A weighted Discrepancy Bound of quasi-Monte Carlo Importance Sampling

Josef Dick‡, Daniel Rudolf‡, Houying Zhu‡

January 25, 2019

Abstract

Importance sampling Monte-Carlo methods are widely used for the approximation of expectations with respect to partially known probability measures. In this paper we study a deterministic version of such an estimator based on quasi-Monte Carlo. We obtain an explicit error bound in terms of the star-discrepancy for this method.

Keywords: Importance sampling, Monte Carlo method, quasi-Monte Carlo

Classification. Primary: 62F15; Secondary: 11K45.

1 Introduction

In statistical physics and Bayesian statistics it is desirable to compute expected values

$$E_\pi(f) = \int_{\mathbb{R}^d} f(x) \, d\pi(x)$$  \hspace{1cm} (1)

with $f: \mathbb{R}^d \to \mathbb{R}$ and a partially known probability measure $\pi$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Here $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel $\sigma$-algebra and partially known means that there is an unnormalized density $u: \mathbb{R}^d \to [0, \infty)$ (with respect to the Lebesgue measure) and $\int_{\mathbb{R}^d} u(x) \, dx \in (0, \infty)$, such that

$$\pi(A) = \frac{\int_A u(x) \, dx}{\int_{\mathbb{R}^d} u(y) \, dy}, \quad A \in \mathcal{B}(\mathbb{R}^d).$$  \hspace{1cm} (2)
Probability measures of this type are met in numerous applications. For example, for the density of a Boltzmann distribution one has
\[ u(x) = \exp(-\beta H(x)), \quad x \in \mathbb{R}^d, \]
with inverse temperature \( \beta > 0 \) and Hamiltonian \( H: \mathbb{R}^d \to \mathbb{R} \). The density of a posterior distribution is also of this form. Given observations \( y \in \mathcal{Y} \), likelihood function \( \ell(y \mid x) \) and prior probability density \( p \), with respect to the Lebesgue measure on \( \mathbb{R}^d \),
\[ u(x) = \ell(y \mid x) p(x), \quad x \in \mathbb{R}^d. \]
In this setting \( \mathbb{R}^d \) is considered as parameter- and \( \mathcal{Y} \) as observable-space. In both examples, the normalizing constant is in general unknown.

In the present work we only consider unnormalized densities \( u \) which are zero outside of the unit cube \([0,1]^d\). Hence we restrict ourself to \( u: [0,1]^d \to [0,\infty) \), i.e., \( \pi \) is a probability measure on \([0,1]^d\), and \( f: [0,1]^d \to \mathbb{R} \). To stress the dependence on the unnormalized density in (1), define
\[ S(f,u) := \frac{\int_{[0,1]^d} f(x) u(x) \, dx}{\int_{[0,1]^d} u(y) \, dy} = E_\pi(f) \]
for \( f \) and \( u \) belonging to some class of functions. It is desirable to have algorithms which approximately compute \( S(f,u) \) by only having access to function values of \( f \) and \( u \) without knowing the normalizing constant a priori. A straightforward strategy to do so provides an importance sampling Monte Carlo approach. It works as follows.

**Algorithm 1. Monte Carlo importance sampling:**

1. Generate a sample of an i.i.d. sequence of random variables \( X_1, \ldots, X_n \) with \( X_i \sim \mathcal{U}([0,1]^d) \) and call the result \( x_1, \ldots, x_n \).
2. Compute
\[ M_n(f,u) := \frac{\sum_{j=1}^n f(x_j) u(x_j)}{\sum_{j=1}^n u(x_j)}. \]

Under the minimal assumption that \( S(f,u) \) is finite, a strong law of large numbers argument guarantees that the importance sampling estimator \( M_n(f,u) \) is well-defined, cf. [16, Chapter 9, Theorem 9.2]. For uniformly bounded \( f \) and finite sup \( u/\inf u \) an explicit error bound of the mean square error is proven in [14, Theorem 2].

Surprisingly, there is not much known about a deterministic version of this method. The idea is to substitute the uniformly in \([0,1]^d\) distributed i.i.d. sequence
\[ ^1 \text{By } \mathcal{U}([0,1]^d) \text{ we denote the uniform distribution on } [0,1]^d. \]
by a carefully chosen deterministic point set. Carefully chosen in the sense that the point set \( P_n = \{x_1, \ldots, x_n\} \subset [0,1]^d \) has “small” star-discrepancy, that is,

\[
D_{\lambda_d}(P_n) := \sup_{x \in [0,1]^d} \left| \frac{1}{n} \sum_{j=1}^{n} 1_{[0,x)}(x_j) - \lambda_d([0,x)) \right|
\]

is “small”. Here, the set \([0,x) = \prod_{i=1}^{d} [0,x_i)\) denotes an anchored box in \([0,1]^d\) with \(x = (x_1, \ldots, x_d)\) and \(\lambda_d([0,x)) = \prod_{i=1}^{d} x_i\) is the \(d\)-dimensional Lebesgue measure of \([0,x)\). This leads to a quasi-Monte Carlo importance sampling method.

**Algorithm 2. Quasi-Monte Carlo importance sampling:**

1. Generate a point set \( P_n = \{x_1, \ldots, x_n\} \) with “small” star discrepancy \( D_{\lambda_d}(P_n) \).

2. Compute

\[
Q_n(f,u) = \frac{\sum_{j=1}^{n} f(x_j)u(x_j)}{\sum_{j=1}^{n} u(x_j)}. \tag{3}
\]

Our main result, stated in Theorem 3, is an explicit error bound for the estimator \(Q_n\) of the form

\[
|S(f,u) - Q_n(f,u)| \leq 4 \frac{\|f\|_{H^1} \|u\|_D}{\int_{[0,1]^d} u(x)dx} D_{\lambda_d}(P_n). \tag{4}
\]

Here \(f\) must be differentiable, such that \(\|f\|_{H^1}\), defined in (7) below, is finite. As a regularity assumption on \(u\) it is assumed that \(\|u\|_D\), defined in (9) below, is also finite.

The estimate of (4) is proven by two results which might be interesting on its own. The first is a Koksma-Hlawka inequality in terms of a weighted star-discrepancy, see Theorem 1. The second is a relation between this quantity and the classical star-discrepancy, see Theorem 2. To illustrate the quasi-Monte Carlo importance sampling procedure and the error bound we provide an example in Section 3 where (4) is applicable.

**Related Literature.** The Monte Carlo importance sampling procedure from Algorithm 1 is well studied. In [14], Novak and Mathé prove that it is optimal on a certain class of tuples \((f,u)\). However, recently this Monte Carlo approach attracted considerable attention, let us mention here [1, 4]. In particular, in [1] upper error bounds not only for bounded functions \(f\) are provided and the relevance of the method for inverse problems is presented.

Another standard approach the approximation of \(E_{\pi}(f)\) are Markov chain Monte Carlo methods. For details concerning error bounds we refer to [11, 12, 13, 17, 19, 20, 21] and the references therein. Combinations of importance sampling and Markov chain Monte Carlo are for example analyzed in [18, 24, 22].

The quasi-Monte Carlo importance sampling procedure of Algorithm 2 is, to our knowledge, less well studied. An asymptotic convergence result is stated in [9, Theorem 1] and promising numerical experiments are conducted in [10]. A
related method, a randomized deterministic sampling procedure according to the unnormalized distribution $\pi$, is studied in [23]. Recently, [3] explore the efficiency of using QMC inputs in importance sampling for Archimedean copulas where significant variance reduction is obtained for a case study.

A quasi-Monte Carlo approach to Bayesian inversion was used in [5] and in [6]. The latter paper uses a combination of quasi-Monte Carlo and the multi-level method. The computation of the likelihood function involves solving a partial differential equation, but otherwise the problem is of the same form as described in the introduction.

2 Weighted Star-discrepancy and error bound

Recall that $[0, x)$ for $x \in [0, 1]^d$ are boxes anchored at 0. As a measure of “closeness” between the empirical distribution $\frac{1}{n} \sum_{i=1}^{n} 1_{[0,x)}(x_i)$ of a point set $P_n = \{x_1, \ldots, x_n\}$ to $\lambda_d([0, x))$ we consider the star-discrepancy $D_{\lambda_d}(P_n)$. A straightforward extension of this quantity taking the probability measure $\pi$ on $[0, 1]^d$ into account is the following weighted discrepancy.

**Definition 1 (Weighted Star-discrepancy).** For a given point set $P_n = \{x_1, \ldots, x_n\} \subset [0, 1]^d$ and weight vector $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$, which might depend on $P_n$ and satisfies $\sum_{i=1}^{n} w_i = 1$, define the weighted star-discrepancy by

$$D_{\pi}(w, P_n) = \sup_{x \in [0, 1]^d} \left| \sum_{i=1}^{n} w_i 1_{[0,x)}(x_i) - \pi([0, x)) \right|.$$ 

**Remark 1.** If $\pi$ is the Lebesgue measure on $[0, 1]^d$ and the weight vector is $w = (1/n, \ldots, 1/n)$, then $D_{\lambda_d}(P_n) = D_{\pi}(w, P_n)$ for any point set $P_n$. For general $\pi$ with unnormalized density $u: [0, 1]^d \to [0, \infty)$, allowing the representation (2), we focus on the weight vector

$$w_i^u := w_i(u, P_n) := \frac{u(x_i)}{\sum_{j=1}^{n} u(x_j)}, \quad i = 1, \ldots, n. \quad (5)$$

Here let us emphasize that $w^u := (w_1^u, \ldots, w_n^u)$ depends on $u$ and $P_n$.

2.1 Integration Error and weighted Star-discrepancy

With standard techniques one can prove a Koksma-Hlawka inequality according to $D_{\pi}(w, P_n)$. For details we refer to [7], [8, Section 2.3] and [15, Chapter 9]. A similar inequality of a quasi-Monte Carlo importance sampler can be found in [2, Corollary 1].

Let $[d] := \{1, \ldots, d\}$ and $L_2([0, 1]^d)$ be the space of square integrable functions with respect to the Lebesgue measure. Define the reproducing kernel $K: [0, 1]^d \times [0, 1]^d \to [0, 1]$ by $K(x, y) := \prod_{i=1}^{d} (1 + \min\{1 - x_i, 1 - y_i\})$. By $H_2 = H_2(K)$ we denote the corresponding reproducing kernel Hilbert space, which consists of
differentiable functions with respect to all variables with first partial derivatives being in $L_2([0, 1]^d)$. For $f, g \in H_2$ the inner product is given by

$$\langle f, g \rangle = \sum_{v \subseteq [d]} \int_{[0, 1]^{|v|}} \frac{\partial^{|v|}}{\partial x_v} f(x_v; 1) \frac{\partial^{|v|}}{\partial x_v} g(x_v; 1) \, dx_v,$$

where for $v \subseteq [d]$ and $x = (x_1, \ldots, x_d)$ we write $x_v = (x_j)_{j \in v}$ and $(x_v; 1) = (z_1, \ldots, z_d)$ with $z_j = x_j$ if $j \in V$ and $z_j = 1$ if $j \not\in v$. Thus, $H_2$ consists of functions which are differentiable according to all variables with first partial derivatives being in $L_2([0, 1]^d)$. Note that, for $v \subseteq [d]$ holds

$$\frac{\partial^{|v|}}{\partial x_v} K((x_v; 1), y) = (-1)^{|v|} 1_{[y_v, 1]}(x_v),$$

where $[y_v, 1] = \prod_{i \in v} [y_i, 1]$ with $y = (y_1, \ldots, y_d) \in [0, 1]^d$. Thus, the reproducing property of the reproducing kernel Hilbert space can be rewritten as

$$f(y) = \sum_{v \subseteq [d]} \int_{[y_v, 1]} (-1)^{|v|} \frac{\partial^{|v|}}{\partial x_v} f(x_v; 1) \, dx_v. \quad (6)$$

Further, we define the space $H_1$ of differentiable functions $f: [0, 1]^d \to \mathbb{R}$ with finite norm

$$\|f\|_{H_1} := \sum_{v \subseteq [d]} \int_{[0, 1]^{|v|}} \left| \frac{\partial^{|v|}}{\partial x_v} f(x_v; 1) \right| \, dx_v, \quad (7)$$

where for $v = \emptyset$ we have $\int_{[0, 1]^{|v|}} \left| \frac{\partial^{|v|}}{\partial x_v} f(x_v; 1) \right| \, dx_v = |f(1)|$. We also define the semi-norm

$$\|f\|_{\bar{H}_1} := \sum_{\emptyset \neq v \subseteq [d]} \int_{[0, 1]^{|v|}} \left| \frac{\partial^{|v|}}{\partial x_v} f(x_v; 1) \right| \, dx_v. \quad (8)$$

It is obvious that $\|f\|_{\bar{H}_1} \leq \|f\|_{H_1}$.

We have the following relation between the integration error in $H_1$ and the weighted discrepancy.

**Theorem 1** (Koskma-Hlawka inequality). *Let $\pi$ be a probability measure of the form (2) with unnormalized density $u: [0, 1]^d \to [0, \infty)$. Then, for $P_n = \{x_1, \ldots, x_n\} \subset [0, 1]^d$, arbitrary weight vector $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ with $\sum_{i=1}^d w_i = 1$, and for all $f \in H_1$ we have

$$\left| S(f, u) - \sum_{i=1}^n w_i f(x_i) \right| \leq \|f\|_{\bar{H}_1} D_n(w, P_n).$$

**Proof.** Define the quadrature error $e(f, P_n) := \int_{[0, 1]^d} f(x) \, d\pi(x) - \sum_{i=1}^n w_i f(x_i)$ of the approximation of $E_\pi(f) = S(f, u)$ by $\sum_{i=1}^n w_i f(x_i)$. Define the function $\tilde{f} = f - f(1)$. Then $\tilde{f}(1) = 0$, $e(f, P_n) = e(\tilde{f}, P_n)$ and $\|f\|_{\bar{H}_1} = \|\tilde{f}\|_{H_1}$.
For
\[ h(x) := \int_{[0,1]^d} K(x, y) \, d\pi(y) - \sum_{i=1}^n w_i K(x, x_i), \]
and \( v \subseteq [d] \) we have \( \frac{\partial^{[v]}}{\partial x_v} h(z_v; 1) = (-1)^{|v|} \left( \pi([0, (z_v; 1)]) - \sum_{i=1}^n w_i 1_{[0,z_v]}(x_i,v) \right) \). A straightforward calculation, see also for instance [7, formula (3)], shows by using (6) that
\[ e(\tilde{f}, P_n) = \sum_{v \subseteq [d]} \int_{[0,1]^{|v|}} \frac{\partial^{[v]}}{\partial x_v} \tilde{f}(z_v; 1)(-1)^{|v|} \left( \pi([0, (z_v; 1)]) - \sum_{i=1}^n w_i 1_{[0,z_v]}(x_i,v) \right) \, dz \]
\[ = \langle \tilde{f}, h \rangle. \]
Finally, by \( \left| \frac{\partial^{[v]}}{\partial x_v} h(z_v; 1) \right| \leq D_x(w, P_n) \) we have
\[ |e(f, P_n)| = |e(\tilde{f}, P_n)| \leq \|f\|_{H_1} D_x(w, P_n) = \|f\|_{\tilde{H}_1} D_\pi(w, P_n), \]
which finishes the proof. \( \square \)

An immediate consequence of the theorem with \( w^u \) from (5) and \( Q_n \) from (3) is the error bound
\[ |S(f, u) - Q_n(f, u)| \leq \|f\|_{H_1} D_x(w^u, P_n). \]
Here the dependence on \( u \) on the right-hand side is hidden in \( D_x(w^u, P_n) \) through \( w^u \) and \( \pi \). The intuition is, that under suitable assumptions on \( u \) the weighted star-discrepancy can be bounded by the classical star-discrepancy of \( P_n \).

### 2.2 Weighted and classical Star-discrepancy

In this section we provide a relation between the classical star-discrepancy \( D_{\lambda_d}(P_n) \) and the weighted star-discrepancy \( D_\pi(w^u, P_n) \).

**Theorem 2.** Let \( \pi \) be a probability measure of the form (2) with unnormalized density function \( u: [0,1]^d \to [0, \infty) \). Then, for any point set \( P_n = \{x_1, \ldots, x_n\} \) in \([0,1]^d\), we have
\[ D_\pi(w^u, P_n) \leq 4D_{\lambda_d}(P_n) \frac{\|u\|_D}{\int_{[0,1]^d} u(x) \, dx}, \]
where
\[ \|u\|_D = \sup_{z \in [0,1]^d} u(z) + \sup_{z \in [0,1]^d} \|u(T_z \cdot)\|_{\tilde{H}_1}, \]
with \( T_z: [0,1]^d \to [0,z] \) and \( T_z(x_1, \ldots, x_d) = (z_1 x_1, \ldots, z_d x_d) \) for \( z \in [0,1]^d \).
Proof. For the given point set \( P_n \subset [0,1]^d \) and unnormalized density \( u \) recall that \( w^u \) is defined in (5). To shorten the notation define \( \|u\|_1 := \int_{[0,1]^d} u(y) dy \). Then, for \( z \in [0,1]^d \) we have
\[
\left| \sum_{j=1}^{n} w_j^n 1_{[0,z]}(x_j) - \pi([0,z]) \right| = \left| \frac{\sum_{j=1}^{n} u(x_j) 1_{[0,z]}(x_j)}{\|u\|_1} - \frac{\int_{[0,z]} u(x) dx}{\|u\|_1} \right|
\leq \frac{\sum_{j=1}^{n} u(x_j) 1_{[0,z]}(x_j)}{\|u\|_1} \left| \|u\|_1 - \frac{n}{\sum_{i=1}^{n} u(x_i)} \right|
+ \frac{1}{\|u\|_1} \left| \frac{n}{\sum_{i=1}^{n} u(x_i)} \sum_{i=1}^{n} u(x_i) 1_{[0,z]}(x_i) - \int_{[0,z]} u(x) dx \right|
\leq \frac{2}{\|u\|_1} \sup_{z \in [0,1]^d} \left| \frac{1}{\sum_{i=1}^{n} u(x_i)} \sum_{i=1}^{n} u(x_i) 1_{[0,z]}(x_i) - \int_{[0,z]} u(x) dx \right|.
\]

For \( z \in [0,1]^d \) denote \( P^z = P_n \cap [0,z] \) and let \( |P^z| \) be the cardinality of \( P^z \). Define
\[
I_1(z) := \frac{\int_{[0,z]} u(x) dx}{\lambda_d([0,z])} \left| \frac{|P^z|}{n} - \lambda_d([0,z]) \right|,
\]
\[
I_2(z) := \frac{|P^z|}{n} \left| \frac{1}{|P^z|} \sum_{x \in P^z} u(x) - \frac{\int_{[0,z]} u(x) dx}{\lambda_d([0,z])} \right|,
\]
and note that
\[
\left| \frac{1}{\sum_{i=1}^{n} u(x_i)} \sum_{i=1}^{n} u(x_i) 1_{[0,z]}(x_i) - \int_{[0,z]} u(x) dx \right|
= \frac{|P^z|}{n} \left| \frac{1}{|P^z|} \sum_{x \in P^z} u(x) - \frac{\int_{[0,z]} u(x) dx}{\lambda_d([0,z])} \right| \leq I_1(z) + I_2(z).
\]

Estimation of \( I_1(z) \): An immediate consequence of the definition of \( I_1(z) \) is
\[
I_1(z) \leq \frac{\int_{[0,z]} u(x) dx}{\lambda_d([0,z])} D_{\lambda_d}(P_n) \leq D_{\lambda_d}(P_n) \sup_{x \in [0,z]} u(x).
\]

Estimation of \( I_2(z) \): With the transformation \( T_z : [0,1]^d \rightarrow [0,z] \) defined in the theorem one has \( \frac{\int_{[0,z]} u(x) dx}{\lambda_d([0,z])} = \int_{[0,1]^d} u(T_z x) dx \). Let
\[
Q := T_z^{-1} P^z = \{ (z_1^{-1} x_1, \ldots, z_d^{-1} x_d) \mid x \in P^z \} \subset [0,1]^d
\]
and observe that \( |P^z| = |Q| \). Then
\[
I_2(z) = \frac{|P^z|}{n} \left| \frac{1}{|Q|} \sum_{x \in Q} u(T_z x) - \int_{[0,1]^d} u(T_z x) dx \right| \leq \frac{|P^z|}{n} D_{\lambda_d}(Q) \| u(T_z \cdot) \|_{H_1},
\]

7
where the last inequality follows from Theorem 1 with \( w = (1/n, \ldots, 1/n) \) and constant unnormalized density. Further,

\[
\frac{|Pz|}{n} D_{\lambda_d}(Q) = \frac{|Pz|}{n} \sup_{y \in [0,1]^d} \left| \frac{1}{|Q|} \sum_{x \in Q} 1_{[0,y)}(x) - \lambda_d([0,y)) \right|
\]

\[
= \sup_{y \in [0,1]^d} \left| \frac{1}{n} \sum_{x \in P_n} 1_{[0,y)}(x) - |Q| \lambda_d([0,y)) \right|
\]

\[
= \sup_{y \in [0,1]^d} \left| \frac{1}{n} \sum_{x \in P_n} 1_{Tz([0,y))}(x) - |Q| \lambda_d([0,y)) \right|
\]

\[
\leq \sup_{y \in [0,1]^d} \left| \frac{1}{n} \sum_{x \in P_n} 1_{Tz([0,y))}(x) - \lambda_d(Tz([0,y))) \right|
\]

\[
+ \sup_{y \in [0,1]^d} \left| \lambda_d(Tz([0,y))) - |Q| \lambda_d([0,y)) \right|
\].

By the fact that \( Tz([0,y)) \) is again a box anchored at 0 and

\[
\sup_{y \in [0,1]^d} \left| \lambda_d(Tz([0,y))) - |Q| \lambda_d([0,y)) \right| = \sup_{y \in [0,1]^d} \lambda_d([0,y)) \left| \lambda_d([0,z)) - \frac{|Pz|}{n} \right|
\]

\[
\leq \left| \lambda_d([0,z)) - \frac{|Pz|}{n} \right|
\],

we have

\[
I_2(z) \leq 2 \| u(Tz \cdot) \|_H_1 \left| \lambda_d([0,z)) - \frac{|Pz|}{n} \right| \leq 2 \| u(Tz \cdot) \|_H_1 D_{\lambda_d}(P_n).
\]

Hence we have

\[
\sup_{z \in [0,1]^d} \left| \sum_{j=1}^n w_j 1_{[0,z)}(x_j) - \pi([0,z)) \right| \leq 2 \sup_{z \in [0,1]^d} I_1(z) + I_2(z)
\]

\[
\leq 2D_{\lambda_d}(P_n) \sup_{z \in [0,1]^d} \left( \sup_{x \in [0,z]} u(x) + 2 \| u(Tz \cdot) \|_H_1 \right),
\]

which implies the result.

In particular, the theorem implies that whenever \( \| u \|_D \) is finite and \( D_{\lambda_d}(P_n) \) goes to zero as \( n \) goes to infinity, also \( D_{\pi}(w^\ast, P_n) \) goes to zero for increasing \( n \) with the same rate of convergence.

### 2.3 Explicit error bound

An immediate consequence of the results of the previous two sections is the following explicit error bound of the quasi-Monte Carlo importance sampling method of Algorithm 2.
Theorem 3. Let \( \pi \) be a probability measure of the form (2) with unnormalized density \( u: [0, 1]^d \to [0, \infty) \). Then, for any point set \( P_n = \{x_1, \ldots, x_n\} \) in \([0, 1]^d\), \( f \in H_1 \) and \( Q_n \) from (3) we obtain

\[
|S(f, u) - Q_n(f, u)| \leq 4 \frac{\|f\|_{H_1} \|u\|_D}{\int_{[0, 1]^d} u(x)dx} D_{\lambda}(P_n),
\]

with \( \|u\|_D \) from Theorem 2.

Under the regularity assumption that \( \|u\|_D \) is finite, the error bound tells us that the classical star-discrepancy determines the rate of convergence on how fast \( Q_n(f, u) \) goes to \( S(f, u) \).

3 Illustrating Example

Define the \( d \)-simplex by \( \Delta_d := \{x \in [0, 1]^d: \sum_{i=1}^d x_i \leq 1\} \) and consider the (slightly differently formulated) unnormalized density \( u: [0, 1]^d \to [0, 1) \) of the Dirichlet distribution with parameter vector \( \alpha \in (1, \infty)^{d+1} \) given by

\[
u(x; \alpha) = \begin{cases} (1 - \sum_{i=1}^d x_i)^{\alpha_d + 1} \prod_{i=1}^d x_i^{\alpha_i - 1}, & x \in \Delta_d, \\ 0, & x \not\in \Delta_d. \end{cases}
\]

The Dirichlet distribution is the conjugate prior of the multinomial distribution: Assume that we observed some data \( y = (y_1, \ldots, y_{d+1}) \in [0, 1]^{d+1} \), which we model as a realization of a multinomial distributed random variable with unknown parameter vector \( x = (x_1, \ldots, x_d) \in [0, 1]^d \). With \( n \in \mathbb{N} \) this leads to a likelihood function \( \ell(y | x) = \frac{n!}{y_1! \cdots y_{d+1}!} (1 - \sum_{i=1}^d x_i)^{y_{d+1}} \prod_{i=1}^d x_i^{y_i} \). For a prior distribution with unnormalized density \( u(x; \beta) \) and \( \beta \in (1, \infty)^{d+1} \) we obtain a posterior measure with unnormalized density \( u(x; \beta + y) \).

The normalizing constant of \( u \) can be computed explicitly, it is known that

\[
\int_{[0, 1]^d} u(x; \alpha)dx = \frac{\prod_{i=1}^{d+1} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{d+1} \alpha_i)}.
\]

To have a feasible setting for the application of Theorem 1 and Theorem 2 we need to show that \( \|u\|_D \) is finite. This is not immediately clear, since in \( \|u\|_D \) we take the supremum over \( z \in [0, 1]^d \). The following lemma is useful.

Lemma 1. Let \( v \subseteq [d] \) and recall that we write \( k_v = (k_i)_{i \in v} \). Define \((k_v; 0; k_{d+1}) = (r_1, \ldots, r_{d+1})\) with \( r_j = k_j \) if \( j \in v \), \( r_j = 0 \) if \( j \not\in v \) and \( r_j = k_{d+1} \) if \( j = d + 1 \). Assume that \( \alpha_i \geq 2 \) for \( 1 \leq i \leq d \) and \( \alpha_{d+1} \geq d \). Then

\[
\frac{\partial |v|}{\partial x_v} u(x; \alpha) = \sum_{k_{d+1} = |v| - \sum_{i \in v} k_i} \sum_{k_v \in [0, 1]^{|v|}} c_{v, k_v, k_{d+1}} u(x; (k_v; 0; k_{d+1}))
\]

with \( c_{v, k_v, k_{d+1}} = (-1)^{k_{d+1}} \prod_{j=1}^{k_{d+1}} (\alpha_{d+1} - j) \prod_{i \in v} (\alpha_i - 1)^{k_i} \).
Proof. The statement follows by induction over the cardinality of $v$. For $|v| = 0$, i.e., $v = \emptyset$ both sides of (13) are equal to $u(x, \alpha)$.

Assume $|v| = 1$, i.e., for some $s \in [d]$ we have $v = \{s\}$. Then

$$
\frac{\partial}{\partial x_s} u(x, \alpha) = (\alpha_s - 1)u(x, \alpha - e_s) - (\alpha_{d+1} - 1)u(x, \alpha - e_{d+1}),
$$

with $e_s = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^{d+1}$ where the $i$th entry is "1". On the other hand

$$
\sum_{k_s \in \{0,1\}} c_{\{s\},k_s,k_{d+1}} u(x, \alpha - (k_s, 0; k_{d+1}))
= c_{\{s\},0,1} u(x, \alpha - e_d) + c_{\{s\},1,0} u(x, \alpha - e_s).
$$

By the fact that $c_{\{s\},0,1} = -(\alpha_{d+1} - 1)$ and $c_{\{s\},1,0} = (\alpha_s - 1)$ the claim is proven for $|v| = 1$.

Now assume that (13) is true for any $v \subseteq [d]$ with $|v| \leq \ell < d$. Let $v \subseteq [d]$ with $|v| = \ell$ be an arbitrary subset and let $r \in [d]$ with $r \notin v$. Then we prove that the result also holds for $\overline{v} = v \cup \{r\}$. We have

$$
\frac{\partial |\overline{v}|}{\partial x_{\overline{v}}} u(x, \alpha) = \frac{\partial |v|}{\partial x_v} u(x, \alpha)
= \sum_{k_r \in \{0,1\}^{\overline{v}}} c_{v,k_r,k_{d+1}} \frac{\partial}{\partial x_r} u(x, \alpha - (k_r, 0; k_{d+1})).
$$

Observe that

$$
\frac{\partial}{\partial x_r} u(x, \alpha - (k_r, 0; k_{d+1}))
= (\alpha_r - 1)u(x, \alpha - (k_r, 0; k_{d+1}) - e_r) - (\alpha_{d+1} - k_{d+1} - 1)u(x, \alpha - (k_r, 0; k_{d+1} + 1))
= \sum_{k_r \in \{0,1\}} (\alpha_r - 1)^{k_r}(-1)^{1-k_r}(\alpha_{d+1} - k_{d+1} - 1)^{1-k_r} u(x, \alpha - (k_r, 0; \tilde{k}_{d+1}) - k_r e_r),
$$

where $\tilde{k}_{d+1} := k_{d+1} + 1 - k_r$. Further, note that

$$
c_{\tilde{v},k_r,k_{d+1}} (\alpha_r - 1)^{k_r}(-1)^{1-k_r}(\alpha_{d+1} - k_{d+1} - 1)^{1-k_r}
= (-1)^{\tilde{k}_{d+1} - 1 - k_r} \prod_{j=1}^{\tilde{k}_{d+1}} (\alpha_{d+1} - j)(\alpha_{d+1} - k_{d+1} - 1)^{1-k_r} \prod_{i \in \overline{v}} (\alpha_i - 1)^{k_i} (\alpha_r - 1)^{k_r}
= (-1)^{\tilde{k}_{d+1}} \prod_{j=1}^{\tilde{k}_{d+1}} (\alpha_{d+1} - j) \prod_{i \in \overline{v}} (\alpha_i - 1)^{k_i} = c_{\tilde{v},k_r,k_{d+1}}.
$$
Hence, by using \( \bar{k}_{d+1} := k_{d+1} + 1 - k_r \) we obtain

\[
\frac{\partial^{|v|}}{\partial x_v} u(x, \alpha) = \sum_{k_v \in \{0,1\}^{|v|}} \sum_{k_r \in \{0,1\}} c_{\bar{k}_r, \bar{k}_d+1} u(x, \alpha - (k_v; 0; \bar{k}_{d+1}) - k_r e_r)
\]

\[
= \sum_{k_v \in \{0,1\}^{|v|}} \sum_{k_r \in \{0,1\}} c_{\bar{k}_r, \bar{k}_d+1} u(x, \alpha - (k_v; 0; \bar{k}_{d+1}))
\]

and the proof is finished. \( \square \)

An immediate consequence of the previous lemma and a chain rule argument we have for arbitrary \( v \subseteq [d], z \in [0,1]^d \) and \( T_z \), defined as in Theorem 2, that

\[
\frac{\partial^{|v|}}{\partial x_v} u(T_z x, \alpha) = \prod_{i \in v} z_i \sum_{k_v \in \{0,1\}^{|v|}} \sum_{k_r \in \{0,1\}} c_{v,k_v,k_{d+1}} u(T_z x, \alpha - (k_v; 0; k_{d+1})).
\]

For \( \alpha_i \geq 2 \) with \( 1 \leq i \leq d, \alpha_{d+1} \geq d \) and arbitrary \( x, z \in [0,1]^d \), holds \( u(T_z x, \alpha - (k_v; 0; k_{d+1})) \leq 1 \), where \( v \subseteq [d], k_v \in \{0,1\}^{|v|} \) and \( k_{d+1} \in [d] \). Then, it follows that \( \frac{\partial^{|v|}}{\partial x_v} u(T_z x, \alpha) \mid \leq C_{d,\alpha}^{(1)} \), with a constant \( C_{d,\alpha}^{(1)} \) depending on \( d \) and \( \alpha \).

Hence, for another constant \( C_{d,\alpha}^{(2)} \) holds \( \|u(T_z \cdot, \alpha)\|_{H_1} \leq C_{d,\alpha}^{(2)} < \infty \) uniformly in \( z \in [0,1]^d \). Finally, by the fact that \( u(x) \leq 1 \) we obtain the following corollary.

**Corollary 1.** For \( \alpha_i \geq 2 \) with \( 1 \leq i \leq d \) and \( \alpha_{d+1} \geq d \) we have for \( u(x, \alpha) \) defined in \( (11) \) that there is a constant \( C_{d,\alpha} \) such that

\[
\frac{\|u(\cdot, \alpha)\|_D}{\int_{[0,1]^d} u(x, \alpha)dx} \leq C_{d,\alpha} < \infty.
\]

This verifies that the application of Theorem 1 and Theorem 2 is justified. For \( w^\alpha \) given by \( (5) \) we obtain

\[
\left| S(f, u(\cdot, \alpha)) - \sum_{i=1}^n f(x_i) u(x_i, \alpha) \right| \leq 4\|f\|_{H_1} C_{d,\alpha} D_{\lambda_d}(P_n).
\]

Consider \( f_\gamma : [0,1]^d \to [0,1] \) with \( \gamma \in (1,\infty)^d \) given by \( f_\gamma(x) = 2^{-d} \prod_{i=1}^d x_i^{\gamma_i} \). Then, by \( (12) \) we have

\[
S(f_\gamma, u(\cdot, \alpha)) = \frac{1}{2^d} \cdot \frac{\int_{[0,1]^d} u(x, \alpha_1 + \gamma_1, \ldots, \alpha_d + \gamma_d, \alpha_{d+1})dx}{\int_{[0,1]^d} u(x, \alpha)dx}
\]

\[
= \frac{1}{2^d} \cdot \frac{\prod_{i=1}^d \Gamma(\alpha_i + \gamma_i)}{\prod_{i=1}^d \Gamma(\alpha_i)} \cdot \frac{\Gamma(\sum_{i=1}^{d+1} \alpha_i)}{\Gamma(\alpha_{d+1} + \sum_{i=1}^d (\alpha_i + \gamma_i))}
\]
and \(\|f_\gamma\|_{H_1} = 1\). Since we know \(S(f_\gamma, u(\cdot, \alpha))\) we can run the quasi-Monte Carlo importance sampling algorithm and plot the error for different \(d\) and fixed \(\alpha\) and \(\gamma\).

**Numerical experiments.** Let \(\gamma = (1, \ldots, 1) \in \mathbb{R}^d\) and \(\alpha = (2, \ldots, 2, d) \in \mathbb{R}^d\). Here the true expectation of \(f_\gamma\) according to the distribution determined by \(u(\cdot, \alpha)\) can be further simplified to \(S(f_\gamma, u(\cdot, \alpha)) = (3d^2 - 1)!/(4d^2 - 1)!\). Since for large \(d\) this value is very small we plot the normalized error. For a given point set \(P_n\) it is defined by

\[
\text{error}(P_n) = \left| 1 - \frac{Q_n(f_\gamma, u(\cdot, \alpha))}{S(f, u(\cdot, \alpha))} \right|,
\]

and can be computed exactly. Let \(H_n\) the first \(n\) points of the Halton sequence and note that it is known that \(D_{\lambda_d}(H_n) \leq O\left(\frac{(\log n)^d}{n}\right)\). By \(S_n\) we denote the first \(n\) points of the Sobol sequence. For details to those standard quasi-Monte Carlo point sets we refer to [8]. We obtain the following plots for \(d = 2, 4, 6\).

![Figure 1: Plot of the normalized error (14) of \(Q_n(f_\gamma, u(\cdot, \alpha))\) based on the Halton sequence \(H_n\) for \(d = 2, 4, 6\).](image1)

![Figure 2: Plot of the normalized error (14) of \(Q_n(f_\gamma, u(\cdot, \alpha))\) based on the Sobol sequence \(S_n\) for \(d = 2, 4, 6\).](image2)

**Acknowledgment**

D. Rudolf is supported by the Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, the Campus laboratory AIMS and the DFG within the project 389483880.

**References**

[1] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso, and A. Stuart, *Importance Sampling: Intrinsic Dimension and Computational Cost*, Statist. Sci. 32.
[2] Ch. Aistleitner and J. Dick, *Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality*, Acta Arith. **167** (2015), 143–171.

[3] P. Arbenz, M. Cambou, M. Hofert, C. Lemieux, and Y. Taniguchi, *Importance sampling and stratification for copula models*, Contemporary Computational Mathematics - a celebration of the 80th birthday of Ian Sloan (J. Dick, F. Y. Kuo, H. Woźniakowski, eds.), Springer-Verlag, 2018.

[4] S. Chatterjee and P Diaconis, *The sample size required in importance sampling*, Ann. Appl. Probab. **28** (2018), 1099–1135.

[5] J. Dick, R. N. Gantner, Q. T. Le Gia, and C. Schwab, *Higher order Quasi-Monte Carlo integration for Bayesian Estimation*, ArXiv e-prints (2016).

[6] ______, *Multilevel higher-order quasi-Monte Carlo Bayesian estimation*, Math. Models Methods Appl. Sci. **27** (2017), 953–995.

[7] J. Dick, A. Hinrichs, and F. Pillichshammer, *Proof Techniques in Quasi-Monte Carlo Theory*, J. Complexity **31** (2015), 327–371.

[8] J. Dick and F. Pillichshammer, *Digital nets and sequences: Discrepancy theory and quasi-Monte Carlo integration*, Cambridge University Press, Cambridge, 2010.

[9] M. Gerber and N. Chopin, *Sequential Quasi-Monte Carlo*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **77** (2015), 509–579.

[10] W. Hörmann and J. Leydold, *Quasi importance sampling*, Preprint, Available at http://epub.wu.ac.at/1394/ (2005).

[11] G. Jones, *On the Markov chain central limit theorem*, Probab. Surv. **1** (2004), 299–320.

[12] A. Joulin and Y. Ollivier, *Curvature, concentration and error estimates for Markov chain Monte Carlo*, Ann. Probab. **38** (2010), no. 6, 2418–2442.

[13] K. Latuszyński, B. Miasojedow, and W. Niemiro, *Nonasymptotic bounds on the estimation error of MCMC algorithms*, Bernoulli **19** (2013), 2033–2066.

[14] P. Mathé and E. Novak, *Simple Monte Carlo and the Metropolis algorithm*, J. Complexity **23** (2007), no. 4-6, 673–696.

[15] E. Novak and H. Woźniakowski, *Tractability of multivariate problems. Vol. 2: Standard information for functionals*, EMS Tracts in Mathematics, vol. 12, European Mathematical Society (EMS), Zürich, 2010.

[16] A. Owen, *Monte Carlo theory, methods and examples*, 2013, in preparation.

[17] D. Paulin, *Mixing and concentration by Ricci curvature*, Journal of Functional Analysis **270** (2016), no. 5, 1623 – 1662.
[18] V. Roy, A. Tan, and J. Flegal, *Estimating standard errors for importance sampling estimators with multiple Markov chains*, ArXiv e-prints 1509.06310 (2015).

[19] D. Rudolf, *Explicit error bounds for lazy reversible Markov chain Monte Carlo*, J. Complexity 25 (2009), no. 1, 11–24.

[20] ______, *Error bounds of computing the expectation by Markov chain Monte Carlo*, Monte Carlo Methods Appl. 16 (2010), 323–342.

[21] ______, *Explicit error bounds for Markov chain Monte Carlo*, Dissertationes Math. 485 (2012), 93 pp.

[22] D. Rudolf and B. Sprungk, *On a Metropolis-Hastings importance sampling estimator*, ArXiv e-prints (2018).

[23] B. Vandewoestyne and R. Cools, *On the convergence of quasi-random sampling/importance resampling*, Math. Comput. Simulation 81 (2010), no. 3, 490–505.

[24] M. Vihola, J. Helske, and J. Franks, *Importance sampling type correction of Markov chain Monte Carlo and exact approximations*, ArXiv e-prints 1609.02541 (2016).