Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Population Quasi-Monte Carlo

Chaofan Huang, V. Roshan Joseph & Simon Mak

To cite this article: Chaofan Huang, V. Roshan Joseph & Simon Mak (2022): Population Quasi-Monte Carlo, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2022.2034637

To link to this article: https://doi.org/10.1080/10618600.2022.2034637

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2022, VOL. 00, NO. 0, 1–14
https://doi.org/10.1080/10618600.2022.2034637

Population Quasi-Monte Carlo

Chaofan Huang[a], V. Roshan Joseph[a], and Simon Mak[b]

[a]H. Milton Stewart School of Industrial and Systems Engineering, Gerogia Institute of Technology, Atlanta, GA; [b]Department of Statistical Science, Duke University, Durham, NC

ABSTRACT

Monte Carlo methods are widely used for approximating complicated, multidimensional integrals for Bayesian inference. Population Monte Carlo (PMC) is an important class of Monte Carlo methods, which adapts a population of proposals to generate weighted samples that approximate the target distribution. When the target distribution is expensive to evaluate, PMC may encounter computational limitations since it requires many evaluations of the target distribution. To address this, we propose a new method, Population Quasi-Monte Carlo (PQMC), which integrates Quasi-Monte Carlo ideas within the sampling and adaptation steps of PMC. A key novelty in PQMC is the idea of importance support points resampling, a deterministic method for finding an "optimal" subsample from the weighted proposal samples. Moreover, within the PQMC framework, we develop an efficient covariance adaptation strategy for multivariate normal proposals. Finally, a new set of correction weights is introduced for the weighted PMC estimator to improve the efficiency from the standard PMC estimator. We demonstrate the improved empirical performance of PQMC over PMC in extensive numerical simulations and a friction drilling application. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received December 2020
Revised December 2021

KEYWORDS
Bayesian computation; Importance sampling; Monte Carlo; Quasi-Monte Carlo; Resampling; Support points

1. Introduction

A fundamental challenge in Bayesian inference is the evaluation of integrals involving some multi-dimensional posterior distribution π that cannot be analytically derived. Monte Carlo methods are often used in practice, and in particular, Markov chain Monte Carlo (MCMC) because of its flexibility and ease of implementation (Robert and Casella 2013). In recent decades, there has been renewed interest in exploring an alternative class of methods called iterated Importance Sampling (IS), which allows for parallel implementation and easy assessment of approximation error (Bugallo et al. 2017; Elvira et al. 2017). However, it is difficult to identify a good proposal distribution that mimics the target posterior π, especially when π is complicated and/or high-dimensional. On the other hand, Monte Carlo methods may encounter computational limitation in applications where the posterior distribution is expensive to evaluate (Joseph et al. 2019). To address the aforementioned challenges, we propose a novel Population Quasi-Monte Carlo (PQMC) framework, which integrates Quasi-Monte Carlo (QMC) ideas (Niederreiter 1992) within the Population Monte Carlo (PMC) method (Cappé et al. 2004) – a popular iterated IS method – for improved sampling efficiency.

Each iteration of PMC consists of three steps. First, from each of the K proposals $\{q_k\}_{k=1}^K$, it *samples* J particles, $x_{k,j} \sim q_k$. Next, it *weighs* the KJ particles $\{x_{k,j}\}_{k=1\,j=1}^{K\quad J}$ to correct for the mismatch between the proposal and target distribution. Last, it *adapts* the K proposals for selectively exploring only the relevant sample space; this adaptation idea can be traced back to Oh and Berger (1993), West (1993), and Givens and Raftery (1996). PMC repeats the *sampling, weighting*, and *adaptation* steps for T iterations, yielding a total of $N = TKJ$ weighted samples for approximating π. PMC is closely related to Sequential Monte Carlo (SMC) but with the same target distribution in all T iterations (see Chen 2003; Moral, Doucet, and Jasra 2006; Cappe, Godsill, and Moulines 2007 and for a review of SMC).

Resampling techniques are widely employed in PMC and SMC for adapting proposals' centers (Cappé et al. 2004; Elvira et al. 2017) to avoid weight degeneracy, as most of the weights could concentrate on only a few particles and thus, resulting in an estimator with large variance (Kong, Liu, and Wong 1994). By eliminating particles with insignificant weights and duplicating the important samples, resampling allows for efficient exploration of the promising regions, albeit with additional randomness. The simplest way to do this is multinomial resampling (Gordon, Salmond, and Smith 1993), which samples with replacement according to the weights. Several improved resampling methods have been proposed for variance reduction, including stratified (Kitagawa 1996), residual (Liu and Chen 1998), systematic (Carpenter, Clifford, and Fearnhead 1999), optimal (Fearnhead and Clifford 2003), and optimal transport resampling (Reich 2013). In one dimension, the ordered stratified resampling yields a set of randomized "low discrepancy" resamples, which reduces variation and yields desired resampling properties (Li et al. 2021). The multi-dimensional generalization of the ordering

CONTACT V. Roshan Joseph roshan@gatech.edu H. Milton Stewart School of Industrial and Systems Engineering, Gerogia Institute of Technology, Atlanta, GA 30332.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

utilizes the projection onto the space-filling Hilbert curve (Gerber and Chopin 2015; Gerber, Chopin, and Whiteley 2019), a continuous measure-preserving mapping from $[0,1]$ to $[0,1]^p$ where $p > 1$ is the dimension of the parameter space. However, because of the nonsmoothness of the Hilbert curve, points far away on the Hilbert curve could be close in $[0,1]^p$. Therefore, the "low discrepancy" resamples on the Hilbert curve might not be all well spread out in $[0,1]^p$. Hence, we propose the importance support points (ISP) resampling, which instead makes use of the support points in Mak and Joseph (2018) to find the set of space-filling resamples in the original parameter space that "optimally" represents the weighted proposal samples in terms of the energy distance (Székely and Rizzo 2013). This could retain the most information from the KJ simulated samples, and thus, may require fewer samples to get the same precision, which is especially beneficial for applications with expensive posterior.

Following Fearnhead (2005) and Gerber and Chopin (2015), we also leverage QMC ideas in the *sampling* step by generating a set of randomized low discrepancy points for the proposal samples $\{x_{k,j}\}_{k=1\,j=1}^{K\quad J}$. By combining QMC proposals with ISP resampling, the proposed PQMC framework can be shown to provide faster empirical convergence rate over PMC, thus, making PQMC a useful tool for tackling problems with expensive posteriors.

In addition, we propose two other improvements to PMC. First, by taking advantage of ISP resampling, we derive a computationally efficient adaptation scheme for the covariance parameter when elliptical proposal distribution is used, addressing one limitation of the existing PMC literature that the covariance is often kept fixed from beginning. This is important because a poorly chosen covariance could significantly hurt the efficiency of PMC. Second, by employing adaptation, the samples in the latter stages are often "better" quality, and thus, we also develop a weighted estimator accounting for the quality of the samples to further minimize the variance of the estimator.

The article is organized as follows. Section 2 first reviews QMC and then introduces the importance support points (ISP). Section 3 proposes the ISP resampling and compares it with the state-of-the-art Hilbert curve ordered resampling. Section 4 discusses the novel Population Quasi-Monte Carlo (PQMC), which makes use of the proposed ISP resampling and QMC proposals. Section 5 presents several simulation studies demonstrating the improvements of PQMC over the existing PMC methods. Section 6 illustrates the usefulness of PQMC on friction drilling model calibration application, where the posterior is computationally expensive. We conclude the article with some remarks in Section 7.

2. Importance Support Points

We first provide a brief review of QMC and then introduce the importance support points, which is fundamental to our proposed importance support points (ISP) resampling.

2.1. Quasi-Monte Carlo

Quasi-Monte Carlo (QMC) is traditionally used for numerical integration of an integrand ϕ with respect to the p-dimensional unit hypercube $[0,1]^p$, that is

$$\int_{[0,1]^p} \phi(x)dx \approx \frac{1}{N}\sum_{n=1}^{N}\phi(x_n)\,. \tag{1}$$

In standard Monte Carlo, the N evaluation points $\{x_n\}_{n=1}^N$ are sampled uniformly on $[0,1]^p$. It is well known that, by the Central Limit Theorem, its integration error converges at a rate of $\mathcal{O}(N^{-1/2})$. QMC aims to improve this rate by carefully choosing a set of well-spread out points that fill the p-dimensional hypercube in an even and uniform way. This measure of sample uniformity is typically referred to as a *discrepancy measure* in the QMC literature. One well-known discrepancy measure for sample $\{x_n\}_{n=1}^N$ on $[0,1]^p$ is the *star discrepancy* (Niederreiter 1992; Aistleitner and Dick 2015),

$$D_N^*(\{x_n\}_{n=1}^N) = \sup_{A\in\mathcal{A}^*}\left|\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}_A(x_n) - \lambda(A)\right|, \tag{2}$$

where λ is the Lebesgue measure and $\mathcal{A}^* = \{A : A = \prod_{i=1}^p [0,a_i],\ 0 < a_i < 1\}$. The star discrepancy measures the maximum difference between the empirical cumulative distribution of the sample $\{x_n\}_{n=1}^N$ and the desired uniform distribution on $[0,1]^p$. A small star discrepancy suggests a more uniform sample on $[0,1]^p$, and vice versa.

The Koksma–Hlawka inequality connects integration error in (1) to the star discrepancy,

$$\left|\frac{1}{N}\sum_{n=1}^{N}\phi(x_n) - \int_{[0,1)^p}\phi(x)dx\right| \le V_{\mathrm{HK}}(\phi)D_N^*(\{x_n\}_{n=1}^N)\,, \tag{3}$$

where $V_{\mathrm{HK}}(\phi)$ is the total variation of ϕ in the sense of Hardy and Krause for measuring smoothness of ϕ (Niederreiter 1992). Equation (3) shows that a smaller integration error can be achieved if the samples have lower star discrepancy. Several methods have been proposed in QMC to generate low discrepancy samples in *uniform hypercube* that can achieve an convergence rate of $\mathcal{O}(N^{-1+\epsilon})$ for some $\epsilon > 0$ (Niederreiter 1992) under mild smoothness assumption on ϕ, showing its comparative performance over Monte Carlo rate of $\mathcal{O}(N^{-1/2})$. Recent developments have focused on *randomized* QMC methods, which provide randomized low discrepancy samples, with each sample marginally distributed as Uniform$[0,1]^p$. Randomized QMC allows for unbiased integral estimates, and provides relief from the curse-of-dimensionality for high-dimensional sampling (Dick, Kuo, and Sloan 2013). We will later make use of randomized QMC for generating proposal samples within PQMC. See Niederreiter (1992), Dick, Kuo, and Sloan (2013), and Owen (2013) for the detailed review of QMC.

The star discrepancy of $\{x_n\}_{n=1}^N$ can be generalized with respect to any *nonuniform* normalized Borel measure μ in $[0,1]^p$ (Aistleitner and Dick 2015),

$$D_N^*(\{x_n\}_{n=1}^N;\mu) = \sup_{A\in\mathcal{A}^*}\left|\frac{1}{N}\sum_{n=1}^{N}\mathbb{1}_A(x_n) - \mu(A)\right|. \tag{4}$$

The Koksma–Hlawka inequality also holds (Aistleitner and Dick 2015, Theorem 1),

$$\left|\frac{1}{N}\sum_{n=1}^{N}\phi(x_n) - \int_{[0,1)^p}\phi(x)d\mu(x)\right| \leq V_{\mathrm{HK}}(\phi)D_N^*(\{x_n\}_{n=1}^N;\mu)\,, \tag{5}$$

However, finding low discrepancy samples from a general probability measure μ is difficult. The inverse transform method is a potential solution, as "Satz 2" of Hlawka and Mück (1972) shows that $D_N^*(\{F_\mu^{-1}(x_n)\}_{n=1}^N;\mu) \leq cD_N^*(\{x_n\}_{n=1}^N)$ under some mild conditions on F_μ. However, this requires a closed form for the inverse distribution function F_μ, which is typically not available in Bayesian problems. Thus, we next introduce the support points (Mak and Joseph 2018) and our proposed extension *importance support points* that can both generate low discrepancy samples from a general distribution.

2.2. Importance Support Points

We first review the definition of support points in Mak and Joseph (2018):

Definition 1 (Support Points; Mak and Joseph, 2018). Let $Y \sim F$, a distribution function on $\emptyset \neq \mathcal{X} \subseteq \mathbb{R}^p$ with finite means. The support points $\{\xi_i\}_{i=1}^n$ of F are

$$\begin{aligned}\{\xi_i\}_{i=1}^n &\in \arg\min_{x_1,\ldots,x_n\in\mathcal{X}} \mathcal{E}(F_n,F)\\ &= \arg\min_{x_1,\ldots,x_n\in\mathcal{X}} \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\|x_i - Y\|_2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|_2\,,\end{aligned} \tag{6}$$

where $\mathcal{E}(F_n,F)$ is the energy distance (Székely and Rizzo 2004, 2013) between F and F_n. F_n is the empirical distribution function (edf) of any n-point $\{x_i\}_{i=1}^n \subseteq \mathcal{X}$.

For most distribution functions F, the expectation in (6) cannot be analytically computed, and thus, in implementation, we would optimize the Monte Carlo approximation

$$\{\xi_i\}_{i=1}^n = \arg\min_{x_1,\ldots,x_n\in\mathcal{X}} \frac{2}{nM}\sum_{i=1}^{n}\sum_{m=1}^{M}\|x_i - y_m\|_2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|_2\,, \tag{7}$$

where $\{y_m\}_{m=1}^M$ (with $M > n$) are sampled from F. The first term in (7) forces the support points $\{\xi_i\}_{i=1}^n$ to mimic the samples $\{y_m\}_{m=1}^M$ from F, while the second term forces these points to be as far apart from each other as possible. The latter is often referred to as a "space-filling property" in experimental design (Johnson, Moore, and Ylvisaker 1990). The problem in (7) can be efficiently solved via the convex-concave procedure (Yuille and Rangarajan 2002), and is implemented in the R package `support` (Mak 2019b). However, (7) shows that the quality of the support points depends critically on the samples $\{y_m\}_{m=1}^M$ from F. From previous discussion on QMC, if samples $\{y_m\}_{m=1}^M$ have low discrepancy measure, then the support points $\{\xi_i\}_{i=1}^n$ computed from (7) can be better representative points for F, but generating low discrepancy samples directly from any nonuniform distribution is difficult.

To overcome this issue, we introduce the importance support points (ISP) that uses Importance Sampling (Robert and Casella 2013) to circumvent sampling directly from F.

Definition 2 (Importance Support Points). Let π be the pdf of the distribution function F. Let $Y \sim q$, where q is the pdf of an importance distribution defined on the same support $\mathcal{X}$. The importance support points $\{\xi_i\}_{i=1}^n$ of F with respect to the importance distribution q are

$$\{\xi_i\}_{i=1}^n \in \arg\min_{x_1,\ldots,x_n\in\mathcal{X}} \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}_q[w(Y)\|x_i - Y\|_2] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|_2\,, \tag{8}$$

where $w(\cdot) = \pi(\cdot)/q(\cdot)$ is the importance weight (or Radon–Nikodym derivative).

The expectation in (8) is with respect to the importance distribution q, so we only require samples $\{y_m\}_{m=1}^M$ simulated from q for Monte Carlo approximation,

$$\{\xi_i\}_{i=1}^n \in \arg\min_{x_1,\ldots,x_n\in\mathcal{X}} \frac{2}{n}\sum_{i=1}^{n}\sum_{m=1}^{M}\bar{w}_m\|x_i - y_m\|_2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|_2\,, \tag{9}$$

where $\bar{w}_m = (\pi(y_m)/q(y_m))/(\sum_{l=1}^M \pi(y_l)/q(y_l))$ are the normalized importance weights. We can choose q to be a simple distribution (e.g., uniform) where low discrepancy samples can be easily simulated. Corollary 1 of Aistleitner and Dick (2015) justifies the use of QMC samples in Importance Sampling. The formulation in (9) applies to the pdf that is known up to an unknown constant of proportionality. It can be further generalized to optimally compact any large set of weighted samples $\{(y_m,\bar{w}_m)\}_{m=1}^M$ to a few unweighted representative points $\{\xi_i\}_{i=1}^n$ in terms of their *weighted* energy distance. This weighted energy criterion is also recently explored in Huling and Mak (2020), but their objective is to optimize for the *weights* given fixed samples. Last, the optimization algorithm for solving (9) is analogous to solving support points (7) in Mak and Joseph (2018), and we refer readers to Appendix A, supplementary materials for the detailed algorithm.

Figure 1 shows $n = 100$ support points for the two-dimensional axe-shaped, banana-shaped, and mixture normal distributions. Top panels shows the support points from 10,000 MCMC samples obtained using the R package `adaptMCMC` (Scheidegger 2018). The Markov chain is run for 15,000 iterations, and the first 5000 samples is discarded as burn-in. Bottom panels shows the proposed ISPs generated from (9) with importance distribution $q = \mathrm{Uniform}[0,1]^2$, using 10,000 Sobol' points (Joe and Kuo 2003) (generated from the R package `randtoolbox` (Christophe and Petr 2019)) as low discrepancy importance samples. When the MCMC explores the distribution well as in the axe-shaped distribution,

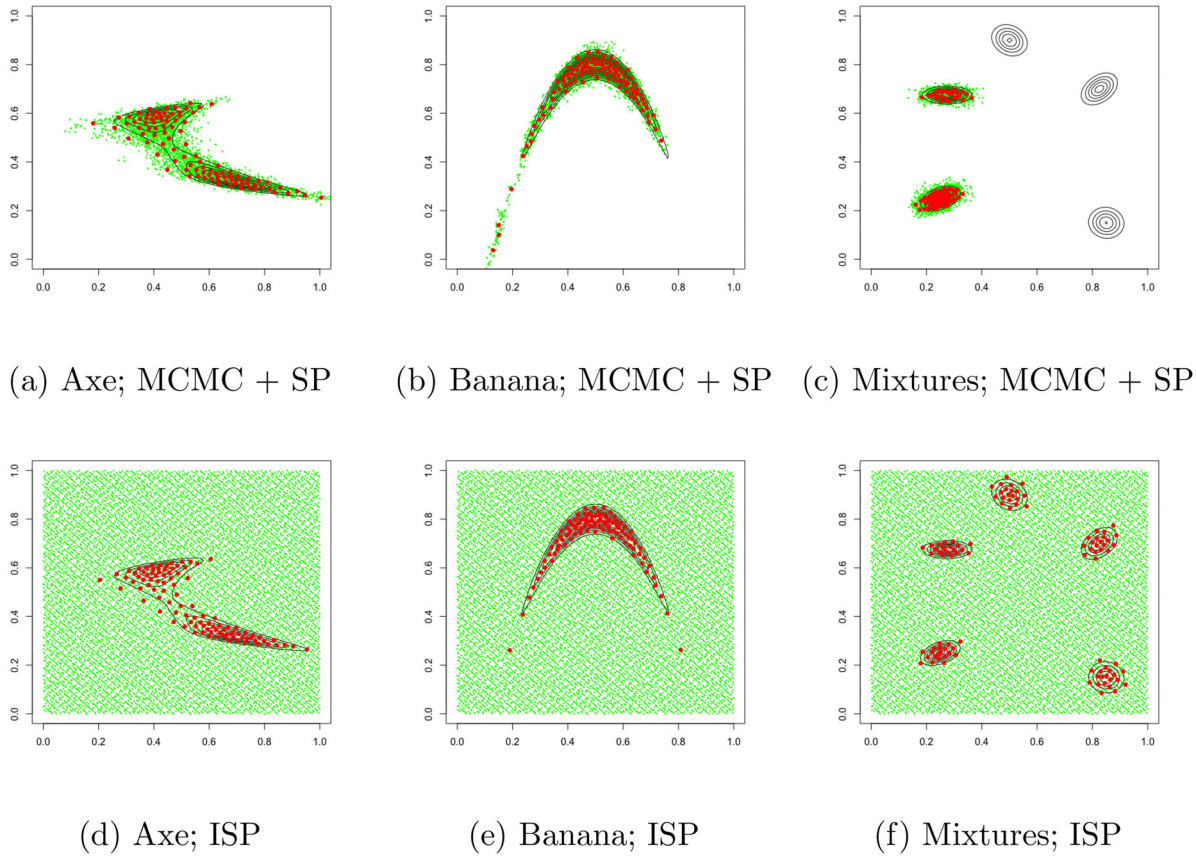


Figure 1. $n = 100$ support points (red dots) generated from 10,000 MCMC samples (green diamonds) and 10,000 importance samples (green diamonds) for axe-shaped, banana-shaped, and mixture normal distributions. Lines represent density contours.

the support points from MCMC samples are as good as the ISPs. However, for the banana-shaped distribution, the support points from MCMC samples cannot reflect the underlying symmetry structure. The problem is more severe for the mixture normal, as poor mixing on multimodal distribution is a known issue of standard MCMC. ISP shows substantial improvement over support points generated by the MCMC samples, by making use of the density information in F for the discrepancy correction.

The choice of the importance distribution q is critical for the success of ISPs. A robust choice would be uniform distribution over a region that covers the support of π as shown in Figure 1, but sufficient samples would also need to be placed in high-probability regions. The effective sample size, $N_e = [\sum_{m=1}^{M} \bar{w}_m^2]^{-1}$, is one popular diagnostic metric that can be used for evaluating the quality of importance samples.

3. Importance Support Points Resampling

With the proposed ISPs in hand, we now present a new importance support points (ISP) resampling procedure which we later integrate within PQMC. We then demonstrate the improvement of ISP resampling over existing resampling schemes via simulation.

3.1. ISP Resampling Algorithm

Let $\{(y_m, \bar{w}_m)\}_{m=1}^{M}$ be any normalized weighted samples, weights satisfying $\sum_{m=1}^{M} \bar{w}_m = 1$, and denote its edf as $F_M^w(y) = \sum_{m=1}^{M} \bar{w}_m \mathbb{1}(y_{m,1} \le y_1, \ldots, y_{m,p} \le y_p)$. A good resampling scheme should return a set of unweighted samples $\{\xi_i\}_{i=1}^{n}$, with $\xi_i \in \{y_m\}_{m=1}^{M}$, such that the discrepancy between the resamples edf $\tilde{F}_n(y) = n^{-1} \sum_{i=1}^{n} \mathbb{1}(\xi_{i,1} \le y_1, \ldots, \xi_{i,p} \le y_p)$ and F_M^w is small. A popular criterion for quantifying the difference between any two multivariate edfs is the aforementioned energy distance (Székely and Rizzo 2004, 2013). Here, we propose the ISP resampling, a deterministic resampling method that aims to find the n-point resamples $\{\xi_i^*\}_{i=1}^{n}$ which minimize the energy distance to the original weighted samples $\{(y_m, \bar{w}_m)\}_{m=1}^{M}$,

$$\begin{aligned} \{\xi_i^*\}_{i=1}^{n} &\in \arg \min_{x_1,\ldots,x_n \in \{y_m\}_{m=1}^{M}} \mathcal{E}(\tilde{F}_n, F_M^w) \\ &= \arg \min_{x_1,\ldots,x_n \in \{y_m\}_{m=1}^{M}} \frac{2}{n} \sum_{i=1}^{n} \sum_{m=1}^{M} \bar{w}_m \|x_i - y_m\|_2 \\ &\quad - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|x_i - x_j\|_2 , \end{aligned} \tag{10}$$

where $\tilde{F}_n$ is the edf of any n-point resamples. Equation (10) is the same optimization problem of the ISPs in (9) but with constraints that $\xi_i^* \in \{y_m\}_{m=1}^{M}$, making it a discrete optimization problem that is much harder to solve.

We present Algorithm 1, a quadratic time complexity sequential procedure that approximately solves for the ISP resamples $\{\xi_i^*\}_{i=1}^{n}$ in (10). The procedure starts with the one-point-at-a-time *greedy initialization*, sequentially adding a new point ξ_i^* from $\{y_m\}_{m=1}^{M}$ that minimizes the energy distance

Algorithm 1: Importance Support Points Resampling (10).

Input: the set of normalized weighted samples $\{(y_m, \bar{w}_m)\}_{m=1}^M$.

Greedy Initialization: conditional on having $\{\xi_j^*\}_{j=1}^{i-1}$, obtain ξ_i^* by

$$\xi_i^* = \arg\min_{x\in\{y_m\}_{m=1}^M} \frac{2}{i}\sum_{m=1}^M \bar{w}_m\|x-y_m\|_2 - \frac{2}{i^2}\sum_{j=1}^{i-1}\|x-\xi_j^*\|_2 \,. \tag{11}$$

Solve (11) for $i = 1, \ldots, n$ sequentially to obtain the initial resamples $\{\xi_i^*\}_{i=1}^n$.

Point Refinement: for $i = 1, \ldots, n$, fixing $\{\xi_j^*\}_{j\neq i}$, refine ξ_i^* by

$$\xi_i^* = \arg\min_{x\in\{y_m\}_{m=1}^M} \frac{2}{n}\sum_{m=1}^M \bar{w}_m\|x-y_m\|_2 - \frac{2}{n^2}\sum_{\substack{j=1\\ j\neq i}}^{n}\|x-\xi_j^*\|_2 \,. \tag{12}$$

Repeat above until the energy distance defined in (10) converges.

Return: $\{\xi_i^*\}_{i=1}^n$, the set of ISP resamples.

between $\{\xi_j^*\}_{j=1}^{i-1} \cup \{\xi_i^*\}$ and $\{(y_m, \bar{w}_m)\}_{m=1}^M$. This one-point-at-a-time greedy procedure is popularly used in experimental design construction from a set of candidate points (Kennard and Stone 1969; Joseph et al. 2019), but it yields a local optimum solution. Thus, we adapt the resamples by *point refinement*: update ξ_i^* via solving (12) to improve the energy distance while fixing the other resamples $\{\xi_j^*\}_{j\neq i}$. In our simulation, less than 10 repetitions of the point refinement step is enough for convergence. The proposed procedure splits the n variables optimization problem in (10) to n smaller problems each with only one variable, making it feasible to solve in polynomial time. Both computational and memory complexity of Algorithm 1 is $\mathcal{O}(M^2)$ from computing and storing the pairwise Euclidean distance of $\{y_m\}_{m=1}^M$.

3.2. Theoretical Properties

We now discuss the deterministic error bounds by the ISP resamples $\{\xi_i^*\}_{i=1}^n$ with edf $\tilde{F}_n^*$. We start with the Koksma–Hlawka-like bound that connects the squared integration error with the energy distance $\mathcal{E}(\tilde{F}_n, F_M^w)$.

Lemma 1. Let $\mathcal{H}$ be the native space induced by the radial kernel $\Phi(\cdot) = -\|\cdot\|_2$ on $\mathcal{X}$. For any integrand $\phi \in \mathcal{H}$, we have the following bound for the squared integration error,

$$\left(\int_{\mathcal{X}} \phi(x)d\tilde{F}_n(x) - \int_{\mathcal{X}} \phi(x)dF_M^w(x)\right)^2 \leq C_\phi \mathcal{E}(\tilde{F}_n, F_M^w) \,, \tag{13}$$

where $C_\phi \geq 0$ is a constant depending on only the integrand ϕ.

The proof of Lemma 1 follows directly from Theorem 4 of Mak and Joseph (2018). The squared integration error introduced by the resamples, another important measure of goodness for resampling schemes (Hol, Schon, and Gustafsson 2006; Li et al. 2021), is upper bounded by a term proportional to the energy distance, again justifying the use of energy distance for choosing a good resampling scheme. Moreover, we have $\mathcal{E}(\tilde{F}_n^*, F_M^w) \leq \mathcal{E}(\tilde{F}_n, F_M^w)$ by definition of the ISP resampling in (10), showing the theoretical advantage of ISP resamples in controlling the squared integration error in (13). Next, we investigate the asymptotic convergence rate of the squared integration error using the ISP resamples.

Theorem 1. Assume the same conditions in Lemma 1 and Theorem 5 of Mak and Joseph (2018), we have, for some constant $\alpha > 1$,

$$\left(\int_{\mathcal{X}} \phi(x)d\tilde{F}_n(x) - \int_{\mathcal{X}} \phi(x)dF_M^w(x)\right)^2 = \mathcal{O}\left(n^{-1}(\log n)^{-(\alpha-1)}\right) . \tag{14}$$

The proof of Theorem 1 follows directly from Theorem 5 of Mak and Joseph (2018). Theorem 1 demonstrate the faster convergence of using the deterministic ISP resamples over the Monte Carlo rate of $\mathcal{O}(n^{-1})$. Moreover, from our simulation studies presented in Section 3.3, we observe a faster empirical convergence rate by the ISP resamples, which is aligned with the conjecture by Mak and Joseph (2018) that the rate in (14) is not tight.

3.3. Comparison to Existing Resampling Methods

Since the introduction of resampling by Rubin (1987) for iterated IS procedure, many randomized resampling schemes have been proposed in the literature. The simplest approach is multinomial resampling: draw each resample ξ_i independently from the weighted empirical distribution F_M^w. In contrast to the ISP resamples ξ_i^* (which are deterministic), in randomized resampling schemes, each ξ_i is a random variable. For any integrand ϕ, the resamples $\{\xi_i\}_{i=1}^n$ then yield an unbiased estimator,

$$\begin{aligned}\mathbb{E}\left[\int_{\mathcal{X}} \phi(x)d\tilde{F}_n(x)\middle| F_M^w\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \phi(\xi_i)\middle| F_M^w\right] \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{m=1}^M \bar{w}_m\phi(y_m) \\ &= \int_{\mathcal{X}} \phi(x)dF_M^w(x) \,. \end{aligned} \tag{15}$$

The existing randomized resampling schemes, such as multinomial, stratified, residual, systematic, and etc., are all unbiased (Douc and Cappe 2005; Li et al. 2021). It follows that the expected square integration error between $\tilde{F}_n$ and F_M^w is

$$\mathbb{E}\left[\left(\int_{\mathcal{X}} \phi(x)d\tilde{F}_n(x) - \int_{\mathcal{X}} \phi(x)dF_M^w(x)\right)^2\middle| F_M^w\right] = \mathrm{var}\left[\frac{1}{n}\sum_{i=1}^n \phi(\xi_i)\middle| F_M^w\right] , \tag{16}$$

reducing to a conditional variance that we want to minimize. We focus on the comparison of the ISP resampling to the stratified

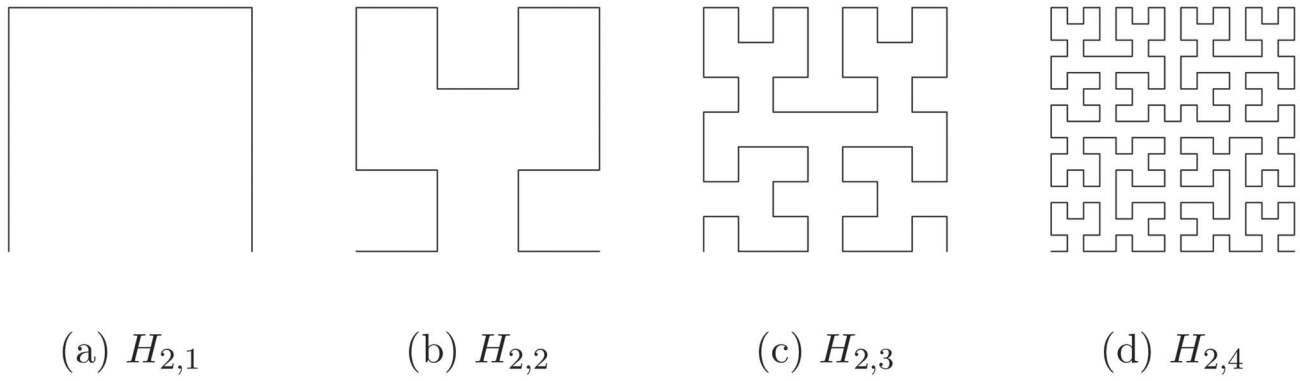


Figure 2. The Hilbert Curve in $p = 2$ dimension is the limit of the curve $H_{p,o}$ as $o \to \infty$.

resampling, which has been shown to achieve the smallest conditional variance in (16) with integrand $\phi(x) = x$ on the one dimensional sorted samples, that is, $y_1 \leq \cdots \leq y_M \in \mathbb{R}$ (Li et al. 2021). Stratified resampling draws $\{\xi_i\}_{i=1}^n$ by

$$\xi_i = F_M^{w\,-1}\left(\frac{i-1+U_i}{n}\right),\ U_i \overset{\text{iid}}{\sim} \text{Uniform}[0,1],\ \forall i = 1,\ldots,n, \tag{17}$$

where $F_M^{w\,-1}$ is the inverse distribution function of F_M^w. By partitioning the space and drawing one resamples from each stratum in (17), stratified resampling can return a set of randomized low discrepancy (space-filling) resamples. Moreover, Theorem 1 of Li et al. (2021) proves that the ordered stratified resampling in one-dimensional case also minimizes the expected squared energy distance and the earth mover distance (Wasserstein metric) between $\tilde{F}_n$ and F_M^w. It can achieve an integration rate of $\mathcal{O}(n^{-2})$ for the conditional variance with respect to any Lipschitz integrand ϕ (Li et al. 2021).

The key to achieve low discrepancy resamples via stratified resampling depends on the ordered particles, but the ordering in multivariate setting is not obvious until the introduction of the space-filling Hilbert curve in the context of resampling (Gerber and Chopin 2015). A p-dimensional Hilbert curve is a continuous function $H : [0,1] \to [0,1]^p$ that is surjective, Hölder continuous with exponent $1/p$, and measure-preserving (He and Owen 2016; Gerber, Chopin, and Whiteley 2019; Li et al. 2021). It is defined as the limit of a sequence of functions, and see Figure 2 for the illustration in $p = 2$ dimensions. Proposition 2 of Gerber, Chopin, and Whiteley (2019) shows that there exists a one-to-one Borel measurable function $h : [0,1]^p \to [0,1]$ such that $H(h(x)) = x$ for all $x \in [0,1]^p$. Intuitively, $h(x)$ is computing the one-dimensional projection of x onto the Hilbert curve. Thus, the multivariate weighted samples $\{(y_m, \bar{w}_m)\}_{m=1}^M$ can be sorted by their projection onto the Hilbert curve $\{h(y_m)\}_{m=1}^M$. By the Hölder continuity, the Hilbert curve sorted resampling essentially first partitions $\{(y_m, \bar{w}_m)\}_{m=1}^M$ into n strata (spatially nonoverlapping) in the hypercube, and then select one resample from each stratum to yield a set of n-point low discrepancy (spatially spread out) resamples in multi-dimension. Theorem 2 of Li et al. (2021) shows that the conditional variance in (16) for Lipschitz integrand with n resamples can be bounded by $\mathcal{O}(n^{-(1+2/p)})$ with Hilbert curve sorted stratified resampling, the best bound that can be obtained for any randomized resampling scheme with/without any ordering, but its improvement over the Monte Carlo rate of $\mathcal{O}(n^{-1})$ diminishes as dimension p increases.

In the implementation of Hilbert curve sorting, we need to first transform the samples $\{(y_m, \bar{w}_m)\}_{m=1}^M$ into unit hypercube. Following Li et al. (2021), we first find the hypercube that covers $\{(y_m, \bar{w}_m)\}_{m=1}^M$ by using the range of each dimension, and then rescale the hypercube to $[0,1]^p$. As we can see from the banana-shaped distribution in Figure 1, such cubified transformation would not work well since most space in the hypercube are not important with density arbitrarily closed to 0. On the other hand, a discretized approximation of Hilbert curve (Figure 2) is used in practice, but a too fine approximation could (i) result in high computational cost and (ii) well spread out points in $[0,1]$ might be close in $[0,1]^p$ since the curve is very nonsmooth (He and Owen 2016). Thus, we follow Li et al. (2021) to use Hilbert curve with $o = 8$ defined in Figure 2. We implement our R version of Hilbert curve according to the C code provided in Skilling (2004) with the R package `gmp` to handle the large bit integer (e.g., $2^{po} - 1$ for $H_{p,o}$).

Consider first the visualization of ISP resampling compared to the baseline multinomial resampling and the state-of-the-art Hilbert curve sorted stratified resampling on a two-dimensional example. Figure 3 shows the 100-point resamples for the earlier mixture of five normals distribution using three different resampling schemes. We can see that the ISP resamples are not only space-filling, but also better captures the shape of the target distribution compared to existing methods.

Consider next a quantitative comparison of different resampling schemes, including residual and Hilbert curve sorted systematic resampling. Let $\pi = \mathcal{N}(0, I_p)$ be the target distribution and $q = \mathcal{N}(0, \sqrt{2}I_p)$ be the importance distribution. Further let $\{y_m\}_{m=1}^M$ be the 1000 inverse Sobol' points from q (i.e., the inverse transformation of q on Sobol' points), with corresponding weights $\bar{w}_m$ set as $\pi(y_m)/q(y_m)$ with normalization. We then obtain 100 resamples from $\{(y_m, \bar{w}_m)\}_{m=1}^M$. The middle panel of Figure 4 shows the mean squared integration error in log over 100 independent sets of resamples, which is an Monte Carlo estimator of the expected squared integration error in (16) with $\phi(x) = x$. Empirically, ISP resampling significantly outperforms the other four resampling schemes up to 20 dimensions. However, ISP resampling appears to suffer from small effective sample sizes as dimension increases. Note that ISP resampling is deterministic, so every run yields the

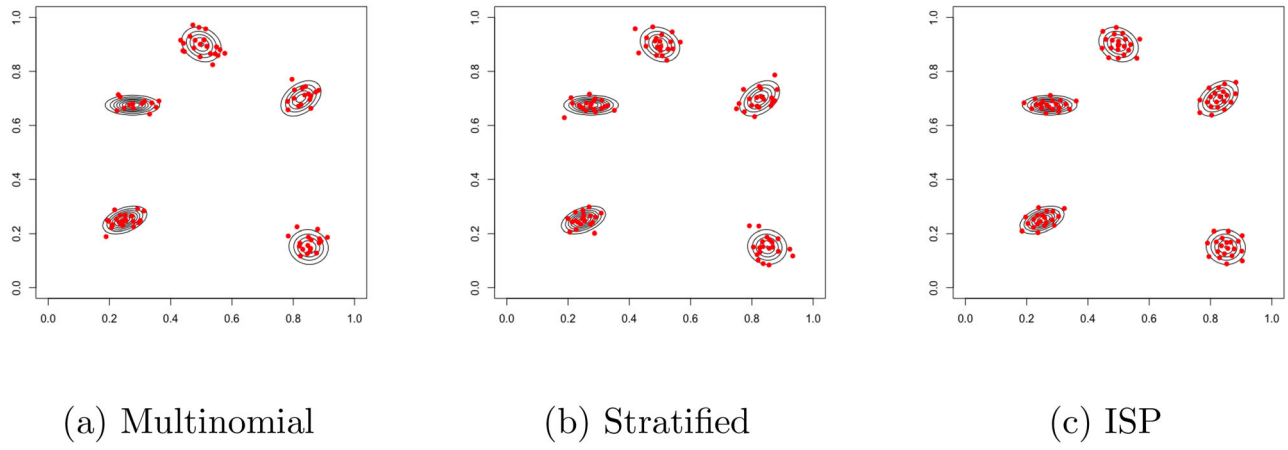


Figure 3. $n = 100$ multinomial, Hilbert curve sorted stratified, and ISP resamples from 10,000 Sobol' points for the earlier mixture normal example.

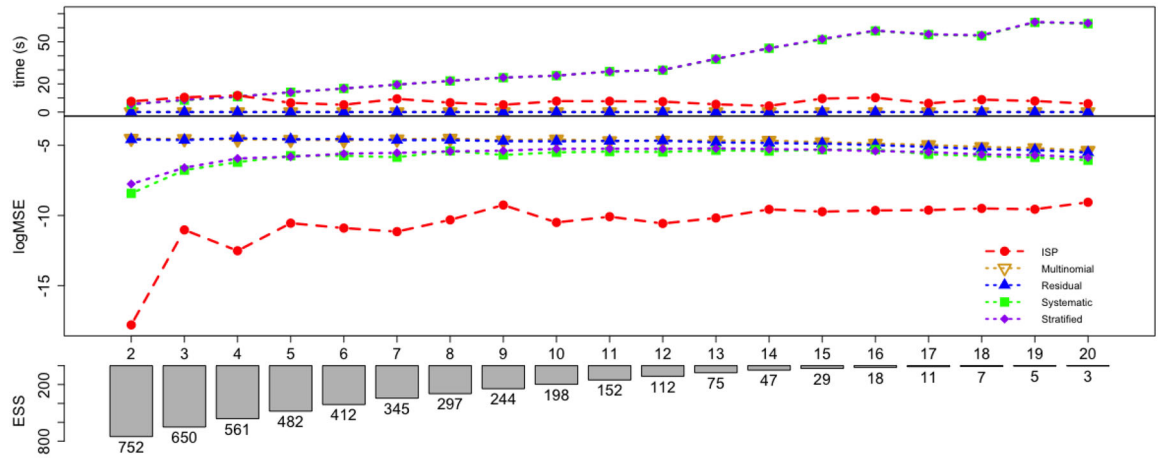


Figure 4. Comparison in computational time and squared integration error in log (log-MSE) of the five resampling schemes on problems ranging from dimension 2 to 20 averaged over 100 independent runs. ESS stands for the effective sample size.

same set of resamples, but we still run it 100 times for the time comparison shown in the top panel of Figure 4. Regarding computation time, ISP resampling does incur a longer runtime than both multinomial and residual resampling, but on average, the runtime is still within around 10 sec and does not grow much as dimensionality increase. Though both Hilbert curve sorted stratified and systematic resampling have a complexity of $\mathcal{O}(M \log M)$ (Gerber and Chopin 2015), we see an increasing computational cost as the dimension gets larger in practice, where the major cost resulted from computing the projection onto the Hilbert curve in high-dimension. For moderate sample size M, these results suggest that ISP resampling may enjoy improved integration performance and reduced computation time over Hilbert curve sorted resampling approach, particularly as dimensionality increases.

4. Population Quasi-Monte Carlo

In this section, we discuss the integration of ISP resampling within the proposed Population Quasi-Monte Carlo (PQMC) framework. For reference, Algorithm 2 outlines the standard PMC procedure, using normal proposal distributions, static global covariance (i.e., all proposals share the same covariance matrix Σ), and the deterministic mixture weighting strategy (19) from Elvira et al. (2017). For this procedure, Elvira et al. (2019) shows that its mixture weights can achieve a smaller variance for the PMC estimator than the standard weights $w_{k,j} = \gamma(x_{k,j})/q_k(x_{k,j})$, but at the cost of higher complexity of $\mathcal{O}(K^2 J)$. We introduce below the proposed adjustment to *sampling* and *adaptation* steps using QMC ideas, with the complete PQMC framework detailed in Algorithm 3.

4.1. Quasi-Monte Carlo Proposals

Consider first the improvement of the *sampling* step in PMC using randomized low discrepancy proposal samples for $\{x_{k,j}^{(t)}\}_{k=1}^{K}{}_{j=1}^{J}$. From the earlier discussion on QMC, this yields an unbiased estimator with a faster convergence rate than standard Monte Carlo proposals. Such an improved convergence rate also holds for the importance samples $\{(x_{k,j}^{(t)}, \bar{w}_{k,j}^{(t)})\}_{k=1}^{K}{}_{j=1}^{J}$ after *weighting* step (see Theorem 2 of Gerber and Chopin (2015)). For the multivariate normal proposal distribution considered here, we can generate randomized low discrepancy samples via inverse transform approach. First, randomized low discrepancy samples $\{u_j^k\}_{j=1}^{J}$ are generated on the uniform hypercube $[0,1]^p$. We use in our implementation Owen-style scrambled Sobol' points (Owen 1998) from the R package `randtoolbox`

Algorithm 2: PMC with Normal Proposals and Static Global Covariance.

Target Distribution: $\pi = \gamma/Z$ where Z is an unknown normalizing constant.
Initialization: set parameters for the initial proposals $\{q_k^{(1)} = \mathcal{N}(\cdot|\mu_k^{(1)}, \Sigma)\}_{k=1}^K$.
for $t = 1, \ldots, T$ **do**

• **Sampling:** draw J samples from each proposal $k = 1, \cdots, K$

$$x_{k,j}^{(t)} \sim q_k^{(t)}, \quad j = 1, \cdots, J, \quad k = 1, \cdots, K, \tag{18}$$

so that a total of KJ samples are simulated.
• **Weighting:** compute deterministic mixture weights

$$w_{k,j}^{(t)} = \frac{\gamma(x_{k,j}^{(t)})}{K^{-1}\sum_{i=1}^K q_i^{(t)}(x_{k,j}^{(t)}|\mu_i^{(t)}\Sigma)}, \quad j = 1, \cdots, J, \quad k = 1, \cdots, K, \tag{19}$$

and normalize them by

$$\bar{w}_{k,j}^{(t)} = \frac{w_{k,j}^{(t)}}{\sum_{i=1}^K \sum_{l=1}^J w_{i,l}^{(t)}}. \tag{20}$$

• **Adaptation:** perform resampling to update the proposal centers

$$\mu_k^{(t+1)} \overset{\text{iid}}{\sim} \sum_{k=1}^K \sum_{j=1}^J \bar{w}_{k,j}^{(t)} \delta_{x_{k,j}^{(t)}}, \quad k = 1, \cdots, K. \tag{21}$$

end
Return: $\{(x_{k,j}^{(t)}, w_{k,j}^{(t)})\}_{t=1}^T{}_{k=1}^K{}_{j=1}^J$ where $w_{k,j}^{(t)}$ are the unnormalized weights.

(Christophe and Petr 2019). Next, the samples $\{x_{k,j}\}_{j=1}^J$ for the k-th proposal are obtained by

$$x_{k,j} = \mu_k + \Sigma^{1/2}[\Phi^{-1}(u_{j1}^k), \ldots, \Phi^{-1}(u_{jp}^k)]^T. \tag{25}$$

The transformed samples can be shown to be randomized low discrepancy samples with respect to $\mathcal{N}(\mu_k, \Sigma)$ ("Satz 2"; Hlawka and Mück 1972). Here, the superscript k is to emphasize that different sets of $\{u_j^k\}_{j=1}^J$ are simulated for different proposals. A similar QMC proposal sampling idea via inverse transformation is also employed in Sequential Quasi-Monte Carlo (Gerber and Chopin 2015) and later extended by Li et al. (2021) to allow for multiple samples ($J > 1$) from each proposal. We adopt the same $J > 1$ setting here, since this provides an empirical improvement for standard PMC (Elvira et al. 2017).

4.2. Proposal Adaptation

For the proposal distribution, we consider a multivariate normal distribution, which is widely used in PMC (Elvira et al. 2017). There are two associated parameters for adaptation: the proposal centers and the covariance matrices. In the literature,

Algorithm 3: PQMC with Normal Proposals and Adaptive Global Covariance.

Target Distribution: $\pi = \gamma/Z$ where Z is the normalizing constant;
Initialization: set parameters for the initial proposals $\{q_k^{(1)} = \mathcal{N}(\cdot|\mu_k^{(1)}, \Sigma^{(1)})\}_{k=1}^K$.
for $t = 1, \ldots, T$ **do**

• **Sampling:** draw J scrambled Sobol' points $\{u_j^k\}_{j=1}^J$, and compute

$$x_{k,j}^{(t)} = \mu_k^{(t)} + (\Sigma^{(t)})^{1/2}[\Phi^{-1}(u_{j1}^k), \ldots, \Phi^{-1}(u_{jp}^k)]^T, \quad j = 1, \ldots, J, \quad k = 1, \cdots, K \tag{22}$$

so that a total of KJ samples are simulated.
• **Weighting:** compute the deterministic weights in (19) and normalize the weights by (20) as in PMC to obtain $\{(x_{k,j}^{(t)}, \bar{w}_{k,j}^{(t)})\}_{k=1}^K{}_{j=1}^J$.
• **Adaptation:** update $\{\mu_k^{(t+1)}\}_{k=1}^K$ via ISP resampling by solving

$$\arg \min_{\xi_1, \ldots, \xi_K \in \{x_{k,j}^{(t)}\}_{k=1}^K{}_{j=1}^J} \frac{2}{K} \sum_{i=1}^K \sum_{k=1}^K \sum_{j=1}^J \bar{w}_{k,j}^{(t)} \|\xi_i - x_{k,j}^{(t)}\|_2 - \frac{1}{K^2} \sum_{i=1}^K \sum_{l=1}^K \|\xi_i - \xi_l\|_2 . \tag{23}$$

Update the global covariance matrix $\Sigma^{(t+1)}$ via lookback adaptation:

$$\Sigma^{(t+1)} = \sum_{k=1}^K \sum_{j=1}^J \bar{w}_{k,j}^{(t)} \frac{q_k^{(t)}(x_{k,j}^{(t)}|\mu_k^{(t)}, \Sigma^{(t)})}{\sum_{i=1}^K q_i^{(t)}(x_{k,j}^{(t)}|\mu_i^{(t)}, \Sigma^{(t)})} \times (x_{k,j}^{(t)} - \mu_k^{(t)})(x_{k,j}^{(t)} - \mu_k^{(t)})^T . \tag{24}$$

end
Return: $\{(x_{k,j}^{(t)}, w_{k,j}^{(t)})\}_{t=1}^T{}_{k=1}^K{}_{j=1}^J$ where $w_{k,j}^{(t)}$ are the unnormalized weights.

resampling is a popular method for adapting proposal centers (Cappé et al. 2004; Elvira et al. 2017). In the new PQMC approach, we replace this traditional resampling step with the proposed ISP resampling approach, which was shown to have theoretical and empirical advantages from Section 3. In particular, ISP resamples allow for reduced integration error from simulation studies, which suggests that this modified resampling procedure can further reduce the additional randomness introduced by the resampling adaptation step. The space-filling property of the ISP resamples also allows for efficient exploration of the parameter space. With this, the number of proposals K can be set smaller without much loss of information after resampling, which then reduces computational cost in the deterministic weight (19) computation. Since ISP resampling is deterministic, it introduces systematic bias in the resamples, whereas existing randomized resampling schemes are unbiased. Such bias is not problematic in PQMC, as proposal centers do not contribute directly to the PQMC estimator. The adaptation

is only to update the proposals such that we can retain information from the samples from previous steps to better approximate the target distribution.

Similar to PMC, the adaptation of proposal covariance matrix is also critical for the success of PQMC, as it plays a key role in determining the size of the proposal ellipsoid. We focus on the case with global covariance (i.e., all proposals share the same covariance matrix Σ). In the PMC literature, the choice of "optimal" kernel covariance often relies on cross-validation, which can be computational expensive (Elvira et al. 2017). An alternative solution to this, which we adopt below for PQMC, is to find the global covariance matrix $\Sigma^{(t+1)}$ that minimizes the Kullback-Leibler (KL) divergence between the target density π and the next iteration mixture proposal $K^{-1}\sum_{k=1}^{K}\mathcal{N}(x|\mu_k^{(t+1)},\Sigma^{(t+1)})$. Letting S_+^p be the set of $p\times p$ symmetric positive-definite matrices, this optimization takes the form

$$\begin{aligned}\Sigma^{(t+1)} &= \arg\min_{C\in S_+^p} \mathrm{KL}\left(\pi(x)\middle\|\frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(x|\mu_k^{(t+1)},C)\right)\\ &= \arg\min_{C\in S_+^p}\left(\int_{\mathcal{X}}\pi(x)\log\pi(x)dx\right.\\ &\quad\left.-\int_{\mathcal{X}}\pi(x)\log\left[\frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(x|\mu_k^{(t+1)},C)\right]dx\right)\\ &= \arg\max_{C\in S_+^p}\int_{\mathcal{X}}\pi(x)\log\left[\frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(x|\mu_k^{(t+1)},C)\right]dx\\ &\approx \arg\max_{C\in S_+^p}\sum_{k=1}^{K}\sum_{j=1}^{J}\bar{w}_{k,j}^{(t)}\log\left[\frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(x_{k,j}^{(t)}|\mu_k^{(t+1)},C)\right],\end{aligned} \tag{26}$$

where the last equality uses a Monte Carlo approximation using the importance samples $\{(x_{k,j}^{(t)},\bar{w}_{k,j}^{(t)})\}_{j=1\,k=1}^{J\quad K}$ from the tth iteration. This optimization of the global covariance was first proposed by Cappé et al. (2008) for mixture PMC and by Ji and Schmidler (2013) for adaptive MCMC. The optimization of the final expression in (26) can be performed via Expectation-Maximization (EM, Dempster, Laird, and Rubin 1977; Wu 1983). We call this approach the *exact* covariance adaptation; this, however, may still be computational expensive as each EM step requires $\mathcal{O}(K^2J)$ evaluations of the proposal distribution.

To address this, we propose the following *lookback* covariance adaptation, which provides computationally efficient way to update the global covariance in PQMC without additional evaluations of the proposal distribution. The key idea lies in the fact that, after several iterations of PQMC, the samples should converge to regions with high probability. After this, the ISP resamples typically do not vary much from iteration to iteration. The lookback covariance adaptation exploits this by using centers from the previous iteration $\{\mu_k^{(t)}\}_{k=1}^{K}$, giving the optimization formulation

$$\Sigma^{(t+1)} = \arg\max_{C\in S_+^p}\sum_{k=1}^{K}\sum_{j=1}^{J}\bar{w}_{k,j}^{(t)}\log\left[\frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(x_{k,j}^{(t)}|\mu_k^{(t)},C)\right]. \tag{27}$$

A closed-form update can then be obtained by one step of EM,

$$\begin{aligned}\Sigma^{(t+1)} = \sum_{k=1}^{K}\sum_{j=1}^{J}\bar{w}_{k,j}^{(t)}&\frac{\mathcal{N}(x_{k,j}^{(t)}|\mu_k^{(t)},\Sigma^{(t)})}{\sum_{i=1}^{K}\mathcal{N}(x_{k,j}^{(t)}|\mu_i^{(t)},\Sigma^{(t)})}\\ &\times(x_{k,j}^{(t)}-\mu_k^{(t)})(x_{k,j}^{(t)}-\mu_k^{(t)})^T,\end{aligned} \tag{28}$$

with $\Sigma^{(t)}$, the covariance matrix from the previous iteration as the prior. No additional evaluations of the proposal distribution are needed for (28), since all quantities are already computed previously in the weighting steps. While this lookback covariance adaptation would also work with the existing resampling methods, its performance may not be as good since this approach assumes that proposal centers remain relatively stationary as the algorithm converges.

4.3. Weighted Estimator

With the PQMC samples $\{(x_{k,j}^{(t)},w_{k,j}^{(t)})\}_{t=1\,k=1\,j=1}^{T\quad K\quad J}$ in hand, one can then tackle the problem of estimating posterior integral $\mathbb{E}_\pi[\phi(X)]$ for a desired integrand ϕ. One may use the standard PMC estimator for $\mathbb{E}_\pi[\phi(X)]$ with known normalizing constant Z is,

$$\hat{I}^{\mathrm{PMC}} = \frac{1}{Z}\left(\frac{1}{TKJ}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J}w_{k,j}^{(t)}\phi(x_{k,j}^{(t)})\right) = \frac{1}{T}\sum_{t=1}^{T}\hat{I}_t^{\mathrm{PMC}}, \tag{29}$$

where $\hat{I}_t^{\mathrm{PMC}} = \frac{1}{Z}(\frac{1}{KJ}\sum_{k=1}^{K}\sum_{j=1}^{J}w_{k,j}^{(t)}\phi(x_{k,j}^{(t)}))$ is the estimator using only the t-th iteration weighted samples. If Z is unknown, we can replace it by a consistent estimator,

$$\hat{Z}^{\mathrm{PMC}} = \frac{1}{TKJ}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J}w_{k,j}^{(t)}. \tag{30}$$

We can see that the standard PMC estimator can be viewed as a simple average of T estimators each is constructed by the weighted samples from the corresponding iteration.

However, when there is adaptation, the standard PMC estimator is not efficient since better samples are obtained as the algorithm proceeds. The weighted PMC (WPMC) estimator assigns a set of correction weights $\{\alpha^{(t)}\}_{t=1}^{T}$ with constraint that $\sum_{t=1}^{T}\alpha^{(t)}=1$ to the T estimators, allowing to "forget" the poor samples simulated at the early stages (Douc et al. 2007; Portier and Delyon 2018). When the normalizing constant Z is known, the WPMC estimator for $\mathbb{E}_\pi[\phi(X)]$ is

$$\begin{aligned}\hat{I}^{\mathrm{WPMC}} &= \frac{1}{Z}\sum_{t=1}^{T}\alpha^{(t)}\hat{I}_t^{\mathrm{PMC}}\\ &= \frac{1}{Z}\left(\frac{1}{KJ}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J}\alpha^{(t)}w_{k,j}^{(t)}\phi(x_{k,j}^{(t)})\right).\end{aligned} \tag{31}$$

If Z is unknown, we replace it by the following consistent estimator,

$$\hat{Z}^{\mathrm{WPMC}} = \frac{1}{KJ}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{j=1}^{J}\alpha^{(t)}w_{k,j}^{(t)}. \tag{32}$$

Let $N_e^{(t)}$ denotes the effective sample size of the t-th iteration samples. We propose the following correction weights $\{\alpha^{(t)}\}_{t=1}^T$ that approximately minimize the variance of $\hat{I}^{\text{WPMC}}$,

$$\alpha^{(t)} = \frac{N_e^{(t)}}{\sum_{i=1}^T N_e^{(i)}} . \tag{33}$$

The proposed weights are proportional to the effective sample size, assigning larger weights to the estimators that are more reliable. See Appendix B.4, supplementary materials for further justification. This approach is free from the integrand ϕ and does not require knowledge of the normalizing constant. A similar idea of using effective sample size to weight the estimators is also used in the Adaptive Population Importance Sampler (Martino et al. 2015).

4.4. Asymptotic Properties

From the standard (29) and weighted (31) PMC estimator, we can see that only the QMC proposal samples play a role in the estimator. By using a set of randomized QMC points which are marginally distributed as the proposal, the PQMC enjoys the same consistency result of standard PMC methods. Appendix B.3, supplementary materials provides a proof of the consistency for standard/weighted PMC estimator. However, with the use of low discrepancy proposal samples, this should result in a faster convergence rate than the Monte Carlo rate of $\mathcal{O}(N^{-1/2})$. Recent results on importance sampling convergence rates with low discrepancy samples on general measures (Gerber and Chopin 2015, Theorem 2) should be useful for showing this theoretically. We will further explore theoretical properties on convergence rates for ISP resampling and lookback adaptation in a future work.

5. Simulation Results

In this section, we report simulation results comparing the performance of the proposed PQMC method to existing approaches. Further simulations results can be found in Appendix C, supplementary materials, with source codes provided at *https://github.com/BillHuang01/PQMC*.

5.1. Two-Dimensional Example

Consider a two-dimensional mixture of five normals distribution,

$$\pi(x) = \frac{1}{5}\sum_{i=1}^{5} \mathcal{N}(x|\mu_i, \Sigma_i) , \tag{34}$$

with μ_i's and Σ_i's listed in Appendix C.1, supplementary materials. The example is from Elvira et al. (2017) but with proper scaling so the main support of π is in $[0,1]^2$. Figure 3 shows the density contour of this distribution. The mean $\mathbb{E}_\pi[X] = [0.540, 0.535]^T$ and the normalizing constant $Z = 1$ can both be computed analytically, so we can validate the performance of PQMC and PMC. We use mean squared error (MSE) of the estimates as the evaluation metric.

We first compare the proposed PQMC method (Algorithm 3) to the generic PMC approach (Algorithm 2), both with normal proposals and global covariance. For PMC, we consider simple Monte Carlo proposals with multinomial/residual resampling as the baseline comparisons. For PQMC, QMC proposals are used, and we consider both Hilbert curve sorted stratified/systematic resampling and our proposed ISP resampling for proposal center adaptation. For covariance, we use isotropic matrix $\sigma^2 I_2$ for initialization, that is, $\Sigma_k^{(1)} = \sigma^2 I_2$, with $\sigma = 0.1, 0.2, 0.5$. We apply lookback covariance adaptation only for PQMC, with the adapted global covariance kept isotropic for fair comparison, that is, $\Sigma_k^{(t)} = (\sigma^{(t)})^2 I_2$. We run both PMC and PQMC for $T = 10$ iterations but vary K and J while keeping $KJ = 1000$, leading to total $TKJ = 10{,}000$ samples for the weighted PMC estimator (31).

Table 1 shows log-MSEs for the estimation of $\mathbb{E}_\pi[X]$ averaged over 100 independent trials. In general, the PQMC outperforms the baseline PMC significantly, with the proposed ISP resampling yielding the largest improvements for different settings of K, J, and σ. Regarding computation, ISP resampling is generally faster than Hilbert curve sorted resampling in implementation (see also Figure 4). Furthermore, PQMC appears to be quite robust in performance even with a small number of proposals K. Recall that the mixture weighting scheme in (19) requires $\mathcal{O}(K^2 J)$ evaluations of the proposal distribution. By being able to use a small number of proposals K, PQMC can yield significant

Table 1. Log-MSEs for the weighted estimator of $\mathbb{E}_\pi[X]$ with π being the 2D mixture of five normals (34).

K	J	Algorithm	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.5$
25	40	PMC (Multinomial)	−8.03 [−15.27,−5.04]	−8.78 [−13.43,−7.03]	−8.07 [−13.48,−6.87]
25	40	PMC (Residual)	−8.31 [−16.02,−4.97]	−8.77 [−17.91,−6.69]	−8.03 [−15.40,−6.50]
25	40	PQMC (Systematic + Lookback)	−14.55 [−19.18,−12.61]	−14.28 [−20.14,−12.29]	−13.36 [−20.98,−11.54]
25	40	PQMC (Stratified + Lookback)	−14.31 [−18.36,−12.68]	−14.10 [−18.90,−12.06]	−13.41 [−18.27,−11.67]
25	40	PQMC (ISP + Lookback)	**−15.15** [−20.72,−13.62]	**−14.72** [−18.00,−12.63]	**−13.86** [−18.42,−11.96]
50	20	PMC (Multinomial)	−10.04 [−15.04,−7.66]	−8.89 [−13.07,−7.19]	−7.93 [−13.26,−6.31]
50	20	PMC (Residual)	−9.37 [−16.82,−5.46]	−8.83 [−12.54,−6.95]	−7.93 [−12.35,−6.24]
50	20	PQMC (Systematic + Lookback)	−14.50 [−18.63,−12.79]	−14.27 [−19.53,−12.52]	−13.21 [−17.76,−11.21]
50	20	PQMC (Stratified + Lookback)	−14.35 [−18.71,−12.45]	−13.91 [−19.92,−12.39]	−13.15 [−20.12,−11.56]
50	20	PQMC (ISP + Lookback)	**−14.83** [−19.35,−13.25]	**−14.43** [−19.40,−12.57]	**−13.25** [−17.29,−11.80]
100	10	PMC (Multinomial)	−10.31 [−16.22,−8.69]	−8.84 [−16.60,−6.99]	−8.02 [−11.63,−6.30]
100	10	PMC (Residual)	−10.03 [−13.56,−8.59]	−8.89 [−16.32,−7.36]	−7.97 [−11.40,−6.27]
100	10	PQMC (Systematic + Lookback)	−14.04 [−20.66,−12.73]	−13.64 [−17.08,−11.88]	**−13.03** [−18.95,−11.53]
100	10	PQMC (Stratified + Lookback)	−13.77 [−19.05,−12.09]	−13.68 [−19.21,−12.24]	−12.97 [−16.52,−11.50]
100	10	PQMC (ISP + Lookback)	**−14.27** [−19.40,−12.50]	**−13.79** [−18.42,−12.25]	−12.97 [−19.30,−11.05]

NOTE: The MSEs are averaged over 100 independent runs and shown in log with "mean [min,max]". The best results for each σ are highlighted in red bold-face.

reductions in computational cost from proposal evaluations, which may somewhat offset the additional computational burden from ISP resampling. Additional simulation results on the estimation of the normalizing constant and the comparison of the standard and weighted estimator are described in Appendix C.1, supplementary materials.

In this simulation, where the synthetic target distribution (34) is cheap to evaluate, PQMC is indeed more computationally expensive than PMC. However, in many complex Bayesian problems such as the friction drilling application in Section 6, the target distribution is oftentimes expensive. For such problems, the additional runtime from ISP resampling and lookback adaptation in PQMC is overshadowed by the additional number of target distribution evaluations required by PMC to achieve the same level of precision.

5.2. Higher Dimensional Example

Consider a 10-dimensional mixture of three normals distribution,

$$\pi(x) = \frac{1}{3}\sum_{i=1}^{3}\mathcal{N}(x|\mu_i, \Sigma_i)\ , \tag{35}$$

with μ_i's and Σ_i's provided in Appendix C.2, supplementary materials. This example is also from Elvira et al. (2017) but with proper scaling so that the main support of π is inside $[0,1]^{10}$. The mean $\mathbb{E}_\pi[X_j] = 0.550$ for $j = 1,\ldots,10$ and the normalizing constant $Z = 1$. We again use mean squared error (MSE) of the weighted estimator for performance evaluation. The same experimental setup in Section 5.1 is used, except we now run PMC and PQMC with $KJ = 2000$, yielding a total of $TKJ = 20{,}000$ evaluations of the target distribution. Table 2 shows the log-MSEs for the estimation of $\mathbb{E}_\pi[X]$ averaging over 100 independent runs. As before, PQMC with the proposed ISP resampling and lookback adaptation provides significant improvements over other methods, for various settings of K, J, and σ. Appendix C.2, supplementary materials provides further comparisons.

To better study the performance of PQMC as dimension p increases, we change the dimension for the mixture normal distribution in (35) while keeping the same structure for the means and covariances. In particular, we run PMC and PQMC with $K = 50$ and $J = 40$, with an isotropic covariance matrix with $\sigma = 0.2$ at initialization (other settings are the same as before). Figure 5 shows log-MSEs for estimating the normalizing constant $Z = 1$ as dimensionality increases. We can see that PQMC outperforms PMC for all dimensions up to $p = 20$, but the improvement diminishes for larger dimensions. This diminishing improvement in higher dimensions is not too surprising for PQMC, since it is known that QMC methods may suffer from a "curse-of-dimensionality" for high-dimensional sampling problems (Dick, Kuo, and Sloan 2013).

6. Expensive Posterior Application: Friction Drilling

Finally, we present an application of PQMC on a practical engineering problem where the posterior distribution is expensive to evaluate. Miller and Shih (2007) develop a thermomechanical finite element model (FEM) to simulate a friction drilling process for analyzing the relationship between the thrust force (y) and the tool travel distance (x). There is an unknown parameter, the friction coefficient (η), in the FEM that one has to specify to obtain the FEM output. The left panel of Figure 6 shows the FEM outputs of thrust force over the tool travel distance for three different values of the friction coefficient, and we can see that

Table 2. Log-MSEs for the weighted estimator of $\mathbb{E}_\pi[X]$ with π being the 10D mixture of three normals (35).

K	J	Algorithm	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.5$
50	40	PMC (Multinomial)	−5.80 [−8.86,−2.88]	−10.86 [−12.62,−9.24]	−6.95 [−8.41,−5.84]
50	40	PMC (Residual)	−6.05 [−8.24,−4.48]	−11.09 [−12.87,−9.73]	−6.91 [−8.65,−5.44]
50	40	PQMC (Systematic + Lookback)	−11.65 [−14.10,−10.41]	−11.65 [−13.30,−10.37]	−11.45 [−13.00,−9.93]
50	40	PQMC (Stratified + Lookback)	−9.65 [−13.78,−5.23]	−11.74 [−13.40,−10.36]	−11.41 [−13.08,−9.64]
50	40	PQMC (ISP + Lookback)	**−11.89** [−13.53,−10.34]	**−12.07** [−14.16,−10.67]	**−11.97** [−13.26,−10.87]
100	20	PMC (Multinomial)	−6.31 [−9.27,−3.65]	−11.34 [−13.02,−10.32]	−6.90 [−8.26,−5.52]
100	20	PMC (Residual)	−6.70 [−9.15,−5.09]	−11.23 [−12.83,−10.23]	−6.97 [−8.50,−5.71]
100	20	PQMC (Systematic + Lookback)	−11.84 [−13.46,−10.33]	−12.03 [−13.80,−10.60]	−11.77 [−13.58,−10.55]
100	20	PQMC (Stratified + Lookback)	−11.81 [−13.39,−10.51]	−11.99 [−14.29,−10.66]	−11.82 [−13.57,−10.75]
100	20	PQMC (ISP + Lookback)	**−12.15** [−13.57,−11.03]	**−12.19** [−13.77,−10.96]	**−11.93** [−13.21,−10.79]

NOTE: The MSEs are averaged over 100 independent trials and shown in log with "mean [min,max]". The best results for each σ are highlighted in red bold-face.

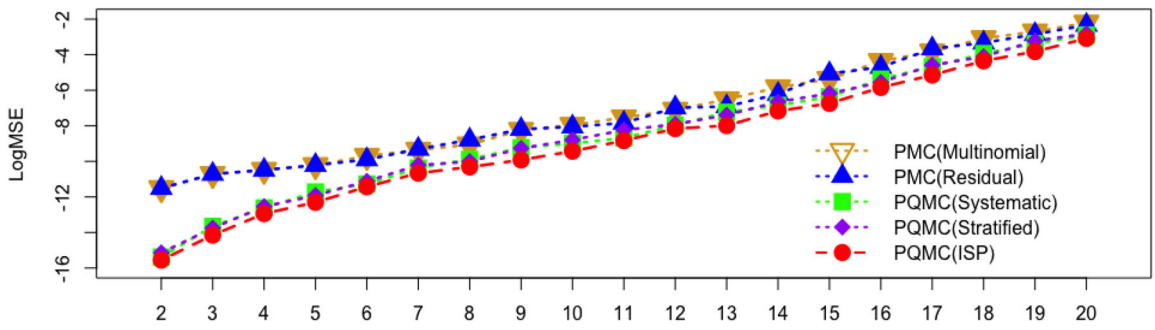
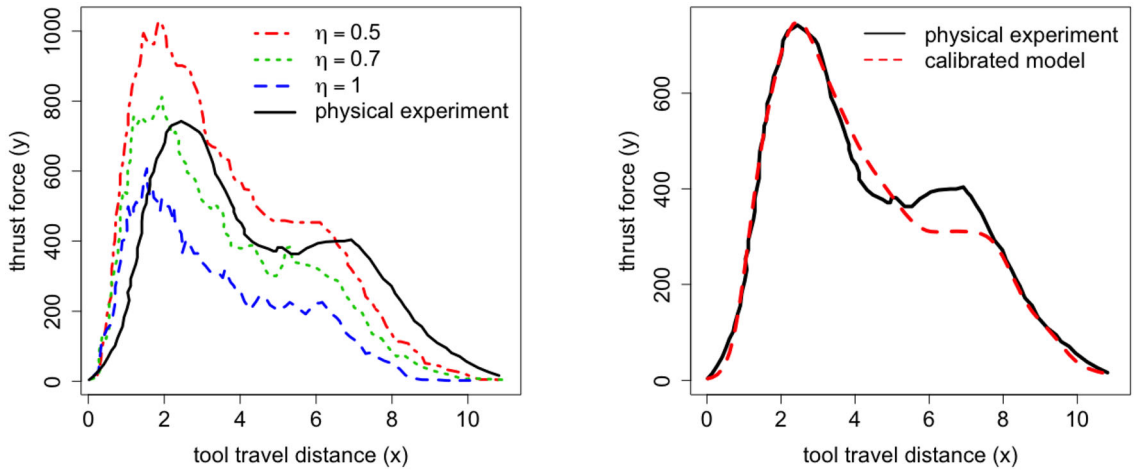
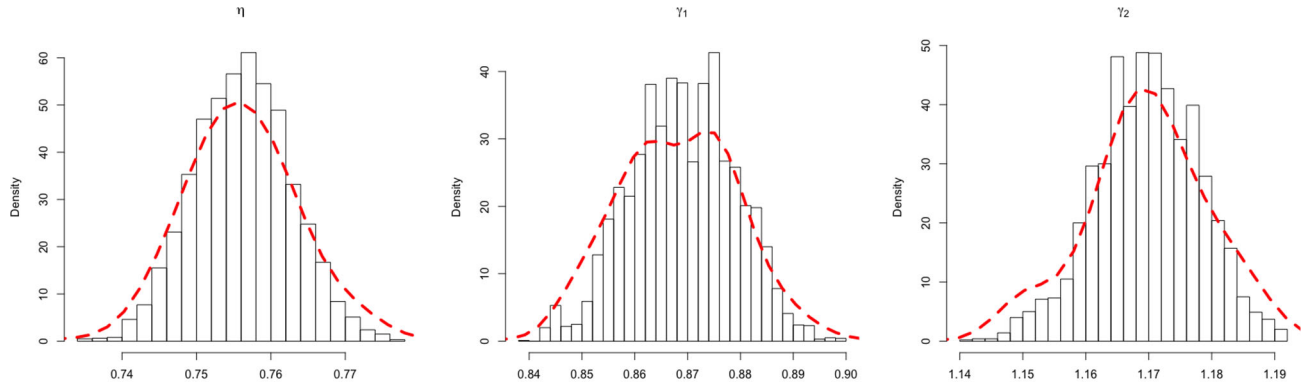


Figure 5. Log-MSEs for the weighted estimator of Z on the $p = 2,\ldots,20$ dimensional mixture of three normals. The MSEs are averaged over 100 independent trials.



(a) FEM (b) Calibrated

Figure 6. Left panel: FEM outputs at three values of the friction coefficient versus the physical experiment output. Right panel: Calibrated model output at the posterior means from PQMC versus the physical experiment output.



(a) η (b) γ_1 (c) γ_2

Figure 7. Histograms of 5000 MCMC samples versus densities of 637 PQMC samples.

none of the FEM output aligns well with the physical experiment output. A further investigation shows that due to the deflection of the sheet at the initial contact with the tool, the tool travel in the physical experiment is less than the tool travel inputted to the FEM, causing the discrepancy. However, fixing this in the FEM code is difficult and computationally expensive. Joseph and Yan (2015) propose an engineering-driven statistical adjustment that can reduce the discrepancy in a more efficient way.

Following the steps described in Section 5 of Joseph et al. (2019), we introduce two adjustment parameters γ_1 and γ_2, leading to the following log posterior distribution,

$$\log p(\eta, \gamma_1, \gamma_2|y) = \text{Const.} - \frac{N}{2}\log\left(\sum_{i=1}^{N}[y_i - \hat{g}(\gamma_1 x_i^{\gamma_2}; \eta)]^2\right) + \log p(\eta) + \log p(\gamma_1) + \log p(\gamma_2). \quad (36)$$

Here, $\hat{g}(\cdot;\eta)$ is the Gaussian process (GP) prediction (Santner, Williams, and Notz 2018) emulating the expensive FEM model $g(\cdot;\eta)$. The details of this Bayesian model are presented in Appendix D, supplementary materials. Using the GP fit from the R package `GPfit` (MacDonald, Ranjan, and Chipman 2015), each posterior evaluation requires more than 10 sec on a 2 cores 2.3 GHz laptop. Thus, given a computational budget, only a handful of posterior evaluations can be performed.

We run the PQMC (Algorithm 3) for $T = 7$ iterations with $K = 13$ proposals and $J = 7$ samples drawn from each proposal, thus, leading to $TKJ = 637$ evaluations of the posterior distribution. The initial centers are the 13 lattice points over $[0.5, 1] \times [0.5, 1] \times [0.75, 1.25]$, which covers the key regions of the prior. The initial proposal covariance matrix is set as $0.2^2 I_3$, since the minimax measure of the 13 centers over the region is 0.3 (computed using the R package `minimaxdesign` (Mak 2019a)). The adapted covariance matrix is kept as isotropic for the sampling procedure. Figure 6 shows the predictions from the calibrated model (right plot) at the posterior mean ($\eta = 0.76, \gamma_1 = 0.87, \gamma_2 = 1.17$), computed using the weighted PQMC estimator. We can see that the calibrated model aligns much better with the actual physical experiment output, com-

pared to the three uncalibrated FEM model outputs (left plot). The performance comparison of PQMC (Algorithm 3) to PMC (Algorithm 2) on this friction drilling example are included in the Appendix D, supplementary materials.

The posterior uncertainty of the three parameters $(\eta, \gamma_1, \gamma_2)$ is also of interest, since it provides a measure of uncertainty for the calibrated FEM model. To verify the uncertainty from the PQMC weighted samples, we generate 5000 MCMC samples using the random walk Metropolis algorithm. A normal proposal distribution is used with covariance $0.01^2 I_3$. The Markov chain is initialized at the point $[0.75, 0.88, 1.18]^T$, which is the posterior mean of $(\eta, \gamma_1, \gamma_2)$ reported in Joseph et al. (2019). Given we start the Markov chain at a promising region, no samples are discarded as burn-in. Figure 7 shows the histograms of these MCMC samples, along with the marginal densities of the PQMC weighted samples. We can see that the densities from PQMC closely mimic the MCMC histograms, which suggest that the proposed approach provides a good quantification of uncertainty for the three calibration parameters, using a reduced sample size of 637 (compared to 5000 for standard MCMC). Given the expensive nature of the posterior evaluations, the reduction in sample size while retaining the same level precision demonstrates the computational advantage of PQMC under this problem setting.

7. Conclusion

This article proposes a new method called Population Quasi-Monte Carlo (PQMC), which incorporates Quasi-Monte Carlo ideas within the sampling and adaptation steps of the generic Population Monte Carlo procedure. The key novelty is the introduction of the importance support points (ISP) resampling, a deterministic resampling method that yields a set of resamples minimized over the energy distance to the weighted samples. Within the PQMC framework, we also propose a computationally efficient adaptation scheme for the proposal covariance, and suggest a new weighted estimator to further reduce the variance of the estimator. Extensive numerical experiments show: (i) significant improvement of the ISP resampling over the existing resampling schemes, including the state-of-the-art Hilbert curve sorted resampling, for problems up to 20 dimensions, and (ii) faster empirical convergence rate of the PQMC over the generic PMC. Although ISP resampling has a higher computational complexity than multinomial resampling, the additional runtime incurred is oftentimes negligible when the posterior is expensive to evaluate, such as the friction drilling example in Section 6.

Supplementary Materials

R codes are available at *https://github.com/BillHuang01/PQMC*. supplementary.pdf contains additional details and simulation results.

Funding

This research is supported by a U.S. National Science Foundation grant CMMI-1921646, and by an NSF CSSI Frameworks grant 2004571.

References

Aistleitner, C., and Dick, J. (2015), "Functions of Bounded Variation, Signed Measures, and a General Koksma-Hlawka Inequality," *Acta Arithmetica*, 167, 143–171, *http://eudml.org/doc/279219*. [2,3]

Bugallo, M. F., Elvira, V., Martino, L., Luengo, D., Miguez, J., and Djuric, P. M. (2017), "Adaptive Importance Sampling: The Past, the Present, and the Future," *IEEE Signal Processing Magazine*, 34, 60–79. [1]

Cappe, O., Godsill, S. J., and Moulines, E. (2007), "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, 95, 899–924. [1]

Cappé, O., Douc, R., Guillin, A., Marin, J.-M., and Robert, C. P. (2008), "Adaptive Importance Sampling in General Mixture Classes," *Statistics and Computing*, 18, 447–459. [9]

Cappé, O., Guillin, A., Marin, J. M., and Robert, C. P. (2004), "Population Monte Carlo," *Journal of Computational and Graphical Statistics*, 13, 907–929. [1,8]

Carpenter, J., Clifford, P., and Fearnhead, P. (1999), "Improved Particle Filter for Nonlinear Problems," *IEE Proceedings - Radar, Sonar and Navigation*, 146, 2–7. [1]

Chen, Z. (2003), "Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond," *Statistics*, 182, 1–69. [1]

Christophe, D., and Petr, S. (2019), *randtoolbox: Generating and Testing Random Numbers*. R package version 1.30.0. [3,8]

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Series B, 39, 1–38. [9]

Dick, J., Kuo, F. Y., and Sloan, I. H. (2013), "High-Dimensional Integration: The Quasi-Monte Carlo Way," *Acta Numerica*, 22, 133–288. [2,11]

Douc, R., and Cappe, O. (2005), "Comparison of Resampling Schemes for Particle Filtering," in *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005.*, IEEE. [5]

Douc, R., Guillin, A., Marin, J.-M., and Robert, C. P. (2007), "Minimum Variance Importance Sampling via Population Monte Carlo," *Probability and statistics*, 11, 427–447. [9]

Elvira, V., Martino, L., Luengo, D., and Bugallo, M. F. (2017), "Improving Population Monte Carlo: Alternative Weighting and Resampling Schemes," *Signal Processing*, 131, 77–91. [1,7,8,9,10,11]

Elvira, V., Martino, L., Luengo, D., and Bugallo, M. F. (2019), "Generalized Multiple Importance Sampling," *Statistical Science*, 34, 129–155. [7]

Fearnhead, P. (2005), "Using Random Quasi-Monte-Carlo within Particle Filters, with Application to Financial Time Series," *Journal of Computational and Graphical Statistics*, 14, 751–769. [2]

Fearnhead, P., and Clifford, P. (2003), "On-line Inference for Hidden Markov Models via Particle Filters," *Journal of the Royal Statistical Society*, Series B, 65, 887–899. [1]

Gerber, M., and Chopin, N. (2015), "Sequential Quasi Monte Carlo," *Journal of the Royal Statistical Society*, Series B, 77, 509–579. [2,6,7,8,10]

Gerber, M., Chopin, N., and Whiteley, N. (2019), "Negative Association, Ordering and Convergence of Resampling Methods," *The Annals of Statistics*, 47, 2236–2260. [2,6]

Givens, G. H. and Raftery, A. E. (1996), "Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationships," *Journal of the American Statistical Association*, 91, 132–141. [1]

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993), "Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation," *IEE Proceedings F - Radar and Signal Processing*, 140, 107–113. [1]

He, Z., and Owen, A. B. (2016), "Extensible Grids: Uniform Sampling on a Space Filling Curve," *Journal of the Royal Statistical Society*, Series B, 78, 917–931. [6]

Hlawka, E., and Mück, R. (1972), "Über eine Transformation von gleichverteilten Folgen II," *Computing*, 9, 127–138. [3,8]

Hol, J. D., Schon, T. B., and Gustafsson, F. (2006), "On Resampling Algorithms for Particle Filters," in *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, IEEE. [5]

Huling, J. D., and Mak, S. (2020), "Energy Balancing of Covariate Distributions," *arXiv preprint arXiv:2004.13962*. [3]

Ji, C., and Schmidler, S. C. (2013), "Adaptive Markov Chain Monte Carlo for Bayesian Variable Selection," *Journal of Computational and Graphical Statistics*, 22, 708–728. [9]

Joe, S., and Kuo, F. Y. (2003), "Remark on algorithm 659: Implementing Sobol's Quasirandom Sequence Generator," *ACM Transactions on Mathematical Software*, 29, 49–57. [3]

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, 26, 131–148. [3]

Joseph, V. R., Wang, D., Gu, L., Lyu, S., and Tuo, R. (2019), "Deterministic Sampling of Expensive Posteriors Using Minimum Energy Designs," *Technometrics*, 61, 297–308. [1,5,12,13]

Joseph, V. R., and Yan, H. (2015), "Engineering-Driven Statistical Adjustment and Calibration," *Technometrics*, 57, 257–267. [12]

Kennard, R. W., and Stone, L. A. (1969), "Computer Aided Design of Experiments," *Technometrics*, 11, 137–148. [5]

Kitagawa, G. (1996), "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, 5, 1–25. [1]

Kong, A., Liu, J. S., and Wong, W. H. (1994), "Sequential Imputations and Bayesian Missing Data Problems," *Journal of the American Statistical Association*, 89, 278–288. [1]

Li, Y., Wang, W., Deng, K., and Liu, J. S. (2021), "Stratification and Optimal Resampling for Sequential Monte Carlo," *Biometrika*, 109, 181–194. [1,5,6,8]

Liu, J. S., and Chen, R. (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93, 1032–1044. [1]

MacDonald, B., Ranjan, P., and Chipman, H. (2015), "GPfit: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs," *Journal of Statistical Software*, 64, 1–23, *http://www.jstatsoft.org/v64/i12/*. [12]

Mak, S. (2019a), *minimaxdesign: Minimax and Minimax Projection Designs*. R package version 0.1.4. [12]

——— (2019b), *support: Support Points*. R package version 0.1.4. [3]

Mak, S., and Joseph, V. R. (2018), "Support Points," *The Annals of Statistics*, 46, 2562–2592. [2,3,5]

Martino, L., Elvira, V., Luengo, D., and Corander, J. (2015), "An Adaptive Population Importance Sampler: Learning From Uncertainty," *IEEE Transactions on Signal Processing*, 63, 4422–4437. [10]

Miller, S. F., and Shih, A. J. (2007), "Thermo-Mechanical Finite Element Modeling of the Friction Drilling Process," *Journal of Manufacturing Science and Engineering*, 129, 531–538. [11]

Moral, P. D., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo Samplers," *Journal of the Royal Statistical Society,* Series B, 68, 411–436. [1]

Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Philadelphia: SIAM. [1,2]

Oh, M.-S., and Berger, J. O. (1993), "Integration of Multimodal Functions by Monte Carlo Importance Sampling," *Journal of the American Statistical Association*, 88, 450–456. [1]

Owen, A. B. (1998), "Scrambling Sobol'and Niederreiter–Xing Points," *Journal of Complexity*, 14, 466–489. [7]

——— (2013), *Monte Carlo Theory, Methods and Examples*. [2]

Portier, F., and Delyon, B. (2018), "Asymptotic Optimality of Adaptive Importance Sampling," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc. [9]

Reich, S. (2013), "A Nonparametric Ensemble Transform Method for Bayesian Inference," *SIAM Journal on Scientific Computing*, 35, A2013–A2024. [1]

Robert, C., and Casella, G. (2013), *Monte Carlo Statistical Methods*, New York: Springer. [1,3]

Rubin, D. B. (1987), "The Calculation of Posterior Distributions by Data Augmentation: Comment: A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing Information are Modest: The SIR Algorithm," *Journal of the American Statistical Association*, 82, 543–546. [5]

Santner, T. J., Williams, B. J., and Notz, W. I. (2018), *The Design and Analysis of Computer Experiments*, New York: Springer. [12]

Scheidegger, A. (2018), *adaptMCMC: Implementation of a Generic Adaptive Monte Carlo Markov Chain Sampler*. R package version 1.3. [3]

Skilling, J. (2004), "Programming the Hilbert Curve," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (Volume 707 of American Institute of Physics Conference Series). [6]

Székely, G. J., and Rizzo, M. L. (2004), "Testing for Equal Distributions in High Dimension," *InterStat*, 5, 1249–1272. [3,4]

——— (2013), "Energy Statistics: A Class of Statistics Based on Distances," *Journal of Statistical Planning and Inference*, 143, 1249–1272. [2,3,4]

West, M. (1993), "Approximating Posterior Distributions by Mixture," *The Journal of the Royal Statistical Society,* Series B, 55, 409–422. [1]

Wu, C. F. J. (1983), "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, 11, 95–103. [9]

Yuille, A. L., and Rangarajan, A. (2002), "The Concave-Convex Procedure (CCCP)," in *Advances in Neural Information Processing Systems* (Vol. 14), eds. T. G. Dietterich, S. Becker and Z. Ghahramani, Cambridge MA: MIT Press. [3]