Check for updates

Sampling and Change of Measure for Monte Carlo Integration on Simplices

Chenxiao Song[1] · Reiichiro Kawai[2]

Abstract

Simplices are the fundamental domain when integrating over convex polytopes. The aim of this work is to establish a novel framework of Monte Carlo integration over simplices, throughout from sampling to variance reduction. Namely, we develop a uniform sampling method on the standard simplex consisting of two independent procedures and construct theories on change of measure on each of the two independent elements in the developed sampling technique with a view towards variance reduction by importance sampling. We provide illustrative figures and numerical results to support our theoretical findings and demonstrate the strong potential of the developed framework for effective implementation and acceleration of Monte Carlo integration over simplices.

Keywords Numerical integration · Simplex · Dirichlet law · Monte Carlo methods · Variance reduction

Mathematics Subject Classification 65C05 · 65D30 · 65C10 · 62H10

1 Introduction

Integrals over polytopes have found important applications in various areas such as computational geometry, finite element methods, statistics and optimization. Integrating over simplices is of particular interest in the literature, since every convex polytope can be decomposed into finitely many simplices. For instance, various formulas are developed for numerical

---

✉ Reiichiro Kawai
raykawai@g.ecc.u-tokyo.ac.jp

Chenxiao Song
g.song@maths.usyd.edu.au

[1] School of Mathematics and Statistics, The University of Sydney, Sydney, NSW 2006, Australia

[2] Graduate School of Arts and Sciences/Mathematics and Informatics Center, The University of Tokyo, Tokyo, Japan

integration in early years [9, 11, 12, 30, 32], to name a few. Polynomials are often the focus of such formulas, as the principle of classical algorithms involves the approximation of an integrand, followed by exact integration of polynomials. More recently, an exact formula is proposed in [21] for integrating homogeneous polynomials on a simplex, of which an extension is studied in [18] to cover a wider class, including non-homogeneous polynomials. A closed-form expression is derived in [20] for the integral of an arbitrary polynomial on a full-dimensional simplex on the basis of integration formulas on the standard simplex. Moreover, a recursive formula is constructed in [22] to improve the accuracy and computing speed when integrating polynomials. For non-polynomial integrands, quadrature formulas are developed in [5] with the aid of functional and derivative values of the integrand up to a fixed order at the vertices of a simplex. Approximate formulas are built in [10] for integrating functions over hyperplane sections of the standard simplex.

Various numerical techniques have also been developed for integrating over simplices in the literature. For instance, an adaptive numerical cubature algorithm is constructed in [8] for approximating an integral over a collection of simplices. As for the exact integration of polynomials over simplicial regions, computational complexity and various algorithms are discussed in [2], where integrating an arbitrary polynomial over a general rational simplex is shown to be NP-hard. An algorithm is presented in [27] for computing the exact value of the integral of a polynomial with the degree of a polynomial and the dimension of a simplex fixed, and is proved to have polynomial-time complexity. Quasi-Monte Carlo methods are applied in [25, 26] to evaluate an integral over simplices by transforming suitable low-discrepancy sequences on the unit hypercube to the simplex. The quasi-Monte Carlo tractability is studied in [3] in integrating functions over the product of copies of the standard simplex.

Monte Carlo methods have been studied extensively for numerical integration over the unit hypercube along with a variety of variance reduction techniques for accelerating the convergence of the central limit theorem, whereas not much attention has been given to Monte Carlo integration, to say nothing of variance reduction techniques, on more challenging domains, such as polytopes. In this work, we aim to build a novel framework of Monte Carlo integration over simplices, from the beginning (random number generation) to the end (variance reduction). To be precise, we develop a uniform sampling technique over the standard simplex and then examine theories on change of measure with a view towards importance sampling.

First, the proposed uniform sampling technique (Theorem 3.2) is novel and efficient in the sense that it only consists of two independent components and wastes no realizations as the sample is projected towards the origin, rather than acceptance-rejection after shifting the sample in the direction perpendicular to the canonical hyperplane [7], by which some realizations end up outside the standard simplex and thus must be thrown out. Built on the uniform sampling technique, we next develop two distinct frameworks (Theorems 4.2 and 4.3) on change of measure on the canonical hyperplane in combination with yet another separate change of measure on the projection in the proposed uniform sampling technique. We demonstrate strong potential of both two frameworks in reducing the estimator variance by sending the mass of the relevant probability law towards more important sections of the standard simplex, such as a single vertex, a single surface, and even multiple components at once.

The rest of this paper is set out as follows. In Sect. 2, we formulate Monte Carlo integration over simplices and justify our focus on the standard simplex. In Sect. 3, after summarizing background materials on the Dirichlet law and its sampling, we develop a uniform sampling method on the standard simplex, along with a brief review of existing sampling methods. We construct in Sect. 4 theories on change of measure on the projection (Sect. 4.1) in combination

with change of measure on the canonical hyperplane in two ways (Sects. 4.2.1 and 4.2.2). In addition to illustrative figures throughout, we present an extensive collection of numerical examples in Sect. 5 to demonstrate the effectiveness of the proposed framework of Monte Carlo integration over simplices. To maintain the flow of the paper, we collect all proofs in the Appendix.

2 Problem Formulation

We first summarize the notation that will be used throughout. We denote by $|\cdot|$ and $\|\cdot\|$ the magnitude and the Euclidean (or a suitable matrix) norm, respectively. We denote by $\mathrm{Leb}(D)$, $\mathrm{int}(D)$, ∂D, $\overline{D}$, $\mathcal{B}(D)$, respectively, the Lebesgue area, the interior, the boundary, the closure and the Borel σ-field of the set D. We let $\overset{\mathcal{L}}{=}$ and $\overset{\mathcal{L}}{\to}$ denote the identity and convergence in law. We use the notation $\nabla_{\mathbf{x}}$ and $\mathrm{Hess}_{\mathbf{x}}$ for the gradient and the Hessian matrix with respect to the multivariate variable $\mathbf{x}$. We denote by $\mathbb{1}_d$ and 0_d, respectively, the vector with all unit-valued components and the zero-valued vector, both in $\mathbb{R}^d$.

The aim of the present work is to establish a novel framework of Monte Carlo integration over the d-dimensional standard simplex [2, 20, 27]:

$$\mu := \int_{\mathcal{X}_d} \Psi(\mathbf{x})d\mathbf{x}, \tag{2.1}$$

where we denote by $\mathcal{X}_d := \{\mathbf{x} \in [0,1]^d : \langle \mathbf{x}, \mathbb{1}_d \rangle \leq 1\}$ the standard d-simplex, and by Ψ a real-valued function on $\mathbb{R}^d$. For later use, we denote by $\mathcal{Y}_d := \{\mathbf{y} \in [0,1]^d : \langle \mathbf{y}, \mathbb{1}_d \rangle = 1\}$ the canonical $(d-1)$-simplex in $\mathbb{R}^d$. Among a few other ways of calling those, we follow [2] to call $\mathcal{X}_d$ the standard d-simplex and $\mathcal{Y}_d$ the canonical $(d-1)$-simplex.

In what follows, we develop theories and methodologies on the standard d-simplex $\mathcal{X}_d$ and the canonical $(d-1)$-simplex $\mathcal{Y}_d$ without paying particular attention to general simplices, for the reason that integration over general simplices can be reformulated as an integration over those standardized simplices $\mathcal{X}_d$ and $\mathcal{Y}_d$ through affine transformation. Suppose one is interested in the integral $\int_{\mathcal{S}} \Psi_0(\mathbf{s})d\mathbf{s}$, where v is a real-valued function on $\mathbb{R}^d$ and $\mathcal{S} := \{\theta_0 \mathbf{v}_0 + \cdots + \theta_d \mathbf{v}_d : \langle \boldsymbol{\theta}, \mathbb{1}_{d+1} \rangle \leq 1$ and $\theta_k \geq 0$ for all $k\}$ with affinely independent vectors $\mathbf{v}_0, \cdots, \mathbf{v}_d$ in $\mathbb{R}^d$. A general d-simplex $\mathcal{S}$ with the set of vertices $\{\mathbf{v}_k\}_{k \in \{0,1,\cdots,d\}}$ can be mapped onto the standard d-simplex $\mathcal{X}_d$ by the affine transformation $T(\mathbf{x}) := A\mathbf{x} + \mathbf{v}_0$ for $\mathbf{x} \in \mathcal{X}_d$, where $A := (\mathbf{v}_1 - \mathbf{v}_0, \mathbf{v}_2 - \mathbf{v}_0, \cdots, \mathbf{v}_d - \mathbf{v}_0)$ is an invertible matrix in $\mathbb{R}^{d \times d}$. Given the function v and the affine transform T, one can set $\Psi(\mathbf{x}) = \Psi_0(T(\mathbf{x}))$ to reformulate the integral $\int_{\mathcal{S}} \Psi_0(\mathbf{s})d\mathbf{s}$ over a general d-simplex as follows:

$$\int_{\mathcal{S}} \Psi_0(\mathbf{s})d\mathbf{s} = |\det(J_T)| \int_{\mathcal{X}_d} \Psi_0(T(\mathbf{x}))d\mathbf{x} = |\det(A)| \int_{\mathcal{X}_d} \Psi(\mathbf{x})d\mathbf{x}, \tag{2.2}$$

which is then based on the original form (2.1) over the standard d-simplex, where J_T denotes the Jacobian matrix of the affine transformation T. Here, we have $|\det(A)| > 0$ since the matrix A is, by definition, invertible. Clearly, the base can be chosen arbitrary from the vertices $\{\mathbf{v}_k\}_{k \in \{0,1,\cdots,d\}}$ as $T(\mathbf{x}) = A\mathbf{x} + \mathbf{v}_k$ with $A = (\mathbf{v}_1 - \mathbf{v}_k, \cdots, \mathbf{v}_{k-1} - \mathbf{v}_k, \mathbf{v}_{k+1} - \mathbf{v}_k, \cdots, \mathbf{v}_d - \mathbf{v}_k)$ for any $k \in \{0, 1, \cdots, d\}$. We refer the reader to, for instance, [27].

Example 2.1 For illustrative purposes, we consider an integration of the constant function $\Psi_0(\mathbf{s}) = 2$ over the 2-simplex $\mathcal{S}$ with vertices $\mathbf{v}_0 = (2, 3)$, $\mathbf{v}_1 = (1, 1)$, and $\mathbf{v}_2 = (-1, 2)$. With the affine transformation

$$T(\mathbf{x}) = A\mathbf{x} + \mathbf{v}_0 = (\mathbf{v}_1 - \mathbf{v}_0, \mathbf{v}_2 - \mathbf{v}_0)\mathbf{x} + \mathbf{v}_0 = \begin{bmatrix} -1 & -3 \\ -2 & -1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

the identity (2.2) reads

$$\int_{\mathcal{S}} \Psi_0(\mathbf{s}) d\mathbf{s} = |\det(J_T)| \int_{\mathcal{X}_2} \Psi_0(T(\mathbf{x})) d\mathbf{x} = 5 \, (2\mathrm{Leb}(\mathcal{X}_2)) = 5,$$

where we have applied $|\det(J_T)| = |\det(A)| = 5$ and $\mathrm{Leb}(\mathcal{X}_d) = 1/d!$ for all $d \in \mathbb{N}$. Through elementary geometry, the area of the triangle formed by $\mathbf{v}_0$, $\mathbf{v}_1$, and $\mathbf{v}_2$ is $5/2$, thus yielding $\int_{\mathcal{S}} \Psi_0(\mathbf{s}) d\mathbf{s} = 2 \times (5/2) = 5$ as well. □

Upon the interpretation of the integral (2.1) as the expectation of the random variable $\Psi(X)$, that is,

$$\mu = \int_{\mathcal{X}_d} \Psi(\mathbf{x}) d\mathbf{x} = \mathbb{E}[\Psi(X)], \tag{2.3}$$

where X is a uniform random vector on the standard d-simplex $\mathcal{X}_d$, we first develop a sampling technique of the uniform random vector X (Sect. 3), and then based on the the sampling technique, tailor theories on change of measure for accelerating the convergence of the central limit theorem

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} \Psi(X_k) - \mu\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathrm{Var}(\Psi(X_1))\right), \quad n \to +\infty, \tag{2.4}$$

by reducing the estimator variance $\mathrm{Var}(\Psi(X_1))$ by changing the Lebesgue measure $d\mathbf{x}$ in (2.3) (Sect. 4). In addition, an estimator with reduced variance is more efficient even for any finite sample size n, not only asymptotically as $n \to +\infty$. In short, a primary advantage of Monte Carlo integration over other deterministic methods is that its convergence rate $1/\sqrt{n}$ in the number of realizations is free of the problem dimension d on the basis of the central limit theorem (2.4). Our developments are thus not focused on very low dimensional problems, but keep high dimensional ones well within reach. In the context of numerical integration over simplices, this point is of particular significance as the problem dimension is usually at least 3.

3 Sampling on the Standard Simplex

We now begin by constructing a uniform sampling technique over the standard d-simplex $\mathcal{X}_d$. The proposed technique is built on the further decomposition of the original single integral or, equivalently, the expectation on a single uniform random vector X in (2.3) into the following double integral or, equivalently, the expectation on two independent random elements (Theorem 3.2):

$$\mu = \int_{\mathcal{X}_d} \Psi(\mathbf{x}) d\mathbf{x} = \int_0^1 \left[\int_{\mathcal{Y}_d} \Psi(v^{1/d}\mathbf{y}) d\mathbf{y} \right] dv = \mathbb{E}\left[\Psi(V^{1/d}Y) \right], \tag{3.1}$$

where V is a uniform random variable on $(0, 1)$ and Y is a uniform random vector on the canonical $(d-1)$-simplex $\mathcal{Y}_d$. That is, the inner integral with respect to $d\mathbf{y}$ represents the expectation on the uniform law on the canonical $(d-1)$-simplex $\mathcal{Y}_d$, whereas the outer integral with respect to dv is simply on the uniform law on $(0, 1)$. Due to the product form of the two Lebesgue measures, the corresponding two random elements V and Y are independent.

In what follows, as the canonical hyperplane, we refer to either the canonical $(d-1)$-simplex $\mathcal{Y}_d$ (as a section of the standard d-simplex $\mathcal{X}_d$) or a collection of (not necessarily uniform) realizations over the canonical simplex, depending on the context. Moreover, we call the scaling operation by the random variable $V^{1/d}$ the projection of the random vector Y on the canonical hyperplane.

3.1 The Dirichlet Law

The probability law on the canonical $(d-1)$-simplex $\mathcal{Y}_d$ is called the Dirichlet law if it admits a probability density function in the form

$$p(\mathbf{y}; \boldsymbol{\alpha}) := \frac{\Gamma(\sum_{k=1}^d \alpha_k)}{\prod_{k=1}^d \Gamma(\alpha_k)} \prod_{k=1}^d y_k^{\alpha_k - 1}, \tag{3.2}$$

for $\mathbf{y} := (y_1, \cdots, y_d) \in \mathcal{Y}_d$ and $\boldsymbol{\alpha} := (\alpha_1, \cdots, \alpha_d) \in (0, +\infty)^d$. We henceforth write $\mathrm{Dir}(\boldsymbol{\alpha})$ for the Dirichlet law with the probability density function (3.2) with parameter $\boldsymbol{\alpha}$, and refer to $\boldsymbol{\alpha}$ as the Dirichlet parameter, for the sake of convenience. It is well known that, among other ways (Remark 3.1), the random vector $Y \sim \mathrm{Dir}(\boldsymbol{\alpha})$ can be generated as

$$Y \stackrel{\mathcal{L}}{=} \left(\frac{M_1}{M_1 + M_2 + \cdots + M_d}, \frac{M_2}{M_1 + M_2 + \cdots + M_d}, \cdots, \frac{M_d}{M_1 + M_2 + \cdots + M_d} \right), \tag{3.3}$$

where $\{M_k\}_{k \in \{1, \cdots, d\}}$ is a sequence of mutually independent gamma random variables with $M_k \sim \mathrm{Gamma}(\alpha_k, 1)$ for $k \in \{1, \cdots, d\}$, each of which admits probability density function $x^{\alpha_k - 1} e^{-x}/\Gamma(\alpha_k)$ on $(0, +\infty)$. By setting $\boldsymbol{\alpha} = \mathbb{1}_d$, the Dirichlet law reduces to $\mathrm{Dir}(\mathbb{1}_d)$ where the probability density function (3.2) reduces to $p(\mathbf{y}; \mathbb{1}_d) = (d-1)!$, which represents the uniform law on the canonical $(d-1)$-simplex $\mathcal{Y}_d$ since the density is flat on the domain. Accordingly, the representation (3.3) with $\boldsymbol{\alpha} = \mathbb{1}_d$ provides a uniform sampling technique on the canonical hyperplane, as

$$\begin{aligned} Y &\stackrel{\mathcal{L}}{=} \left(\frac{E_1}{E_1 + E_2 + \cdots + E_d}, \cdots, \frac{E_d}{E_1 + E_2 + \cdots + E_d} \right) \\ &\stackrel{\mathcal{L}}{=} \left(\frac{\ln(U_1)}{\ln(U_1) + \cdots + \ln(U_d)}, cdots, \frac{\ln(U_d)}{\ln(U_1) + \cdots + \ln(U_d)} \right), \end{aligned} \tag{3.4}$$

where $\{E_k\}_{k \in \{1, \cdots, d\}}$ is now a sequence of iid standard exponential random variables, since $\mathrm{Gamma}(1, 1)$ is nothing but the standard exponential distribution. The second identity in law holds by employing the concept of inverse transform sampling $E_1 \stackrel{\mathcal{L}}{=} -\ln(1 - U_1) \stackrel{\mathcal{L}}{=} -\ln(U_1)$. We do not provide a full description of the Dirichlet law and related topics but refer the reader to, for instance, [1, 19, 23, 24] for detail.

Remark 3.1 The Dirichlet random vector can also be represented using beta or inverted beta (also called Type II beta) random variables. Though not employed in the present work, we provide a brief summary of those representations for the sake of completeness. The following identities in law hold true for the random vector $Y \sim \mathrm{Dir}(\boldsymbol{\alpha})$:

$$Y \stackrel{\mathcal{L}}{=} \left(B_1, B_2(1 - B_1), B_3(1 - B_1)(1 - B_2), \cdots, B_{d-1} \prod_{k=1}^{d-2}(1 - B_k), \prod_{k=1}^{d-1}(1 - B_k) \right)$$

$$\stackrel{\mathcal{L}}{=} \left(\frac{B_1'}{1 + B_1'}, \frac{B_2'}{(1 + B_1')(1 + B_2')}, \cdots, B_{d-1}' \prod_{k=1}^{d-1}\frac{1}{1 + B_k'}, \prod_{k=1}^{d-1}\frac{1}{1 + B_k'} \right),$$

where $\{B_k\}_{k\in\{1,\cdots,d-1\}}$ is a sequence of mutually independent beta random variables with $B_k \sim \mathrm{Beta}(\alpha_k, \sum_{j=k+1}^{d}\alpha_j)$, and $\{B_k'\}_{k\in\{1,\cdots,d-1\}}$ is a sequence of mutually independent inverted beta random variables with $B_k' \sim \mathrm{IBeta}(\alpha_k, \sum_{j=k+1}^{d}\alpha_j)$. On an implementation level, those two representations do not fully differ from each other as the inverted beta random variable is often generated using the beta random variable. Sampling based on beta or inverted beta random variables is generally considered less efficient than based on gamma random variables (3.3), for the reason that one often uses two independent gamma random variables to generate a single beta random variable. □

3.2 Projecting the Dirichlet Random Vector

The standard d-simplex $\mathcal{X}_d$ is the pyramidal region between the canonical hyperplane and the origin 0_d in the unit hypercube $(0, 1)^d$. Given this structure, we here develop a sampling technique on the standard simplex based on the projection of the canonical hyperplane (Sect. 3.1) towards the origin 0_d so that the standard d-simplex $\mathcal{X}_d$ is filled. The following result acts as the theoretical base for this goal, with its proofs deferred to the Appendix.

Theorem 3.2 *It holds that for every $c \in [0, 1]$ and $\mathbf{y} \in \mathcal{Y}_d$,*

$$\mathbb{P}(\langle \mathbf{y}, Z \rangle \leq c) = c^d,$$

if and only if $Z = (V^{1/d}, V^{1/d}, \cdots, V^{1/d})$, where V is a uniform random variable on $(0, 1)$.

With Theorem 3.2 in hand, we are ready to construct a uniform sampling technique on the standard d-simplex $\mathcal{X}_d$. Consider a (not necessarily standard) d-simplex $\mathcal{X}_d(c)$, defined by $\mathcal{X}_d(c) := \{\mathbf{x} \in [0, c]^d : \langle \mathbf{x}, \mathbb{1}_d \rangle \leq c\}$ for $c \in [0, 1]$. It is clear that $\mathcal{X}_d(c) \subseteq \mathcal{X}_d(1) = \mathcal{X}_d$ and $\mathrm{Leb}(\mathcal{X}_d(c))/\mathrm{Leb}(\mathcal{X}_d) = c^d$ for all $c \in [0, 1]$. Hence, a necessary condition for a random vector X in $\mathbb{R}^d$ to be uniformly distributed on the standard d-simplex $\mathcal{X}_d$ is the identity $\mathbb{P}(X \in \mathcal{X}_d(c)) = \mathbb{P}(\langle X, \mathbb{1}_d \rangle \leq c) = c^d$ for all $c \in [0, 1]$. Thanks to Theorem 3.2, this necessary condition is satisfied by setting $X = V^{1/d}Y$, where $V \sim U(0, 1)$ and where Y is a random vector on the canonical $(d-1)$-simplex $\mathcal{Y}_d$, since then $\langle Y, Z \rangle = V^{1/d}\langle Y, \mathbb{1}_d \rangle = V_{1/d}$ due to the constraint $\langle \mathbf{y}, \mathbb{1}_d \rangle = 1$ for all $\mathbf{y} \in \mathcal{Y}_d$. If, moreover, $Y \sim \mathrm{Dir}(\mathbb{1}_d)$, that is, uniformly distributed on the canonical hyperplane, the random vector $V^{1/d}Y$ (after the projection by the random variable $V^{1/d}$ towards the origin 0_d) is uniformly distributed on the standard d-simplex $\mathcal{X}_d$. We note that the support of the random vector $V^{1/d}Y$ is exactly the standard d-simplex $\mathcal{X}_d$, since every point on the canonical $(d-1)$-simplex $\mathcal{Y}_d$ is projected in a direction towards the origin 0_d by a scale between 0 and 1, rather than in a direction perpendicular to the canonical hyperplane. In other words, every single point on the canonical hyperplane ends up inside the standard d-simplex $\mathcal{X}_d$ after the projection, that is, no realizations are rejected.

For the reader's convenience, we summarize the developed uniform sampling technique on the standard d-simplex $\mathcal{X}_d$:

(I) Generate a random vector $Y \sim \mathrm{Dir}(\mathbb{1}_d)$ on the canonical $(d-1)$-simplex $\mathcal{Y}_d$.

(II) Generate a uniform random variable $V \sim U(0, 1)$.

(III) Return the random vector $V^{1/d} Y$.

Recall that step (I) can be implemented via the representation (3.4). It is worth stressing that this operation becomes no more complex for a higher problem dimension. Even for the case $d = 100$, the required operation remains elementary (generating 100 iid exponential random variables, and summing and dividing those). This uniform sampling technique can then play a central role in computing the rightmost integral (3.1) by standard Monte Carlo methods on the basis of the central limit theorem:

$$\sqrt{n} \left(\frac{1}{n} \sum_{k=1}^{n} \Psi(V_k^{1/d} Y_k) - \mu \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \text{Var}(\Psi(V_1^{1/d} Y_1)) \right), \tag{3.5}$$

as $n \to +\infty$, where $\{Y_k\}_{k\in\mathbb{N}}$ is a sequence of iid random vectors with common law $\text{Dir}(\mathbb{1}_d)$ and $\{V_k\}_{k\in\mathbb{N}}$ is a sequence of iid uniform random variables on $(0, 1)$.

3.3 Comparison with Existing Sampling Methods

For the sake of comparison and completeness, we describe some existing sampling methods on the standard d-simplex $\mathcal{X}_d$. We begin with an existing work [7] in a similar yet different manner (not projecting but shifting), as follows:

(A) Generate a Dirichlet random vector $Y \sim \text{Dir}(\mathbb{1}_d)$ on the canonical $(d-1)$-simplex $\mathcal{Y}_d$.

(B) Generate a uniform random variable $V \sim U(0, 1)$.

(C) If the random vector $Y - V \mathbb{1}_d/d$ lies inside the standard d-simplex $\mathcal{X}_d$, then accept and return it. If not, reject it and go back to (A).

This method is based on the parallel shifting of the canonical hyperplane Y by the random vector $V \mathbb{1}_d/d$ in the direction perpendicular to the canonical hyperplane, unlike the projection by the random variable V towards the origin 0_d of our procedure (I)-(III). The computing cost of (A)-(C) may look the same as (I)-(III) at first glance, whereas some portion of the sample is rejected at the step (C). In fact, it is known [7] that the acceptance rate here is only $1/d$, meaning that quite a large portion of the sample will be thrown out, unless $d = 1$, which is however too trivial to be relevant. (To be fair, we note that this method is not meant to sample from the simplex alone, but from a more complex pyramidal set whose base is the intersection of a simplex with the faces of a unit hypercube.) In Fig. 1, we plot results of 2000 iid runs for comparison between the two sampling methods. Although each of Fig. 1a and b looks like a uniform sample over the standard 3-simplex $\mathcal{X}_3$, there remain approximately only a third (668 in this particular experiment) of the 2000 points after parallel shifting by the vector $V \mathbb{1}_3/3$ (Fig. 1b), whereas all the 2000 points stay inside after projection by the scalar $V^{1/3}$ (Fig. 1a). Hence, the developed uniform sampling technique (Sect. 3.2) requires significantly less sampling cost.

Another known method is based on the so-called uniform spacings [4, Theorem 2.1]. Here, as the grid $S_k := U_{(k)} - U_{(k-1)}$ for $k \in \{1, \cdots, d+1\}$ with $U_{(0)} := 0$ and $U_{(d+1)} := 1$, where $U_{(1)} < \cdots < U_{(d)}$ denotes the order statistics of d iid uniform random variables on $(0, 1)$, the d-dimensional random vector $(S_1, \cdots, S_d)$ is uniformly distributed on the standard d-simplex $\mathcal{X}_d$. The computing cost required here consists of the generation of d iid uniform random variables, sorting of those d numbers (to construct the order statistics) and then subtraction d times (to construct the uniform spacings). Unless the problem dimension d is very high and an inefficient sorting algorithm is employed, the cost is typically lower than that of ours that consists more steps, such as the generation of $(d+1)$ uniform random

(a) Projection towards the origin $(V^{1/d}Y)$ (b) Perpendicular shifting $(Y - V\mathbb{1}_d/d)$
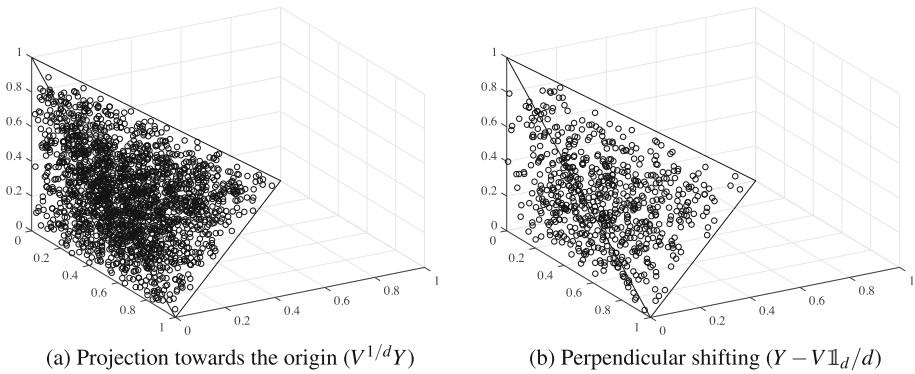
Fig. 1 Plots of results of 2000 iid runs of the two uniform sampling methods on the standard 3-simplex $\mathcal{X}_3$ ($d = 3$): **a** projecting the canonical 2-simplex by the random variable $V^{1/3}$ via (I)-(III), and **b** shifting the canonical 2-simplex by the random vector $V\mathbb{1}_3/3$ via (A)-(C)

variables ($\{U_k\}_{k \in \{1, \cdots, d\}}$ and V), taking a natural logarithm, adding and dividing d times (to construct the fractions (3.4)), and then multiplication of the canonical hyperplane by the powered scalar $V^{1/d}$ componentwise. We however do not employ this uniform sampling method for the reason that the required sorting procedure does not fit well to the framework of changing the underlying measure that we develop in what follows.

Yet another sampling method for the uniform law on the standard simplex $\mathcal{X}_d$ is available by removing any component of a Dirichlet random vector (say, $(Y_1, Y_2, \cdots, Y_{d+1})$) on the canonical d-simplex $\mathcal{Y}_{d+1}$ (for instance, resulting in $(Y_2, \cdots, Y_{d+1})$ if the first component Y_1 is chosen to be dropped), rather than treating the projection V differently from the canonical $(d-1)$-simplex $\mathcal{Y}_d$ on the basis of Theorem 3.2. The framework that we develop in what follows may well be somehow tailored to this method, while we do not go in this direction in the present work, because the canonical $(d-1)$-simplex $\mathcal{Y}_d$ alone can often be the object of interest (for instance, in copositive programming [6]) and thus change of measure on the canonical simplex $\mathcal{Y}_d$ independently from its projection V is beneficial.

4 Change of Measure on the Standard Simplex

Built on the uniform sampling technique (Sect. 3), we next develop theories on change of measure on the standard d-simplex $\mathcal{X}_d$ with a view towards variance reduction by importance sampling in Monte Carlo integration of the integral (3.1), that is,

$$\mu = \int_{\mathcal{X}_d} \Psi(\mathbf{x})d\mathbf{x} = \int_0^1 \left[\int_{\mathcal{Y}_d} \Psi(v^{1/d}\mathbf{y})d\mathbf{y} \right] dv = \mathbb{E}\left[\Psi(V^{1/d}Y) \right]. \qquad (4.1)$$

We note that it suffices to deal with those two measures separately thanks to the product form of the two measures $d\mathbf{y}$ and dv or, equivalently, independence of the corresponding two random elements V and Y in the representation (4.1). We stress that our approaches below do not yield an estimator with variance larger than that of the crude estimator (4.1), since the original measure is a member of the relevant parametric family.

4.1 Change of Measure on the Projection

We begin with change of the outer Lebesgue measure dv on the unit interval $(0, 1)$ in the representation (4.1), which plays a role of projecting realizations on the canonical $(d - 1)$-simplex $\mathcal{Y}_d$ towards the origin 0_d (Theorem 3.2). In order to form the proposal law, we employ the so-called bypass function [14, 15], a typical choice of which is an exponential bypass function f (and its tail mass F and inverse F^{-1}):

$$f(w; \lambda) = \lambda e^{-\lambda w}, \quad F(w; \lambda) = e^{-\lambda w}, \quad F^{-1}(v; \lambda) = -\frac{1}{\lambda} \ln(v),$$

$$F(F^{-1}(v; \lambda); \lambda_0) = v^{\lambda_0/\lambda}, \quad \Lambda_0 = (0, +\infty), \tag{4.2}$$

for $(v, w) \in (0, 1) \times (0, +\infty)$. With such a suitable bypass function chosen, the integral (4.1) can be parameterized further with $\lambda_0, \lambda \in \Lambda_0$, as follows:

$$
\begin{aligned}
\mu &= \int_0^1 \left[\int_{\mathcal{Y}_d} \Psi(v^{1/d}\mathbf{y}) d\mathbf{y} \right] dv \\
&= \int_D \left[\int_{\mathcal{Y}_d} \Psi\left((F(w; \lambda_0))^{1/d}\mathbf{y} \right) d\mathbf{y} \right] f(w; \lambda_0) dw \\
&= \int_D \frac{f(w; \lambda_0)}{f(w; \lambda)} \left[\int_{\mathcal{Y}_d} \Psi\left((F(w; \lambda_0))^{1/d}\mathbf{y} \right) d\mathbf{y} \right] f(w; \lambda) dw \\
&= \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{\mathcal{Y}_d} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d}\mathbf{y} \right) d\mathbf{y} \right] dv \tag{4.3} \\
&= \mathbb{E}\left[\frac{f(F^{-1}(V; \lambda); \lambda_0)}{f(F^{-1}(V; \lambda); \lambda)} \Psi\left((F(F^{-1}(V; \lambda); \lambda_0))^{1/d}Y \right) \right], \tag{4.4}
\end{aligned}
$$

where we have changed variables $v = F(w; \lambda_0)$ followed by $w = F^{-1}(v; \lambda)$ along the way and have placed square brackets inside to stress that only the outer integral is dealt with. In the expectation (4.4), it holds that $V \sim U(0, 1)$ and $Y \sim \text{Dir}(\mathbb{1}_d)$, corresponding to the two respective measures dv and $p(\mathbf{y}; \mathbb{1}_d)d\mathbf{y}$, and moreover that they are independent, thanks to the product form of those two measures in (4.3) even after changing variables.

In fact, instead of the exponential bypass function (4.2) (which we employ in all numerical illustrations (Figs. 2, 4 and 6), and all numerical examples in Sect. 5), one may come up with, for instance, a Gaussian bypass function $f(w; \lambda) = \phi(w - \lambda)$, with a little more computation required than (4.2), due to the lack of closed-form expressions of its distribution function Φ and inverse Φ^{-1}. We do not proceed in this direction but refer the reader to [15, Section 5] for details on other bypass functions. In short, the choice of bypass function is rather arbitrary as long as Assumption 4.1 below is satisfied.

Assumption 4.1 We choose in advance an open set $\Lambda_0 \subseteq \mathbb{R}$ with $\text{Leb}(\Lambda_0) > 0$, a family $\{f(\cdot; \lambda) : \lambda \in \Lambda_0\}$ of probability density functions on the domain $D(\subset \mathbb{R})$ and a family $\{F(\cdot; \lambda); \lambda \in \Lambda_0\}$ of functions on D in such a way that

(a) The support D of the probability density function $f(\cdot; \lambda)$ is open and independent of the parameter λ;
(b) For almost every $w \in D$ (with respect to dw), the function $f(w; \cdot)$ is twice continuously differentiable on Λ_0;
(c) For every $\lambda \in \Lambda_0$ and $B \in \mathcal{B}(D)$, it holds that $\int_D \mathbb{1}(F(w; \lambda) \in B) f(w; \lambda) dw = \text{Leb}(B)$;
(d) For every $\lambda \in \Lambda_0$, the inverse $F^{-1}(\cdot; \lambda)$ is well defined and continuous on $(0, 1)$;

(a) $\lambda = 1.0$ (b) $\lambda = 1.5$ (c) $\lambda = 0.5$
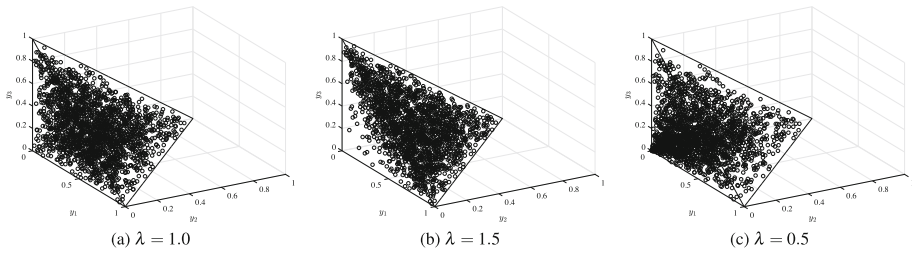
Fig. 2 Typical 2000 iid realizations of the random vector $(V^{\lambda_0/\lambda})^{1/d}Y$ (that is, the uniform random vector Y on the canonical hyperplane projected by the random variable $(V^{\lambda_0/\lambda})^{1/d}$ towards the origin) for 3 different values of the projection parameter λ with $\lambda_0 = 1.0$ fixed, resulting in **a** uniform sampling on the standard d-simplex $\mathcal{X}_d$ (corresponding to Fig. 1 a), and more mass towards **b** the canonical hyperplane and **c** the origin

(e) For every $\lambda \in \Lambda_0$ and $B \in \mathcal{B}(D)$, it holds that $\int_{(0,1)} \mathbb{1}(F^{-1}(v; \lambda) \in B)dv = \int_B f(w; \lambda)dw$;

(f) For almost every $w \in D$ (with respect to dw), it holds that $\lim_{n \to +\infty} \sup_{\lambda \in \partial K_n} f(w; \lambda) = 0$, where $\{K_n\}_{n \in \mathbb{N}}$ is an increasing sequence of compact subsets of the open set Λ_0, satisfying $\cup_{n \in \mathbb{N}} K_n = \Lambda_0$ and $K_n \subsetneq \text{int}(K_{n+1})$.

(g) For almost every $w \in D$ (with respect to dw), the reciprocal $1/f(w; \cdot)$ is twice continuously differentiable and convex on Λ_0.

Assumption 4.1 (c), (d), and (e) indicate that if $V \sim U(0, 1)$ and W is a random variable taking values in D with density $f(\cdot; \lambda)$, then it holds that $F(W; \lambda) \overset{\mathcal{L}}{=} V$, and $Z \overset{\mathcal{L}}{=} F^{-1}(V; \lambda)$, as we have already seen to get to the expressions (4.3) and (4.4). Hereafter, for the sake of convenience, we refer to λ as the projection parameter, as it is concerned with the projection $V^{1/d}$ in the representation (4.1). The conditions (f) and (g) are employed later in Theorems 4.2 and 4.3 for technical purposes.

Evidently, the exponential bypass function (4.2) satisfies Assumption 4.1 with the support $D = (0, +\infty)$. In particular, (f)-(g) hold true due to $\lim_{\lambda \to 0+} f(w; \lambda) = \lim_{\lambda \to +\infty} f(w; \lambda) = 0$ and $1/f(w; \lambda) = \lambda^{-1}e^{\lambda w}$ is twice continuously differentiable and is convex in λ on Λ_0. We provide Fig. 2 to visualize how the projection parameter λ can change the law of the random vector $(F(F^{-1}(V; \lambda); \lambda_0))^{1/3}Y$ on the standard 3-simplex $\mathcal{X}_3$ (because $d = 3$), with $V \sim U(0, 1)$, $Y \sim \text{Dir}(\mathbb{1}_3)$ and $\lambda_0 = 1.0$. We remark that $(F(F^{-1}(V; \lambda); \lambda_0))^{1/3}Y$ is the random vector inside the expectation (4.4) and reduces to $(V^{\lambda_0/\lambda})^{1/3}Y$ based on the exponential bypass function (4.2). In short, by wisely setting the value of the projection parameter λ, one may send the mass of the law (b) away, or (c) towards the origin. Clearly, the parameter set (a) $\lambda = \lambda_0$ reduces the law to the original uniform law on the standard 3-simplex $\mathcal{X}_3$, corresponding to the uniform sampling technique developed in Sect. 3.2.

4.2 Change of Measure on the Canonical Simplex

Now that change of the outer Lebesgue measure has been developed (Sect. 4.1), we next address the inner integral with respect to the Lebesgue measure $d\mathbf{y}$ on the canonical hyperplane in the double integral (4.3) in two distinct ways (Sects. 4.2.1 and 4.2.2).

4.2.1 Change of Measure Within the Dirichlet Law

The first framework we develop here is based on the interpretation of the inner Lebesgue measure $d\mathbf{y}$ on the canonical $(d-1)$-simplex $\mathcal{Y}_d$ as its identical form $p(\mathbf{y}; \mathbb{1}_d)d\mathbf{y}$, that is, an element $\mathrm{Dir}(\mathbb{1}_d)$ in the class of the Dirichlet law $\mathrm{Dir}(\boldsymbol{\alpha})$ for $\boldsymbol{\alpha} \in (0, +\infty)^d$. Further to the integral (4.3) with the projection parameter λ fixed in its domain Λ_0, the Dirichlet parameter $\boldsymbol{\alpha}$ can be incorporated by changing the inner measure $p(\mathbf{y}; \mathbb{1}_d)d\mathbf{y}$, in a similar spirit to [13], as follows:

$$
\begin{aligned}
\mu &= \int_0^1 \left[\int_{\mathcal{Y}_d} \Psi(v^{1/d}\mathbf{y})d\mathbf{y} \right] dv \\
&= \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{\mathcal{Y}_d} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d}\mathbf{y} \right) d\mathbf{y} \right] dv \\
&= \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{\mathcal{Y}_d} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d}\mathbf{y} \right) p(\mathbf{y}; \mathbb{1}_d)d\mathbf{y} \right] dv \\
&= \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{\mathcal{Y}_d} \frac{p(\mathbf{y}; \mathbb{1}_d)}{p(\mathbf{y}; \boldsymbol{\alpha})} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d}\mathbf{y} \right) p(\mathbf{y}; \boldsymbol{\alpha})d\mathbf{y} \right] dv
\end{aligned}
\tag{4.5}
$$

$$
= \mathbb{E}_{\boldsymbol{\alpha}} \left[\frac{f(F^{-1}(V; \lambda); \lambda_0)}{f(F^{-1}(V; \lambda); \lambda)} \frac{p(Y; \mathbb{1}_d)}{p(Y; \boldsymbol{\alpha})} \Psi\left((F(F^{-1}(V; \lambda); \lambda_0))^{1/d}Y \right) \right],
\tag{4.6}
$$

where the likelihood ratio $p(\cdot; \mathbb{1}_d)/p(\cdot; \boldsymbol{\alpha})$ is well defined over the canonical $(d-1)$-simplex $\mathcal{Y}_d$ for all $\boldsymbol{\alpha} \in (0, +\infty)^d$, since the Dirichlet law $\mathrm{Dir}(\boldsymbol{\alpha})$ has the support $\mathcal{Y}_d$ in common. In the representation (4.6), we have denoted by $\mathbb{E}_{\boldsymbol{\alpha}}$ the expectation under which the random vector Y follows the law $\mathrm{Dir}(\boldsymbol{\alpha})$ and remains independent of the random variable $V \sim U(0, 1)$.

We provide Fig. 3 to illustrate the flexibility of the law $\mathrm{Dir}(\boldsymbol{\alpha})$ with respect to the Dirichlet parameter $\boldsymbol{\alpha}$. That is, we here focus on the behavior of the random vector $Y \sim \mathrm{Dir}(\boldsymbol{\alpha})$ on the canonical hyperplane under the expectation operator $\mathbb{E}_{\boldsymbol{\alpha}}$ in (4.6). For each of 6 different values of the Dirichlet parameter $\boldsymbol{\alpha}$ with $d = 3$, we plot 2000 iid realizations from the law $\mathrm{Dir}(\boldsymbol{\alpha})$ on the canonical hyperplane by generating gamma random variables in accordance with the representation (3.3). The parameter set $\boldsymbol{\alpha} = (1.0, 1.0, 1.0)$ (Fig. 3a) corresponds to uniform sampling. Otherwise, the law $\mathrm{Dir}(\boldsymbol{\alpha})$ may tilt its mass towards (b) a vertex, and (c) an edge. Even more flexibly, the mass can be sent towards (d) two edges, and (e) all vertices. Finally, the mass can also stay (f) away from all vertices and edges, thus towards the center of the canonical hyperplane.

In addition, we demonstrate in Fig. 4 how the Dirichlet and projection parameters $(\boldsymbol{\alpha}, \lambda)$ together may alter the uniform law on the standard 3-simplex $\mathcal{X}_3$, by recycling the realizations of the law $\mathrm{Dir}(\boldsymbol{\alpha})$ on the canonical hyperplane of Fig. 3. The two parameters $(\boldsymbol{\alpha}, \lambda)$ together offer quite flexible means of tilting the uniform law, such as towards (a) A single surface, (b) The three vertices alone on the canonical hyperplane, and (c) A single bottom edge.

Now, as opposed to the insensitivity of the first moment (4.6) to the Dirichlet and projection parameters $(\boldsymbol{\alpha}, \lambda)$, its estimator variance (or, equivalently, the second moment since the first moment is invariant) is not independent of those parameters, as follows:
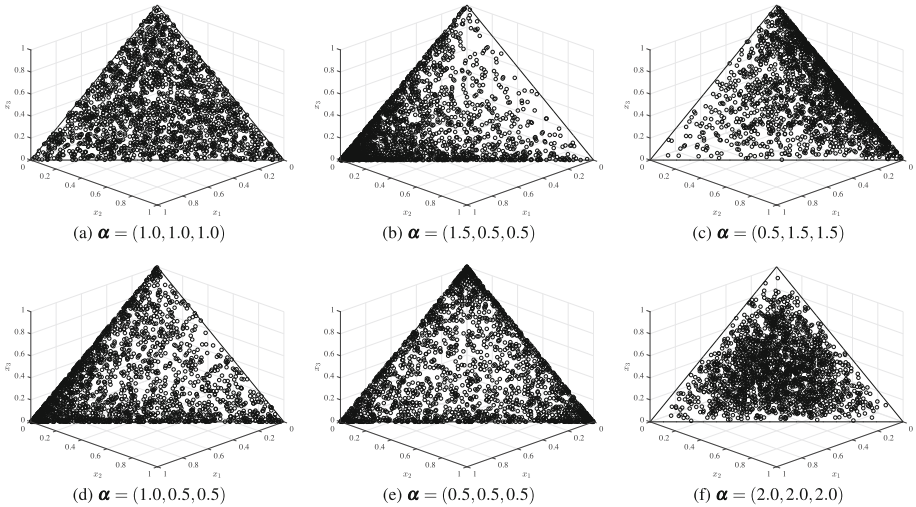
Fig. 3 Typical 2000 iid realizations from the law $\mathrm{Dir}(\boldsymbol{\alpha})$ on the canonical hyperplane for 6 different values of the Dirichlet parameter $\boldsymbol{\alpha}$, resulting in **a** uniform sampling, and more mass towards **b** a vertex, **c** an edge, **d** two edges simultaneously, **e** all vertices simultaneously, and **f** the center of the canonical hyperplane
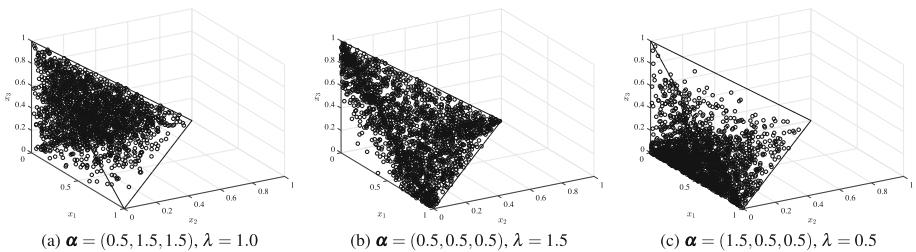
Fig. 4 Typical 2000 iid realizations of the random vector $(F(F^{-1}(V; \lambda); \lambda_0))^{1/d} Y$ on the standard d-simplex $\mathcal{X}_d$ in the expectation (4.6), that is, the random vector $Y \sim \mathrm{Dir}(\boldsymbol{\alpha})$ on the canonical hyperplane, projected by the random variable $(F(F^{-1}(V; \lambda); \lambda_0))^{1/d}$ towards the origin, for 3 different sets of the Dirichlet and projection parameters $(\boldsymbol{\alpha}, \lambda)$ with $\lambda_0 = 1.0$. More mass is present towards **a** a single surface, **b** all three (non-origin) vertices simultaneously, and **c** a single edge

$$
\mathrm{Var}_{\boldsymbol{\alpha}} \left(\frac{f(F^{-1}(V; \lambda); \lambda_0)}{f(F^{-1}(V; \lambda); \lambda)} \frac{p(Y; \mathbb{1}_d)}{p(Y; \boldsymbol{\alpha})} \Psi \left((F(F^{-1}(V; \lambda); \lambda_0))^{1/d} Y \right) \right)
$$

$$
= \int_0^1 \int_{\mathcal{Y}_d} \left| \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \frac{p(\mathbf{y}; \mathbb{1}_d)}{p(\mathbf{y}; \boldsymbol{\alpha})} \Psi \left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d} \mathbf{y} \right) \right|^2 p(\mathbf{y}; \boldsymbol{\alpha}) d\mathbf{y} dv - \mu^2
$$

$$
= \int_0^1 \int_{\mathcal{Y}_d} \frac{f(F^{-1}(v; \lambda_0); \lambda_0)}{f(F^{-1}(v; \lambda_0); \lambda)} \frac{p(\mathbf{y}; \mathbb{1}_d)}{p(\mathbf{y}; \boldsymbol{\alpha})} |\Psi(v^{1/d} \mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv - \mu^2
$$

$$
:= W_a(\boldsymbol{\alpha}, \lambda) - \mu^2 \tag{4.7}
$$

where we have changed variables $v = F(w; \lambda)$ and then $w = F^{-1}(v; \lambda_0)$ along the way. Here, we have denoted by $\mathrm{Var}_{\boldsymbol{\alpha}}$ the variance associated with the expectation $\mathbb{E}_{\boldsymbol{\alpha}}$ in (4.6). We henceforth call W_a the second moment function (of the estimator inside the expectation (4.6). We note that setting $(\boldsymbol{\alpha}, \lambda) = (\mathbb{1}_d, \lambda_0)$ yields $W_a(\mathbb{1}_d, \lambda_0) =$

$\int_0^1 \int_{\mathcal{Y}_d} |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv$, which is nothing but the crude second moment (crude, as no change of measure is employed).

Hence, the convergence of Monte Carlo integration of the representation (4.6) may be accelerated by choosing the Dirichlet and projection parameters $(\boldsymbol{\alpha}, \lambda)$ in such a wise way to yield a smaller estimator variance ($W_a(\boldsymbol{\alpha}, \lambda) < W_a(\mathbb{1}_d, \lambda_0)$) or, more ideally, by finding the minimizer ($\mathrm{argmin}_{(\boldsymbol{\alpha}, \lambda)} W_a(\boldsymbol{\alpha}, \lambda)$) over a suitable parameter domain. To describe a tractable structure of the second moment function $W_a(\boldsymbol{\alpha}, \lambda)$, we define the following two parameter sets based on the representation (4.7):

$$A(\lambda) := \mathrm{int} \bigcup_B \left\{ B \subseteq (0, +\infty)^d : \int_0^1 \int_{\mathcal{Y}_d} \frac{f(F^{-1}(v; \lambda_0); \lambda_0)}{f(F^{-1}(v; \lambda_0); \lambda)} \right.$$

$$\left. \times \sup_{\boldsymbol{\alpha} \in \overline{B}} \left[\max\{1, \|\nabla_{\boldsymbol{\alpha}}\|, \|\mathrm{Hess}_{\boldsymbol{\alpha}}\|\} \frac{p(\mathbf{y}; \mathbb{1}_d)}{p(\mathbf{y}; \boldsymbol{\alpha})} \right] |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv < +\infty \right\}, \quad (4.8)$$

$$\Lambda_a(\boldsymbol{\alpha}) := \mathrm{int} \bigcup_B \left\{ B \subseteq \Lambda_0 : \int_0^1 \int_{\mathcal{Y}_d} \sup_{\lambda \in \overline{B}} \left[\max\left\{1, |(\partial/\partial\lambda)|, |(\partial^2/\partial\lambda^2)|\right\} \frac{f(F^{-1}(v; \lambda_0); \lambda_0)}{f(F^{-1}(v; \lambda_0); \lambda)} \right] \right.$$

$$\left. \times \frac{p(\mathbf{y}; \mathbb{1}_d)}{p(\mathbf{y}; \boldsymbol{\alpha})} |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv < +\infty \right\}, \quad (4.9)$$

where we have written $\max\{1, \|\nabla_{\mathbf{x}}\|, \|\mathrm{Hess}_{\mathbf{x}}\|\} b(\mathbf{x}) := \max\{|b(\mathbf{x})|, \|\nabla b(\mathbf{x})\|, \|\mathrm{Hess}(b(\mathbf{x}))\|\}$ for the sake of brevity.

Theorem 4.2 *(i) Let $\lambda \in \Lambda_0$. If $\mathrm{Leb}(A(\lambda)) > 0$, then the matrix $\mathrm{Hess}_{\boldsymbol{\alpha}}(W_a(\boldsymbol{\alpha}, \lambda))$ is positive semi-definite on the domain $A(\lambda)$.*

(ii) Let $\boldsymbol{\alpha} \in (0, +\infty)^d$. If $\mathrm{Leb}(\Lambda_a(\boldsymbol{\alpha})) > 0$, then the function $W_a(\boldsymbol{\alpha}, \cdot)$ is twice continuously differentiable and strictly convex on the domain $\Lambda_a(\boldsymbol{\alpha})$. Moreover, if $\mathrm{int}\{\lambda \in \Lambda_0 : W_a(\boldsymbol{\alpha}, \lambda) < +\infty\} = \Lambda_a(\boldsymbol{\alpha})$, then $\lambda^*(\boldsymbol{\alpha}) := \mathrm{argmin}_{\lambda \in \Lambda_a(\boldsymbol{\alpha})} W_a(\boldsymbol{\alpha}, \lambda)$ exists uniquely in $\Lambda_a(\boldsymbol{\alpha})$ satisfying $(\partial/\partial\lambda)W_a(\boldsymbol{\alpha}, \lambda)|_{\lambda=\lambda^*(\boldsymbol{\alpha})} = 0$.

The second moment function $W_a(\boldsymbol{\alpha}, \lambda)$ is convex in each argument, whereas it does not seem jointly convex in the both Dirichlet and projection parameters $(\boldsymbol{\alpha}, \lambda)$. Similar partial convexity has been encountered in [17], where the second moment is convex in either importance sampling or control variates alone, but not in both. At any rate, the partial convexity provides a theoretical basis of searching the relevant domain for a minimizer in one parameter without worrying about local minima, with the other parameter fixed. We refer the reader to existing methodologies for parameter search, such as [13].

4.2.2 Change of Measure Through Inverse Transform

Apart from the first framework developed in Sect. 4.2.1, we next construct change of measure on the canonical hyperplane through inverse transform of the exponential law by the uniform law. Recall first the second identity in law in the representation (3.4), which provides an expression of the uniform random vector on the canonical $(d-1)$-simplex $\mathcal{Y}_d$ using iid uniform random variables on the unit hypercube $(0, 1)^d$, resulting in the following reformulation:

$$\mu = \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{\mathcal{Y}_d} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d}\mathbf{y}\right) p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} \right] dv$$

$$= \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{(0,1)^d} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d}h(\mathbf{u})\right) d\mathbf{u} \right] dv, \quad (4.10)$$

by change of variables via the function $h : (0, 1)^d \to \mathcal{Y}_d$, given by

$$h(\mathbf{u}) := \left(\frac{\ln(u_1)}{\sum_{k=1}^d \ln(u_k)}, \frac{\ln(u_2)}{\sum_{k=1}^d \ln(u_k)}, \cdots, \frac{\ln(u_d)}{\sum_{k=1}^d \ln(u_k)} \right), \quad \mathbf{u} := (u_1, u_2, \cdots, u_d) \in (0, 1)^d.$$

In this section, we change the Lebesgue measure $d\mathbf{u}$ on the unit hypercube $(0, 1)^d$ in the representation (4.10). To this end, we again employ the bypass function as in Sect. 4.1 without repeating similar details in order to avoid overloading the section. Here, in a similar manner to Assumption 4.1, we denote by $g(\cdot; \boldsymbol{\theta})$ the bypass function on the support D (this time, a subset of $\mathbb{R}^d$) for $\boldsymbol{\theta} \in \Theta_0$ and suppose that the function G and its inverse G^{-1} satisfy $\int_D \mathbb{1}(G(\mathbf{z}; \boldsymbol{\theta}) \in B) g(\mathbf{z}; \boldsymbol{\theta}) d\mathbf{z} = \mathrm{Leb}(B)$ and $\int_{(0,1)^d} \mathbb{1}(G^{-1}(\mathbf{u}; \boldsymbol{\theta}) \in B) d\mathbf{u} = \int_B g(\mathbf{z}; \boldsymbol{\theta}) d\mathbf{z}$ for all $\boldsymbol{\theta} \in \Theta_0$ and $B \in \mathcal{B}(D)$. As in Assumption 4.1, we assume the reciprocal $1/g(\mathbf{u}; \cdot)$ is convex for almost every $\mathbf{u} \in (0, 1)^d$. Hereafter, for the sake of convenience and clarity, we also refer to $\boldsymbol{\theta}$ as the bypass parameter. By introducing a suitable bypass function to (4.10), it holds, in a similar manner to (4.5), that

$$
\begin{aligned}
\mu &= \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{(0,1)^d} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d} h(\mathbf{u}) \right) d\mathbf{u} \right] dv, \\
&= \int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \left[\int_{(0,1)^d} \frac{g(G^{-1}(\mathbf{u}; \boldsymbol{\theta}); \boldsymbol{\theta}_0)}{g(G^{-1}(\mathbf{u}; \boldsymbol{\theta}); \boldsymbol{\theta})} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d} h(G(G^{-1}(\mathbf{u}; \boldsymbol{\theta}); \boldsymbol{\theta}_0)) \right) d\mathbf{u} \right] dv
\end{aligned}
\tag{4.11}
$$

$$
= \mathbb{E}\left[\frac{f(F^{-1}(V; \lambda); \lambda_0)}{f(F^{-1}(V; \lambda); \lambda)} \frac{g(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta}_0)}{g(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta})} \Psi\left((F(F^{-1}(V; \lambda); \lambda_0))^{1/d} h(G(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta}_0)) \right) \right],
\tag{4.12}
$$

where we have changed variables $\mathbf{u} = G(\mathbf{z}; \boldsymbol{\theta}_0)$ and then $\mathbf{z} = G^{-1}(\mathbf{u}; \boldsymbol{\theta})$ along the way. Let us stress that, unlike change of measure within the Dirichlet law (4.5), the proposal measure here in (4.11) remains the Lebesgue measure $d\mathbf{u}$ on the unit hypercube $(0, 1)^d$. That is, unlike the parameterized expectation operator $\mathbb{E}_{\boldsymbol{\alpha}}$ needs to be prepared in (4.6), Monte Carlo integration (4.12) can be performed on the uniform random vector U on $(0, 1)^d$, irrespective of the bypass parameter $\boldsymbol{\theta}$.

The estimator variance of Monte Carlo integration (4.12), however, depends on the bypass parameter $\boldsymbol{\theta}$, as it is given by:

$$
\begin{aligned}
&\mathrm{Var}\left(\frac{f(F^{-1}(V; \lambda); \lambda_0)}{f(F^{-1}(V; \lambda); \lambda)} \frac{g(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta}_0)}{g(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta})} \Psi\left((F(F^{-1}(V; \lambda); \lambda_0))^{1/d} h(G(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta}_0)) \right) \right) \\
&= \int_0^1 \int_{(0,1)^d} \left| \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \frac{g(G^{-1}(\mathbf{u}; \boldsymbol{\theta}); \boldsymbol{\theta}_0)}{g(G^{-1}(\mathbf{u}; \boldsymbol{\theta}); \boldsymbol{\theta})} \Psi\left((F(F^{-1}(v; \lambda); \lambda_0))^{1/d} h(G(G^{-1}(\mathbf{u}; \boldsymbol{\theta}); \boldsymbol{\theta}_0)) \right) \right|^2 d\mathbf{u} dv - \mu^2 \\
&= \int_0^1 \int_{(0,1)^d} \frac{f(F^{-1}(v; \lambda_0); \lambda_0)}{f(F^{-1}(v; \lambda_0); \lambda)} \frac{g(G^{-1}(\mathbf{u}; \boldsymbol{\theta}_0); \boldsymbol{\theta}_0)}{g(G^{-1}(\mathbf{u}; \boldsymbol{\theta}_0); \boldsymbol{\theta})} |\Psi(v^{1/d} h(\mathbf{u}))|^2 d\mathbf{u} dv - \mu^2 =: W_b(\boldsymbol{\theta}, \lambda) - \mu^2,
\end{aligned}
\tag{4.13}
$$

where we have changed variables $\mathbf{u} = G(\mathbf{z}; \boldsymbol{\theta})$ and then $\mathbf{z} = G^{-1}(\mathbf{u}; \boldsymbol{\theta}_0)$, as well as $v = F(w; \lambda)$ and then $w = F^{-1}(v; \lambda_0)$, without any intervention between the two lines along the way.

To describe in Theorem 4.3 below that the second moment function $W_b(\boldsymbol{\theta}, \lambda)$ is finite valued with a tractable structure, we define the parameter sets based on the representation (4.11):

$$
\Theta_1(\lambda) := \mathrm{int} \bigcup_B \left\{ B \subseteq \Theta_0 : \int_0^1 \int_{(0,1)^d} \frac{f(F^{-1}(v;\lambda_0);\lambda_0)}{f(F^{-1}(v;\lambda_0);\lambda)} \sup_{\theta \in \overline{B}} \left[\max\{1, \|\nabla_\theta\|, \|\mathrm{Hess}_\theta\|\} \frac{g(G^{-1}(\mathbf{u};\theta_0);\theta_0)}{g(G^{-1}(\mathbf{u};\theta_0);\theta)} \right] \right.
$$
$$
\left. \times |\Psi(v^{1/d} h(\mathbf{u}))|^2 d\mathbf{u} dv < +\infty \right\}, \tag{4.14}
$$

$$
\Lambda_b(\theta) := \mathrm{int} \bigcup_B \left\{ B \subseteq \Lambda_0 : \int_0^1 \int_{(0,1)^d} \sup_{\lambda \in \overline{B}} \left[\max\{1, |(\partial/\partial\lambda)|, |(\partial^2/\partial\lambda^2)|\} \frac{f(F^{-1}(v;\lambda_0);\lambda_0)}{f(F^{-1}(v;\lambda_0);\lambda)} \right] \right.
$$
$$
\left. \times \frac{g(G^{-1}(\mathbf{u};\theta_0);\theta_0)}{g(G^{-1}(\mathbf{u};\theta_0);\theta)} |\Psi(v^{1/d} h(\mathbf{u}))|^2 d\mathbf{u} dv < +\infty \right\}, \tag{4.15}
$$

where we have again employed the notation $\max\{1, \|\nabla_\mathbf{x}\|, \|\mathrm{Hess}_\mathbf{x}\|\} b(\mathbf{x}) := \max\{|b(\mathbf{x})|,$ $\|\nabla_\mathbf{x} b(\mathbf{x})\|, \|\mathrm{Hess}_\mathbf{x}(b(\mathbf{x}))\|\}$ for the sake of brevity. The next result states partial convexity of the second moment function $W_b(\theta, \lambda)$. As for $W_a(\alpha, \lambda)$ of Sect. 4.2.1, it does not seem convex jointly in the bypass and projection parameters (θ, λ).

Theorem 4.3 *(i) Let* $\lambda \in \Lambda_0$. *If* $\mathrm{Leb}(\Theta_1(\lambda)) > 0$, *then the function* $W_b(\cdot, \lambda)$ *is twice continuously differentiable and strictly convex on the domain* $\Theta_1(\lambda)$. *Moreover, if* $\mathrm{int}\{\theta \in \Theta_0 : W_b(\theta, \lambda) < +\infty\} = \Theta_1(\lambda)$, *then* $\theta^\star(\lambda) := \mathrm{argmin}_{\theta \in \Theta_1(\lambda)} W_b(\theta, \lambda)$ *exists uniquely in* $\Theta_1(\lambda)$ *satisfying* $\nabla_\theta W_b(\theta, \lambda)|_{\theta=\theta^\star(\lambda)} = 0_d$.

(ii) Let $\theta \in \Theta_0$. *If* $\mathrm{Leb}(\Lambda_b(\theta)) > 0$, *then the function* $W_b(\theta, \cdot)$ *is twice continuously differentiable and strictly convex on the domain* $\Lambda_b(\theta)$. *Moreover, if* $\mathrm{int}\{\lambda \in \Lambda_0 : W_b(\theta, \lambda) < +\infty\} = \Lambda_b(\theta)$, *then* $\lambda^\star(\theta) := \mathrm{argmin}_{\lambda \in \Lambda_b(\theta)} W_b(\theta, \lambda)$ *exists uniquely in* $\Lambda_b(\theta)$ *satisfying* $(\partial/\partial\lambda) W_b(\theta, \lambda)|_{\lambda=\lambda^\star(\theta)} = 0$.

We illustrate in Fig. 5 how the bypass parameter θ changes the law of the random vector $h(G(G^{-1}(U; \theta); \theta_0))$ inside the expectation (4.12) on the canonical hyperplane. As for the projection (Sect. 4.1), we employ the (multivariate) exponential bypass function with independent components, that is, for $\theta = (\theta_1, \cdots, \theta_d) \in (0, +\infty)^d$,

$$
g(\mathbf{z}; \theta) = \prod_{k=1}^d \theta_k e^{-\theta_k z_k}, \quad G(\mathbf{z}; \theta) = \left(e^{-\theta_1 z_1}, \cdots, e^{-\theta_d z_d} \right),
$$
$$
G^{-1}(\mathbf{u}; \theta) = \left(-\frac{1}{\theta_1} \ln(u_1), \cdots, -\frac{1}{\theta_d} \ln(u_d) \right), \tag{4.16}
$$

on the open support $(0, +\infty)^d$ irrespective of the bypass parameter θ. Then, with $\theta_0 = (\theta_0, \cdots, \theta_0)$ fixed, we have

$$
G(G^{-1}(\mathbf{u}; \theta); \theta_0) = \left(u_1^{\theta_0/\theta_1}, \cdots, u_d^{\theta_0/\theta_d} \right),
$$
$$
h(G(G^{-1}(\mathbf{u}; \theta); \theta_0)) = \left(\frac{(\theta_0/\theta_1) \ln(u_1)}{\sum_{k=1}^d (\theta_0/\theta_k) \ln(u_k)}, \cdots, \frac{(\theta_0/\theta_d) \ln(u_d)}{\sum_{k=1}^d (\theta_0/\theta_k) \ln(u_k)} \right). \tag{4.17}
$$

As is clear from the expressions (4.17), the law of the random vector $h(G(G^{-1}(U; \theta); \theta_0))$ depends solely on the (componentwise) ratio of θ and θ_0 and is moreover invariant up to a constant multiple of θ (for instance, θ, 2θ and 3θ do not yield different laws). In a similar manner to verifying Assumption 4.1 based on the univariate exponential bypass function (4.2), one can easily show that the multivariate exponential bypass function (4.16) and (4.17) satisfies a multivariate version of Assumption 4.1. We refer the reader to [15, Section 5.1] for more details. In short, by wisely choosing the bypass parameter θ, one may tilt the law towards (b) a vertex, and (c) an edge. Clearly, the parameter set $\theta = \theta_0$ (Fig. 5a) corresponds to uniform sampling.
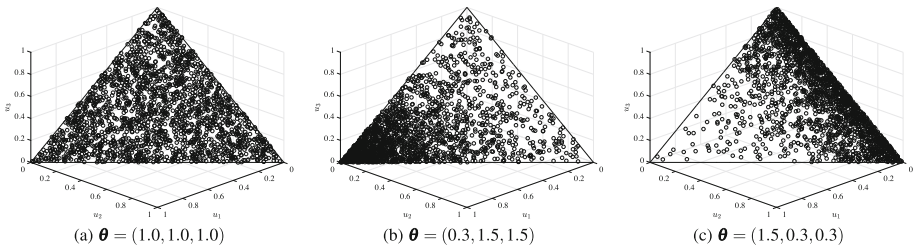
Fig. 5 Typical 2000 iid realizations of the random vector $h(G(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta}_0))$ inside the expectation (4.12) on the canonical hyperplane for 3 different values of the bypass parameter $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_0 = (1.0, 1.0, 1.0)$, resulting in **a** uniform sampling, and more mass towards **b** a vertex and **c** an edge
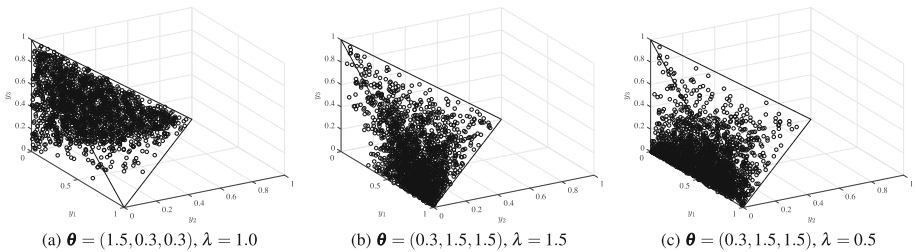
Fig. 6 Typical 2000 iid realizations of the random vector $(F(F^{-1}(V; \lambda); \lambda_0))^{1/d} h(G(G^{-1}(U; \boldsymbol{\theta}); \boldsymbol{\theta}_0))$ on the standard d-simplex $\mathcal{X}_d$ in the expectation (4.2), for 3 different sets of the bypass and projection parameters $(\boldsymbol{\theta}, \lambda)$ with $\boldsymbol{\theta}_0 = (1.0, 1.0, 1.0)$ and $\lambda_0 = 1.0$. More mass is present towards (a) A single surface, (b) A vertex, and (c) An edge

Next, in Fig. 6, we illustrate how the bypass and projection parameters $(\boldsymbol{\theta}, \lambda)$ together tilt the probability law on the standard d-simplex $\mathcal{X}_d$, by recycling the same sets of realizations used for plotting Fig. 5. The multivariate and univariate exponential bypass functions (4.17) and (4.2) are employed as $g(\cdot; \boldsymbol{\theta})$ and $f(\cdot; \lambda)$, respectively. By wisely adjusting the Dirichlet and projection parameters $(\boldsymbol{\theta}, \lambda)$, one can tilt the law towards (a) A single surface, (b) A vertex, and (c) An edge.

Now that the two ways have been constructed for changing the measure rather separately (Sects. 4.2.1 and 4.2.2), let us summarize and compare those, or more precisely, claim superiority of the latter (Sect. 4.2.2) over the former (Sect. 4.2.1) from a practical point of view. Both frameworks are based on the decomposition (4.1) (or, identically, (3.1)) of the base integral $\int_{\mathcal{X}_d} \Psi(\mathbf{x}) d\mathbf{x}$ onto the canonical hyperplane and its projection towards the origin. As for the projection in both two frameworks, among many possible ways of changing its Lebesgue measure (such as via the beta and triangular distributions), we have focused on a way via the bypass function (4.3) in common.

Hence, the two frameworks only differ as to how the measure on the canonical hyperplane is treated. In the former (Sect. 4.2.1), on the one hand, we interpret the uniform law $d\mathbf{y}$ as the law $\mathrm{Dir}(\mathbb{1}_d)$ (with probability density function $p(\mathbf{y}; \mathbb{1}_d)$) and then change it within the class of the Dirichlet law $\mathrm{Dir}(\boldsymbol{\alpha})$ (with probability density function $p(\mathbf{y}; \boldsymbol{\alpha})$). That is, every time the Dirichlet parameter is altered (say, from $\boldsymbol{\alpha}_{k-1}$ to $\boldsymbol{\alpha}_k$), the random number generator needs to be updated as well (from $\mathrm{Dir}(\boldsymbol{\alpha}_{k-1})$ to $\mathrm{Dir}(\boldsymbol{\alpha}_k)$). This point has been indicated by the parameterization of the expectation operator $\mathbb{E}_{\boldsymbol{\alpha}}$ in the expression (4.6). On the other hand, in the latter framework (Sect. 4.2.2), we again start with the interpretation of the uniform law $d\mathbf{y}$ as the law $\mathrm{Dir}(\mathbb{1}_d)$, but then further represent the law $\mathrm{Dir}(\mathbb{1}_d)$ by the

representation (3.4) using iid uniform random variables, which we again tilt via a bypass function (as for, yet separately from, the projection in Sect. 4.2). Given that both measures on the canonical hyperplane and its projection are changed via the bypass function in the latter framework, the random number generator remains unchanged even when the bypass and projection parameters $(\boldsymbol{\theta}, \lambda)$ are updated, as indicated by no parameterization in the expectation operator $\mathbb{E}$ in the expression (4.12), and moreover has only to be (a repeated use of) standard uniform $U(0, 1)$ throughout the experiment.

This invariance of the random number generator in the latter framework plays a crucial role in the context of adaptive implementation of Monte Carlo averaging and simultaneous parameter search, which we describe in brief. Based upon two sequences $\{V_k\}_{k \in \mathbb{N}}$ and $\{U_k\}_{k \in \mathbb{N}}$, respectively, of iid uniform random vectors on the unit hypercube $(0, 1)^d$ (not on the canonical hyperplane) and of iid uniform random variables on $(0, 1)$, the martingale strong law of large numbers (not the ordinary strong law of large numbers for iid random variables, like (3.5)) asserts the almost sure convergence

$$\frac{1}{n} \sum_{k=1}^{n} \frac{f(F^{-1}(V_k; \lambda_{k-1}); \lambda_0)}{f(F^{-1}(V_k; \lambda_{k-1}); \lambda_{k-1})} \frac{g(G^{-1}(U_k; \boldsymbol{\theta}_{k-1}); \boldsymbol{\theta}_0)}{g(G^{-1}(U_k; \boldsymbol{\theta}_{k-1}); \boldsymbol{\theta}_{k-1})}$$
$$\times \Psi\left((F(F^{-1}(V_k; \lambda_{k-1}); \lambda_0))^{1/d} h(G(G^{-1}(U_k; \boldsymbol{\theta}_{k-1}); \boldsymbol{\theta}_0)) \right) \to \mu, \quad (4.18)$$

as $n \to +\infty$, while the bypass and projection parameters $(\boldsymbol{\theta}, \lambda)$ keep updating, for instance, by stochastic approximation [16, 17] as gradient decent of the second moment function $W_b(\boldsymbol{\theta}, \lambda)$ through the expression (4.13):

$$\begin{cases} \boldsymbol{\theta}_k = \prod_{\Theta_1(\lambda_{k-1})} \left[\frac{f(F^{-1}(V_k; \lambda_0); \lambda_0)}{f(F^{-1}(V_k; \lambda_0); \lambda_{k-1})} \left(\nabla_{\boldsymbol{\theta}} \frac{g(G^{-1}(U_k; \boldsymbol{\theta}_0); \boldsymbol{\theta}_0)}{g(G^{-1}(U_k; \boldsymbol{\theta}_0); \boldsymbol{\theta})} \right) \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{k-1}} |\Psi(V_k^{1/d} h(U_k))|^2 \right], \\ \lambda_k = \prod_{\Lambda_b(\boldsymbol{\theta}_{k-1})} \left[\left(\frac{d}{d\lambda} \frac{f(F^{-1}(V_k; \lambda_0); \lambda_0)}{f(F^{-1}(V_k; \lambda_0); \lambda)} \right) \Big|_{\lambda = \lambda_{k-1}} \frac{g(G^{-1}(U_k; \boldsymbol{\theta}_0); \boldsymbol{\theta}_0)}{g(G^{-1}(U_k; \boldsymbol{\theta}_0); \boldsymbol{\theta}_{k-1})} |\Psi(V_k^{1/d} h(U_k))|^2 \right], \end{cases}$$
$$(4.19)$$

where we have denoted by $\prod_B[\mathbf{x}]$ the metric projection of the point $\mathbf{x}$ onto the set B. Upon implementation of the search algorithm (4.19), there remain a few points to be cleared, such as how to find and update the search domains $\Theta_1(\lambda_{k-1})$ and $\Lambda_b(\boldsymbol{\theta}_{k-1})$ in the metric projections, and whether or not the iteration converges in the absence of joint convexity of the second moment function in both parameters. We however do not go in the direction towards parameter search but leave the relevant topics as future research.

In summary, it is an advantage of the latter framework (Sect. 4.2.2), at least from the perspective of implementation, that the averaging operation (4.18) proceeds along the parameter updating (4.19), where common random sequences $\{V_k\}_{k \in \mathbb{N}}$ and $\{U_k\}_{k \in \mathbb{N}}$ can be applied to both lines (4.18) and (4.19). In contrast, such adaptive implementation is impossible in the former framework (Sect. 4.2.1), where the random number generator needs to be altered every time the Dirichlet parameter is updated.

5 Examples

In this section, we examine four problems (Examples 5.1, 5.2, 5.3 and 5.4) to demonstrate the effectiveness of the established Monte Carlo integration along with change of measure over simplices (as well as exemplify a practical problem in Example 5.5 for which the proposed method is not fully valid due to its infinite estimator variance). Although we have just claimed

practical superiority of one framework (Sect. 4.2.2) over the other (Sect. 4.2.1) in terms of the parameter search, we here do not pay attention to this point but focus on the effectiveness of two ways of changing the underlying measure given every relevant parameter has already been fixed. Recall that λ, $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ denote, respectively, the projection parameter (Sect. 4.1), the Dirichlet parameter (Sect. 4.2.1) and the bypass parameter (Sect. 4.2.2). We note that the parameter sets $\boldsymbol{\alpha} = (1.0, \cdots, 1.0)$ and $\boldsymbol{\theta} = (1.0, \cdots, 1.0)$ with $\lambda = 1.0$ recover the original crude estimator (4.1), presented in every example for comparison purposes.

Example 5.1 We start with a simple example to demonstrate that the expectations (4.6) and (4.12) yield the constant μ of ultimate interest, that is, changing the measure does not affect the value of an integral. To this end, consider an integral of the form $\int_{\mathcal{X}_d} h(\langle \mathbf{x}, \mathbb{1}_d \rangle) d\mathbf{x} (= \int_{\mathcal{X}_d} h(x_1 + \cdots + x_d) d\mathbf{x})$, where h is a continuously differentiable function on the unit interval $(0, 1)$, that is, $\Psi(\mathbf{x}) = h(\langle \mathbf{x}, \mathbb{1}_d \rangle)$ in the formulation (2.1). We have chosen this simple problem setting because this integral on the standard d-simplex $\mathcal{X}_d$ can be reformulated as an integral on the unit interval via the formula $\int_{\mathcal{X}_d} h(\langle \mathbf{x}, \mathbb{1}_d \rangle) d\mathbf{x} = \int_0^1 v^{d-1} h(v) dv / (d-1)!$, which provides a convenient means for numerical comparison. It is worth mentioning that this formula does not require the integrand to be polynomial.

Consider the integral of the non-polynomial function $\Psi(\mathbf{x}) = e^{x_1 + x_2 + x_3}$ on the standard 3-simplex $\mathcal{X}_3$:

$$\mu = \int_{\mathcal{X}_3} e^{x_1 + x_2 + x_3} d\mathbf{x} = \frac{1}{2} \int_0^1 z^2 e^z dz = \frac{e-2}{2} \approx 0.359141.$$

We present in Tables 1 and 2 the averages and standard deviations of 100 iid empirical means. Each of those 100 iid experiments is obtained by Monte Carlo integration of the integrand $e^{x_1 + x_2 + x_3}$ on the standard 3-simplex $\mathcal{X}_3$ based on 10^5 iid realizations, without approximation by polynomial, where the average and the standard deviation are computed in accordance with the formulas:

$$\frac{1}{100} \sum_{k_1=1}^{100} \frac{1}{10^5} \sum_{k_2=1}^{10^5} \Psi(X_{k_1, k_2}),$$

$$\left[\frac{1}{100-1} \sum_{k_1=1}^{100} \left(\frac{1}{10^5} \sum_{k_2=1}^{10^5} \Psi(X_{k_1, k_2}) - \frac{1}{100} \sum_{k=1}^{100} \frac{1}{10^5} \sum_{k_2=1}^{10^5} \Psi(X_{k, k_2}) \right)^2 \right]^{1/2}, \quad (5.1)$$

for an array of iid samples $\{X_{k_1, k_2}\}_{k_1 \in \{1, \cdots, 100\}, k_2 \in \{1, \cdots, 10^5\}}$ on $\mathcal{X}_3$. We note that the empirical standard deviations presented here are not for illustrating variance reduction but for validation of the proposed change of measure. In estimating the value $\int_{\mathcal{X}_3} e^{x_1 + x_2 + x_3} d\mathbf{x} \approx 0.359141$ by Monte Carlo integration, we examine both representations (4.6) and (4.12), each with 9 distinct sets of the parameters $(\boldsymbol{\alpha}, \lambda)$ and $(\boldsymbol{\theta}, \lambda)$ with $\boldsymbol{\theta}_0 = (1.0, 1.0, 1.0)$ and $\lambda_0 = 1.0$, some of which are taken from Figs. 4 and 6.

The 95%-confidence intervals for all the examined 18 parameter sets here contain the true value $\int_{\mathcal{X}_d} e^{x_1 + x_2 + x_3} d\mathbf{x} \approx 0.359141$. Among those 18 confidence intervals, the widest one is $[0.358318, 0.359540]$ when $\boldsymbol{\theta} = (1.5, 0.3, 0.3)$ and $\lambda = 0.5$, whereas the narrowest one is $[0.359080, 0.359160]$ under uniform sampling, that is, when $\boldsymbol{\alpha} = \boldsymbol{\theta} = (1.0, 1.0, 1.0)$ with $\lambda = 1.0$. We add that the 90%-confidence intervals, which are obviously even narrower, contain the true value in all the 18 parameter sets. □

Table 1 Averages of 100 iid empirical means in estimating the value $\int_{\mathcal{X}_3} e^{x_1+x_2+x_3} d\mathbf{x} \approx 0.359141$

λ	α			θ		
	(1.0, 1.0, 1.0)	(0.5, 1.5, 1.5)	(0.5, 0.5, 0.5)	(1.0, 1.0, 1.0)	(1.5, 0.3, 0.3)	(0.3, 1.5, 1.5)
1.0	0.359120	0.358993	0.359182	0.359120	0.358998	0.359020
1.5	0.359183	0.359055	0.359225	0.359183	0.359058	0.359089
0.5	0.359105	0.358848	0.359190	0.359105	0.358929	0.358869

Each of those 100 iid empirical means is obtained by Monte Carlo integration based on 10^5 iid realizations

Table 2 Standard deviations of the 100 iid empirical means corresponding to Table 1

λ	α			θ		
	(1.0, 1.0, 1.0)	(0.5, 1.5, 1.5)	(0.5, 0.5, 0.5)	(1.0, 1.0, 1.0)	(1.5, 0.3, 0.3)	(0.3, 1.5, 1.5)
1.0	0.000204	0.001953	0.000899	0.000204	0.002522	0.001946
1.5	0.000353	0.001896	0.000983	0.000353	0.002667	0.001927
0.5	0.000801	0.002499	0.001145	0.000801	0.003115	0.002437

Table 3 Averages of 100 iid empirical means $(\times 10^{-7})$ in estimating the value $\int_{\mathcal{X}_{10}} e^{x_1+\cdots+x_{10}} d\mathbf{x} \approx 6.86255 \times 10^{-7}$

λ	α			θ		
	(1.0, ⋯, 1.0)	(1.5, ⋯, 1.5)	(0.5, ⋯, 0.5)	(1.0, ⋯, 1.0)	(1.5, ⋯, 1.5)	(0.5, ⋯, 0.5)
1.0	6.86265	6.85804	6.86429	6.86265	6.86327	6.87051
1.5	6.86290	6.85936	6.86657	6.86290	6.86135	6.86844
0.5	6.86423	6.86205	6.86746	6.86423	6.87087	6.87977

Each of those 100 iid empirical means is obtained by Monte Carlo integration based on 10^5 iid realizations

Example 5.2 Next, to demonstrate the effectiveness of the developed frameworks in terms of problem dimension, we continue the problem setting of Example 5.1, but here in 10 dimensions,

$$\mu = \int_{\mathcal{X}_{10}} e^{x_1+\cdots+x_{10}} d\mathbf{x} = \frac{1}{9!} \int_0^1 z^9 e^z dz \approx 6.86255 \times 10^{-7}.$$

We present in Tables 3 and 4 the averages and standard deviations of 100 iid empirical means, again in accordance with the formulas (5.1). As in Example 5.1, each of the 100 iid empirical means is obtained by Monte Carlo integration of the integrand $e^{x_1+\cdots+x_{10}}$ on the standard 10-simplex $\mathcal{X}_{10}$ using 10^5 iid realizations, without approximation by polynomial. For simplicity, we set the vectors of identical components $(1.0, \cdots, 1.0)$, $(1.5, \cdots, 1.5)$, and $(0.5, \cdots, 0.5)$ to the Dirichlet $\boldsymbol{\alpha}$ and bypass parameters $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_0 = (1.0, \cdots, 1.0)$ and $\lambda_0 = 1.0$. We note that the empirical standard deviations presented here are not for illustrating variance reduction but for validation of the proposed change of measure.

Even in this high dimensional problem, the 95%-confidence intervals for all the examined 18 parameter sets contain the true value $\int_{\mathcal{X}_{10}} e^{x_1+\cdots+x_{10}} d\mathbf{x} \approx 6.86255 \times 10^{-7}$, with the widest one $[6.84440 \times 10^{-7}, 6.89734 \times 10^{-7}]$ and the narrowest one $[6.86231 \times 10^{-7}, 6.86300 \times 10^{-7}]$. We report that unlike $d = 3$ in Example 5.1, one of the 18 examined parameter

Table 4 Standard deviations ($\times 10^{-7}$) of the 100 iid empirical means corresponding to Table 3

	α			θ		
λ	$(1.0, \cdots, 1.0)$	$(1.5, \cdots, 1.5)$	$(0.5, \cdots, 0.5)$	$(1.0, \cdots, 1.0)$	$(1.5, \cdots, 1.5)$	$(0.5, \cdots, 0.5)$
1.0	0.00177	0.04206	0.04880	0.00177	0.10404	0.08839
1.5	0.01121	0.05048	0.05853	0.01121	0.09916	0.10726
0.5	0.01418	0.05248	0.05828	0.01418	0.13504	0.10165

sets fails slightly to capture the true value within its 90%-confidence interval $[6.86305 \times 10^{-7}, 6.98142 \times 10^{-7}]$, when $\theta = (0.5, \cdots, 0.5)$ and $\lambda = 0.5$. It is worth stressing that, as discussed in Sect. 3.2, the computing cost required here remains comparable to the case $d = 3$ in Example 5.1, as the essential difference only lies in the number of random elements involved while the operation remains elementary irrespective of dimension. □

Now that our Monte Carlo integration proves effective on non-polynomial integrands, we henceforth focus on polynomial integrands. We next demonstrate and compare the effectiveness of the two frameworks (Sects. 4.2.1 and 4.2.2). Namely, the primary objective from here on is to prove the potential of those two frameworks in reducing the estimator variance for the acceleration of Monte Carlo integration. As the integrand is polynomial, we make use of the well-known Stroud formula [31] for comparison purposes:

$$\int_{\mathcal{X}_d} x_1^{a_1} x_2^{a_2} \cdots x_d^{a_d} d\mathbf{x} = \frac{\prod_{k=1}^d a_k!}{(d + \sum_{k=1}^d a_k)!}, \tag{5.2}$$

provided that $\{a_k\}_{k \in \{1, \cdots, d\}}$ is a sequence of non-negative integers, as a direct consequence of the fact that the function (3.2) is a probability density function. Clearly, the integral $\int_{\mathcal{X}_d} \Psi(\mathbf{x}) d\mathbf{x}$ of a general polynomial Ψ can be expressed as a linear combination of the formula (5.2).

For each experiment in both numerical examples below, we present empirical variances on a single run of sample size 10^5, that is,

$$\frac{1}{10^5 - 1} \sum_{k_2=1}^{10^5} \left(\Psi(X_{1,k_2}) - \frac{1}{10^5} \sum_{k=1}^{10^5} \Psi(X_{1,k}) \right)^2, \tag{5.3}$$

which should not be confused with the empirical standard deviation (5.1) based on 100 iid runs, each of sample size 10^5 as in Examples 5.1 and 5.2. We present tables of empirical variances under various sets of the relevant parameters for illustration of the effectiveness of the proposed change of measure. For convenience, we attach the superscript $\star$ to the lowest variance among the examined parameter sets. We stress the the lowest variances are only the lowest among the examined parameter sets and are highly unlikely to be optimal over the parameter domain.

Example 5.3 Consider the following integral on the standard 3-simplex $\mathcal{X}_3$:

$$\mu = \int_{\mathcal{X}_3} (x_1^2 + x_2^2 + x_3^2) d\mathbf{x} = \frac{1}{20}.$$

We begin with the first framework (Sect. 4.2.1). In Table 5, we present empirical variances under 15 different sets of the Dirichlet and projection parameters (α, λ). The parameter set

Table 5 Empirical variances ($\times 10^{-2}$) on a single run of sample size 10^5 under 15 different sets of the Dirichlet and projection parameters ($\boldsymbol{\alpha}, \lambda$) with $\lambda_0 = 1.0$ when integrating the polynomial $\Psi(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2$ over the standard 3-simplex $\mathcal{X}_3$

λ	$\boldsymbol{\alpha}$				
	(1.0, 1.0, 1.0)	(1.5, 1.5, 1.5)	(2.0, 2.0, 2.0)	(0.8, 0.8, 0.8)	(0.5, 0.5, 0.5)
1.0	2.4411	15.016	86.198	1.9854	4.9765
0.5	9.8840	33.724	184.08	9.0837	14.009
1.5	0.7010	10.616	59.433	0.3216*	2.8595

The value 0.3216 with superscript $\star$ is the lowest among the 15 parameter sets

Table 6 Empirical variances ($\times 10^{-2}$) on a single run of sample size 10^5 under 15 different sets of the bypass and projection parameters ($\boldsymbol{\theta}, \lambda$) with $\boldsymbol{\theta}_0 = (1.0, 1.0, 1.0)$ and $\lambda_0 = 1.0$, when integrating the polynomial $\Psi(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2$ on the standard 3-simplex $\mathcal{X}_3$

λ	$\boldsymbol{\theta}$				
	(1.0, 1.0, 1.0)	(1.2, 0.8, 0.8)	(0.8, 1.2, 1.2)	(1.5, 0.3, 0.3)	(0.3, 1.5, 1.5)
1.0	2.4411	4.1522	4.2187	58.430	33.934
0.5	9.8840	12.718	12.935	101.80	65.343
1.5	0.7010*	2.1888	2.1932	48.644	26.530

The value 0.7010 with superscript $\star$ is the lowest among the 15 parameter sets. The values of the bypass parameter $\boldsymbol{\theta}$ here, except for $\boldsymbol{\theta} = (1.0, 1.0, 1.0)$, do not correspond to and are thus not comparative to the values of the Dirichlet parameter $\boldsymbol{\alpha}$ in Table 5

$\boldsymbol{\alpha} = (1.0, 1.0, 1.0)$ and $\lambda = 1.0$ corresponds to uniform sampling. With the projection parameter λ fixed, the experiment with $\boldsymbol{\alpha} = (0.8, 0.8, 0.8)$ presents the lowest estimator variance among the five examined Dirichlet parameters. This is a natural consequence as the Dirichlet law with $\boldsymbol{\alpha} = (0.8, 0.8, 0.8)$ has more mass equally towards all three vertices of the canonical hyperplane (in a similar manner to Fig. 3e), where the integrand $x_1^2 + x_2^2 + x_3^2$ returns its largest value 1. It is worth noting that tilting the law too much may cause a negative effect, as the vertices are the most important, but not only the places of importance. Here, the Dirichlet parameter $\boldsymbol{\alpha} = (0.5, 0.5, 0.5)$ returns a larger variance than $\boldsymbol{\alpha} = (0.8, 0.8, 0.8)$.

Next, we present in Table 6 empirical variances under 15 different sets of the bypass and projection parameters ($\boldsymbol{\theta}, \lambda$) in the second framework (Sect. 4.2.2). We note that the first columns in Tables 5 and 6 are identical because both represent the same set of experiments. With the projection parameter λ fixed, the estimator variance is unfortunately increased by altering the bypass parameter $\boldsymbol{\theta}$, since this change of measure does not seem capable of sending the mass towards all vertices of the canonical hyperplane at once.

The numerical results in Tables 5 and 6 together indicate that with the Dirichlet $\boldsymbol{\alpha}$ or bypass parameter $\boldsymbol{\theta}$ fixed, the lowest estimator variance is found when the projection parameter is $\lambda = 1.5$, by which realizations are projected back towards the canonical hyperplane (in the direction opposite to the origin by $\lambda = 0.5$), as illustrated in Figs. 2b, 4b and 6b. This phenomenon does not deserve a surprise because the integrand $x_1^2 + x_2^2 + x_3^2$ tends to vanish towards the origin. Although the first framework (Sect. 4.2.1) may look more effectiveness as it can deal with all three vertices at once, there is still potential for substantial reduction of the variance in the second framework (Sect. 4.2.2) in practice, particularly in the absence of prior knowledge on important regions. $\qquad\square$

Table 7 Empirical variances ($\times 10^{-2}$) on a single run of sample size 10^5 under 15 different sets of the Dirichlet and projection parameters $(\boldsymbol{\alpha}, \lambda)$ with $\lambda_0 = 1.0$ when integrating the polynomial $\Psi(\mathbf{x}) = (1 - x_1)^4$ over the standard 3-simplex $\mathcal{X}_3$

λ	$\boldsymbol{\alpha}$				
	(1.0, 1.0, 1.0)	(0.8, 1.0, 1.0)	(0.8, 1.2, 1.2)	(0.5, 1.5, 1.5)	(1.0, 0.5, 0.5)
1.0	8.9011	4.9058	3.9558*	11.932	39.733
0.5	14.424	8.9606	7.5102	14.004	52.630
1.5	30.302	24.822	23.144	43.736	83.488

The value 3.9558 with superscript $\star$ is the lowest among the 15 parameter sets

Table 8 Empirical variances ($\times 10^{-2}$) on a single run of sample size 10^5 under 15 different sets of the bypass and projection parameters $(\boldsymbol{\theta}, \lambda)$ with $\boldsymbol{\theta}_0 = (1.0, 1.0, 1.0)$ and $\lambda_0 = 1.0$ when integrating the polynomial $\Psi(\mathbf{x}) = (1 - x_1)^4$ over the standard 3-simplex $\mathcal{X}_3$

λ	$\boldsymbol{\theta}$				
	(1.0, 1.0, 1.0)	(1.0, 0.8, 0.8)	(1.5, 0.3, 0.3)	(1.6, 0.8, 0.8)	(0.3, 1.5, 1.5)
1.0	8.9429	8.4928	39.026	3.7432*	142.42
0.5	14.531	13.117	45.840	6.5266	196.61
1.5	29.425	31.721	74.556	27.666	208.66

The value 3.7432 with superscript $\star$ is the lowest among the 15 parameter sets. The values of the bypass parameter $\boldsymbol{\theta}$ here, except for $\boldsymbol{\theta} = (1.0, 1.0, 1.0)$, do not correspond to and are thus not comparative to the values of the Dirichlet parameter $\boldsymbol{\alpha}$ in Table 7

Example 5.4 Consider the following integral on the standard 3-simplex $\mathcal{X}_3$:

$$\mu = \int_{\mathcal{X}_3} (1 - x_1)^4 d\mathbf{x} = \frac{1}{14}.$$

Clearly, the hyperplane $x_1 = 0$ is the most important region (in a similar manner to Figs. 4a and 6a), because the integrand $(1 - x_1)^4$ attains its largest value 1 over the hyperplane.

We present in Tables 7 and 8 empirical variances under 15 different sets, respectively, in the first and second frameworks (Sects. 4.2.1 and 4.2.2). As can be predicted, the mass can be sent towards the plane $x_1 = 0$ in a systematic manner, by tilting the law towards the edge connecting the two vertices $(0, 1, 0)$ and $(0, 0, 1)$ (in a similar manner to Figs. 3c and 5c), with the uniformity of the projection retained (that is, $\lambda = 1.0$).

It is worth warning that tilting the law further in the same direction may not necessarily improve the estimator variance. For instance, by tilting the mass further (for instance, with $\boldsymbol{\alpha} = (0.5, 1.5, 1.5)$ in the same direction as $\boldsymbol{\alpha} = (0.8, 1.2, 1.2)$), we may end up with (in fact, substantially) larger variances. □

Example 5.5 We close the present study with an illustrative problem with an unbounded yet integrable function on a general twisted simplex. Consider the integrand $\Psi_0(\mathbf{s}) = \|\mathbf{s} - \mathbf{v}_0\|^{-2}$ on a general 3-simplex $\mathcal{S}$ ($d = 3$) with the four distinct vertices $\mathbf{v}_0 = (0, 10, 10)^\top$, $\mathbf{v}_1 = (0, 1, 0)^\top$, $\mathbf{v}_2 = (-0.5, 0, 0)^\top$ and $\mathbf{v}_3 = (0.5, 0, 0)^\top$, which occurs often in electrostatic problems. In accordance with the transform formula (2.2) with $\mathbf{s} \leftarrow A\mathbf{x} + \mathbf{v}_0$ and the projection

principle (Theorem 3.2), the integral can be reformulated, as follows:

$$\int_{\mathcal{S}} \Psi_0(\mathbf{s})d\mathbf{s} = |\det(A)| \int_{\mathcal{X}_3} \frac{1}{\|A\mathbf{x}\|^2}d\mathbf{x} = |\det(A)| \int_0^1 v^{-2/3}dv \int_{\mathcal{Y}_3} \frac{1}{\|A\mathbf{y}\|^2}d\mathbf{y} \approx 0.1551,$$
(5.4)

where the matrix A and its determinant are given by

$$A = [\mathbf{v}_1 - \mathbf{v}_0, \ \mathbf{v}_2 - \mathbf{v}_0, \ \mathbf{v}_3 - \mathbf{v}_0] = \begin{bmatrix} 0 & -0.5 & 0.5 \\ -9 & -10 & -10 \\ -10 & -10 & -10 \end{bmatrix}, \quad |\det(A)| = 10.$$

As far as Monte Carlo integration is concerned, there is an issue of infinite estimator variance due to singularity at the origin, as follows:

$$\int_{\mathcal{S}} (\Psi_0(\mathbf{s}))^2 d\mathbf{s} = |\det(A)| \int_{\mathcal{X}_3} \frac{1}{\|A\mathbf{x}\|^4}d\mathbf{x} = |\det(A)| \int_0^1 v^{-4/3}dv \int_{\mathcal{Y}_3} \frac{1}{\|A\mathbf{y}\|^4}d\mathbf{y} = +\infty.$$
(5.5)

In other words, the strong law of large numbers remains true for Monte Carlo integration (2.3) due to a finite first moment (5.4), whereas the central limit theorem (2.4) fails to hold due to infinite variance (5.5). In the context of this particular example, the so-called partial averaging can serve as a remedy, as soon as noticed. That is, in the decomposition (5.4), the first integral $\int_0^1 v^{-2/3}dv (= 3)$, which is the source of infinite variance, requires no numerical approximation at all. It then remains to implement Monte Carlo integration for the second integral $\int_{\mathcal{Y}_3} \|A\mathbf{y}\|^{-2}d\mathbf{y}$ on the canonical simplex $\mathcal{Y}_3$.

Before continuing, let us briefly touch on the transform using a different base (from $\mathbf{v}_0$), for instance, $\mathbf{s} \leftarrow B\mathbf{x} + \mathbf{v}_1$ with $B = [\mathbf{v}_0 - \mathbf{v}_1, \ \mathbf{v}_2 - \mathbf{v}_1, \ \mathbf{v}_3 - \mathbf{v}_1]$, with which the second moment can be reformulated as $\int_{\mathcal{S}} (\Psi_0(\mathbf{s}))^2 d\mathbf{s} = |\det(B)| \int_{\mathcal{X}_3} \|B(\mathbf{x} - (1, 0, 0)^\top)\|^{-4}d\mathbf{x}$. It is easy to observe non-integrability due to a quartic explosion towards the vertex $(1, 0, 0)^\top$ (for instance, via a further transform on $\mathcal{X}_3$ using the vertex $(1, 0, 0)^\top$), in a similar manner to the explosion towards the origin in (5.5). Hence, we do not adopt such other transforms but proceed with the original formulation (5.4) and (5.5) based on the transform using the base $\mathbf{v}_0$.

To closely look into infinite variance, observe that the probability ϵ on the non-standard simplex $\mathcal{X}_d(\epsilon^{1/d})$, that contains the origin, causes explosion of the second moment as slowly as $\int_{\mathcal{X}_3 \setminus \mathcal{X}_3(\epsilon^{1/3})} \|A\mathbf{x}\|^{-4}d\mathbf{x} = \int_{\epsilon^{1/3}}^1 v^{-4/3}dv \int_{\mathcal{Y}_3} \|A\mathbf{y}\|^{-4}d\mathbf{y} \sim C\epsilon^{-1/9}$, as $\epsilon \to 0+$, with $C = 3 \int_{\mathcal{Y}_3} \|A\mathbf{y}\|^{-4}d\mathbf{y}$. Hence, on the one hand, a majority (more precisely, 9 out of 10) of the crude running averages (Fig. 7a) look fairly convergent in Monte Carlo integration for $|\det(A)| \int_{\mathcal{X}_3} \|A\mathbf{x}\|^{-2}d\mathbf{x}$. Those stable trajectories are necessarily underestimating the value 0.1551, because most uniform realizations $\{X_k\}_{k \in \mathbb{N}}$ generated on $\mathcal{X}_3$ are away from the origin. On the other hand, one trajectory has exhibited an explosive behavior on its way, caused by a single realization very close to the origin. In general practice, a typical experiment cannot detect infinite variance explicitly but sends implicit signals at best in the form of extreme fluctuations, like this bumpy trajectory.

In reality, even if the theoretical variance is infinite, Monte Carlo integration is often still implementable without a serious issue, since the empirical variance (such as (5.1) and (5.3)) is necessarily finite. With this point in mind, we have searched for the minimizer by numerical approximation as $\lambda^\star(\theta^\star) = 0.3333$ and $\theta^\star(\lambda^\star) = (1.037, 1.048, 1.043)^\top$, based on the representation (4.13). In short, with the parameter value $\lambda^\star(\theta^\star) = 0.3333$ (and $\lambda_0 = 1$),
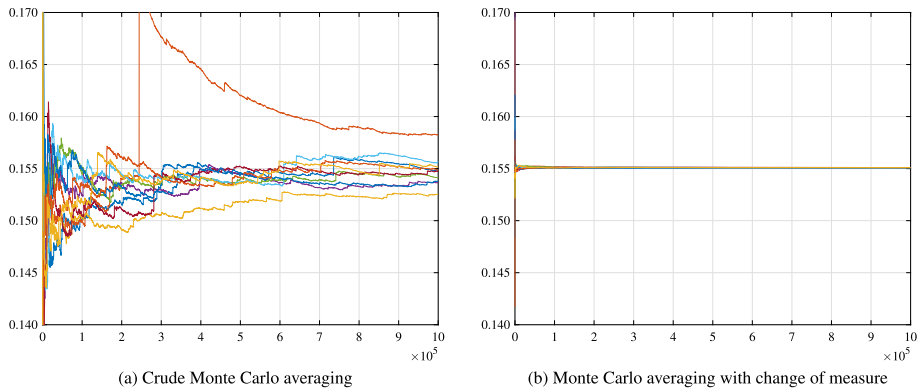
Fig. 7 Typical 10 iid trajectories of the running average (that is, $n^{-1} \sum_{k=1}^{n} \Psi(X_k)$ against n) **a** without and **b** with change of measure

the (transformed) projection $V^{\lambda_0/\lambda}$ (according to (4.2)) tends to stay away from the origin. The minimizer $\theta^\star(\lambda^\star)$ here is fairly close to $\theta_0 = \mathbb{1}_3$ with no surprise, because Monte Carlo integration for the integral $\int_{\mathcal{Y}_3} \|Ay\|^{-4} dy$ is extremely stable without a variance reduction technique. In Fig. 7b, we have presented 10 typical trajectories of the running average based on the expression (4.12) with the parameters $\lambda^\star(\theta^\star)$ and $\theta^\star(\lambda^\star)$ applied form the outset of the experiment. To be compatible with the numerical results presented in the previous examples, we have also run 100 iid experiments, each of sample size 10^5, from which we have obtained empirical mean 0.1548 and variance 3.009×10^{-5} for the crude implementation (in line with Fig. 7a), while mean 0.1551 and variance 1.115×10^{-9} for change of measure with the parameters $\lambda^\star(\theta^\star)$ and $\theta^\star(\lambda^\star)$ (as for Fig. 7b).

The proposed method has significantly stabilized Monte Carlo integration by roughly a factor of 27000 in terms of of the empirical variances (as the ratio $(3.009 \times 10^{-5})/(1.115 \times 10^{-9})$)), despite bringing up the empirical variance may seem a little awkward when the theoretical variance is originally infinite. On the contrary, interestingly, the present framework, involving a change of measure on the projection via the exponential bypass function (4.2), has effectively mitigated the issue of infinite variance associated with the power-law explosion (5.5), in fact, resulting in the induction of finite theoretical variance. That is to say, with the exponential bypass function (4.2), the second moment $W_b(\theta, \lambda)$ defined in (4.13) can be expressed (with $\theta = \theta_0$ fixed for the sake of simplicity), as follows:

$$W_b(\theta_0, \lambda) = |\det(A)| \int_0^1 \frac{f(F^{-1}(v; \lambda_0); \lambda_0)}{f(F^{-1}(v; \lambda_0); \lambda)} v^{-4/d} dv \int_{(0,1)^d} \frac{1}{|h(\mathbf{u})|^4} d\mathbf{u}$$

$$= |\det(A)| \int_0^1 \frac{\lambda_0}{\lambda} v^{1-4/d-\lambda/\lambda_0} dv \int_{(0,1)^d} \frac{1}{|h(\mathbf{u})|^4} d\mathbf{u},$$

where the projection term reduces to

$$\int_0^1 \frac{\lambda_0}{\lambda} v^{1-4/d-\lambda/\lambda_0} dv = \frac{1}{\lambda/\lambda_0(2 - 4/d - \lambda/\lambda_0)} \geq \frac{1}{(1 - 2/d)^2}, \qquad (5.6)$$

with the minimum $(1 - 2/d)^{-2}$ attained uniquely when $\lambda/\lambda_0 = 1 - 2/d$. In the present context with $d = 3$ and $\lambda_0 = 1$, all the above indicates that change of measure with the exponential bypass function (4.2) has lead the second moment $W_b(\theta_0, \lambda)$ to be finite-valued

with $\lambda \in (0, 2/3)$. No change of measure (that is, $\lambda = 1$) is outside this interval and thus has originally resulted in infinite variance (5.5). As a consequence, the parameter search, such as (4.19), would cause no trouble as soon as the parameter λ gets in the open interval $(0, 2/3)$, not because the empirical variance is necessarily finite no matter what, but because the theoretical variance is indeed finite there.

Even further, optimal change of measure, that is, with $\lambda = 1 - 2/d = 1/3(= \lambda^\star(\theta^\star))$ in line with (5.6), has resulted in the so-called perfect important sampling in the decomposition form (5.4). Here, the integral on the projection has been transformed, as follows:

$$\int_0^1 \frac{f(F^{-1}(v; \lambda); \lambda_0)}{f(F^{-1}(v; \lambda); \lambda)} \frac{1}{(F(F^{-1}(v; \lambda); \lambda_0))^{2/d}} dv = \int_0^1 \frac{\lambda_0}{\lambda} v^{\lambda_0/\lambda - 1} \frac{1}{(v^{\lambda_0/\lambda})^{2/d}} dv = 3 \int_0^1 dv,$$

where the rightmost integral corresponds to Monte Carlo integration of constant 1 with respect to the projection dv, meaning that the infinite variance in its original form has now been (not quite reduced but) completely eliminated. The trajectories presented in Fig. 7b are thus extremely stable with no surprise, since Monte Carlo integration for the remaining integral $\int_{\mathcal{Y}_3} \|A\mathbf{y}\|^{-2} d\mathbf{y}$ is intrinsically easy.

As such, the proposed framework has showcased its capability to transform an initially infinite variance into a finite one. Furthermore, it excels in identifying perfect importance sampling, even in cases where one might overlook such specific structures beforehand. Although the achieved success may seem problem-dependent at first glance, let us stress its significant value because this kind of problem settings are typical (and even central) in various fields of application. □

6 Concluding Remarks

In this paper, we have established a novel framework of Monte Carlo integration over simplices, from sampling to variance reduction. We have first developed a uniform sampling technique over the standard simplex. This technique is built on the decomposition of the uniform law on the standard simplex into the uniform law on the canonical hyperplane and its projection towards the origin in the form of two independent random elements. Its implementation is quite simple and wastes no computing cost, unlike acceptance-rejection sampling. We have next constructed theories on change of measure in integration over simplices with a view towards variance reduction by importance sampling. For the projection, we have employed the so-called bypass function to change its uniform law. To address the canonical hyperplane, we have developed two ways of changing the measure. One is to stay within the Dirichlet law, while the other is, as for the projection, the bypass function on the uniform random variables appearing in a representation of the Dirichlet random vector. Throughout, we have provided figures and numerical examples to support our theoretical developments, as well as to claim great potential of the proposed framework of sampling and acceleration of Monte Carlo integration over simplices.

We close this study by highlighting future research directions. As described in Remark 3.1 and Sect. 3.3, there exist other methods for generating the uniform sample on the simplices $\mathcal{X}_d$ and $\mathcal{Y}_d$, for which the change of measure can also be developed and investigated as appropriate. As discussed towards the end of Sect. 4, the second framework (changing the measure via the bypass function on both canonical simplex and projection) has further potential for adaptive implementation of Monte Carlo averaging and parameter search on common random elements. Other types of variance reduction techniques, such as antithetic variates, control variates [17] and stratified sampling [29], should be effective, to a large extent, in the context

of Monte Carlo integration over simplices, even under a batching procedure [28]. Finally, it would certainly be worthwhile to conduct an exhaustive numerical study in a wide variety of relevant problem settings in application. Those topics would be interesting future directions of research deserving of their own separate investigation.

Data availability The data that support the findings of this study are available upon reasonable request from the authors.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

A Proofs

We collect all proofs here. Complete proofs entail rather lengthy arguments of somewhat routine nature. To avoid overloading the Appendix, we skip non-essential details in some instances. In what follows, we denote by $A^{\otimes 2}$ the outer product of the matrix A.

Proof of Theorem 3.2 Note that $\langle \mathbf{y}, \mathbb{1}_d \rangle = 1$ since $\mathbf{y} \in \mathcal{Y}_d$. First, suppose $\mathbb{P}(\langle \mathbf{y}, Z \rangle \leq c) = c^d$ for all $c \in [0, 1]$ and $\mathbf{y} \in \mathcal{Y}_d$. Then, for every $c \in [0, 1]$ and $\mathbf{y} \in \mathcal{Y}_d$, we have

$$c^d = \mathbb{P}(\langle \mathbf{y}, Z \rangle \leq c) = \mathbb{P}(y_1 Z_1 + y_2 Z_2 + \cdots + y_d Z_d \leq c)$$
$$= \mathbb{P}(y_1(Z_1 - Z_d) + \cdots + y_{d-1}(Z_{d-1} - Z_d) + Z_d \leq c), \tag{A.1}$$

where the last equality holds by the constraint $\langle \mathbf{y}, \mathbb{1}_d \rangle = 1$. This asserts the identity in law $\mathcal{L}(y_1(Z_1 - Z_d) + \cdots + y_{d-1}(Z_{d-1} - Z_d) + Z_d) = \mathcal{L}(V^{1/d})$, since $\mathbb{P}(V^{1/d} \leq c) = \mathbb{P}(V \leq c^d) = c^d$ if $V \sim U(0, 1)$. We next show that $Z = (V^{1/d}, V^{1/d}, \cdots, V^{1/d})$ is a necessary condition for this identity in law to hold true, by considering the special case where $y_k = 1$ for some $k \in \{1, \cdots, d\}$. It suffices to consider two cases. First, if $k = d$ and thus $\mathbf{y} = (0, \cdots, 0, 1)$, it follows directly that $\mathcal{L}(Z_d) = \mathcal{L}(V^{1/d})$. Next, if $k \in \{1, \cdots, d-1\}$, we have $\mathcal{L}(y_k(Z_k - Z_d) + Z_d) = \mathcal{L}(Z_k) = \mathcal{L}(V^{1/d})$. Therefore, it is necessary to have $Z_k \equiv V^{1/d}$ for the identity (A.1) to hold true for all $\mathbf{y} \in \mathcal{Y}_d$.

Conversely, suppose $V \sim U(0, 1)$ and $Z = (V^{1/d}, V^{1/d}, \cdots, V^{1/d})$. Then, for every $c \in [0, 1]$ and $\mathbf{y} \in \mathcal{Y}_d$, we have

$$\mathbb{P}(\langle \mathbf{y}, Z \rangle \leq c) = \mathbb{P}\left(V^{1/d}\langle \mathbf{y}, \mathbb{1}_d \rangle \leq c\right) = \mathbb{P}\left(V^{1/d} \leq c\right) = c^d,$$

where the second equality holds due to the constraint $\langle \mathbf{y}, \mathbb{1}_d \rangle = 1$. The proof is now complete.
□

Proof We omit the proof of Theorem 4.3, as the results follow in a similar manner to Theorem 4.2 (ii). For relevant derivations, we refer the reader to [14–16] as well. In what follows, we prove Theorem 4.2 with the aid of the notation

$$M(v; \lambda) := \frac{f(F^{-1}(v; \lambda_0); \lambda_0)}{f(F^{-1}(v; \lambda_0); \lambda)}, \quad H(\mathbf{y}; \boldsymbol{\alpha}) := \frac{p(\mathbf{y}; \mathbb{1}_d)}{p(\mathbf{y}; \boldsymbol{\alpha})}.$$

(i) For $\mathbf{y} := (y_1, \cdots, y_d) \in \mathcal{Y}_d$ and $\boldsymbol{\alpha} \in (0, +\infty)^d$, define $Q_a(\mathbf{y}; \boldsymbol{\alpha})$ and $Q_b(\boldsymbol{\alpha})$ in $\mathbb{R}^{d \times d}$ by

$$Q_a(\mathbf{y}; \boldsymbol{\alpha}) := \begin{bmatrix} \psi(\alpha_1) - \psi(\langle \mathbb{1}_d, \boldsymbol{\alpha} \rangle) - \ln(y_1) \\ \vdots \\ \psi(\alpha_d) - \psi(\langle \mathbb{1}_d, \boldsymbol{\alpha} \rangle) - \ln(y_d) \end{bmatrix}^{\otimes 2}, \quad Q_b(\boldsymbol{\alpha}) := Q_c(\boldsymbol{\alpha}) - \psi_1(\langle \mathbb{1}_d, \boldsymbol{\alpha} \rangle) \mathbb{1}_d^{\otimes 2},$$

$$(A.2)$$

where $Q_c(\boldsymbol{\alpha}) := \mathrm{diag}(\psi_1(\alpha_1), \cdots, \psi_1(\alpha_d))$ and where ψ and ψ_1 denote the digamma and trigamma functions, respectively. Then, for every $\boldsymbol{\alpha} \in A(\lambda)$, it holds that

$$\mathrm{Hess}_{\boldsymbol{\alpha}}(W_a(\boldsymbol{\alpha}, \lambda)) = \int_0^1 \int_{\mathcal{Y}_d} M(v; \lambda) \mathrm{Hess}_{\boldsymbol{\alpha}}(H(\mathbf{y}; \boldsymbol{\alpha})) |\Psi(v^{1/d} \mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv$$

$$= \int_0^1 M(v; \lambda) \left[\int_{\mathcal{Y}_d} Q_a(\mathbf{y}; \boldsymbol{\alpha}) H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d} \mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} \right.$$

$$\left. + Q_b(\boldsymbol{\alpha}) \int_{\mathcal{Y}_d} H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d} \mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} \right] dv,$$

where the dominated convergence theorem justifies the interchange of the Hessian and the integral due to the integrability condition imposed in the definition (4.8). It suffices to investigate the inner integrals with respect to the Lebesgue measure $d\mathbf{y}$, since the likelihood ratio $M(v; \lambda)$ is non-negative for almost all $v \in (0, 1)$.

The first inner integral is positive semi-definite since the matrix $Q_a(\mathbf{y}; \boldsymbol{\alpha})$ is positive semi-definite by definition (A.2) and the integrand $H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d} \mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d)$ is non-negative for almost every $(v, \mathbf{y}) \in (0, 1) \times \mathcal{Y}_d$. It remains to show that the second integral, more precisely, the matrix $Q_b(\boldsymbol{\alpha})$ is positive semi-definite. We show this by proving that its inverse is positive semi-definite. To this end, observe that

$$\left\langle \mathbf{z}, Q_b^{-1}(\boldsymbol{\alpha}) \mathbf{z} \right\rangle = \left\langle \mathbf{z}, Q_c^{-1}(\boldsymbol{\alpha}) \mathbf{z} \right\rangle + \frac{\psi_1(\langle \mathbb{1}_d, \boldsymbol{\alpha} \rangle)}{1 - \psi_1(\langle \mathbb{1}_d, \boldsymbol{\alpha} \rangle) \langle \mathbb{1}_d, Q_c^{-1}(\boldsymbol{\alpha}) \mathbb{1}_d \rangle} \left\langle \mathbb{1}_d, Q_c^{-1}(\boldsymbol{\alpha}) \mathbf{z} \right\rangle^2,$$

for $\mathbf{z} \in \mathbb{R}^d$. Since $Q_c^{-1}(\boldsymbol{\alpha})$ is positive semi-definite and the trigamma function ψ_1 is strictly positive, the necessary and sufficient condition for the claim to hold true is that the denominator is strictly positive, that is, $1 - \psi_1(\langle \mathbb{1}_d, \boldsymbol{\alpha} \rangle) \langle \mathbb{1}_d, Q_c^{-1}(\boldsymbol{\alpha}) \mathbb{1}_d \rangle > 0$, which can be rewritten as $1/\psi_1(\alpha_1) + \cdots + 1/\psi_1(\alpha_d) < 1/\psi_1(\langle \mathbb{1}_d, \boldsymbol{\alpha} \rangle)$. This indeed holds true for all $\boldsymbol{\alpha} \in (0, +\infty)^d$, since the function $1/\psi_1(x)$ is positive, increasing, convex on $(0, +\infty)$, and $\lim_{x \to 0+}(1/\psi_1(x)) = 0$, that is, superadditive.

(ii) Fix a point λ in the domain $\Lambda_a(\boldsymbol{\alpha})$ and let $\{\lambda_k\}_{k \in \mathbb{N}}$ be a sequence of points in the domain $\Lambda_a(\boldsymbol{\alpha})$ converging to λ. By Assumption 4.1 (b), the function $M(v; \cdot)$ is twice continuously differentiable for almost every $v \in (0, 1)$. The integrability condition imposed in the definition (4.9) ensures that the families $\{(\partial^q / \partial \lambda^q) M(\cdot; \lambda); \lambda \in \Lambda_a(\boldsymbol{\alpha}), q \in \{0, 1, 2\}\}$ of functions are dominated by a non-negative function, independent of λ, integrable with respect to the measure $H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d} \mathbf{y})|^2 d\mathbf{y} dv$ on $\mathcal{Y}_d \times (0, 1)$. It thus holds by the dominated convergence

theorem that for every $q \in \{0, 1, 2\}$,

$$
\lim_{k \to +\infty} \int_0^1 \int_{(0,1)^d} \left[\frac{\partial^q}{\partial \lambda^q} M(v; \lambda_k) \right] H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv
$$
$$
= \int_0^1 \int_{(0,1)^d} \lim_{k \to +\infty} \left[\frac{\partial^q}{\partial \lambda^q} M(v; \lambda_k) \right] H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv,
$$

which shows that the function $W_a(\boldsymbol{\alpha}, \cdot)$ is twice continuously differentiable.

Since the set $\Lambda_a(\boldsymbol{\alpha})$ is open, it holds by the mean value theorem that for each $q \in \{0, 1\}$, there exists $\epsilon > 0$ satisfying $\lambda \pm \epsilon \in \Lambda_a(\boldsymbol{\alpha})$ and

$$
\text{Hess}_{\boldsymbol{\alpha}}\left(W_a(\boldsymbol{\alpha}, \lambda)\right) = \int_0^1 \int_{\mathcal{Y}_d} M(v; \lambda) \text{Hess}_{\boldsymbol{\alpha}}\left(H(\mathbf{y}; \boldsymbol{\alpha})\right) |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv
$$
$$
= \int_0^1 M(v; \lambda) \left[\int_{\mathcal{Y}_d} Q_a(\mathbf{y}; \boldsymbol{\alpha}) H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} \right.
$$
$$
\left. + Q_b(\boldsymbol{\alpha}) \int_{\mathcal{Y}_d} H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} \right] dv,
$$

where λ_ϵ is an intermediate point on the line segment joining two points $\lambda \pm \epsilon$ both in the set $\Lambda_a(\boldsymbol{\alpha})$, again due to twice continuous differentiability of the function $M(v; \cdot)$. Here, the last inequality holds by the integrability condition imposed in the definition (4.9). Hence, with the aid of the dominated convergence theorem, we get

$$
\frac{\partial^2}{\partial \lambda^2} W_a(\boldsymbol{\alpha}, \lambda) = \int_0^1 \int_{(0,1)^d} \left[\frac{\partial^2}{\partial \lambda^2} M(v; \lambda) \right] H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d) d\mathbf{y} dv,
$$

where the remaining integrand $H(\mathbf{y}; \boldsymbol{\alpha}) |\Psi(v^{1/d}\mathbf{y})|^2 p(\mathbf{y}; \mathbb{1}_d)$ is clearly neither negative nor degenerate. Hence, with the aid of the convexity of the reciprocal $1/f(v; \cdot)$ by Assumption 4.1 (g), we get the desired strict convexity of the function $W_a(\boldsymbol{\alpha}, \cdot)$ on the domain $\Lambda_a(\boldsymbol{\alpha})$.

Finally, let $\Lambda_1(\boldsymbol{\alpha}) := \text{int}\{\lambda \in \Lambda_0 : W_a(\boldsymbol{\alpha}, \lambda) < +\infty\}$, and let $\{L_n\}_{n \in \mathbb{N}}$ be an increasing sequence of compact subsets of $\Lambda_1(\boldsymbol{\alpha})$, satisfying $\cup_{n \in \mathbb{N}} L_n = \Lambda_1(\boldsymbol{\alpha})$ and $L_n \subsetneq \text{int}(L_{n+1})$ for all $n \in \mathbb{N}$. The second moment $W_a(\boldsymbol{\alpha}, \lambda)$ explodes as $\lambda \to \partial \Lambda_0 \cup \partial \Lambda_1(\boldsymbol{\alpha})$, since the probability density function $f(\cdot; \lambda)$ in the denominator in the expression (4.7) tends to zero for almost every $v \in (0, 1)$, due to Assumption 4.1 (f). Hence, the uniqueness and first-order necessary optimality of the point $\lambda^*(\boldsymbol{\alpha})$ holds true by the strict convexity of $W_a(\boldsymbol{\alpha}, \cdot)$. $\qquad \square$

References

1. Balakrishnan, N., Nevzorov, V.B.: A Primer on Statistical Distributions. John Wiley & Sons Inc, New York (2003)
2. Baldoni, V., Berline, N., De Loera, J.A., Köppe, M., Vergne, M.: How to integrate a polynomial over a simplex. Math. Comput. **80**(273), 297–325 (2011)
3. Basu, K.: Quasi-Monte Carlo tractability of high dimensional integration over products of simplices. J. Complex. **31**, 817–834 (2015)
4. Devroye, L.: Non-Uniform Random Variate Generation. Springer-Verlag, New York (1986)
5. Di Tommaso, F., Zerroudi, B.: On some numerical integration formulas on the d-dimensional simplex. Mediterr. J. Math. **17**(5), 142 (2020)
6. Dür, M.: Copositive programming—a survey. Recent Advances in Optimization and its Applications in Engineering, pp. 3–20. Springer Berlin Heidelberg, Berlin (2010)
7. Fieldsend, J.E.: A short note on the efficient random sampling of the multi-dimensional pyramid between a simplex and the origin lying in the unit hypercube. Technical Report 419, Department of Computer Science, University of Exeter (2005)

8. Genz, A., Cools, R.: An adaptive numerical cubature algorithm for simplices. ACM Trans. Math. Softw. **29**(3), 297–308 (2003)
9. Grundmann, A., Möller, H.M.: Invariant integration formulas for the n-simplex by combinatorial methods. SIAM J. Numer. Anal. **15**(2), 282–290 (1978)
10. Guessab, A., Semisalov, B.: Numerical integration using integrals over hyperplane sections of simplices in a triangulation of a polytope. BIT Numer. Math. **58**(3), 613–660 (2018)
11. Hammer, P.C., Stroud, A.H.: Numerical integration over simplexes. Math. Tables and Other Aids Comput. **10**(55), 137–139 (1956)
12. Hillion, P.: Numerical integration on a triangle. Int. J. Numer. Meth. Eng. **11**, 797–815 (1977)
13. Kawai, R.: Optimal importance sampling parameter search for Lévy processes via stochastic approximation. SIAM J. Numer. Anal. **47**(1), 293–307 (2009)
14. Kawai, R.: Acceleration on adaptive importance sampling with sample average approximation. SIAM J. Sci. Comput. **39**(4), A1586–A1615 (2017)
15. Kawai, R.: Adaptive importance sampling Monte Carlo simulation for general multivariate probability laws. J. Comput. Appl. Math. **319**, 440–459 (2017)
16. Kawai, R.: Optimizing adaptive importance sampling by stochastic approximation. SIAM J. Sci. Comput. **40**(4), A2774–A2800 (2018)
17. Kawai, R.: Adaptive importance sampling and control variates. J. Math. Anal. Appl. **483**(1), 123608 (2020)
18. Khosravifard, M., Esmaeili, M., Saidi, H.: Extension of the Lasserre–Avrachenkov theorem on the integral of multilinear forms over simplices. Appl. Math. Comput. **212**(1), 94–99 (2009)
19. Kotz, S., Balakrishnan, N., Johnson, N.L.: Continuous Multivariate Distributions. Models and Applications. Wiley, New York (2000)
20. Lasserre, J.B.: Simple formula for integration of polynomials on a simplex. BIT Numer. Math. **61**, 523–533 (2021)
21. Lasserre, J.B., Avrachenkov, K.E.: The Multi-Dimensional Version of $\int_a^b x^p dx$. Am. Math. Mon. **108**(2), 151–154 (2001)
22. Lin, S., Xie, Z.: A new recursive formula for integration of polynomial over simplex. Appl. Math. Comput. **376**, 125140 (2020)
23. MacKay, D.J.C.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press, Cambridge (2003)
24. Ng, K.W., Tian, G.-L., Tang, M.-L.: Dirichlet and Related Distributions: Theory, Methods and Applications. Wiley, Wiley Series in Probability and Statistics (2011)
25. Pillards, T., Cools, R.: A theoretical view on transforming low-discrepancy sequences from a cube to a simplex. Monte Carlo Methods Appl. **10**(3–4), 511–529 (2004)
26. Pillards, T., Cools, R.: Transforming low-discrepancy sequences from a cube to a simplex. J. Comput. Appl. Math. **174**, 29–42 (2005)
27. Rouigueb, A., Maiza, M., Tkourt, A., Cherchour, I.: Integration of polynomials over n-dimensional simplices. In 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 157–163 (2019)
28. Song, C., Kawai, R.: Batching adaptive variance reduction. ACM Trans. Model. Comput. Simul. **33**(3), 1–24 (2023)
29. Song, C., Kawai, R.: Dynamic finite-budget allocation of stratified sampling with adaptive variance reduction by strata. SIAM J. Sci. Comput. **45**(2), A898–A932 (2023)
30. Stroud, A.H.: A fifth degree integration formula for the n-simplex. SIAM J. Numer. Anal. **6**(1), 90–98 (1969)
31. Stroud, A.H.: Approximate Calculation of Multiple Integrals. Prentice-Hall series in automatic computation, Englewood Cliffs, Prentice-Hall (1971)
32. Sylvester, P.: Symmetric quadrature formulae for simplexes. Math. Comput. **24**, 95–100 (1970)