(t, \boldsymbol{b}_i)$ from here onwards.

2.2. Estimation

The *observed* data likelihood for the joint outcome is given by

$$\prod_{i=1}^{n} \left(\int_{-\infty}^{\infty} f(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) f(T_i, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) f(\boldsymbol{b}_i \mid \boldsymbol{\theta}) d\boldsymbol{b}_i \right), \tag{4}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \text{vech}(\boldsymbol{D}), \sigma_1^2, \ldots, \sigma_K^2, \lambda_0(t), \boldsymbol{\gamma}_v^\top, \boldsymbol{\gamma}_y^\top)$ is the collection of unknown parameters that we want to estimate, with $\text{vech}(\boldsymbol{D})$ denoting the half-vectorisation operator that returns the vector of lower-triangular elements of matrix $\boldsymbol{D}$, given by

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_{11} & \cdots & \boldsymbol{D}_{1K} \\ \vdots & \ddots & \vdots \\ \boldsymbol{D}_{1K}^\top & \cdots & \boldsymbol{D}_{KK} \end{pmatrix},$$

and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_K^\top)^\top$, and $\boldsymbol{b}_i = (\boldsymbol{b}_{i1}^\top, \ldots, \boldsymbol{b}_{iK}^\top)^\top$.

As per (Henderson et al., 2000), we exploit the expectation–maximisation (EM) algorithm (Dempster et al., 1977) for fitting the model, by treating the random effects $\boldsymbol{b}_i$ as missing data. Starting from an initial estimate of the parameters, $\hat{\boldsymbol{\theta}}^{(0)}$, the procedure involves iterating between an M-step and an E-step until convergence is achieved. Full details of the M-step are provided elsewhere (Wulfsohn and Tsiatis, 1997; Lin et al., 2002), and remain identical for all approaches considered in this research. In short, all parameters except $\boldsymbol{\gamma}_v$ and $\boldsymbol{\gamma}_y$ are available in closed-form; the parameters in the Cox proportional hazards submodel are estimated by a one-step Newton–Raphson or a quasi-Newton one-step update that is an analogue of the Gauss–Newton method (McLachlan and Krishnan, 2008, p. 8). Standard errors can be approximated after the EM algorithm has converged using the empirical information matrix approximation (Lin et al., 2002), allowing for Wald-like confidence intervals to be estimated and this is the method adopted here. Alternatively, bootstrap estimation can be used, but at increased computational expense (Henderson et al., 2000; Hsieh et al., 2006) — we utilise this approach in Appendix C.

3. *E*-step approaches

At each E-step, it is required that we compute the expected log-likelihood of the *complete* data conditional on the *observed* data and the current estimate of the parameters,

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(m)}) &= \sum_{i=1}^{n} \mathbb{E}\left\{\log f(\boldsymbol{y}_i, T_i, \delta_i, \boldsymbol{b}_i \mid \boldsymbol{\theta})\right\} \\ &= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left\{\log f(\boldsymbol{y}_i, T_i, \delta_i, \boldsymbol{b}_i \mid \boldsymbol{\theta})\right\} f(\boldsymbol{b}_i \mid T_i, \delta_i, \boldsymbol{y}_i; \hat{\boldsymbol{\theta}}^{(m)}) d\boldsymbol{b}_i. \end{aligned}$$

Here, the complete-data likelihood contribution for subject i is given by the integrand of (4).

Consequently, the M-step update involves terms of the form $\mathbb{E}\left[h(\boldsymbol{b}_i) \mid T_i, \delta_i, \boldsymbol{y}_i; \hat{\boldsymbol{\theta}}\right]$, for known functions $h(\cdot)$. It can be shown that this conditional expectation can be written as

$$\mathbb{E}\left[h(\boldsymbol{b}_i) \mid T_i, \delta_i, \boldsymbol{y}_i; \hat{\boldsymbol{\theta}}\right] = \frac{\int_{-\infty}^{\infty} h(\boldsymbol{b}_i) f(\boldsymbol{b}_i \mid \boldsymbol{y}_i; \hat{\boldsymbol{\theta}}) f(T_i, \delta_i \mid \boldsymbol{b}_i; \hat{\boldsymbol{\theta}}) d\boldsymbol{b}_i}{\int_{-\infty}^{\infty} f(\boldsymbol{b}_i \mid \boldsymbol{y}_i; \hat{\boldsymbol{\theta}}) f(T_i, \delta_i \mid \boldsymbol{b}_i; \hat{\boldsymbol{\theta}}) d\boldsymbol{b}_i}, \tag{5}$$

where $f(T_i, \delta_i \mid \boldsymbol{b}_i; \hat{\boldsymbol{\theta}})$ is given by

$$\begin{aligned} f(T_i, \delta_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}) = &\left[\lambda_0(T_i) \exp\left\{\boldsymbol{v}_i^\top \boldsymbol{\gamma}_v + W_{2i}(T_i, \boldsymbol{b}_i)\right\}\right]^{\delta_i} \\ &\times \exp\left\{-\int_0^{T_i} \lambda_0(u) \exp\left\{\boldsymbol{v}_i^\top \boldsymbol{\gamma}_v + W_{2i}(u, \boldsymbol{b}_i)\right\} du\right\}, \end{aligned}$$

and $f(\boldsymbol{b}_i \mid \boldsymbol{y}_i; \hat{\boldsymbol{\theta}})$ is calculated from multivariate normal distribution theory as

$$\boldsymbol{b}_i \mid \boldsymbol{y}_i, \boldsymbol{\theta} \sim N\left(\boldsymbol{A}_i \left\{\boldsymbol{Z}_i^\top \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})\right\}, \boldsymbol{A}_i\right), \tag{6}$$

with $\boldsymbol{A}_i = \left(\boldsymbol{Z}_i^\top \boldsymbol{\Sigma}_i^{-1} \boldsymbol{Z}_i + \boldsymbol{D}^{-1}\right)^{-1}$, where $\boldsymbol{\Sigma}_i = \bigoplus_{k=1}^{K} \sigma_k^2 I_{n_{ik}}$ is a diagonal matrix. Also $\boldsymbol{X}_i = \bigoplus_{k=1}^{K} \boldsymbol{X}_{ik}$ and $\boldsymbol{Z}_i = \bigoplus_{k=1}^{K} \boldsymbol{Z}_{ik}$ are block-diagonal matrices, with $\boldsymbol{X}_{ik} = \left(\boldsymbol{x}_{i1k}^\top, \ldots, \boldsymbol{x}_{in_{ik}k}^\top\right)$ an $(n_{ik} \times p_k)$-design matrix, with the jth row corresponding to the p_k-vector of covariates measured at time t_{ijk}, and where $\bigoplus$ denotes the direct matrix sum.

Without loss of generality, we outline the approaches with respect to the integral of a Lebesgue integrable function $f(\boldsymbol{x})$ on the unit cube $I^d = [0, 1]^d$ in d-dimensions. We define

$$I[f] = \mathbb{E}[f(\boldsymbol{x})] = \int_{I^d} f(\boldsymbol{x}) d\boldsymbol{x}. \tag{7}$$

All of the approaches described naturally generalise to non-uniform distributions. That is, for a d-dimensional non-uniform random variable $\boldsymbol{x} \in \mathbb{R}^d$ with density function $p(\boldsymbol{x}) : \mathbb{R}^d \mapsto \mathbb{R}$, we can define

$$I[f] = \mathbb{E}[f(\boldsymbol{x})] = \int_{\mathbb{R}^d} f(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}.$$

3.1. Monte Carlo

Monte Carlo integration is a probabilistic representation of the integral (Lemieux, 2009). Namely, consider a random sequence $\{\boldsymbol{x}_n\}_{n=1}^N$ independently sampled from $U^d(0, 1)$, then an empirical approximation to (7) is

$$I_N[f] = \frac{1}{N} \sum_{n=1}^{N} f(\boldsymbol{x}_n), \tag{8}$$

which converges *almost surely* to $I[f]$, i.e.

$$\lim_{N \to \infty} I_N[f] \overset{\text{a.s.}}{\to} I[f].$$

The central limit theorem (CLT) can be used to determine that the error of the Monte Carlo integration is of order $O(N^{-1/2})$. It is useful to note that this order is independent of the integral dimension d.

For the E-step in the multivariate joint model problem, this translates to a strategy of first sampling a sequence $\{\boldsymbol{b}_i^n\}_{n=1}^N$ from (6) and, for each subject i, calculating

$$\sum_{n=1}^{N} \frac{h(\boldsymbol{b}_i^n) f(T_i, \delta_i \mid \boldsymbol{b}_i^n; \hat{\boldsymbol{\theta}})}{\sum_{n=1}^{N} f(T_i, \delta_i \mid \boldsymbol{b}_i^n; \hat{\boldsymbol{\theta}})}. \tag{9}$$

3.2. Antithetic variables

Variance reduction techniques for Monte Carlo integration are used to accelerate the convergence (Lemieux, 2009). As noted above, the error of the Monte Carlo integration is of order $O(N^{-1/2})$. By exploiting the CLT, it is possible to show that the constant is equal to the standard deviation of the integrand, $\sigma[f]$ given by

$$\sigma[f] = \left(\int_{I^d} (f(\boldsymbol{x}) - I[f])^2 \, d\boldsymbol{x}\right)^{\frac{1}{2}}.$$

The rationale of variance reduction techniques is to reduce this constant term, thus making the convergence relatively faster.

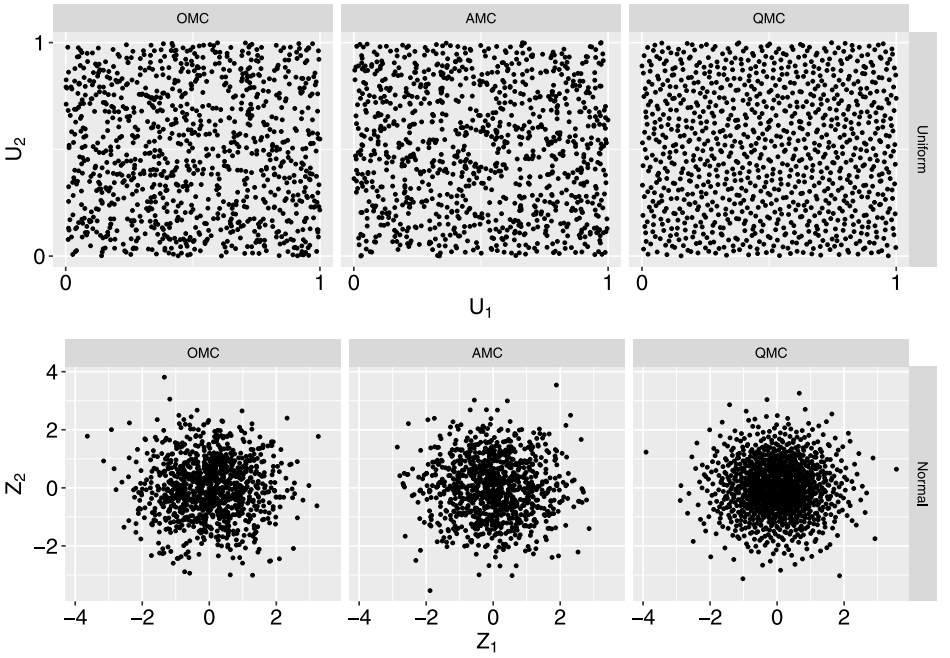

Fig. 1. Bivariate uniform $U^2(0,1)$ (top row) and normal $N_2(0, I_2)$ (bottom row) deviates sampled according to ordinary Monte Carlo (OMC; left column), antithetic Monte Carlo (AMC; middle column), and quasi-Monte Carlo (QMC; right column) using a Sobol sequence with Owen-type scrambling.

If we again consider a random sequence $\{\boldsymbol{x}_n\}_{n=1}^{N/2}$, with N an even number, and also the corresponding antithetic variates $\{\tilde{\boldsymbol{x}}_n\}_{n=1}^{N/2} = \{\mathbf{1} - \boldsymbol{x}_n\}_{n=1}^{N/2}$, then the antithetic empirical approximation to (7) is

$$I_N[f] = \frac{1}{N}\sum_{n=1}^{N/2}\left(f(\boldsymbol{x}_n) + f(\tilde{\boldsymbol{x}}_n)\right). \tag{10}$$

Assuming $\sigma[f] < \infty$, the variance in the antithetic sampling approach is

$$\frac{1}{N}\left[\sigma^2[f] + \operatorname{cov}\left(f(\boldsymbol{x}), f(\tilde{\boldsymbol{x}})\right)\right].$$

Hence, if $\operatorname{cov}\left(f(\boldsymbol{x}), f(\tilde{\boldsymbol{x}})\right) < 0$, then the error term will be smaller than if we had sampled N independent draws of $\boldsymbol{x}_n$ as per ordinary Monte Carlo; see Section 3.1.

For the E-step in the multivariate joint model case, this translates to a strategy of first sampling $\boldsymbol{\Omega} \sim N(0, \boldsymbol{I}_r)$ and obtaining the antithetic pairs $\pm\boldsymbol{\Omega}$, which will both be r-vectors of standard normal samples, that are then transformed to the required form of (6) via

$$\boldsymbol{A}_i\left\{\boldsymbol{Z}_i^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\} \pm \boldsymbol{C}_i\boldsymbol{\Omega},$$

where $\boldsymbol{I}_r$ is the identity matrix of dimension $r = \dim(\boldsymbol{b}_i)$, and $\boldsymbol{C}_i$ is the Cholesky decomposition of $\boldsymbol{A}_i$ such that $\boldsymbol{C}_i\boldsymbol{C}_i^\top = \boldsymbol{A}_i$. Drawing $N/2$ pairs and evaluating (9) yields the antithetic estimate of (5).

3.3. Quasi-Monte Carlo

Ordinary Monte Carlo (OMC) and antithetic Monte Carlo (AMC) are predicated on a probabilistic interpretation. That is, they use random (or rather, pseudo-random) sequences, which ensures convergence of order $O(N^{-1/2})$. Quasi-Monte Carlo (QMC) methods, on the other hand, use quasi-random sequences, which are deterministic (Lemieux, 2009; Caflisch, 1998). For this reason, they are sometimes referred to as low-order deterministic sequences. The motivation of QMC integration is to reduce the order of convergence and number of nodes required, by using nodes that are scattered more uniformly on I^d than pseudo-random points, which, by virtue of independence, often display clusterings. Fig. 1 shows sampling nodes for the distributions $U^2(0, 1)$ and $N_2(0, I_2)$ deviates under OMC, AMC, and QMC (based on the Sobol sequence, described in Section 3.3.2).

Quasi-random sequences yield smaller errors than standard Monte Carlo integration methods, which follows from the Koksma–Hlawka inequality (Koksma, 1942; Hlawka, 1961):

$$|I[f] - I_N[f]| \leq V[f] D_N^*(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N), \tag{11}$$

where $V[f]$ is the variation of $f(\cdot)$ over I^d in the sense of Hardy and Krause (Hardy, 1906). The quantity D_N^* is a measure of the uniformity (called the star discrepancy) of the sequence $\{\boldsymbol{x}_n\}_{n=1}^N$, defined by

$$D_N^*(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \sup_{J \in I^d} \left| \frac{1}{N} \sum_{n=1}^{N} 1_{\{\boldsymbol{x}_n \in J\}} - \mathcal{A}(J) \right|, \tag{12}$$

where $\mathcal{A}(J)$ is the volume of the hyper-rectangular set J in I^d that has one vertex at 0. See Caflisch (1998, §5.4) for an outline of the proof. It can be recognised that (12) is in fact the Kolmogorov–Smirnov test statistic. The inequality in (11) is an upper-bound and, moreover, since $V[f]$ is fixed for a given $f(\cdot)$, the bound is determined by D_N^*. Hence, for any sequence, the error of the approximation is bounded by order $O\left((\log N)^d N^{-1}\right)$. It has been reported that this bound is usually overcautious, and in practice faster convergence is observed.

In the E-step under consideration here, this translates to a strategy of first calculating uniform deterministic sequences, $\{\boldsymbol{x}_i\}_{i=1}^N$, and transforming them to multivariate normal deviates $\boldsymbol{b}_i = \Phi_d^{-1}(\boldsymbol{x}_i)$, where $\Phi_d(\cdot)$ is the d-variate Gaussian cumulative distribution function. Alternative methods of transformation and the discrepancy properties have been explored using the Box–Muller transformation (Ökten and Göncü, 2011). Several deterministic sequences have been proposed in recent years. In the following, we consider two standard sequences: the Halton sequence and the Sobol sequence. Efficient algorithms for implementing both are widely available, which we access through the R package `randtoolbox` (Dutang and Savicky, 2018).

3.3.1. The Halton sequence

The Halton sequence (Halton, 1960) is the multidimensional generalisation of the one-dimensional van der Corput (VDC) sequence (van der Corput, 1935). The nth element of the VDC sequence is constructed by reversing the representation of base-n. In other words, every integer n has a b-adic representation $n = \sum_{l=0}^{L-1} d_l(n) b^l$ (with $0 \leq d_l(n) < b$ the lth digit). The nth number in the VDC sequence is then $g_b(n) = \sum_{l=0}^{L-1} d_l(n) b^{-l-1}$. The Halton sequence extends this approach by considering the p_m-adic expansion of n for the mth dimension, where p_m is the mth prime number.

3.3.2. The Sobol sequence

The Sobol sequence (Sobol, 1967) is an alternative, as well as the most-widely used (Atanassov et al., 2010), low-discrepancy sequence. The mathematical determination of the sequence is based on linear recurrence relationships over the finite field $\mathbb{F}_2 = \{0, 1\}$; hence, it is more involved than the Halton sequence, but a fast algorithm based on the Grey code implementation is available. For details, we refer the reader to Antonov and Saleev (1979). By considering orthogonal projections of multidimensional Sobol sequences, it has been shown that non-uniformity can occur. By scrambling the low-discrepancy sequence – a hybrid of OMC and QMC – this issue can potentially be alleviated (Chi et al., 2005). In particular, a commonly used scrambling method is that of Owen (1998).

4. Simulation study and results

Two simulation studies were conducted, assuming $K = 2$ and $K = 3$ longitudinal outcomes respectively, each with $n = 250, 500$ and 1000 subjects. The submodels, for $k = 1, \ldots, K$, are given as

$$y_{ijk} = (\beta_{0,k} + b_{i0k}) + (\beta_{1,k} + b_{i1k}) t_j + \beta_{2,k} x_{i1} + \beta_{3,k} x_{i2} + \varepsilon_{ijk}, \tag{13}$$

$$\lambda_i(t) = \lambda_0(t) \exp\left\{ \boldsymbol{x}_i \boldsymbol{\gamma}_{\boldsymbol{v}} + \Sigma_{k=1}^{K} \gamma_{yk} (b_{i0k} + b_{i1k} t) \right\}, \tag{14}$$

where $\varepsilon_{ijk} \sim N(0, \sigma_k^2)$ and $K = 2$ and $K = 3$ for the bivariate and trivariate scenarios respectively; we refer to these scenarios as 'Scenario 1' and 'Scenario 2' hereafter.

In both scenarios, longitudinal data were simulated according to a follow-up schedule of 6 time points (at times $0, 1, \ldots, 5$), with each model including subject-and-outcome-specific random-intercepts and random-slopes, whereby $r_k = 2$. The event rate in the simulations was $\approx$ 40%. Independent censoring times were drawn from an exponential distribution with rate 0.05. Any subject with $T_i > 5$ was censored at time $C = 5.1$. For all submodels, we included a pair of covariates $\boldsymbol{x}_i = (x_{i1}, x_{i2})^\top$, where x_{i1} is a continuous covariate independently drawn from $N(0, 1)$ and x_{i2} is a binary covariate independently drawn from $Bin(1, 0.5)$. In the above, $\lambda_0(t) = e^{0.25 - 3.5t}$, $\boldsymbol{b}_i$ has a multivariate normal distribution, $N_{2K}(0, D)$, and D is a specified unstructured $(2K \times 2K)$-covariance matrix (since $r_k = 2$) with (up to) $\sum_i^{2K} i$ unique parameters. For this study, the true parameters were: $\boldsymbol{\beta}_{\boldsymbol{1}}^\top = (0, 1, 1, 1)$, $\boldsymbol{\beta}_{\boldsymbol{2}}^\top = (0, -1, 0, 0.5)$, $\sigma_1^2 = \sigma_2^2 = 0.25$, $\boldsymbol{\gamma}_{\boldsymbol{v}}^\top = (0, 1)$, $\boldsymbol{\gamma}_{\boldsymbol{y}}^\top = (-0.5, 1)$. The random effect covariance matrices were and $D_{1,1} = D_{3,3} = 0.5^2$, $D_{2,2} = D_{4,4} = 0.2^2$, and $D_{1,3} = D_{3,1} = -0.5^3$ with the remaining elements of $D_{i,j}$ set to zero.

A total of 1000 datasets were simulated for each setting using the `simData()` function in the `joineRML` R package (Hickey et al., 2018), and the `mjoint()` function used to fit the models. For each simulated dataset, we fitted the

Table 1
Computing times in seconds (SD in parentheses) under ordinary (OMC), antithetic (AMC) and quasi (QMC) Monte Carlo methods for both simulation scenarios and all sample sizes.

		Method		
		OMC	AMC	QMC
Scenario 1	$n = 250$	129.67 (81.88)	23.83 (24.32)	13.77 (6.62)
	$n = 500$	449.15 (921.31)	108.49 (288.44)	167.55 (339.91)
	$n = 1000$	573.30 (680.37)	132.53 (59.35)	259.62 (122.54)
Scenario 2	$n = 250$	243.83 (193.42)	51.96 (43.90)	24.90 (9.78)
	$n = 500$	570.12 (602.72)	128.25 (141.53)	122.59 (113.68)
	$n = 1000$	692.14 (303.15)	318.31 (96.09)	482.80 (145.89)

Table 2
Coverage rates of regression parameters using ordinary (OMC), antithetic (AMC) and quasi (QMC) Monte Carlo methods for both simulation scenarios: $n = 500$.

Parameter	Scenario 1			Scenario 2		
	OMC	AMC	QMC	OMC	AMC	QMC
$\beta_{0,1}$	0.945	0.942	0.941	0.956	0.956	0.955
$\beta_{1,1}$	0.960	0.959	0.955	0.960	0.959	0.955
$\beta_{2,1}$	0.958	0.959	0.958	0.961	0.961	0.962
$\beta_{3,1}$	0.949	0.949	0.949	0.950	0.949	0.948
$\beta_{0,2}$	0.969	0.964	0.967	0.956	0.959	0.954
$\beta_{1,2}$	0.949	0.936	0.945	0.954	0.945	0.937
$\beta_{2,2}$	0.956	0.957	0.955	0.950	0.951	0.953
$\beta_{3,2}$	0.960	0.961	0.962	0.966	0.968	0.967
$\beta_{0,3}$	–	–	–	0.953	0.954	0.954
$\beta_{1,3}$	–	–	–	0.951	0.952	0.953
$\beta_{2,3}$	–	–	–	0.962	0.961	0.962
$\beta_{3,3}$	–	–	–	0.954	0.956	0.955
γ_{v1}	0.954	0.956	0.955	0.955	0.955	0.956
γ_{v2}	0.957	0.957	0.962	0.932	0.932	0.932
γ_{y1}	0.960	0.961	0.961	0.957	0.958	0.958
γ_{y2}	0.945	0.948	0.946	0.936	0.938	0.937
γ_{y3}	–	–	–	0.967	0.970	0.970

model with OMC (using `control` argument `type` = ‘‘`montecarlo`’’), AMC (`type` = ‘‘`antithetic`’’) and QMC using a Sobol sequence with Owen-type scrambling (`type` = ‘‘`sobol`’’). Full R code to run the simulation studies is available from https://github.com/petephilipson/QMC.

Each of the three methods (using the different E-step routines) was successfully applied to 1000 simulations, which were carried out in parallel on an HPC cluster, consisting of nodes with 64 GB RAM and 14-core 2.4 GHz CPU for each scenario. As expected, the time for the joint model to fit was much faster for the AMC approach relative to the OMC approach in all cases (Fig. 2), with a smaller number of nodes used at the last iteration (Fig. A.3). The proposed QMC method outperforms AMC for small sample sizes in each scenario, and is comparable when $n = 500$. However, AMC is superior to QMC for the largest sample size under consideration here. Convergence times and standard deviations across both scenarios and all three MC approaches are given in Table 1. Pilot results with larger samples sizes indicate that AMC requires fewer nodes than QMC, which, in turn, requires fewer nodes than OMC.

The differences in computational times were primarily explained by the changes in MC nodes due to the dynamic nature of the E-step algorithm described — see Hickey et al. (2018) for details of the software implementation of the algorithm. In other words, the mean number of MCEM algorithm iterations were often similar for each method, but the mean number of nodes used at the last iteration were substantially different. For example, in scenario 1, the mean number of iterations for the OMC, AMC, and QMC approaches was only 33, 20, and 15, respectively, whereas the mean number of nodes used at the last iteration – at which (9) is evaluated – were 3516, 1043 and 703.

Coverage rates were also calculated under each approach for both scenarios. Table 2 and Table 3 display the coverage rates for the regression and variance parameters respectively when $n = 500$, with equivalent results for $n = 250$ and $n = 1000$ given in Table A.5 to Table A.8. The rates are very similar across methods and scenarios, and are all close to the nominal 0.95 level.

5. Application

To compare the three MC methods in a clinical setting we consider an application to the oft-used (Albert and Shih, 2010; Crowther et al., 2013; Andrinopoulou and Rizopoulos, 2016) primary biliary cirrhosis (PBC) data, which is publicly available in a variety of places, including within the `joineRML` R package; further details are given in Hickey et al. (2018).

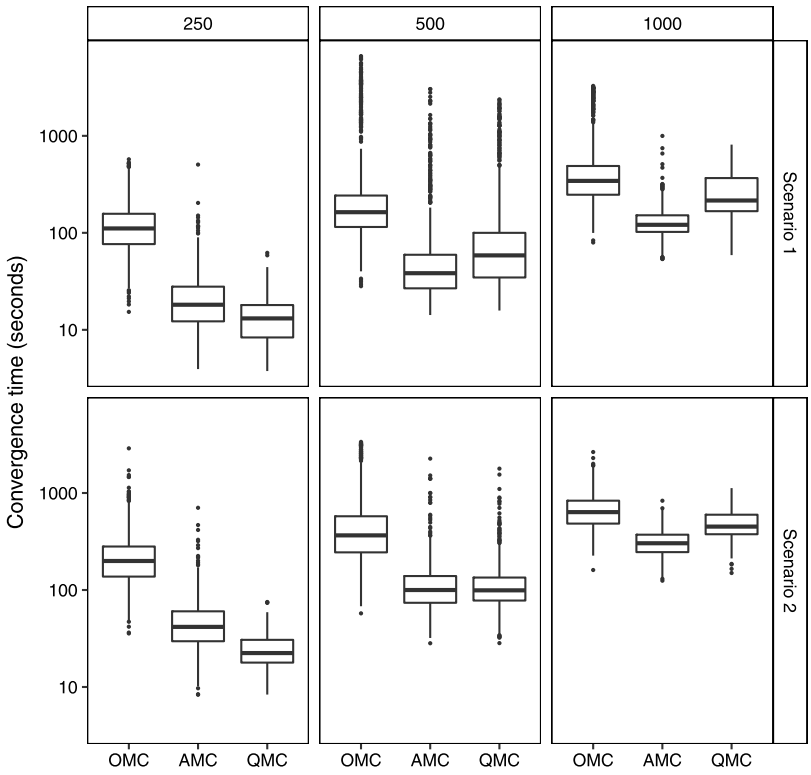

Fig. 2. Distribution of run times over 1000 simulated datasets using each MC method for simulation study 1 (upper row) and simulation study 2 (lower row) at each sample size. Run-times are on the $\log_{10}$ scale.

Table 3
Coverage rates of variance parameters using ordinary (OMC), antithetic (AMC) and quasi (QMC) Monte Carlo methods for both simulation scenarios: $n = 500$.

Parameter	Scenario 1			Scenario 2		
	OMC	AMC	QMC	OMC	AMC	QMC
$D_{1,1}$	0.951	0.956	0.948	0.959	0.960	0.956
$D_{2,2}$	0.946	0.947	0.932	0.954	0.954	0.945
$D_{3,3}$	0.941	0.945	0.935	0.957	0.956	0.950
$D_{4,4}$	0.948	0.944	0.950	0.956	0.952	0.944
$D_{5,5}$	–	–	–	0.951	0.952	0.953
$D_{6,6}$	–	–	–	0.958	0.958	0.959
σ_1^2	0.963	0.964	0.968	0.969	0.967	0.967
σ_2^2	0.952	0.951	0.950	0.956	0.958	0.959
σ_3^2	–	–	–	0.966	0.968	0.966

PBC is a chronic liver disease affecting the bile ducts of the liver, causing liver damage, cirrhosis and, ultimately, death in many cases.

In keeping with the simulation study, we consider both two and three longitudinal biomarkers, although many more are available within this rich data source. In keeping with the simulation study, the models for each biomarker include random effects for both the intercept and slope. In the PBC study, patients were randomly assigned to either the drug D-penicillamine ($n = 158$) or placebo ($n = 154$), here represented by the binary covariate x_{i2} and we include the age of patients at baseline as our continuous covariate, x_{i1}. There were a total of 140 (44.9%) deaths amongst the 312 patients.

As detailed in Hickey et al. (2018), a log transformation is commonly used for one of the biomarkers (serum bilirubin) and another (prothrombin index) requires a transformation based on the residuals from the longitudinal sub-model, and, hence, these transformations are adopted here too. The submodels for $k = 1, 2, 3$ are as given by (13) where $k = 1$ represents the log-transformed serum bilirubin, $k = 2$ corresponds to serum albumin and $k = 3$ is the Box–Cox-transformed prothrombin time. Also, $\boldsymbol{x}_i = (x_{i1}, x_{i2})^\top$ is the vector of baseline covariates and $\boldsymbol{\gamma}_v$ the vector of associated parameters in the survival submodel, and $\boldsymbol{b}_i$ follows a multivariate normal distribution, $N_{2K}(0, D)$.

Table 4
Parameter estimates from each Monte Carlo method (with standard errors in parentheses) for the bivariate application to the PBC data.

Parameter	Method		
	OMC	AMC	QMC
$\beta_{0,1}$	0.495 (0.298)	0.498 (0.299)	0.489 (0.302)
$\beta_{1,1}$	0.188 (0.010)	0.188 (0.010)	0.187 (0.010)
$\beta_{2,1}$	0.001 (0.006)	0.001 (0.006)	0.001 (0.006)
$\beta_{3,1}$	−0.108 (0.117)	−0.107 (0.118)	−0.105 (0.119)
$\beta_{0,2}$	3.952 (0.120)	3.949 (0.121)	3.951 (0.122)
$\beta_{1,2}$	−0.109 (0.005)	−0.109 (0.005)	−0.108 (0.005)
$\beta_{2,2}$	−0.008 (0.002)	−0.008 (0.002)	−0.008 (0.002)
$\beta_{3,2}$	0.037 (0.043)	0.037 (0.044)	0.037 (0.044)
γ_{v1}	0.066 (0.014)	0.065 (0.014)	0.066 (0.015)
γ_{v2}	−0.250 (0.291)	−0.248 (0.291)	−0.246 (0.292)
γ_{y1}	1.020 (0.124)	1.019 (0.124)	1.023 (0.123)
γ_{y2}	−2.415 (0.335)	−2.410 (0.333)	−2.405 (0.333)
$D_{1,1}$	0.993 (0.111)	0.992 (0.111)	0.992 (0.111)
$D_{2,2}$	0.034 (0.005)	0.034 (0.005)	0.033 (0.005)
$D_{3,3}$	0.117 (0.014)	0.117 (0.014)	0.118 (0.014)
$D_{4,4}$	0.005 (0.001)	0.005 (0.001)	0.005 (0.001)
σ_1^2	0.347 (0.002)	0.347 (0.002)	0.347 (0.002)
σ_2^2	0.319 (0.002)	0.319 (0.002)	0.319 (0.002)

Table A.5
Coverage rates of regression parameters under each Monte Carlo method for both simulation scenarios: $n = 250$.

Parameter	Scenario 1			Scenario 2		
	OMC	AMC	QMC	OMC	AMC	QMC
$\beta_{0,1}$	0.949	0.952	0.950	0.964	0.965	0.963
$\beta_{1,1}$	0.957	0.957	0.952	0.968	0.971	0.968
$\beta_{2,1}$	0.961	0.962	0.961	0.963	0.964	0.964
$\beta_{3,1}$	0.948	0.950	0.950	0.971	0.969	0.971
$\beta_{0,2}$	0.952	0.950	0.948	0.964	0.965	0.965
$\beta_{1,2}$	0.952	0.954	0.952	0.960	0.957	0.956
$\beta_{2,2}$	0.953	0.955	0.954	0.968	0.968	0.968
$\beta_{3,2}$	0.957	0.955	0.955	0.963	0.966	0.966
$\beta_{0,3}$	–	–	–	0.966	0.965	0.968
$\beta_{1,3}$	–	–	–	0.967	0.967	0.970
$\beta_{2,3}$	–	–	–	0.969	0.971	0.970
$\beta_{3,3}$	–	–	–	0.967	0.967	0.967
γ_{v1}	0.959	0.960	0.960	0.966	0.966	0.966
γ_{v2}	0.940	0.941	0.940	0.942	0.942	0.942
γ_{y1}	0.960	0.960	0.963	0.968	0.969	0.969
γ_{y2}	0.948	0.948	0.951	0.959	0.959	0.960
γ_{y3}	–	–	–	0.967	0.968	0.967

We fitted the bivariate and trivariate models with OMC, AMC and QMC (using the control arguments described in Section 4) in turn to the PBC data (the code for model fitting is available at https://github.com/petephilipson/QMC. The superiority of QMC over the other two methods seen in the simulation study for small sample size is repeated here. The times for the model fits for $K = 2$ were 200.9, 122.5 and 14.6 s for OMC, AMC and QMC respectively, and for $K = 3$ the corresponding times were 429.0, 256.9 and 12.0 s.

Parameter estimates and standard errors (using the approximate method discussed earlier in Section 2.2) from the three methods are displayed in Table 4 for the bivariate model. Both parameter estimates and associated standard errors are very similar across the three methods. However, we also utilise the bootstrap to quantify the uncertainty in run times since times based on a single model fit may not be indicative of generally quicker performance. The bootstrap results consolidated our findings in this case, namely that QMC provided considerably faster run times than the other two methods for a small sample size; indeed the differences were even more marked in this setting. Results under the bootstrap approach are included in Appendix C; we note in passing that differences between the empirical and bootstrap standard errors were negligible in this instance.

6. Discussion

In this article, we compared traditional and quasi-MC approaches for the numerical integration of random effects in a multivariate joint model. QMC integration is based on choosing uniformly scattered deterministic nodes in a

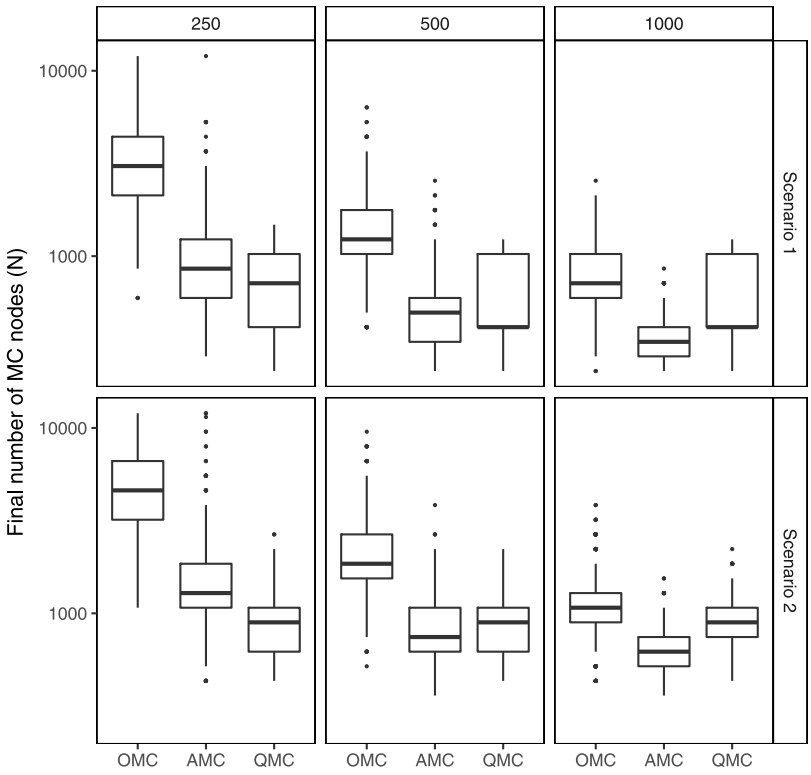

Fig. A.3. Distribution of final number of MC nodes over 1000 simulated datasets using each MC method for simulation study 1 (upper) and simulation study 2 (lower) at each sample size. Nodes are on the $\log_{10}$ scale.

Table A.6
Coverage rates of variance parameters under each Monte Carlo method for both simulation scenarios: $n = 250$.

Parameter	Scenario 1			Scenario 2		
	OMC	AMC	QMC	OMC	AMC	QMC
$D_{1,1}$	0.962	0.962	0.955	0.962	0.961	0.960
$D_{2,2}$	0.953	0.954	0.942	0.966	0.961	0.953
$D_{3,3}$	0.957	0.956	0.952	0.966	0.966	0.952
$D_{4,4}$	0.957	0.954	0.957	0.968	0.969	0.962
$D_{5,5}$	–	–	–	0.964	0.962	0.965
$D_{6,6}$	–	–	–	0.959	0.961	0.962
σ_1^2	0.964	0.962	0.962	0.963	0.963	0.963
σ_2^2	0.966	0.969	0.967	0.974	0.973	0.974
σ_3^2	–	–	–	0.966	0.966	0.967

unit hypercube rather than pseudo-random nodes. There are many proposals on how to calculate these deterministic sequences, including Sobol sequences, Halton sequences, Faure sequences, and various scrambled versions. Despite the promising results here and elsewhere in other applications (e.g. Pan and Thompson, 2007), QMC methods have not widely penetrated the biostatistical methodological field. In the joint model described here, the time-to-event model was semiparametric. Standard approaches to the estimation of this model have been based on the EM algorithm (Henderson et al., 2000; Lin et al., 2002), which converges linearly. Hence, it is necessary to speed up convergence by offsetting computational overheads such as the E-step.

Alongside the setting considered in this work, joint models using parametric time-to-event models (e.g. Crowther et al., 2013) can also exploit QMC methods in alternative fitting algorithms, e.g. Newton–Raphson. Furthermore, QMC could be used in any other setting that typically relies on numerical integration, with examples ranging from competing risks in a classic joint modelling setting (Williamson et al., 2008) to semi-competing risks for clustered survival data (Peng et al., 2018), or with frailty in a copula model (Emura et al., 2017).

In any event, there are practical issues in using QMC methods for fitting joint models. Like the OMC and AMC cases, a suitable number of nodes need to be selected; in our example, we initially choose a small value of $N = 100 \times K$, where

Table A.7
Coverage rates of regression parameters under each Monte Carlo method for both simulation scenarios: $n = 1000$.

Parameter	Scenario 1			Scenario 2		
	OMC	AMC	QMC	OMC	AMC	QMC
$\beta_{0,1}$	0.955	0.946	0.951	0.953	0.951	0.951
$\beta_{1,1}$	0.947	0.948	0.935	0.950	0.946	0.949
$\beta_{2,1}$	0.958	0.959	0.958	0.960	0.958	0.958
$\beta_{3,1}$	0.944	0.947	0.947	0.944	0.943	0.944
$\beta_{0,2}$	0.933	0.931	0.930	0.940	0.939	0.937
$\beta_{1,2}$	0.922	0.880	0.926	0.942	0.915	0.919
$\beta_{2,2}$	0.952	0.952	0.952	0.955	0.958	0.958
$\beta_{3,2}$	0.943	0.949	0.949	0.954	0.954	0.955
$\beta_{0,3}$	–	–	–	0.950	0.954	0.952
$\beta_{1,3}$	–	–	–	0.944	0.938	0.947
$\beta_{2,3}$	–	–	–	0.961	0.961	0.961
$\beta_{3,3}$	–	–	–	0.955	0.952	0.954
γ_{v1}	0.955	0.954	0.954	0.950	0.952	0.951
γ_{v2}	0.941	0.943	0.943	0.942	0.942	0.942
γ_{y1}	0.944	0.946	0.947	0.950	0.956	0.955
γ_{y2}	0.934	0.937	0.937	0.948	0.943	0.945
γ_{y3}	–	–	–	0.951	0.954	0.953

Table A.8
Coverage rates of variance parameters under each Monte Carlo method for both simulation scenarios: $n = 1000$.

Parameter	Scenario 1			Scenario 2		
	OMC	AMC	QMC	OMC	AMC	QMC
$D_{1,1}$	0.947	0.946	0.945	0.955	0.955	0.950
$D_{2,2}$	0.950	0.948	0.932	0.963	0.964	0.948
$D_{3,3}$	0.958	0.959	0.946	0.956	0.958	0.942
$D_{4,4}$	0.945	0.940	0.947	0.953	0.949	0.940
$D_{5,5}$	–	–	–	0.965	0.966	0.966
$D_{6,6}$	–	–	–	0.941	0.939	0.941
σ_1^2	0.954	0.953	0.952	0.950	0.947	0.955
σ_2^2	0.959	0.959	0.958	0.948	0.947	0.949
σ_3^2	–	–	–	0.951	0.951	0.950

K is the number of longitudinal outcomes, as it is computationally inefficient to initialise the algorithm with a large N as the estimates may be far from their true values. In addition, we continue to dynamically increase the number of nodes as $K + \lfloor K/\delta \rfloor$ for some small δ (e.g. $\delta = 3$) following the same methodology used for the OMC and AMC approaches (Ripatti et al., 2002). As described in Hickey et al. (2018), this ensures that Monte Carlo error does not overwhelm any changes to parameter estimates. A limitation of QMC methods is that they do not permit the estimation of MC error. However, using the scrambling methods this can be overcome (Owen, 1998).

Here, we showed that QMC reduced the computational time required to fit multivariate joint models with a small sample size and was comparable to AMC for moderate sample size. Comparison of times should be viewed cautiously, as they may not translate across software implementations. However, in the comparisons described here, the code was identical except for the Monte Carlo simulator, thus it is reasonable to state that the above findings hold.

In future work, we will explore the application of QMC to a wider range of generalised mixed models including multi-outcome models. A natural avenue for research within the framework of the models considered here would be to compare quadrature, MCEM and Laplace methods for choices of K and n in order to make recommendations as to which method should be advocated, and when.

Acknowledgements

This research was funded by a Medical Research Council (MRC), United Kingdom grant (MR/M013227/1) awarded to RKD, and used to fund GLH's position. MJC is part-funded by an MRC New Investigator Research, United Kingdom Grant (MR/P015433/1). The funder had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. Results for the simulation study and bootstrap standard errors for the application were performed on the Oswald HPC cluster at Northumbria University. We thank Jimmy Gibson (HPC Technology Specialist) for his assistance in using Oswald.

Table B.9
Regression parameter estimates from each Monte Carlo method (with standard errors in parentheses) for the trivariate application to the PBC data.

Parameter	Method		
	OMC	AMC	QMC
$\beta_{0,1}$	0.519 (0.309)	0.513 (0.312)	0.501 (0.308)
$\beta_{1,1}$	0.194 (0.012)	0.194 (0.012)	0.191 (0.012)
$\beta_{2,1}$	0.000 (0.006)	0.000 (0.006)	0.000 (0.006)
$\beta_{3,1}$	−0.093 (0.121)	−0.092 (0.121)	−0.093 (0.120)
$\beta_{0,2}$	3.918 (0.125)	3.920 (0.126)	3.922 (0.124)
$\beta_{1,2}$	−0.111 (0.006)	−0.111 (0.006)	−0.109 (0.006)
$\beta_{2,2}$	−0.008 (0.002)	−0.008 (0.002)	−0.008 (0.002)
$\beta_{3,2}$	0.036 (0.044)	0.036 (0.044)	0.035 (0.044)
$\beta_{0,3}$	1.004 (0.060)	1.005 (0.060)	1.006 (0.059)
$\beta_{1,3}$	−0.053 (0.003)	−0.053 (0.003)	−0.053 (0.003)
$\beta_{2,3}$	−0.003 (0.001)	−0.004 (0.001)	−0.003 (0.001)
$\beta_{3,3}$	0.025 (0.024)	0.025 (0.024)	0.025 (0.024)
γ_{v1}	0.065 (0.015)	0.065 (0.015)	0.065 (0.015)
γ_{v2}	−0.218 (0.301)	−0.217 (0.299)	−0.217 (0.294)
γ_{y1}	0.928 (0.145)	0.930 (0.145)	0.933 (0.146)
γ_{y2}	−1.991 (0.451)	−1.990 (0.452)	−1.981 (0.446)
γ_{y3}	−1.727 (1.075)	−1.702 (1.065)	−1.694 (1.067)

Table B.10
Variance component parameter estimates from each Monte Carlo method (with standard errors in parentheses) for the trivariate application to the PBC data.

Parameter	Method		
	OMC	AMC	QMC
$D_{1,1}$	0.992 (0.114)	0.992 (0.114)	0.993 (0.115)
$D_{2,2}$	0.037 (0.006)	0.036 (0.006)	0.036 (0.006)
$D_{3,3}$	0.116 (0.015)	0.116 (0.015)	0.116 (0.015)
$D_{4,4}$	0.005 (0.001)	0.005 (0.001)	0.005 (0.001)
$D_{5,5}$	0.042 (0.006)	0.042 (0.006)	0.042 (0.006)
$D_{6,6}$	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
σ_1^2	0.346 (0.002)	0.346 (0.002)	0.346 (0.002)
σ_2^2	0.319 (0.002)	0.319 (0.002)	0.319 (0.002)
σ_3^2	0.159 (0.001)	0.159 (0.001)	0.158 (0.001)

Appendix A. Simulation study: further results

Coverage rates for the simulation studies conducted with both $n = 250$ and $n = 1000$ are included in Tables A.5 to A.8. As for $n = 500$, we see near-identical coverage rates for each parameter, across the three methods. The final number of MC nodes for all three MC methods, each sample size and both simulation scenarios are included in Fig. A.3. These boxplots show that the final number of nodes is, as expected, strongly linked with the computing time. We also observe that, under QMC, the number of nodes is asymmetric under the first simulation scenario for $n > 500$.

Appendix B. Application: further details

B.1. Trivariate model

Full results from the trivariate joint model applied to the PBC data are included here. Tables B.9 and B.10 give parameter estimates and associated standard errors based on the trivariate model for the PBC data. Once more, we observe negligible differences between the parameter estimates, for both the regression coefficients and the variance components.

Appendix C. Bootstrap standard errors

The superiority of QMC over the other two methods seen in the simulation study is repeated when using 100 bootstrap samples to obtain standard errors in the application to the PBC dataset. Recall, the times for the model fits for $K = 2$ were 200.9, 122.5 and 14.6 s for OMC, AMC and QMC respectively. The median bootstrap sample fitting times (in seconds) were 533.8 (OMC), 324.7.0 (AMC) and 54.0 (QMC), showing a near ten-fold difference between OMC and QMC, and a six-fold difference comparing AMC with QMC. Similarly, for the trivariate model the sample fitting times (in seconds) were 429.0 (OMC), 256.9 (AMC) and 12.0 (QMC). Under bootstrap sampling, the median computing times were 1112, 598.1 and 117.0 s, demonstrating a near ten-fold difference between OMC and QMC, and a five-fold difference between AMC and

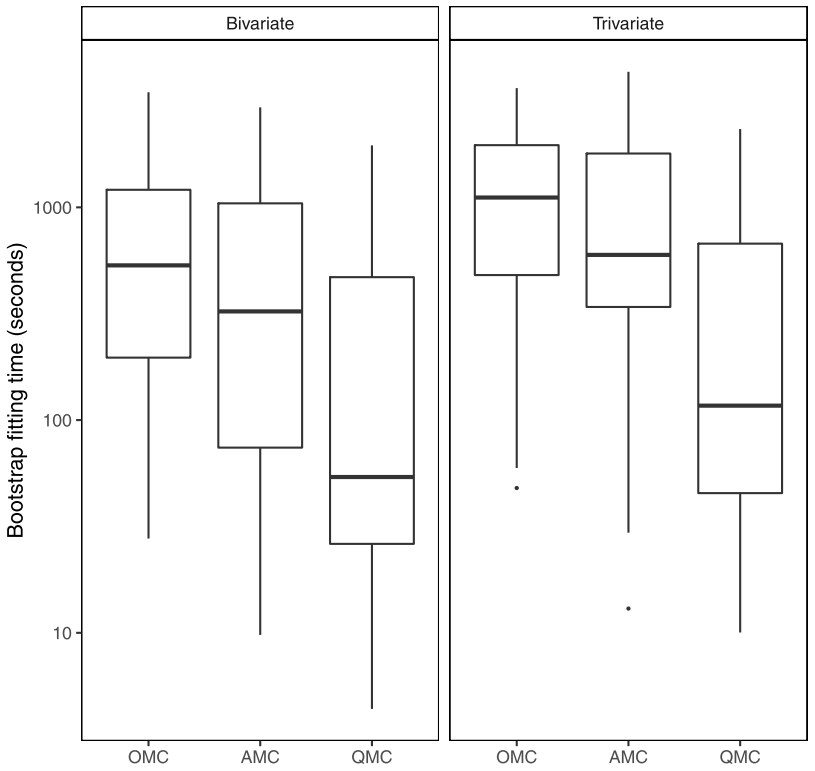

Fig. C.4. Distribution of algorithm multivariate joint model run times across 100 bootstrap samples using different MC integration routines for the E-step. Run-times are on the $\log_{10}$ scale.

QMC. A boxplot showing the distribution of fitting times across bootstrap samples for both models is given in Fig. C.4. The bootstrap samples and subsequent model fits were carried out in parallel on the same HPC cluster as for the simulations.

References

Albert, P.S., Shih, J.H., 2010. An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. Ann. Appl. Stat. 4 (3), 1517–1532.

Andrinopoulou, E.-R., Rizopoulos, D., 2016. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. Stat. Med. 35 (26), 4813–4823.

Antonov, I.A., Saleev, V., 1979. An economic method of computing LPτ-sequences. USSR Comput. Math. Math. Phys. 19 (1), 252–256.

Asar, O., Ritchie, J., Kalra, P., Diggle, P.J., 2015. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. Int. J. Epidemiol. 1–11.

Atanassov, E., Karaivanova, A., Ivanovska, S., 2010. Tuning the generation of Sobol sequence with Owen scrambling. In: Lirkov, I., Margenov, S., Waśniewski, J. (Eds.), Large-Scale Scientific Computing. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 459–466.

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

Chi, H., Beerli, P., Evans, D.W., Mascagni, M., 2005. On the scrambled Sobol sequence. In: International Conference on Computational Science. Springer, Atlanta, GA, USA, pp. 775–782, URL http://dx.doi.org/10.1007/11428862_105.

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

Crowther, M.J., 2018. Merlin-a unified modelling framework for data analysis and methods development in stata. arXiv preprint arXiv:1806.01615.

Crowther, M.J., Abrams, K.R., Lambert, P.C., 2013. Joint modeling of longitudinal and survival data. Stata J. 13 (1), 165–184.

Dempster, A., Laird, N., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Stat. Methodol. 39 (1), 1–38.

Dutang, C., Savicky, P., 2018. Randtoolbox: Generating and testing random numbers. R package version 1.17.1.

Emura, T., Nakatochi, M., Murotani, K., Rondeau, V., 2017. A joint frailty-copula model between tumour progression and death for meta-analysis. Stat. Methods Med. Res. 26 (6), 2649–2666.

Gould, A.L., Boye, M.E., Crowther, M.J., Ibrahim, J.G., Quartey, G., Micallef, S., Bois, F.Y., 2015. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. Stat. Med. 34, 2181–2195.

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2 (1), 84–90.

Hardy, G.H., 1906. On double Fourier series and especially those which represent the double zeta-function with real and incommensurable parameters. Quart. J. Math. 37 (1), 53–79.

Henderson, R., Diggle, P.J., Dobson, A., 2000. Joint modelling of longitudinal measurements and event time data. Biostatistics 1 (4), 465–480.

Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R., 2016. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Med. Res. Methodol. 16 (1), 1–15.

Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R., 2018. JoineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. BMC Med. Res. Methodol. 18 (1), 50.
Hlawka, E., 1961. Funktionen von beschränkter variatiou in der theorie der gleichverteilung. Ann. Mat. Pura Appl. 54 (1), 325–333.
Hsieh, F., Tseng, Y.K., Wang, J.L., 2006. Joint modeling of survival and longitudinal data: Likelihood approach revisited. Biometrics 62 (4), 1037–1043.
Ibrahim, J.G., Chu, H., Chen, L.M., 2010. Basic concepts and methods for joint models of longitudinal and survival data. J. Clin. Oncol. 28 (16), 2796–2801.
Kim, S., 2016. Jointmodel: Semiparametric joint models for longitudinal and counting processes. R package version 1.0. URL https://CRAN.R-project.org/package=JointModel.
Koksma, J., 1942. A general theorem from the theory of uniform distribution modulo 1. Math. B (Zutphen) 1 (7–11), 43.
Lemieux, C., 2009. Monte Carlo and Quasi-Monte Carlo Sampling. Springer, New York, NY.
Lin, H., McCulloch, C.E., Mayne, S.T., 2002. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. Stat. Med. 21, 2369–2382.
Martin, E., Gasparini, A., Crowther, M., 2020. Merlin: Mixed effects regression for linear, non-linear and user-defined models. R package version 0.0.2. URL https://CRAN.R-project.org/package=merlin.
McLachlan, G.J., Krishnan, T., 2008. The EM Algorithm and Extensions, second ed. Wiley-Interscience.
Ökten, G., Göncü, A., 2011. Generating low-discrepancy sequences from the normal distribution: Box-Muller or inverse transform? Math. Comput. Modelling 53 (5–6), 1268–1281.
Owen, A.B., 1998. Scrambling Sobol'and Niederreiter–Xing points. J. Complexity 14 (4), 466–489.
Pan, J., Thompson, R., 2007. Quasi-Monte Carlo estimation in generalized linear mixed models. Comput. Statist. Data Anal. 51 (12), 5765–5775.
Peng, M., Xiang, L., Wang, S., 2018. Semiparametric regression analysis of clustered survival data with semi-competing risks. Comput. Statist. Data Anal. 124, 53–70.
Philipson, P., Sousa, I., Diggle, P.J., Williamson, P., Kolamunnage-Dona, R., Henderson, R., Hickey, G.L., 2017. `JoineR`: Joint modelling of repeated measurements and time-to-event data. R package version 1.2.4. URL https://github.com/petephilipson/joineR/.
Proust-Lima, C., Sene, M., Taylor, J.M.G., Jacqmin-Gadda, H., 2012. Joint latent class models for longitudinal and time-to-event data: a review. Stat. Methods Med. Res. 23 (1), 74–90.
Ripatti, S., Larsen, K., Palmgren, J., 2002. Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. Lifetime Data Anal. 8 (2002), 349–360.
Rizopoulos, D., 2010. `JM`: an R package for the joint modelling of longitudinal and time-to-event data. J. Stat. Softw. 35 (9), 1–33.
Rizopoulos, D., 2012. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. Comput. Statist. Data Anal. 56 (3), 491–501.
Rizopoulos, D., Verbeke, G., Lesaffre, E., 2009. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (3), 637–654.
Sobol, I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. Zh. Vychisl. Mat. Mat. Fiz. 7 (4), 784–802.
Song, X., Davidian, M., Tsiatis, A.A., 2002. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics 3 (4), 511–528.
van der Corput, J.G., 1935. Verteilungsfunktionen. I. Mitt. Proc. Akad. Wet. Amst. 38, 813–821.
Wei, G.C., Tanner, M.A., 1990. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. J. Amer. Statist. Assoc. 85 (411), 699–704.
Williamson, P.R., Kolamunnage-Dona, R., Philipson, P., Marson, A.G., 2008. Joint modelling of longitudinal and competing risks data. Stat. Med. 27, 6426–6438.
Wulfsohn, M., Tsiatis, A.A., 1997. A joint model for survival and longitudinal data measured with error. Biometrics 53 (1), 330–339.
Xu, C., Hadjipantelis, P., Wang, J.-L., 2020. Semi-parametric joint modeling of survival and longitudinal data: The R package JSM. J. Stat. Softw. 93 (2), 1–29.
Xu, J., Zeger, S.L., 2001. Joint analysis of longitudinal data comprising repeated measures and times to events. J. R. Stat. Soc. Ser. C. Appl. Stat. 50 (3), 375–387.