

ANALYSIS OF PREINTEGRATION FOLLOWED BY QUASI-MONTE CARLO INTEGRATION FOR DISTRIBUTION FUNCTIONS AND DENSITIES*

ALEXANDER D. GILBERT[†], FRANCES Y. KUO[†], AND IAN H. SLOAN[†]

Abstract. In this paper, we analyze a method for approximating the distribution function and density of a random variable that depends in a nontrivial way on a possibly high number of independent random variables, each with support on the whole real line. Starting with the integral formulations of the distribution and density, the method involves *smoothing* the original integrand *by preintegration* with respect to one suitably chosen variable, and then applying a suitable quasi-Monte Carlo method to compute the integral of the resulting smoother function. Interpolation is then used to reconstruct the distribution or density on an interval. The preintegration technique is a special case of *conditional sampling*, a method that has previously been applied to a wide range of problems in statistics and computational finance. In particular, the pointwise approximation studied in this work is a specific case of the conditional density estimator previously considered by L'Ecuyer, Puchhammer, and Ben Abdellah *INFORMS J. Comput.*, 34 (2022), pp. 1729–1748]. Our theory provides a rigorous regularity analysis of the preintegrated function, which is then used to show that the errors of the pointwise and interpolated estimators can both achieve nearly first-order convergence. Numerical results support the theory.

Key words. quasi-Monte Carlo methods, density estimation, preintegration, conditional sampling

MSC codes. 65D30, 65D32, 62G07

DOI. 10.1137/21M146658X

1. Introduction. In this paper, we analyze a numerical method for the computation of the cumulative distribution function (cdf) or probability density function (pdf) of a continuous random variable X , where X depends in a nontrivial way on many independent continuous random variables Y_0, Y_1, \dots, Y_d with known pdfs, each of which has support on the whole real line. We write $X = \phi(Y_0, \dots, Y_d)$, where often the computation of ϕ requires significant resources, for example, the solution of a partial differential equation.

Starting with the formulation of the cdf at a given point as an expected value (integral) of an indicator function (see (1.2) below), the method consists of *preintegration* [15, 18] (a process in which one well-chosen variable is first integrated out, with the aim of producing a relatively smooth function of the remaining d variables), followed by a suitable quasi-Monte Carlo (QMC) method to integrate over the remaining variables. This approach overcomes the difficult aspect of the original integration problem, namely, that ϕ occurs in the argument of a nonsmooth function (a jump discontinuity in the case of the cdf; a delta function in the case of the pdf), making the direct application of QMC integration problematic. Interpolation is then used to reconstruct the cdf on a given interval. This is useful for applications where one wishes to build a surrogate of the cdf to be used many times within a larger

*Received by the editors December 23, 2021; accepted for publication (in revised form) August 29, 2022; published electronically February 23, 2023.

<https://doi.org/10.1137/21M146658X>

Funding: The work of the authors was supported by Australian Research Council Discovery project DP21010083.

[†]School of Mathematics and Statistics, UNSW Sydney, Sydney NSW 2052, Australia (alexander.gilbert@unsw.edu.au, f.kuo@unsw.edu.au, i.sloan@unsw.edu.au).

computational problem, e.g., when evaluating ϕ is computationally expensive and so computing pointwise approximations is also expensive.

Smoothing by preintegration can be considered as a special case of *conditional sampling* from the statistics and computational finance literature [1, 2, 4, 12, 13, 19, 22, 29], whereby a given function has its variance reduced by the operation of conditioning on or “hiding” (integrating out) partial information, reducing the problem to a conditional expectation. Preintegration involves integrating out just a single variable, say, Y_0 , which is equivalent to conditioning on the other remaining variables, Y_1, Y_2, \dots, Y_d . In particular, the pointwise approximation method in the present paper is a specific case of the conditional density estimator (CDE) studied in [23].

The principal contribution of the present paper lies in a rigorous error analysis of the method, assuming only properties of the original function ϕ *before* performing preintegration. The first step in the analysis (a nontrivial one) is to establish regularity properties of the preintegrated function, where regularity is meant in the sense of the preintegrated function having all its mixed derivatives of first order being square integrable.

The analysis in [23] is essentially different, in that it started with an assumption of suitable smoothness of the resulting integrand *after* preintegration, not the original function.

The minimal smoothness property assumed in classical QMC analyses on the unit cube is membership of BVHK—the class of functions with bounded variation in the sense of Hardy and Krause. A sufficient condition for membership of BVHK is that f has continuous, integrable mixed derivatives of first order. Our analysis and function space setting [9, 24] include preintegrated functions which, after transformation to the unit cube, do *not* belong to BVHK; see Appendix A.1 for further details.

In this paper we assume that the random variables Y_i for $i = 0, \dots, d$ have as support the whole real line \mathbb{R} . This assumption is essential, since the results do not apply if the Y_i have compact support. In Appendix A.2 we give a simple example on the unit cube illustrating that a single step of preintegration is not sufficient for the preintegrated function to belong to BVHK, nor does it possess the necessary mixed derivative smoothness. To achieve the required first-order mixed derivative smoothness on $[0, 1]^d$, the paper [14] shows that, in general, one must perform preintegration with respect to $d/2$ different variables.

It is now known [10] that for the preintegrated function to have the classical smoothness property assumed in QMC analyses, it is necessary (as assumed here and in [18]) that ϕ be a monotone function of the chosen preintegration variable. If that monotonicity assumption does not hold and $d \geq 2$, then from [10] the preintegrated function has generically one-sided square-root or higher-root singularities, which, as shown in Appendix A.3, precludes membership of BVHK.

1.1. The problem and the approach. Let X be a continuous real-valued random variable, and denote its cdf and pdf by $F := F_X$ and $f := f_X$, respectively, which are defined on an interval $[a, b]$ but which are not known a priori. Suppose that X is a function of $d + 1$ independent random variables $Y_0, Y_1, \dots, Y_d \in \mathbb{R}$,

$$(1.1) \quad X = \phi(Y_0, Y_1, \dots, Y_d),$$

for some computable (but possibly expensive) function $\phi : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$. Suppose also that the density and distribution functions of each Y_i are known and are denoted by ρ_i and Φ_i , respectively, with $\text{supp}(\rho_i) = \mathbb{R}$. Realizations of the random variables X and Y_i are denoted by x and y_i , respectively.

An example of such a random variable X is the value of an Asian option, where

the random variables Y_0, \dots, Y_d are the Brownian motion increments at each time step and are normally distributed. Another example is the linear functional of the solution of a PDE with a log-normal random coefficient, where Y_0, \dots, Y_d are zero-mean normal random variables that correspond to the random parameters in a series expansion of a Gaussian random field.

Our goal is to approximate the cdf F and the pdf f (the derivative of F) on a compact interval $[a, b] \subset \mathbb{R}$, which is done in two steps:

1. Approximate F and f at finitely many points $\{t_m\}_{m=0}^M \subset [a, b]$ using QMC with preintegration (see sections 2.2 and 3, respectively).
2. Reconstruct F and f on $[a, b]$ by interpolating the approximations at the points $\{t_m\}_{m=0}^M$.

The cdf and pdf can each be written as expected values (i.e., potentially high-dimensional integrals) with respect to Y_0, Y_1, \dots, Y_d . For $t \in [a, b]$, we have

$$(1.2) \quad F(t) = \mathbb{E}[\text{ind}(t - X)] = \int_{\mathbb{R}^{d+1}} \text{ind}(t - \phi(y_0, \dots, y_d)) \left(\prod_{i=0}^d \rho_i(y_i) \right) dy_0 \cdots dy_d,$$

$$(1.3) \quad f(t) = \mathbb{E}[\delta(t - X)] = \int_{\mathbb{R}^{d+1}} \delta(t - \phi(y_0, \dots, y_d)) \left(\prod_{i=0}^d \rho_i(y_i) \right) dy_0 \cdots dy_d,$$

where $\text{ind}(\cdot)$ is the indicator function

$$\text{ind}(z) = \begin{cases} 1 & \text{if } z \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

and $\delta(\cdot)$ is the Dirac δ distribution characterized by the properties

$$(1.4) \quad \begin{aligned} &\delta(z) = 0 \quad \text{for all } z \neq 0 \quad \text{and} \\ &\int_{-\infty}^{\infty} g(z) \delta(z) dz = g(0) \quad \text{for all sufficiently smooth functions } g. \end{aligned}$$

Note that if the pdf exists, then it is the derivative of the cdf, i.e., $f = F'$. Indeed, the integral in (1.3) can be interpreted as differentiating (1.2) with respect to t in the distributional sense (recall that δ is the distributional derivative of ind). In the following, we will give more precise conditions we need to impose on the function ϕ such that the integral formulations (1.2) and (1.3) are well defined.

QMC theory alone is unable to tackle integrals such as (1.2) and (1.3) due to the discontinuity introduced by the indicator and Dirac δ functions. This discontinuity means that the integrand fails to belong to the function spaces required for QMC theory. The integrand in the formulation of the pdf (1.3) is not even a function, but rather a distribution that is 0 everywhere except when $\phi(y_0, \dots, y_d) = t$. However, recent work [18] on smoothing by preintegration was successful in handling simple discontinuities caused by an indicator function in the integrand, both practically and theoretically. In this paper, we extend the work to cover distributions involving a δ function, as well as extending the theory for the indicator function.

1.2. Preintegration. To explain the idea of preintegration, consider a simple discontinuous function

$$g(y_0, \dots, y_d) = \text{ind}(\phi(y_0, \dots, y_d)),$$

where the inner function $\phi : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ is sufficiently smooth and satisfies certain technical assumptions with respect to a specially chosen variable, which throughout this paper we shall take to be y_0 . As stated above, a key assumption is that ϕ is strictly increasing in y_0 ,

$$(1.5) \quad \frac{\partial \phi}{\partial y_0} > 0$$

(cf. Assumption 1 below). To perform the preintegration step, we integrate with respect to this special variable y_0 to give a *preintegrated* function

$$(1.6) \quad P_0 g(y_1, \dots, y_d) := \int_{-\infty}^{\infty} g(y_0, \dots, y_d) \rho_0(y_0) dy_0,$$

which is now a d -variate function of the remaining variables y_1, \dots, y_d .

The key point from [18] is that if we fix $(y_1, \dots, y_d) \in \mathbb{R}^d$ and treat $\phi(\cdot, y_1, \dots, y_d)$ as a function of the single variable y_0 , then since ϕ is strictly increasing with respect to y_0 , the discontinuity in g either occurs at a single point, in which case that variable can then be integrated out, or does not occur at all. Thus after preintegration, there is no longer any discontinuity and the result is a d -variate function $P_0 g$ that under suitable conditions is as smooth as the original smooth function ϕ . In this way, after performing the preintegration step, either exactly or numerically, one can use a d -dimensional cubature rule for the remaining dimensions. Due to the smoothness of $P_0 g$, the cubature error will now converge at a faster rate, e.g., close to $\mathcal{O}(1/N)$ for a QMC method using N points.

The smoothing by preintegration step was recently analyzed in [18], which extends the earlier work [15, 16, 17]; see also [10]. In this paper we follow the terminology of preintegration instead of conditioning, since [18] forms the foundation of our theory.

1.3. Related work and other approaches. The recent paper [5] used QMC to construct kernel density estimators and histograms. Their need to balance the variance and bias means the provided error convergence rate deteriorates very rapidly with dimension. The paper [11] replaced the nonsmooth functions in (1.2) and (1.3) by smooth approximations and then applied multilevel Monte Carlo (MC) methods. The paper [6] introduced smooth cdf and pdf estimators for the specific case of a sum of dependent log-normals, with promising numerical results using QMC.

As stated earlier, preintegration is a specific case of conditional sampling or conditioning, a method widely used in the statistics and computational finance literature [1, 2, 4, 12, 13, 19, 22, 29]. The idea to combine a QMC rule with conditioning was first presented in [22], where it was applied to compute probabilities for a stochastic activity network. More recently, [3] used conditional MC for density estimation in the specific case of a sum of random variables. Then, the paper [23] introduced a general CDE and [26] used conditioning for variance reduction for a generalized likelihood ratio density estimator, where both works used MC and QMC integration. The pointwise estimator analyzed in this paper is a special case of the QMC CDE in [23]. That paper also presented a more general CDE, by allowing conditioning on general sets, as well as considering convex combinations of CDEs and using QMC within a generalized likelihood ratio density estimator. To analyze the CDE with QMC, the paper [23] assumed that after conditioning the result belonged to the class BVHK, which allowed the use of the classical Koksma–Hlawka inequality to bound the QMC error.

2. Mathematical background. In this section, we introduce the required background material on preintegration, QMC, and the function spaces that we need.

We start with some notation. Recall that y_0 is the special variable with respect to which we perform preintegration. We denote the remaining d variables by $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ and all of the $d+1$ variables collectively by $\mathbf{y}_{0:d} := (y_0, \dots, y_d) \in \mathbb{R}^{d+1}$ or (y_0, \mathbf{y}) . Similarly, we denote the products of univariate functions $(\rho_i)_{i=0}^d$ by

$$\rho(\mathbf{y}) := \prod_{i=1}^d \rho_i(y_i) \quad \text{and} \quad \rho_{0:d}(y_0, \mathbf{y}) := \prod_{i=0}^d \rho_i(y_i) = \rho_0(y_0) \rho(\mathbf{y}).$$

Let $\mathbb{N}_0 := \{0, 1, 2, \dots\}$ and $\mathbb{N} := \{1, 2, \dots\}$ denote the set of natural numbers with and without zero, respectively. Let $\{0 : d\} := \{0, 1, \dots, d\}$ and define $\{1 : d\}$ analogously. Let $\boldsymbol{\nu} = (\nu_0, \nu_1, \dots, \nu_d) \in \mathbb{N}_0^{d+1}$ be a multi-index, and let $|\boldsymbol{\nu}| := \sum_{i=0}^d \nu_i$ denote its *order* and $\text{supp}(\boldsymbol{\nu}) := \{j \in \{0 : d\} : \nu_j > 0\}$ denote its *support*. Operations and relations between multi-indices are defined componentwise, e.g., for $\boldsymbol{\eta}, \boldsymbol{\nu} \in \mathbb{N}_0^{d+1}$ we write $\boldsymbol{\eta} \leq \boldsymbol{\nu}$ if and only if $\eta_i \leq \nu_i$ for all $i = 0, 1, \dots, d$, and addition is defined by $\boldsymbol{\eta} + \boldsymbol{\nu} = (\eta_i + \nu_i)_{i=0}^d$. For $\mathbf{y}_{0:d} \in \mathbb{R}^{d+1}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^{d+1}$, we denote the *active* variables by $\mathbf{y}_{\boldsymbol{\nu}} := (y_i : \nu_i > 0)_{i=0}^d$ and the *inactive* variables by $\mathbf{y}_{-\boldsymbol{\nu}} := (y_i : \nu_i = 0)_{i=0}^d$. Analogously to the notation (y_0, \mathbf{y}) , we denote the $(d+1)$ -dimensional concatenation of $\nu_0 \in \mathbb{N}_0$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_d) \in \mathbb{N}_0^d$ by $(\nu_0, \boldsymbol{\nu}) = (\nu_0, \nu_1, \nu_2, \dots, \nu_d) \in \mathbb{N}_0^{d+1}$.

2.1. Function spaces. Here we introduce our function space setting. Although we deal with both $(d+1)$ - and d -variate functions throughout this paper, we give definitions only for the $(d+1)$ -variate spaces, since the d -variate spaces can be defined analogously by simply excluding the variable y_0 .

We begin by defining some shorthand notation for mixed partial derivatives. For $i = 0, 1, \dots, d$ and a multi-index $\boldsymbol{\nu} \in \mathbb{N}_0^{d+1}$, let

$$D^i := \frac{\partial}{\partial y_i} \quad \text{and} \quad D^{\boldsymbol{\nu}} = \prod_{i=0}^d \frac{\partial^{\nu_i}}{\partial y_i^{\nu_i}}$$

denote the first-order derivative and the higher-order mixed derivative of order $\boldsymbol{\nu}$, respectively. This notation will also be used for weak derivatives, where the $\boldsymbol{\nu}$ th weak derivative of g is defined to be the distribution $D^{\boldsymbol{\nu}}g$ that satisfies

$$\int_{\mathbb{R}^{d+1}} D^{\boldsymbol{\nu}}g(\mathbf{y}_{0:d}) v(\mathbf{y}_{0:d}) d\mathbf{y}_{0:d} = (-1)^{|\boldsymbol{\nu}|} \int_{\mathbb{R}^{d+1}} g(\mathbf{y}_{0:d}) D^{\boldsymbol{\nu}}v(\mathbf{y}_{0:d}) d\mathbf{y}_{0:d}$$

for all $v \in C_0^\infty(\mathbb{R}^{d+1})$. Here $C_0^\infty(\mathbb{R}^{d+1})$ is the space of infinitely differentiable functions with compact support.

Let $C(\mathbb{R}^{d+1})$ denote the space of continuous functions on \mathbb{R}^{d+1} . For $\boldsymbol{\nu} \in \mathbb{N}_0^{d+1}$ let $C^{\boldsymbol{\nu}}(\mathbb{R}^{d+1})$ denote the space of functions with continuous *mixed* derivatives up to $\boldsymbol{\nu}$:

$$C^{\boldsymbol{\nu}}(\mathbb{R}^{d+1}) := \{g \in C(\mathbb{R}^{d+1}) : D^{\boldsymbol{\eta}}g \in C(\mathbb{R}^{d+1}) \text{ for all } \boldsymbol{\eta} \leq \boldsymbol{\nu}\}.$$

To provide a function space setting for ϕ in (1.1), we introduce a class of Sobolev spaces of dominating mixed smoothness on \mathbb{R}^{d+1} , where the behavior of derivatives as $y_i \rightarrow \pm\infty$ is controlled by functions different from the densities ρ_i . To this end, for $i = 0, 1, \dots, d$, let $\psi_i : \mathbb{R} \rightarrow \mathbb{R}$ be a strictly positive and integrable *weight function*. We denote the whole collection of weight functions by $\boldsymbol{\psi} = (\psi_i)_{i=0}^d$. Also, let $\boldsymbol{\gamma} := \{\gamma_{\mathbf{u}} > 0 : \mathbf{u} \subseteq \{0 : d\}\}$ be a collection of positive real numbers called *weight*

parameters; they model the relative importance of different collections of variables, i.e., γ_u describes the relative importance of the collection of variables $(y_i : i \in u)$. We set $\gamma_\emptyset := 1$.

Then for $\nu \in \mathbb{N}_0^{d+1}$, define the *Sobolev space of dominating mixed smoothness* of order ν , denoted by \mathcal{H}_{d+1}^ν , to be the space of locally integrable functions on \mathbb{R}^{d+1} such that the norm

$$\|g\|_{\mathcal{H}_{d+1}^\nu}^2 := \sum_{\eta \leq \nu} \frac{1}{\gamma_\eta} \int_{\mathbb{R}^{d+1}} |D^\eta g(\mathbf{y}_{0:d})|^2 \psi_\eta(\mathbf{y}_\eta) \rho_{-\eta}(\mathbf{y}_{-\eta}) d\mathbf{y}_{0:d}$$

is finite, where we introduced the shorthand notation

$$\gamma_\eta := \gamma_{\text{supp}(\eta)}, \quad \psi_\eta(\mathbf{y}_\eta) := \prod_{i=0, \eta_i \neq 0}^d \psi_i(y_i) \quad \text{and} \quad \rho_{-\eta}(\mathbf{y}_{-\eta}) := \prod_{i=0, \eta_i=0}^d \rho_i(y_i).$$

Recall from the introduction that we plan to carry out preintegration on a non-smooth function of $d+1$ variables with appropriate properties, with the aim of obtaining a smooth function of d variables. We therefore need (but do not write down) an analogous d -variate Sobolev space \mathcal{H}_d^ν with variables indexed from 1 to d .

An important property of the Sobolev space of *first-order* dominating mixed smoothness, i.e., $\mathcal{H}_d^{\mathbf{1}}$ with $\mathbf{1} := (1, 1, \dots, 1)$, is that it is equivalent to the (unanchored) ANOVA space over the unbounded domain \mathbb{R}^d introduced in [24]. Explicitly, it was shown recently in [9] that if the weight functions ψ_i satisfy

$$(2.1) \quad \int_{-\infty}^{\infty} \frac{\Phi_i(z)(1 - \Phi_i(z))}{\psi_i(z)} dz < \infty \quad \text{for all } i = 1, 2, \dots, d,$$

then $\mathcal{H}_d^{\mathbf{1}}$ and the ANOVA space from [24] are equivalent. This equivalence is crucial for our analysis, because it immediately shows that the bounds on the QMC error from [24] also hold in $\mathcal{H}_d^{\mathbf{1}}$ (see Theorem 2.1 below). Since the preintegration theory in [18] is formulated in \mathcal{H}_d^ν , without this equivalence, there would be a mismatch between the settings for the analysis of preintegration and QMC methods. With the equivalence established, we can from now on work exclusively with the spaces \mathcal{H}_d^ν , and have no need to introduce the ANOVA space from [24]. Note that the condition (2.1) is also assumed throughout [24], where it is required for the QMC error bounds to hold. Examples of common pairings (ρ_i, ψ_i) satisfying (2.1) can be found in [21, Table 3]. Note that ψ_i^2 in [21, 24] is replaced here, and in [9], by ψ_i . We assume (2.1) holds throughout.

2.2. Quasi-Monte Carlo methods. In the classic case of the unit cube, an N -point QMC approximation (see, e.g., [8, 25]) for the integral of a function $g : [0, 1]^d \rightarrow \mathbb{R}$ is given by

$$\frac{1}{N} \sum_{n=0}^{N-1} g(\mathbf{q}_n) \approx \int_{[0,1]^d} g(\mathbf{u}) d\mathbf{u},$$

where here the cubature points $\{\mathbf{q}_n\}_{n=0}^{N-1}$ are deterministically chosen to be well-distributed within $[0, 1]^d$ and to have desirable approximation properties.

In this paper, we consider a simple class of randomized QMC methods called *randomly shifted rank-1 lattice rules*, for which the QMC points are given by

$$(2.2) \quad \mathbf{q}_n = \left\{ \frac{n\mathbf{z}}{N} + \Delta \right\} \quad \text{for } n = 0, 1, \dots, N-1.$$

Here $\mathbf{z} \in \{1, 2, \dots, N-1\}^d$ is the *generating vector*, $\Delta \in [0, 1)^d$ is a uniformly distributed random shift, and $\{\cdot\}$ denotes taking the fractional part of each component.

The benefits of randomly shifting the point set are threefold: (i) the resulting approximation is unbiased; (ii) we can take the average of the approximations from a small number of independent and identically distributed random shifts as the final approximation and use the sample variance to estimate the mean-square error; and (iii) for functions in \mathcal{H}_d^1 , randomly shifted lattice rules with good \mathbf{z} can be constructed efficiently (see below) to achieve nearly $\mathcal{O}(N^{-1})$ convergence of the root-mean-square error (RMSE).

To approximate an integral over an unbounded domain, one must map the point set $\{\mathbf{q}_n\}_{n=0}^{N-1}$ from the unit cube to \mathbb{R}^d . In the case of an integral with respect to a product of densities, as we have in (1.2), we can perform this mapping by applying the inverse cdf componentwise. An N -point QMC approximation for the integral of a function $g: \mathbb{R}^d \rightarrow \mathbb{R}$ is then given by

$$(2.3) \quad Q_{d,N}(g) := \frac{1}{N} \sum_{n=0}^{N-1} g(\Phi^{-1}(\mathbf{q}_n)) \approx \int_{[0,1]^d} g(\Phi^{-1}(\mathbf{u})) d\mathbf{u} = \int_{\mathbb{R}^d} g(\mathbf{y}) \rho(\mathbf{y}) d\mathbf{y},$$

where Φ^{-1} denotes the application of the inverse cdf Φ_i^{-1} in each dimension i , recalling that Φ_i is the cdf of the density ρ_i . For the remainder of the paper we only consider approximating integrals on \mathbb{R}^d , and so we denote the transformed QMC points by

$$(2.4) \quad \tau_n = \Phi^{-1}(\mathbf{q}_n) \in \mathbb{R}^d \quad \text{for } n = 0, 1, \dots, N-1.$$

It was proved in [24] that good generating vectors \mathbf{z} for the approximation (2.3) can be constructed using a component-by-component (CBC) algorithm to achieve almost the optimal convergence rate for the RMSE in a certain *first-order ANOVA space* (which as we have discussed is equivalent to \mathcal{H}_d^1). Below, we restate the error bound from [24], but now in terms of \mathcal{H}_d^1 rather than the equivalent ANOVA space used in [24].

THEOREM 2.1. *Suppose (2.1) holds. Let $\omega \in (1, 2]$ and $c < \infty$ be such that*

$$(2.5) \quad \frac{1}{\pi^2 k^2} \int_{-\infty}^{\infty} \frac{\sin^2(\pi k \Phi_i(y))}{\psi_i(y)} dy \leq \frac{c}{|k|^\omega} \quad \text{for all } k \in \mathbb{Z} \setminus \{0\} \text{ and all } i = 1, \dots, d.$$

Let $N \in \mathbb{N}$ and suppose that \mathbf{z} is a generating vector constructed using the CBC algorithm from [24]. Then for $g \in \mathcal{H}_d^1$, the RMSE (with the expectation taken with respect to the random shift Δ) of the randomly shifted lattice rule approximation (2.3) corresponding to \mathbf{z} satisfies

$$(2.6) \quad \sqrt{\mathbb{E}_\Delta \left[\left\| \int_{\mathbb{R}^d} g(\mathbf{y}) \rho(\mathbf{y}) d\mathbf{y} - Q_{d,N}(g) \right\|^2 \right]} \leq C_{d,\gamma,\lambda} [\phi_{\text{tot}}(N)]^{-1/(2\lambda)} \|g\|_{\mathcal{H}_d^1}$$

for all $\lambda \in (1/\omega, 1]$, with

$$C_{d,\gamma,\lambda} := \left(\sum_{\mathbf{0} \neq \boldsymbol{\eta} \in \{0,1\}^d} \gamma_{\boldsymbol{\eta}}^\lambda [2c\zeta(\omega\lambda)]^{|\boldsymbol{\eta}|} \right)^{1/(2\lambda)},$$

where ϕ_{tot} is the Euler totient function and ζ is the Riemann zeta function.

Proof. Let \mathcal{W}_d denote the ANOVA space from [24]. Theorem 8 from [24] gives the error bound (2.6) for $g \in \mathcal{W}_d$ and with the \mathcal{W}_d -norm on the right. Since [9, Theorem 13] shows that \mathcal{W}_d and \mathcal{H}_d^1 are equivalent under assumption (2.1) on the weight functions, with $\|g\|_{\mathcal{W}_d} \leq \|g\|_{\mathcal{H}_d^1}$, the result is proved. \square

Observe that the convergence rate $N^{-1/(2\lambda)}$ of the RMSE is governed by the parameter $\omega \in (1, 2]$ from (2.5), which in turn depends on the interaction between the pairs (ρ_i, ψ_i) . Ideally, we would like ω to be arbitrarily close to 2, which would allow us to take λ arbitrarily close to $1/2$, giving a convergence rate arbitrarily close to $1/N$. However, this is not always possible. Table 3 in [21] gives values of $r_2 = \omega/2$ such that (2.5) holds for several pairs (ρ_i, ψ_i) , and in particular, it provides common examples for which $\omega \approx 2$. Note that, as before, ψ_i in this paper is ψ_i^2 in [21]. As an example, if each ρ_i is a standard normal density, then one can take either $\psi_i(y) = e^{-|y|}$ or $\psi_i = 1/(1 + |y|)$, resulting in $\omega \approx 2$. Alternatively, one can take a scaled normal density, $\psi_i(y) = e^{-y^2/(2\eta)}$ for $\eta > 1$, giving $\omega = 2(1 - 1/\eta)$, which for η sufficiently large will give a convergence rate close to $1/N$. In general, the essential feature is that ψ_i decays more slowly than ρ_i .

2.3. Lagrange interpolation in one dimension. There are two steps to the cdf and pdf estimation algorithms: pointwise approximation, which we do using a QMC rule after a preintegration step, and then interpolation on the interval $[a, b]$. For the latter step, we use Lagrange interpolation based on Chebyshev points in $[a, b]$.

Let $\{t_m\}_{m=0}^M$ be a collection of distinct points in $[a, b]$ and let V_M denote the set of all polynomials up to degree M on $[a, b]$. The *Lagrange interpolation operator* $L_M : C[a, b] \rightarrow V_M$ is given by

$$(2.7) \quad L_M g := \sum_{m=0}^M g(t_m) \chi_{M,m}, \quad \chi_{M,m}(t) := \prod_{\substack{\ell=0 \\ \ell \neq m}}^M \frac{t - t_\ell}{t_m - t_\ell}.$$

We now state the classical error bounds for Lagrange interpolation based on Chebyshev points from, e.g., [27]. For $\sigma \in \mathbb{N}$, let $W^{\sigma, \infty}[a, b]$ denote the Sobolev space of functions on $[a, b]$ with essentially bounded derivatives up to order σ , which we equip with the norm $\|g\|_{W^{\sigma, \infty}} := \max_{q=0,1,\dots,\sigma} \|g^{(q)}\|_{L^\infty}$. Let $\sigma \in \mathbb{N}$ and suppose that $g \in W^{\sigma+1, \infty}[a, b]$. Then for $M > \sigma$, the error of the Lagrange interpolant based on Chebyshev nodes satisfies

$$(2.8) \quad \|g - L_M g\|_{L^\infty} \leq \frac{4 \|g^{(\sigma+1)}\|_{L^1}}{\pi \sigma (M - \sigma)^\sigma}.$$

The original result [27, Theorem 7.2] was stated in terms of the *total variation* of $g^{(\sigma)}$ on $[a, b]$, which for $g \in W^{\sigma+1, \infty}[a, b]$ is given by $\|g^{(\sigma+1)}\|_{L^1}$. As we will see in section 5.1, under our assumptions on ϕ , the cdf and pdf are smooth enough to take σ up to $d - 1$.

One may also use other methods to approximate F and f on $[a, b]$, such as splines or best polynomial approximation, but we do not pursue those directions here.

3. Smoothing by preintegration. As explained in the introduction, smoothing by preintegration is a method of smoothing a discontinuous or kink function by integrating out a single, specially chosen variable. It is a special case of conditional sampling. For notational convenience we take y_0 to be this special variable. In this section, we formalize the preintegration step for indicator functions by following [18],

and then extend the method to simple distributions involving δ distributions, which will allow us to also apply the preintegration technique to approximate the pdf as formulated in (1.3).

First, we make the following assumptions about the function ϕ in (1.1).

Assumption 1. For $d \geq 1$ and $\boldsymbol{\nu} \in \mathbb{N}_0^d$, let $\phi : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ satisfy

1. $D^0 \phi(y_0, \mathbf{y}) > 0$ for all $(y_0, \mathbf{y}) \in \mathbb{R}^{d+1}$; and
2. for each $\mathbf{y} \in \mathbb{R}^d$, $\phi(y_0, \mathbf{y}) \rightarrow \infty$ as $y_0 \rightarrow \infty$; and
3. $\phi \in \mathcal{H}_{d+1}^{(\nu_0, \boldsymbol{\nu})} \cap C^{(\nu_0, \boldsymbol{\nu})}(\mathbb{R}^{d+1})$, where $\nu_0 := |\boldsymbol{\nu}| + 1$.

Additionally, suppose that $\rho_0 \in C^{|\boldsymbol{\nu}|}(\mathbb{R})$.

It was unresolved from the analysis in [15, 17, 18] whether the *monotonicity* assumption (item 1 above) is necessary. Recently it was proved in [10] that this is indeed necessary: if it fails, then there may remain a singularity after preintegration.

3.1. Smoothing by preintegration for indicator functions. Motivated by the cdf (1.2), for $t \in [a, b]$ we define a discontinuous function $g_t : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ by

$$(3.1) \quad g_t(y_0, \mathbf{y}) := \text{ind}(t - \phi(y_0, \mathbf{y})), \quad y_0 \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^d,$$

where $\phi : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ satisfies Assumption 1. Note that [18] considered functions of the more general form $g(y_0, \mathbf{y}) = \theta(y_0, \mathbf{y}) \text{ind}(\phi(y_0, \mathbf{y}))$, where both θ and ϕ are sufficiently smooth, a formulation that allows for more general discontinuities and kinks. However, since we are here only concerned with computing probabilities using (1.2), the restricted form in (3.1) is sufficient. Also, note that we consider here shifted indicator functions $\text{ind}(t - \cdot)$ instead of $\text{ind}(\cdot)$ as in [18]. This results only in minor changes in the presentation and does not affect any of the theory.

Since in Assumption 1 we assume that $\phi(\cdot, \mathbf{y})$ is strictly increasing with respect to y_0 for fixed $\mathbf{y} \in \mathbb{R}^d$, and also tends to ∞ as $y_0 \rightarrow \infty$, there are only two possible scenarios: either the discontinuity of $g_t(\cdot, \mathbf{y})$ for fixed $t \in [a, b]$ and $\mathbf{y} \in \mathbb{R}^d$ occurs at a unique point in dimension 0, or the discontinuity does not occur at all because $\phi(y_0, \mathbf{y}) > t$ for all $y_0 \in \mathbb{R}$.

To describe the former case more explicitly, for given $t \in [a, b]$, we define the set of $\mathbf{y} \in \mathbb{R}^d$ for which the discontinuity occurs by

$$(3.2) \quad U_t := \{\mathbf{y} \in \mathbb{R}^d : \phi(y_0, \mathbf{y}) = t \text{ for some } y_0 \in \mathbb{R}\},$$

which (unlike in [18]) now depends on the point t . Since $\text{supp}(\rho_0) = \mathbb{R}$, the point y_0 in (3.2) is in the support of ρ_0 . We have the equivalence

$$\mathbf{y} \in U_t \iff t \in \phi(\mathbb{R}, \mathbf{y}),$$

where with a slight abuse of notation we use $\phi(\mathbb{R}, \mathbf{y})$ to denote the image of \mathbb{R} under $\phi(\cdot, \mathbf{y})$. For $t \in [a, b]$ and $\mathbf{y} \in U_t$, the point at which the discontinuity occurs in dimension 0 is denoted by $\xi(t, \mathbf{y})$, i.e., $\xi(t, \mathbf{y}) \in \mathbb{R}$ is the unique real number such that

$$(3.3) \quad \phi(\xi(t, \mathbf{y}), \mathbf{y}) = t.$$

Here uniqueness follows from the monotonicity condition Assumption 1, item 1. Similarly, because $\phi(\cdot, \mathbf{y})$ is increasing, it follows from (3.3) that for $\mathbf{y} \in U_t$,

$$(3.4) \quad \phi(y_0, \mathbf{y}) < t \iff y_0 < \xi(t, \mathbf{y}).$$

The following implicit function theorem adapted from [15] shows that ξ is a well-defined function of both t and \mathbf{y} and implies that ξ “inherits” the smoothness of ϕ .

THEOREM 3.1. Let $d \geq 1$, $\nu \in \mathbb{N}_0^d$, and $[a, b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumption 1, and define

$$(3.5) \quad V := \{(t, \mathbf{y}) \in (a, b) \times \mathbb{R}^d : \phi(y_0, \mathbf{y}) = t \text{ for some } y_0 \in \mathbb{R}\} \subset [a, b] \times \mathbb{R}^d.$$

If V is not empty, then there exists a unique function $\xi \in C^{(\nu_0, \nu)}(\bar{V})$ satisfying (3.3) for all $(t, \mathbf{y}) \in \bar{V}$. Furthermore, for $(t, \mathbf{y}) \in V$ the first-order derivatives are given by

$$(3.6) \quad D^i \xi(t, \mathbf{y}) = \frac{\partial}{\partial y_i} \xi(t, \mathbf{y}) = -\frac{D^i \phi(\xi(t, \mathbf{y}), \mathbf{y})}{D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})} \quad \text{for all } i = 1, \dots, d, \quad \text{and}$$

$$(3.7) \quad \frac{\partial}{\partial t} \xi(t, \mathbf{y}) = \frac{1}{D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})}.$$

Proof. The result $\xi \in C^{(\nu_0, \nu)}(\bar{V})$ follows by applying [15, Theorem 2.3] to the function $\phi(y_0, \mathbf{y}) - t$, with $j = 1$ along with the variables labeled as $x_{i+1} = y_i$ for $i = 0, 1, \dots, d$ and $x_{d+2} = t$. Since the proof of [15, Theorem 2.3] is based on a local argument about an arbitrary point \mathbf{x} , the restriction $x_{d+2} = t \in (a, b)$, instead of \mathbb{R} , does not affect the result. Additionally, the original proof was conducted for the isotropic smoothness spaces, but it can easily be extended to the dominating mixed smoothness space $C^{(\nu_0, \nu)}(\bar{V})$. Finally, differentiating (3.3) with respect to y_i leads to the formula (3.6). Similarly, differentiating (3.3) with respect to t implies (3.7). \square

With P_0 the preintegration operator defined by (1.6), we now apply preintegration to the function g_t defined by (3.1) for $t \in [a, b]$, obtaining

$$P_0 g_t(\mathbf{y}) = \int_{-\infty}^{\infty} \text{ind}(t - \phi(y_0, \mathbf{y})) \rho_0(y_0) dy_0.$$

From the definition of $\xi(t, \mathbf{y})$ in (3.3) and the relation (3.4), for $\mathbf{y} \in U_t$ we can write

$$(3.8) \quad P_0 g_t(\mathbf{y}) = \int_{-\infty}^{\xi(t, \mathbf{y})} \rho_0(y_0) dy_0 = \Phi_0(\xi(t, \mathbf{y})),$$

while for $\mathbf{y} \in \mathbb{R}^d \setminus U_t$ we have $P_0 g_t \equiv 0$. In both cases there is no longer any discontinuity.

The main result from [18, Theorem 3] showed that if ϕ satisfies Assumption 1, along with some extra technical conditions in Assumption 2 below, then the preintegrated function is *as smooth as* ϕ . The technical conditions are required to control all of the terms that arise when differentiating $P_0 g_t$ using the chain rule.

As a first illustration, for any $i \in \{1 : d\}$ we have

$$(3.9) \quad D^i [P_0 g_t(\mathbf{y})] = D^i [\Phi_0(\xi(t, \mathbf{y}))] = \rho_0(\xi(t, \mathbf{y})) D^i \xi(t, \mathbf{y}) = -\frac{\rho_0(\xi(t, \mathbf{y})) D^i \phi(\xi(t, \mathbf{y}), \mathbf{y})}{D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})},$$

where we used (3.6). This motivates the general form of functions in Assumption 2 below. Our assumption is formulated differently from [18] because we need to account for the t dependence in this paper and we also aim to give a tight estimate on the number of terms that arise from the differentiation. This allows us in Theorem 3.2 below to extend [18, Theorem 3] by providing an explicit bound on the norm. We use \mathbf{e}_i to denote a multi-index whose i th component is 1 and all other components are 0.

Assumption 2. Let $d \geq 1$, $\boldsymbol{\nu} \in \mathbb{N}_0^d$, $[a, b] \subset \mathbb{R}$, and suppose that ϕ and ρ_0 satisfy Assumption 1. Recall the definitions of U_t , ξ and V in (3.2), (3.3), and (3.5), respectively. Given $q \in \mathbb{N}_0$ and $\boldsymbol{\eta} \leq \boldsymbol{\nu}$ satisfying $|\boldsymbol{\eta}| + q \leq |\boldsymbol{\nu}| + 1$, we consider functions $h_{q,\boldsymbol{\eta}}: \bar{V} \rightarrow \mathbb{R}$ of the form

$$(3.10) \quad \begin{cases} h_{q,\boldsymbol{\eta}}(t, \mathbf{y}) = h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t, \mathbf{y}) := \frac{(-1)^r \rho_0^{(\beta)}(\xi(t, \mathbf{y})) \prod_{\ell=1}^r D^{\boldsymbol{\alpha}_\ell} \phi(\xi(t, \mathbf{y}), \mathbf{y})}{[D^{\mathbf{0}} \phi(\xi(t, \mathbf{y}), \mathbf{y})]^{r+q}}, \\ \text{with } r \in \mathbb{N}_0, \boldsymbol{\alpha} = (\boldsymbol{\alpha}_\ell)_{\ell=1}^r, \boldsymbol{\alpha}_\ell \in \mathbb{N}_0^{d+1} \setminus \{\mathbf{e}_0, \mathbf{0}\}, \beta \in \mathbb{N}_0 \text{ satisfying} \\ r \leq 2|\boldsymbol{\eta}| + q - 1, \alpha_{\ell,0} \leq |\boldsymbol{\eta}| + q, \beta \leq |\boldsymbol{\eta}| + q - 1, \beta \mathbf{e}_0 + \sum_{\ell=1}^r \boldsymbol{\alpha}_\ell = (r + q - 1, \boldsymbol{\eta}). \end{cases}$$

We assume that all such functions $h_{q,\boldsymbol{\eta}}$ satisfy

$$(3.11) \quad \lim_{\mathbf{y} \rightarrow \partial U_t} h_{q,\boldsymbol{\eta}}(t, \mathbf{y}) = 0 \quad \text{for all } t \in [a, b],$$

and there is a constant $B_{q,\boldsymbol{\eta}}$ such that

$$(3.12) \quad \sup_{t \in [a, b]} \int_{U_t} |h_{q,\boldsymbol{\eta}}(t, \mathbf{y})|^2 \psi_{\boldsymbol{\eta}}(\mathbf{y}_{\boldsymbol{\eta}}) \rho_{-\boldsymbol{\eta}}(\mathbf{y}_{-\boldsymbol{\eta}}) d\mathbf{y} \leq B_{q,\boldsymbol{\eta}} < \infty.$$

It is worthwhile to briefly discuss Assumption 2. As we will see in the theorem below, the functions $h_{q,\boldsymbol{\eta}}$ occur when differentiating a preintegrated function using the multivariate chain and product rules. Loosely speaking, the parameter q relates to differentiating with respect to t , whereas $\boldsymbol{\eta}$ relates to differentiating with respect to \mathbf{y} . The conditions (3.11) and (3.12) then ensure that the derivatives are well-behaved enough for the preintegrated function to be sufficiently smooth. It follows from Assumption 1 and Theorem 3.1 that each function $h_{q,\boldsymbol{\eta}}$ of the form (3.10) is continuous on \bar{V} . For an appropriate choice of $\{\psi_i\}$, Assumption 2 will hold for functions ϕ for which the preintegrated function $P_0 g_t$ as in (3.8) (and also (3.16) below) is unbounded and has unbounded derivatives $h_{q,\boldsymbol{\eta}}$. In this case, since $P_0 g_t$ is unbounded, after mapping back to $[0, 1]^d$ it does not belong to BVHK.

Having introduced preintegration and our key assumptions, we now state the main smoothing by preintegration theorem for functions of the form (3.1). It is a refined version of [18, Theorem 3].

THEOREM 3.2. *Let $d \geq 1$, $\boldsymbol{\nu} \in \mathbb{N}_0^d$, and $[a, b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumptions 1 and 2 for $q=0$ and all $\mathbf{0} \neq \boldsymbol{\eta} \leq \boldsymbol{\nu}$. Then for $t \in [a, b]$, the function*

$$g_t(y_0, \mathbf{y}) := \text{ind}(t - \phi(y_0, \mathbf{y})) \quad \text{satisfies} \quad P_0 g_t \in \mathcal{H}_d^{\boldsymbol{\nu}} \cap C^{\boldsymbol{\nu}}(\mathbb{R}^d),$$

with its $\mathcal{H}_d^{\boldsymbol{\nu}}$ -norm bounded uniformly in t ,

$$(3.13) \quad \sup_{t \in [a, b]} \|P_0 g_t\|_{\mathcal{H}_d^{\boldsymbol{\nu}}} \leq \left(1 + \sum_{\mathbf{0} \neq \boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{(8^{|\boldsymbol{\eta}|-1} (|\boldsymbol{\eta}| - 1)!)^2 B_{\mathbf{0}, \boldsymbol{\eta}}}{\gamma_{\boldsymbol{\eta}}} \right)^{1/2} < \infty.$$

Proof. From (3.8) the preintegrated function can be written as

$$P_0 g_t(\mathbf{y}) = \begin{cases} \Phi_0(\xi(t, \mathbf{y})) & \text{if } \mathbf{y} \in U_t, \\ 0 & \text{if } \mathbf{y} \in \mathbb{R}^d \setminus U_t. \end{cases}$$

If $U_t = \emptyset$, then $P_0 g_t \equiv 0$ on \mathbb{R}^d , and the result holds trivially. If $U_t \neq \emptyset$, then for any $\boldsymbol{\eta} \in \mathbb{N}_0^d$ with $\mathbf{0} \neq \boldsymbol{\eta} \leq \boldsymbol{\nu}$, we first prove by induction on $|\boldsymbol{\eta}| \geq 1$ that the $\boldsymbol{\eta}$ th derivative of $P_0 g_t$ is given by

$$(3.14) \quad D^{\boldsymbol{\eta}}[P_0 g_t(\mathbf{y})] = \begin{cases} \sum_{j=1}^{J_{0,\boldsymbol{\eta}}} h_{0,\boldsymbol{\eta}}^{[j]}(t, \mathbf{y}) & \text{if } \mathbf{y} \in U_t \quad \text{with } J_{0,\boldsymbol{\eta}} \leq 8^{|\boldsymbol{\eta}|-1}(|\boldsymbol{\eta}| - 1)!, \\ 0 & \text{if } \mathbf{y} \in \mathbb{R}^d \setminus U_t, \end{cases}$$

where each function $h_{0,\boldsymbol{\eta}}^{[j]}$ is of the form (3.10) with $q = 0$.

For the base case $\boldsymbol{\eta} = \mathbf{e}_i$ with any $i \in \{1 : d\}$, we take $r = 1$, $\boldsymbol{\alpha}_1 = \mathbf{e}_i$, $\beta = 0$, and $J_{0,\mathbf{e}_i} = 1$ to recover the single function (3.9). Suppose next that (3.14) holds for some $\boldsymbol{\eta} \in \mathbb{N}_0^d$ with $|\boldsymbol{\eta}| \geq 1$ and consider any $i \in \{1 : d\}$ and $\mathbf{y} \in U_t$. We have

$$(3.15) \quad \begin{aligned} D^i D^{\boldsymbol{\eta}}[P_0 g_t(\mathbf{y})] &= \sum_{j=1}^{J_{0,\boldsymbol{\eta}}} D^i h_{0,\boldsymbol{\eta}}^{[j]}(t, \mathbf{y}) = \sum_{j=1}^{J_{0,\boldsymbol{\eta}}} \sum_{k=1}^{K_{0,\boldsymbol{\eta}}} h_{0,\boldsymbol{\eta}+\mathbf{e}_i}^{[j,k]}(t, \mathbf{y}) \\ &= \sum_{j'=1}^{J_{0,\boldsymbol{\eta}+\mathbf{e}_i}} h_{0,\boldsymbol{\eta}+\mathbf{e}_i}^{[j']} (t, \mathbf{y}). \end{aligned}$$

In the second equality we used Lemma B.1 in Appendix B with $q = 0$, which states that each function $D^i h_{0,\boldsymbol{\eta}}^{[j]}$ can be written as a sum of $K_{0,\boldsymbol{\eta}} \leq 8^{|\boldsymbol{\eta}|} - 3$ functions of the form (3.10) with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta} + \mathbf{e}_i$. We enumerated these functions with the notation $h_{0,\boldsymbol{\eta}+\mathbf{e}_i}^{[j,k]}$ and then relabeled all functions for different combinations of indices j and k with the notation $h_{0,\boldsymbol{\eta}+\mathbf{e}_i}^{[j']}$. The total number of functions satisfies

$$J_{0,\boldsymbol{\eta}+\mathbf{e}_i} = J_{0,\boldsymbol{\eta}} K_{0,\boldsymbol{\eta}} \leq 8^{|\boldsymbol{\eta}|-1}(|\boldsymbol{\eta}| - 1)!(8^{|\boldsymbol{\eta}|} - 3) \leq 8^{|\boldsymbol{\eta}+\mathbf{e}_i|-1}(|\boldsymbol{\eta} + \mathbf{e}_i| - 1)!,$$

as required. This completes the induction proof for (3.14).

Since, for all $\boldsymbol{\eta} \leq \boldsymbol{\nu}$, every function $h_{0,\boldsymbol{\eta}}^{[j]}(t, \cdot)$ in (3.14) is continuous on U_t , it follows that $P_0 g_t \in C^{\boldsymbol{\nu}}(U_t)$. Also, $P_0 g_t \equiv 0$ on $\mathbb{R}^d \setminus U_t$ is clearly smooth, and so we just need the derivatives to be continuous across the boundary ∂U_t . Indeed, the assumption (3.11) implies that $D^{\boldsymbol{\eta}}[P_0 g_t(\mathbf{y})] \rightarrow 0$ as $\mathbf{y} \rightarrow \partial U_t$ for all $\boldsymbol{\eta} \leq \boldsymbol{\nu}$. Hence, it follows by an adaptation of [18, Lemma 9] that $P_0 g_t \in C^{\boldsymbol{\nu}}(\mathbb{R}^d)$.

It remains to show that $P_0 g_t \in \mathcal{H}_d^{\boldsymbol{\nu}}$ by estimating its norm. We have

$$\begin{aligned} \|P_0 g_t\|_{\mathcal{H}_d^{\boldsymbol{\nu}}}^2 &= \sum_{\boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{1}{\gamma_{\boldsymbol{\eta}}} \int_{\mathbb{R}^d} |D^{\boldsymbol{\eta}}[P_0 g_t(\mathbf{y})]|^2 \psi_{\boldsymbol{\eta}}(\mathbf{y}_{\boldsymbol{\eta}}) \rho_{-\boldsymbol{\eta}}(\mathbf{y}_{-\boldsymbol{\eta}}) d\mathbf{y} \\ &= \int_{U_t} |\Phi_0(\xi(t, \mathbf{y}))|^2 \rho(\mathbf{y}) d\mathbf{y} + \sum_{\mathbf{0} \neq \boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{1}{\gamma_{\boldsymbol{\eta}}} \int_{U_t} \left| \sum_{j=1}^{J_{0,\boldsymbol{\eta}}} h_{0,\boldsymbol{\eta}}^{[j]}(t, \mathbf{y}) \right|^2 \psi_{\boldsymbol{\eta}}(\mathbf{y}_{\boldsymbol{\eta}}) \rho_{-\boldsymbol{\eta}}(\mathbf{y}_{-\boldsymbol{\eta}}) d\mathbf{y} \\ &\leq 1 + \sum_{\mathbf{0} \neq \boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{J_{0,\boldsymbol{\eta}}}{\gamma_{\boldsymbol{\eta}}} \sum_{j=1}^{J_{0,\boldsymbol{\eta}}} \int_{U_t} |h_{0,\boldsymbol{\eta}}^{[j]}(t, \mathbf{y})|^2 \psi_{\boldsymbol{\eta}}(\mathbf{y}_{\boldsymbol{\eta}}) \rho_{-\boldsymbol{\eta}}(\mathbf{y}_{-\boldsymbol{\eta}}) d\mathbf{y} \\ &\leq 1 + \sum_{\mathbf{0} \neq \boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{(8^{|\boldsymbol{\eta}|-1}(|\boldsymbol{\eta}| - 1)!)^2 B_{0,\boldsymbol{\eta}}}{\gamma_{\boldsymbol{\eta}}} < \infty, \end{aligned}$$

where we used the assumption (3.12) with $q = 0$. This completes the proof. \square

We remark that it would suffice to have $\nu_0 := |\boldsymbol{\nu}|$ and $\rho_0 \in C^{|\boldsymbol{\nu}|-1}(\mathbb{R})$ in Assumption 1 for Theorem 3.2 to hold. In other words, we have assumed an extra order

of regularity on ϕ and ρ_0 with respect to y_0 beyond that required for the cdf. The extra regularity is needed for the corresponding theorem for the pdf; see Theorem 3.3 below.

The bound on the norm (3.13) with constants $\{B_{0,\eta}\}$ (e.g., obtained from information on specific problems) can be used to choose the weight parameters $\{\gamma_\eta\}$ to model the relative importance of subsets of variables. This would in turn allow for a complete error analysis that is explicit in the dependence on dimension. Performing this analysis for specific problems will be pursued in future work.

3.2. Smoothing by preintegration for Dirac δ distributions. In this section, we show that the same smoothing by preintegration theory also works for distributions that are constructed by a Dirac δ function, which will allow us to also estimate the pdf as formulated in (1.3).

For $t \in [a, b]$, consider a distribution of the form

$$g_t(y_0, \mathbf{y}) = \delta(t - \phi(y_0, \mathbf{y})),$$

where $\delta(\cdot)$ is the Dirac δ function as characterized by (1.4) and $\phi: \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ satisfies Assumption 1.

For $t \in [a, b]$, $\mathbf{y} \in U_t$, and assuming $U_t \neq \emptyset$, we have $t \in \phi(\mathbb{R}, \mathbf{y})$. Let $\xi(t, \mathbf{y})$ be the unique point of discontinuity in dimension 0 as in (3.3). Applying the preintegration operator (1.6) to the distribution g_t and using the change of variables $z = \phi(y_0, \mathbf{y})$ so that $y_0 = \xi(z, \mathbf{y})$, we obtain

$$\begin{aligned} P_0 g_t(\mathbf{y}) &= \int_{-\infty}^{\infty} \delta(t - \phi(y_0, \mathbf{y})) \rho_0(y_0) dy_0 = \int_{\phi(\mathbb{R}, \mathbf{y})} \delta(t - z) \rho_0(\xi(z, \mathbf{y})) \frac{\partial}{\partial z} \xi(z, \mathbf{y}) dz \\ (3.16) \quad &= \int_{\phi(\mathbb{R}, \mathbf{y})} \delta(t - z) \rho_0(\xi(z, \mathbf{y})) \frac{1}{D^0 \phi(\xi(z, \mathbf{y}), \mathbf{y})} dz = \frac{\rho_0(\xi(t, \mathbf{y}))}{D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})}, \end{aligned}$$

where we used (3.3), (3.7), and the definition of the $\delta(\cdot)$ function (1.4). For $\mathbf{y} \in \mathbb{R}^d \setminus U_t$ and so $t \notin \phi(\mathbb{R}, \mathbf{y})$, we have $\delta(t - z) = 0$ for all $z \in \phi(\mathbb{R}, \mathbf{y})$, and hence $P_0 g_t(\mathbf{y}) = 0$. As expected, (3.16) is the derivative of (3.8) with respect to t .

With a similar proof to Theorem 3.2, we now show that the preintegrated distribution $P_0 g_t$ is also smooth, with a different bound on its norm.

THEOREM 3.3. *Let $d \geq 1$, $\boldsymbol{\nu} \in \mathbb{N}_0^d$, and $[a, b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumptions 1 and 2 for $q = 1$ and all $\boldsymbol{\eta} \leq \boldsymbol{\nu}$. Then for $t \in [a, b]$, the distribution*

$$g_t(y_0, \mathbf{y}) := \delta(t - \phi(y_0, \mathbf{y})) \quad \text{satisfies} \quad P_0 g_t \in \mathcal{H}_d^\nu \cap C^\nu(\mathbb{R}^d),$$

with its \mathcal{H}_d^ν -norm bounded uniformly in t ,

$$(3.17) \quad \sup_{t \in [a, b]} \|P_0 g_t\|_{\mathcal{H}_d^\nu} \leq \left(\sum_{\boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{(8^{|\boldsymbol{\eta}|} |\boldsymbol{\eta}|!)^2 B_{1,\boldsymbol{\eta}}}{\gamma_\boldsymbol{\eta}} \right)^{1/2} < \infty.$$

Proof. From (3.16) the preintegrated distribution can be written as

$$P_0 g_t(\mathbf{y}) = \begin{cases} \frac{\rho_0(\xi(t, \mathbf{y}))}{D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})} & \text{if } \mathbf{y} \in U_t, \\ 0 & \text{if } \mathbf{y} \in \mathbb{R}^d \setminus U_t. \end{cases}$$

The proof follows the same strategy as the proof of Theorem 3.2, but now with $q = 1$. Again, $P_0 g_t \equiv 0$ for the case $U_t = \emptyset$, so the result holds trivially.

For $\boldsymbol{\eta} \leq \boldsymbol{\nu}$, we first prove by induction on $|\boldsymbol{\eta}|$ that the $\boldsymbol{\eta}$ th derivative of $P_0 g_t$ is given by (3.14) with $q = 1$ (instead of 0), where each $h_{1,\boldsymbol{\eta}}^{[j]}$ is of the form (3.10) with $q = 1$ and $J_{1,\boldsymbol{\eta}} \leq 8^{|\boldsymbol{\eta}|} |\boldsymbol{\eta}|!$. For the base case $\boldsymbol{\eta} = \mathbf{0}$, we take $r = 0$ (and so there are no α_ℓ terms), $\beta = 0$, and $J_{1,\mathbf{0}} = 1$ to recover the single function $\rho_0(\xi(t, \mathbf{y})) / D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})$. For the inductive step, Lemma B.1 with $q = 1$ implies that (3.15) holds with $q = 1$. Since in this case $K_{1,\boldsymbol{\eta}} \leq 8^{|\boldsymbol{\eta}|} |\boldsymbol{\eta}| - 3$, the total number of functions satisfies

$$J_{1,\boldsymbol{\eta}+\mathbf{e}_i} = J_{1,\boldsymbol{\eta}} K_{1,\boldsymbol{\eta}} \leq 8^{|\boldsymbol{\eta}|} |\boldsymbol{\eta}|! (8^{|\boldsymbol{\eta}|} |\boldsymbol{\eta}| + 3) \leq 8^{|\boldsymbol{\eta}+\mathbf{e}_i|} |\boldsymbol{\eta}+\mathbf{e}_i|!,$$

as required. This completes the induction proof for (3.14) with $q = 1$.

Since each $h_{1,\boldsymbol{\eta}}^{[j]}(t, \cdot) \in C(U_t)$ and by (3.11), it follows that $P_0 g_t \in C^\nu(U_t)$ with $D^\boldsymbol{\eta} P_0 g_t(\mathbf{y}) \rightarrow 0$ as $\mathbf{y} \rightarrow \partial U_t$ for all $\boldsymbol{\eta} \leq \boldsymbol{\nu}$. Hence, [18, Lemma 9] again implies that $P_0 g_t \in C^\nu(\mathbb{R}^d)$.

Finally, it remains to show that $P_0 g_t \in \mathcal{H}_d^\nu$. The norm of $P_0 g_t$ is given by

$$\begin{aligned} \|P_0 g_t\|_{\mathcal{H}_d^\nu}^2 &= \sum_{\boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{1}{\gamma_\boldsymbol{\eta}} \int_{U_t} \left| \sum_{j=1}^{J_{1,\boldsymbol{\eta}}} h_{1,\boldsymbol{\eta}}^{[j]}(\mathbf{y}) \right|^2 \psi_\boldsymbol{\eta}(\mathbf{y}_\boldsymbol{\eta}) \rho_{-\boldsymbol{\eta}}(\mathbf{y}_{-\boldsymbol{\eta}}) d\mathbf{y} \\ &\leq \sum_{\boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{J_{1,\boldsymbol{\eta}}}{\gamma_\boldsymbol{\eta}} \sum_{j=1}^{J_{1,\boldsymbol{\eta}}} \int_{U_t} |h_{1,\boldsymbol{\eta}}^{[j]}(\mathbf{y})|^2 \psi_\boldsymbol{\eta}(\mathbf{y}_\boldsymbol{\eta}) \rho_{-\boldsymbol{\eta}}(\mathbf{y}_{-\boldsymbol{\eta}}) d\mathbf{y} \leq \sum_{\boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{(8^{|\boldsymbol{\eta}|} |\boldsymbol{\eta}|!)^2 B_{1,\boldsymbol{\eta}}}{\gamma_\boldsymbol{\eta}} < \infty, \end{aligned}$$

where we used the assumption (3.12) with $q = 1$. This completes the proof. \square

4. Distribution function and density estimators. In this section we briefly outline the QMC with preintegration algorithms for approximating the cdf and pdf. First, note that the cdf and pdf can be written as d -dimensional integrals after carrying out the preintegration step. Explicitly, by Fubini's theorem we can write the representation (1.2) for the cdf as

$$(4.1) \quad F(t) = \int_{\mathbb{R}^d} P_0(\text{ind}(t - \phi(\cdot, \mathbf{y}))) \rho(\mathbf{y}) d\mathbf{y} = \int_{U_t} \Phi_0(\xi(t, \mathbf{y})) \rho(\mathbf{y}) d\mathbf{y},$$

where in the last step we have substituted in the simplified formula (3.8) for a preintegrated indicator function. Similarly, using the representation (1.3) along with (3.16) we can write

$$(4.2) \quad f(t) = \int_{\mathbb{R}^d} P_0(\delta(t - \phi(\mathbf{y}))) \rho(\mathbf{y}) d\mathbf{y} = \int_{U_t} \frac{\rho_0(\xi(t, \mathbf{y}))}{D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})} \rho(\mathbf{y}) d\mathbf{y}.$$

4.1. Pointwise approximation. As a start, we consider approximating F and f pointwise at $t \in [a, b]$. Applying an N -point QMC rule (2.3) to the d -dimensional integrals (4.1) and (4.2), we obtain the approximations \hat{F}_N and \hat{f}_N as follows:

$$(4.3) \quad F(t) \approx \hat{F}_N(t) := Q_{d,N}(\Phi_0(\xi(t, \cdot))) = \frac{1}{N} \sum_{n=0}^{N-1} \Phi_0(\xi(t, \boldsymbol{\tau}_n)),$$

$$(4.4) \quad f(t) \approx \hat{f}_N(t) := Q_{d,N}\left(\frac{\rho_0(\xi(t, \cdot))}{D^0 \phi(\xi(t, \cdot), \cdot)}\right) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\rho_0(\xi(t, \boldsymbol{\tau}_n))}{D^0 \phi(\xi(t, \boldsymbol{\tau}_n), \boldsymbol{\tau}_n)}.$$

We are particularly interested in using randomly shifted lattice points (2.2) with (2.4), but the description in this section applies equally to other QMC rules.

Since a randomly shifted lattice rule $Q_{d,N}$ is an unbiased estimator of the d -dimensional integral, it follows that the estimators \hat{F}_N and \hat{f}_N are also *unbiased*. However, this assumes that we can compute the point of discontinuity $\xi(t, \cdot)$ exactly, which is not generally true. In practice, we must often approximate this $\xi(t, \cdot)$ by some numerical approximation, as performed in [18, 23] for the case where \mathbf{Y} is a multivariate normal vector. This leads to *biased* estimators \tilde{F}_N and \tilde{f}_N . We now detail how to construct these biased estimators efficiently.

First, consider approximating the cdf F at $t \in [a, b]$. For each transformed QMC point τ_n we must (i) find the point of discontinuity $\xi(t, \tau_n)$, and then (ii) evaluate Φ_0 at this point. In practice, these two actions must be performed numerically; however, we stress that we only need to work with the univariate function

$$\phi_{0,n} := \phi(\cdot, \tau_n),$$

which can be evaluated efficiently for multiple inputs if we “precompute” and store the contribution of τ_n to $\phi_{0,n}$. As a trivial example to demonstrate this, if we have a product function $\phi(\cdot, \tau_n) = p_0(\cdot) \prod_{i=1}^d p_i(\tau_{n,i})$, then evaluating ϕ in general has a cost of $\mathcal{O}(d)$, but if we precompute and store the product involving τ_n , then we can evaluate $\phi_{0,n}$ for K different inputs with a cost of $\mathcal{O}(K + d)$ instead of $\mathcal{O}(Kd)$. We assume that $\phi'_{0,n} = D^0 \phi(\cdot, \tau_n)$ can be evaluated directly, and we also precompute and store the contribution of τ_n to $\phi'_{0,n}$.

To find the point of discontinuity, we use a numerical root-finding algorithm, e.g., Newton’s method. Since $\phi \in C^{(\nu_0, \nu)}(\mathbb{R}^{d+1})$ with $\nu_0 = |\nu| + 1$, we have $\phi_{0,n} \in C^{\nu_0}(\mathbb{R})$ for each τ_n . If $|\nu| \geq 1$, then $\phi_{0,n} \in C^2(\mathbb{R})$ and Newton’s method converges quadratically, so in practice only a few iterations are required. Alternatively, if the higher-order derivatives of $\phi_{0,n}$ can be computed explicitly, then a higher-order Householder method can instead be used. We denote the numerical approximation of ξ by $\tilde{\xi}$.

If ρ_0 is a Gaussian distribution, then fast and accurate approximations of its cdf Φ_0 are readily available. Otherwise, if we cannot evaluate Φ_0 easily, then we approximate the one-dimensional integral $\Phi_0(y_0) = \int_{-\infty}^{y_0} \rho_0(z) dz$ by a quadrature rule. In both cases we denote the approximation of Φ_0 by $\tilde{\Phi}_0$.

Approximating the pdf f at $t \in [a, b]$ is similar: (i) obtaining the point $\xi(t, \tau_n)$ is the same, while (ii) evaluating the ratio $\rho_0/\phi'_{0,n}$ instead of the one-dimensional integral for Φ_0 is slightly simpler.

The QMC with approximate preintegration estimators of the cdf F and pdf f are

$$(4.5) \quad \tilde{F}_N(t) := \frac{1}{N} \sum_{n=0}^{N-1} \tilde{\Phi}_0(\tilde{\xi}(t, \tau_n)),$$

$$(4.6) \quad \tilde{f}_N(t) := \frac{1}{N} \sum_{n=0}^{N-1} \frac{\rho_0(\tilde{\xi}(t, \tau_n))}{D^0 \phi(\tilde{\xi}(t, \tau_n), \tau_n)}.$$

Algorithms 4.1 and 4.2 give explicit implementations of (4.5) and (4.6).

4.2. Cost of pointwise approximation. First, in the special case where the point of discontinuity $\xi(t, \cdot)$ and the one-dimensional integral Φ_0 can be computed analytically, we have $\text{cost}(\tilde{F}_N(t)) = \mathcal{O}(N)$ and $\text{cost}(\tilde{f}_N(t)) = \mathcal{O}(N)$. However, as mentioned above, this is not the typical case in practice, and we must approximate these quantities by numerical root-finding and quadrature methods.

Algorithm 4.1 Pointwise cdf estimator.

Given $t \in [a, b]$, $N \in \mathbb{N}$, and $\{\tau_n\}_{n=0}^{N-1}$ a transformed d -dimensional QMC point set:

- 1: Initialize: $\tilde{F}_N(t) \leftarrow 0$
- 2: **for** $n = 0, 1, \dots, N-1$ **do**
- 3: Precompute the contribution of τ_n to $\phi_{0,n} = \phi(\cdot, \tau_n)$ and $\phi'_{0,n} = D^0 \phi(\cdot, \tau_n)$
- 4: Compute the point of discontinuity $\tilde{\xi}(t, \tau_n)$
- 5: Approximate the one-dimensional integral $\tilde{\Phi}_0(\tilde{\xi}(t, \tau_n)) \approx \int_{-\infty}^{\tilde{\xi}(t, \tau_n)} \rho_0(z) dz$
- 6: Sum: $\tilde{F}_N(t) \leftarrow \tilde{F}_N(t) + \tilde{\Phi}_0(\tilde{\xi}(t, \tau_n))$
- 7: **end for**
- 8: Average: $\tilde{F}_N(t) \leftarrow \tilde{F}_N(t)/N$

Algorithm 4.2 Pointwise pdf estimator.

Given $t \in [a, b]$, $N \in \mathbb{N}$, and $\{\tau_n\}_{n=0}^{N-1}$ a transformed d -dimensional QMC point set:

- 1: Initialize: $\tilde{f}_N(t) \leftarrow 0$
- 2: **for** $n = 0, 1, \dots, N-1$ **do**
- 3: Precompute the contribution of τ_n to $\phi_{0,n} = \phi(\cdot, \tau_n)$ and $\phi'_{0,n} = D^0 \phi(\cdot, \tau_n)$
- 4: Compute the point of discontinuity $\tilde{\xi}(t, \tau_n)$
- 5: Sum: $\tilde{f}_N(t) \leftarrow \tilde{f}_N(t) + \frac{\rho_0(\tilde{\xi}(t, \tau_n))}{D^0 \phi(\tilde{\xi}(t, \tau_n), \tau_n)}$
- 6: **end for**
- 7: Average: $\tilde{f}_N(t) \leftarrow \tilde{f}_N(t)/N$

To analyze the cost of the pointwise approximations $\tilde{F}_N(t)$ and $\tilde{f}_N(t)$ in Algorithms 4.1 and 4.2, we assume that the number of evaluations of $\phi_{0,n}$ and $\phi'_{0,n}$ in the root-finding method to compute $\tilde{\xi}(t, \tau_n)$ for each n in step 4 is bounded by K_{root} , which is assumed to be independent of n . For the cdf approximation in Algorithm 4.1, we also assume that $\text{cost}(\rho_0) = \mathcal{O}(1)$, and that the number of quadrature points to compute the one-dimensional integral in step 5 is bounded by K_{quad} , also independently of n .

Then to more concretely illustrate why it is important to precompute the contribution of τ_n to $\phi_{0,n}$ and $\phi'_{0,n}$, we make the following assumption about the difference in cost in evaluating ϕ and $D^0 \phi$ compared with the univariate functions $\phi_{0,n}$ and $\phi'_{0,n}$:

$$(4.7) \quad \begin{cases} \text{cost}(\phi) = \$ (d), & \text{cost}(\phi_{0,n}) = \$ (1) \text{ with precomputed contribution of } \tau_n, \\ \text{cost}(D^0 \phi) = \$ (d), & \text{cost}(\phi'_{0,n}) = \$ (1) \text{ with precomputed contribution of } \tau_n, \end{cases}$$

for some nondecreasing function $\$: \mathbb{N} \rightarrow \mathbb{N}$.

The cost of Algorithms 4.1 and 4.2 is then

$$\begin{aligned} \text{cost}(\tilde{F}_N(t)) &= \mathcal{O}(N [\$(d) + K_{\text{root}} \$(1) + K_{\text{quad}}]) \quad \text{and} \\ \text{cost}(\tilde{f}_N(t)) &= \mathcal{O}(N [\$(d) + K_{\text{root}} \$(1)]). \end{aligned}$$

For large d , this will be much more efficient than a naive implementation without precomputed contribution of τ_n , which would have

$$\text{cost}(\tilde{F}_N(t)) = \mathcal{O}(N [K_{\text{root}} \$(d) + K_{\text{quad}}]) \quad \text{and} \quad \text{cost}(\tilde{f}_N(t)) = \mathcal{O}(N K_{\text{root}} \$(d)).$$

4.3. Approximating the cdf and pdf on an interval. Now we outline the full QMC with preintegration method for approximating the cdf and pdf on $[a, b]$, obtained by applying Lagrange interpolation L_M based on points $\{t_m\}_{m=0}^M \subset [a, b]$ to the pointwise estimators \hat{F}_N and \hat{f}_N . We denote the approximations by

$$(4.8) \quad \hat{F}_{N,M} := L_M(\hat{F}_N) = L_M\left(Q_{d,N}(\Phi_0(\xi(\bullet, \cdot)))\right),$$

$$(4.9) \quad \hat{f}_{N,M} := L_M(\hat{f}_N) = L_M\left(Q_{d,N}\left(\frac{\rho_0(\xi(\bullet, \cdot))}{D^0\phi(\xi(\bullet, \cdot), \cdot)}\right)\right),$$

where the QMC rule $Q_{d,N}$ acts on a function with respect to \cdot whereas Lagrange interpolation L_M acts on \bullet . As discussed in section 2.3 we will use Chebyshev points, but the description below allows for any set of distinct interpolation nodes.

In practice, we must approximate the point of discontinuity by $\xi \approx \tilde{\xi}$, and for the cdf also the one-dimensional integral by $\tilde{\Phi}_0 \approx \Phi_0$. This leads to the biased estimators

$$(4.10) \quad \tilde{F}_{N,M} := L_M(\tilde{F}_N) = L_M\left(Q_{d,N}(\tilde{\Phi}_0(\tilde{\xi}(\bullet, \cdot)))\right),$$

$$(4.11) \quad \tilde{f}_{N,M} := L_M(\tilde{f}_N) = L_M\left(Q_{d,N}\left(\frac{\rho_0(\tilde{\xi}(\bullet, \cdot))}{D^0\phi(\tilde{\xi}(\bullet, \cdot), \cdot)}\right)\right).$$

Recall from the definition of the Lagrange interpolation operator (2.7) that to construct the estimators $\tilde{F}_{N,M}$ and $\tilde{f}_{N,M}$, we must compute the pointwise approximations \tilde{F}_N and \tilde{f}_N at all of the interpolation nodes $\{t_m\}_{m=0}^M$. One way to implement the estimator $\tilde{F}_{N,M}$ as in (4.10) is to simply run Algorithm 4.1 for each t_m for $m = 0, 1, \dots, M$ with $\text{cost}(\tilde{F}_{N,M}) = (M+1) \times \text{cost}(\tilde{F}_N)$. However, since we can use the same QMC rule for each interpolation node t_m , it is more efficient to instead *vectorize* Algorithm 4.1 and utilize precomputed contributions of each point τ_n so that we only have to deal with $M+1$ univariate functions. Similar arguments can also be made for the cdf estimator $\tilde{f}_{N,M}$. Explicit algorithms detailing how to construct the estimators $\tilde{F}_{N,M}$ and $\tilde{f}_{N,M}$ are given in Algorithms 4.3 and 4.4.

4.4. Cost of full cdf and pdf estimators. Following the analysis of the cost of the pointwise estimators in section 4.2, we again assume the cost model (4.7) and assume that the number of evaluations of the univariate functions in the root-finding method and the number of quadrature points for computing the one-dimensional integral are bounded by K_{root} and K_{quad} , respectively, which are additionally assumed to be independent of n and m . The cost of Algorithms 4.3 and 4.4 is then

$$\begin{aligned} \text{cost}(\tilde{F}_{N,M}) &= \mathcal{O}(N[\$(d) + M K_{\text{root}} \$(1) + M K_{\text{quad}}]), \\ \text{cost}(\tilde{f}_{N,M}) &= \mathcal{O}(N[\$(d) + M K_{\text{root}} \$(1)]). \end{aligned}$$

To once again illustrate the importance of the precomputation step, we note that a naïve implementation that simply evaluates ϕ at all of its components each time would have $\text{cost}(\tilde{f}_{N,M}) = \mathcal{O}(N M K_{\text{root}} \$(d))$.

5. Error analysis.

5.1. Regularity of F and f . In order to utilize the results on the error for interpolation on $[a, b]$ from section 2.3, we need to know quantitatively how smooth

Algorithm 4.3 cdf estimator.

Given $M \in \mathbb{N}$, $\{t_m\}_{m=0}^M \subset [a, b]$, $N \in \mathbb{N}$, and $\{\tau_n\}_{n=0}^{N-1}$ a transformed d -dimensional QMC point set:

- 1: Initialize: $\tilde{F}_N(t_m) \leftarrow 0$ for each $m = 0, 1, \dots, M$
- 2: **for** $n = 0, 1, \dots, N-1$ **do**
- 3: Precompute the contribution of τ_n to $\phi_{0,n} = \phi(\cdot, \tau_n)$ and $\phi'_{0,n} = D^0 \phi(\cdot, \tau_n)$
- 4: **for** $m = 0, 1, \dots, M$ **do**
- 5: Compute the point of discontinuity $\tilde{\xi}(t_m, \tau_n)$
- 6: Approximate the one-dimensional integral $\tilde{\Phi}_0(\tilde{\xi}(t_m, \tau_n)) \approx \int_{-\infty}^{\tilde{\xi}(t_m, \tau_n)} \rho_0(z) dz$
- 7: Sum: $\tilde{F}_N(t_m) \leftarrow \tilde{F}_N(t_m) + \tilde{\Phi}_0(\tilde{\xi}(t_m, \tau_n))$
- 8: **end for**
- 9: **end for**
- 10: Average: $\tilde{F}_N(t_m) \leftarrow \tilde{F}_N(t_m)/N$ for each $m = 0, 1, \dots, M$
- 11: Interpolate: $\tilde{F}_{N,M} \leftarrow \sum_{m=0}^M \tilde{F}_N(t_m) \chi_{M,m}$

Algorithm 4.4 pdf estimator.

Given $M \in \mathbb{N}$, $\{t_m\}_{m=0}^M \subset [a, b]$, $N \in \mathbb{N}$, and $\{\tau_n\}_{n=0}^{N-1}$ a transformed d -dimensional QMC point set:

- 1: Initialize: $\tilde{f}_N(t_m) \leftarrow 0$ for each $m = 0, 1, \dots, M$
- 2: **for** $n = 0, 1, \dots, N-1$ **do**
- 3: Precompute the contribution of τ_n to $\phi_{0,n} = \phi(\cdot, \tau_n)$ and $\phi'_{0,n} = D^0 \phi(\cdot, \tau_n)$
- 4: **for** $m = 0, 1, \dots, M$ **do**
- 5: Compute the point of discontinuity $\tilde{\xi}(t_m, \tau_n)$
- 6: Sum: $\tilde{f}_N(t_m) \leftarrow \tilde{f}_N(t_m) + \frac{\rho_0(\tilde{\xi}(t_m, \tau_n))}{D^0 \phi(\tilde{\xi}(t_m, \tau_n), \tau_n)}$
- 7: **end for**
- 8: **end for**
- 9: Average: $\tilde{f}_N(t_m) \leftarrow \tilde{f}_N(t_m)/N$ for each $m = 0, 1, \dots, M$
- 10: Interpolate: $\tilde{f}_{N,M} \leftarrow \sum_{m=0}^M \tilde{f}_N(t_m) \chi_{M,m}$

the cdf F and pdf f are with respect to t . Clearly this smoothness will depend on the smoothness of the original transformation ϕ from (1.1). Since in Assumption 1 we assume that ϕ is $|\nu| + 1$ times differentiable with respect to variable y_0 and $\rho_0 \in C^{|\nu|}(\mathbb{R})$, we can expect a similar level of smoothness for F and f .

To see the dependence on t more explicitly, recall that the formulas (1.2) for cdf and (1.3) for the pdf can be formulated as d -dimensional integrals as in (4.1) and (4.2), respectively. From these formulas, it is then clear that the smoothness of F and f depends on the smoothness of ξ , which in turn depends on the smoothness of ϕ . In particular, Theorem 3.1 implies that ξ is as smooth (with respect to t) as ϕ (with respect to y_0).

Note that the assumptions we make here on the smoothness of ϕ and ρ_0 are the same as those required for the preintegration step, i.e., we do not need any further smoothness assumptions.

THEOREM 5.1. *Let $d \geq 1$, $\nu \in \mathbb{N}_0^d$, and $[a, b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumptions 1 and 2 for $\eta = \mathbf{0}$ and all $q = 1, 2, \dots, |\nu| + 1$. Assume additionally*

that $U_t = \mathbb{R}^d$ for all $t \in [a, b]$. Then $F \in W^{|\nu|+1, \infty}[a, b]$ and $f \in W^{|\nu|, \infty}[a, b]$, and for $q = 1, \dots, |\nu| + 1$ the derivatives are bounded by

$$\|F^{(q)}\|_{L^\infty} = \|f^{(q-1)}\|_{L^\infty} \leq 3^{q-1} (q-1)! B_{q, \mathbf{0}}^{1/2}.$$

Proof. We prove that the cdf satisfies $F \in W^{|\nu|+1, \infty}[a, b]$. Then since $f = F'$, the result for the pdf follows immediately.

Consider the derivative of order $q \leq |\nu| + 1$. First, differentiating (4.1) with respect to t and applying the Leibniz rule [18, Theorem 4] q times, we obtain

$$(5.1) \quad F^{(q)}(t) = \int_{\mathbb{R}^d} \frac{\partial^q}{\partial t^q} \Phi_0(\xi(t, \mathbf{y})) \rho(\mathbf{y}) d\mathbf{y}.$$

Recall that we have

$$(5.2) \quad \frac{\partial}{\partial t} \Phi_0(\xi(t, \mathbf{y})) = \frac{\rho_0(\xi(t, \mathbf{y}))}{D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})}.$$

We now prove by induction on $q \geq 1$ that

$$(5.3) \quad \frac{\partial^q}{\partial t^q} \Phi_0(\xi(t, \mathbf{y})) = \sum_{j=1}^{J_{q, \mathbf{0}}} h_{q, \mathbf{0}}^{[j]}(t, \mathbf{y}) \quad \text{with} \quad J_{q, \mathbf{0}} \leq 3^{q-1} (q-1)!,$$

where each function $h_{q, \mathbf{0}}^{[j]}$ is of the form (3.10) with $\boldsymbol{\eta} = \mathbf{0}$. The base step $q = 1$ holds for the single function (5.2) with $r = 0$ (no $\boldsymbol{\alpha}$), $\beta = 0$, and $J_{1, \mathbf{0}} = 1$. Suppose next that (5.3) holds for some $q \geq 1$. Then we have

$$\frac{\partial}{\partial t} \left(\frac{\partial^q}{\partial t^q} \Phi_0(\xi(t, \mathbf{y})) \right) = \sum_{j=1}^{J_{q, \mathbf{0}}} \frac{\partial}{\partial t} h_{q, \mathbf{0}}^{[j]}(t, \mathbf{y}) = \sum_{j=1}^{J_{q, \mathbf{0}}} \sum_{k=1}^{K_{q, \mathbf{0}}} h_{q+1, \mathbf{0}}^{[j, k]}(t, \mathbf{y}) = \sum_{j'=1}^{J_{q+1, \mathbf{0}}} h_{q+1, \mathbf{0}}^{[j']} (t, \mathbf{y}).$$

In the second equality we used Lemma B.2 in Appendix B with $\boldsymbol{\eta} = \mathbf{0}$, which states that each function $\frac{\partial}{\partial t} h_{q, \mathbf{0}}^{[j]}$ can be written as a sum of $K_{q, \mathbf{0}} \leq 3q - 1$ functions of the form (3.10), with q replaced by $q + 1$. We enumerated these functions with the notation $h_{q+1, \mathbf{0}}^{[j, k]}$ and then relabeled all functions for different combinations of indices j and k with the notation $h_{q+1, \mathbf{0}}^{[j']}$. The total number of functions satisfies

$$J_{q+1, \mathbf{0}} = J_{q, \mathbf{0}} K_{q, \mathbf{0}} \leq 3^{q-1} (q-1)! (3q-1) \leq 3^q q!,$$

as required. This completes the induction proof for (5.3).

Thus we have

$$f^{(q-1)}(t) = F^{(q)}(t) = \sum_{j=1}^{J_{q, \mathbf{0}}} \int_{\mathbb{R}^d} h_{q, \mathbf{0}}^{[j]}(t, \mathbf{y}) \rho(\mathbf{y}) d\mathbf{y}$$

and

$$\|f^{(q-1)}\|_{L^\infty} = \|F^{(q)}\|_{L^\infty} \leq \sum_{j=1}^{J_{q, \mathbf{0}}} \sup_{t \in [a, b]} \int_{\mathbb{R}^d} |h_{q, \mathbf{0}}^{[j]}(t, \mathbf{y})| \rho(\mathbf{y}) d\mathbf{y} \leq 3^{q-1} (q-1)! B_{q, \mathbf{0}}^{1/2},$$

where we used assumption (3.12) with $\boldsymbol{\eta} = \mathbf{0}$ and the Cauchy-Schwarz inequality. Hence $F \in W^{|\nu|+1, \infty}[a, b]$ and also $f = F' \in W^{|\nu|, \infty}[a, b]$. \square

Remark 5.2. In Theorem 5.1, we have assumed for simplicity that $U_t = \mathbb{R}$ for all $t \in [a, b]$, which implies that for each $\mathbf{y} \in \mathbb{R}^d$ and $t \in [a, b]$ there is some $y_0 \in \mathbb{R}$ such that $\phi(y_0, \mathbf{y}) = t$. This can be viewed as a restriction on the interval $[a, b]$.

5.2. Error of cdf and pdf estimators. In this subsection we analyze the error of the unbiased estimators from section 4. First, we prove bounds for the RMSE of the pointwise estimators \widehat{F}_N and \widehat{f}_N . Then we bound the *root-mean-integrated-square error (RMISE)* of the full estimators $\widehat{F}_{N,M}$ and $\widehat{f}_{N,M}$ on $[a, b]$. Recall that the expectation in the RMSE and RMISE, which we denote by \mathbb{E}_Δ , is taken with respect to the random shift Δ in the lattice rule. In this section, we assume that the generating vector is constructed using the CBC algorithm from [24].

THEOREM 5.3 (pointwise RMSE). *Let $d \geq 1$, $\nu = \mathbf{1} \in \mathbb{N}^d$, and $[a, b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumptions 1 and 2 for all $\eta \in \{0, 1\}^d$ with $q = 0$ for the cdf case and $q = 1$ for the pdf case. Let $Q_{d,N}$ be a CBC-constructed randomly shifted lattice rule as in (2.3). Then, for $t \in [a, b]$, the estimators $\widehat{F}_N(t)$ and $\widehat{f}_N(t)$ as given in (4.3) and (4.4) satisfy, for all $\lambda \in (1/\omega, 1]$,*

$$(5.4) \quad \sqrt{\mathbb{E}_\Delta [|F(t) - \widehat{F}_N(t)|^2]} \leq C_{F,\lambda} \phi_{\text{tot}}(N)^{-1/(2\lambda)},$$

$$(5.5) \quad \sqrt{\mathbb{E}_\Delta [|f(t) - \widehat{f}_N(t)|^2]} \leq C_{f,\lambda} \phi_{\text{tot}}(N)^{-1/(2\lambda)},$$

where, with ω and c as in Theorem 2.1,

$$C_{F,\lambda} := \left(\sum_{\mathbf{0} \neq \eta \in \{0,1\}^d} \gamma_\eta^\lambda [2c\zeta(\omega\lambda)]^{|\eta|} \right)^{\frac{1}{2\lambda}} \left(1 + \sum_{\mathbf{0} \neq \eta \in \{0,1\}^d} \frac{(8^{|\eta|-1}(|\eta|-1)!)^2 B_{0,\eta}}{\gamma_\eta} \right)^{\frac{1}{2}},$$

$$C_{f,\lambda} := \left(\sum_{\mathbf{0} \neq \eta \in \{0,1\}^d} \gamma_\eta^\lambda [2c\zeta(\omega\lambda)]^{|\eta|} \right)^{\frac{1}{2\lambda}} \left(\sum_{\eta \in \{0,1\}^d} \frac{(8^{|\eta|}|\eta|!)^2 B_{1,\eta}}{\gamma_\eta} \right)^{\frac{1}{2}}.$$

Proof. First for the cdf estimator, using (4.1) and the definition (4.3) of \widehat{F}_N we can write the mean-square error as

$$\mathbb{E}_\Delta [|F(t) - \widehat{F}_N(t)|^2] = \mathbb{E}_\Delta \left[\left| \int_{\mathbb{R}^d} \Phi_0(\xi(t, \mathbf{y})) \rho(\mathbf{y}) d\mathbf{y} - Q_{d,N}(\Phi_0(\xi(t, \cdot))) \right|^2 \right].$$

Then since ϕ and ρ_0 satisfy Assumptions 1 and 2 for all $\eta \in \{0, 1\}^d$ with $q = 0$, we can apply Theorem 3.2 to show that the preintegrated function $\Phi_0(\xi(t, \cdot))$ belongs to \mathcal{H}_d^1 and its norm is bounded by (3.13) with $\nu = \mathbf{1}$. Substituting this norm bound into the CBC error bound (2.6) and taking the square root proves the desired result.

The result for the pdf estimator follows by essentially the same argument, but with $q = 1$ and using the norm bound (3.17) in Theorem 3.3 instead of Theorem 3.2.

Next, we bound the RMISE on $[a, b]$. For the cdf estimator $\widehat{F}_{N,M}$ the mean-integrated square error (MISE) is formulated as

$$\mathbb{E}_\Delta [\|F - \widehat{F}_{N,M}\|_{L^2}^2] = \mathbb{E}_\Delta \left[\int_a^b |F(t) - \widehat{F}_{N,M}(t)|^2 dt \right],$$

and similarly for $\widehat{f}_{N,M}$.

THEOREM 5.4. Let $d \geq 1$, $\boldsymbol{\nu} = \mathbf{1} \in \mathbb{N}^d$, and $[a, b] \subset \mathbb{R}$. Suppose ϕ and ρ_0 satisfy

$$\begin{cases} \text{Assumption 1; and} \\ \text{Assumption 2 for } \boldsymbol{\eta} = \mathbf{0} \text{ and all } q \leq d+1; \text{ and} \\ \text{Assumption 2 for all } \boldsymbol{\eta} \in \{0, 1\}^d \text{ with } q = 0 \text{ for the cdf case; and} \\ \text{Assumption 2 for all } \boldsymbol{\eta} \in \{0, 1\}^d \text{ with } q = 1 \text{ for the pdf case.} \end{cases}$$

Suppose also that $U_t = \mathbb{R}^d$ for all $t \in [a, b]$. Let $Q_{d,N}$ be a CBC-constructed randomly shifted lattice rule as in (2.2) and let L_M be the Lagrange interpolation operator on $[a, b]$ based on Chebyshev points as in (2.7). Then for $\sigma \in \mathbb{N}$, specified below, and $M > \sigma$, the estimators $\widehat{F}_{N,M}$ and $\widehat{f}_{N,M}$ in (4.8) and (4.9) satisfy, for all $\lambda \in (1/\omega, 1]$,

$$(5.6) \quad \sqrt{\mathbb{E}_{\Delta} [\|F - \widehat{F}_{N,M}\|_{L^2}^2]} \leq C_{F,\lambda,\sigma} \left(\phi_{\text{tot}}(N)^{-1/(2\lambda)} + M^{-\sigma} \right) \quad \text{for } \sigma \leq d,$$

$$(5.7) \quad \sqrt{\mathbb{E}_{\Delta} [\|f - \widehat{f}_{N,M}\|_{L^2}^2]} \leq C_{f,\lambda,\sigma} \left(\phi_{\text{tot}}(N)^{-1/(2\lambda)} + M^{-\sigma} \right) \quad \text{for } \sigma \leq d-1,$$

where $C_{F,\lambda,\sigma} := \sqrt{2(b-a)} \max(C_{F,\lambda}, C_{F,\sigma})$, $C_{f,\lambda,\sigma} := \sqrt{2(b-a)} \max(C_{f,\lambda}, C_{f,\sigma})$, with $C_{F,\lambda}$, $C_{f,\lambda}$ as in Theorem 5.3, $\omega \in (1, 2]$ as in Theorem 2.1, and

$$\begin{aligned} C_{F,\sigma} &:= \frac{4(b-a)}{\pi} [3(\sigma+1)]^\sigma (\sigma-1)! B_{\sigma+1,\mathbf{0}}^{1/2}, \\ C_{f,\sigma} &:= \frac{4(b-a)}{\pi} [3(\sigma+1)]^{\sigma+1} (\sigma-1)! B_{\sigma+2,\mathbf{0}}^{1/2}. \end{aligned}$$

Proof. First, consider the cdf estimator $\widehat{F}_{N,M}$. We can split the MISE into the QMC and interpolation components

$$(5.8) \quad \mathbb{E}_{\Delta} [\|F - \widehat{F}_{N,M}\|_{L^2}^2] \leq 2\mathbb{E}_{\Delta} [\|F - \widehat{F}_N\|_{L^2}^2] + 2\mathbb{E}_{\Delta} [\|\widehat{F}_N - L_M \widehat{F}_N\|_{L^2}^2].$$

The first term in (5.8) can be bounded by the pointwise error as follows. By Fubini's theorem we may swap the expected value with respect to Δ and the integral over $[a, b]$ to obtain

$$(5.9) \quad \mathbb{E}_{\Delta} [\|F - \widehat{F}_N\|_{L^2}^2] = \int_a^b \mathbb{E}_{\Delta} [|F(t) - \widehat{F}_N(t)|^2] dt \leq (b-a) C_{F,\lambda}^2 \phi_{\text{tot}}(N)^{-1/\lambda},$$

where we have substituted in the bound (5.4).

For the second term in (5.8) we will use the Lagrange interpolation error bound (2.8) by first adapting the proof of Theorem 5.1 to show that $\widehat{F}_N \in W^{\sigma+1,\infty}[a, b]$ for $\sigma \leq d$ and all random shifts Δ . Differentiating (4.3) with respect to t then substituting in the formula (5.3) for $q = \sigma + 1$ gives

$$(5.10) \quad \widehat{F}_N^{(\sigma+1)}(t) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\partial^{\sigma+1}}{\partial t^{\sigma+1}} \Phi_0(\xi(t, \boldsymbol{\tau}_n^\Delta)) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{J_{\sigma+1,\mathbf{0}}} h_{\sigma+1,\mathbf{0}}^{[j]}(t, \boldsymbol{\tau}_n^\Delta),$$

with $J_{\sigma+1,0} \leq 3^\sigma \sigma!$, where we emphasized the explicit dependence of each transformed QMC point on the random shift with the superscript Δ . Hence, we can apply (2.8) to obtain

$$\begin{aligned}
 \mathbb{E}_\Delta [\|\widehat{F}_N - L_M \widehat{F}_N\|_{L^2}^2] &\leq (b-a) \mathbb{E}_\Delta [\|\widehat{F}_N - L_M \widehat{F}_N\|_{L^\infty}^2] \\
 &\leq (b-a) \left(\frac{4}{\pi \sigma (M-\sigma)^\sigma} \right)^2 \mathbb{E}_\Delta [\|\widehat{F}_N^{(\sigma+1)}\|_{L^1}^2] \\
 (5.11) \quad &\leq (b-a) \left(\frac{4(\sigma+1)^\sigma}{\pi \sigma} \right)^2 \mathbb{E}_\Delta [\|\widehat{F}_N^{(\sigma+1)}\|_{L^1}^2] M^{-2\sigma},
 \end{aligned}$$

where we used the easily verified inequality $M-\sigma \geq M/(\sigma+1)$ for all $M \geq \sigma+1$.

To bound the expected value in (5.11), we use the formula (5.10), the Cauchy–Schwarz inequality for integral and sum, and Fubini’s theorem to obtain

$$\begin{aligned}
 \mathbb{E}_\Delta [\|\widehat{F}_N^{(\sigma+1)}\|_{L^1}^2] &\leq \frac{(b-a) J_{\sigma+1,0}}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{J_{\sigma+1,0}} \int_a^b \int_{[0,1]^d} |h_{\sigma+1,0}^{[j]}(t, \tau_n^\Delta)|^2 d\Delta dt \\
 &= \frac{(b-a) J_{\sigma+1,0}}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{J_{\sigma+1,0}} \int_a^b \int_{\mathbb{R}^d} |h_{\sigma+1,0}^{[j]}(t, \mathbf{y})|^2 \rho(\mathbf{y}) d\mathbf{y} dt \\
 (5.12) \quad &\leq (b-a)^2 J_{\sigma+1,0}^2 B_{\sigma+1,0} \leq [(b-a) 3^\sigma \sigma!]^2 B_{\sigma+1,0},
 \end{aligned}$$

where we made a change of variables and then used the upper bound (3.12), which is assumed to hold for all $q = \sigma+1 \leq |\boldsymbol{\nu}|+1 = d+1$, as well as the bound $J_{\sigma+1,0} \leq 3^\sigma \sigma!$. Substituting (5.12) into (5.11), we conclude that

$$(5.13) \quad \mathbb{E}_\Delta [\|\widehat{F}_N - L_M \widehat{F}_N\|_{L^2}^2] \leq (b-a) C_{F,\sigma}^2 M^{-2\sigma}$$

with $C_{F,\sigma}$ as defined in the theorem. Substituting (5.9) and (5.13) into (5.8), we obtain the RMISE bound (5.6) for the cdf estimator.

The result for the pdf estimator follows by essentially the same argument. The key difference is that we must instead bound the norm $\|\widehat{f}_N^{(\sigma+1)}\|_{L^1}$ for $\sigma+1 \leq |\boldsymbol{\nu}| = d$. Similar to the relationship $f = F'$, it follows from (5.2) that \widehat{f}_N is the derivative with respect to t of \widehat{F}_N . Thus

$$\widehat{f}_N^{(\sigma+1)}(t) = \widehat{F}_N^{(\sigma+2)}(t) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\partial^{\sigma+2}}{\partial t^{\sigma+2}} \Phi_0(\xi(t, \tau_n^\Delta)) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{J_{\sigma+2,0}} h_{\sigma+2,0}^{[j]}(t, \tau_n^\Delta)$$

with $J_{\sigma+2,0} \leq 3^{\sigma+1}(\sigma+1)!$. Following the same steps as before, we obtain

$$\mathbb{E}_\Delta [\|\widehat{f}_N^{(\sigma+1)}\|_{L^1}^2] \leq [(b-a) 3^{\sigma+1}(\sigma+1)!]^2 B_{\sigma+2,0},$$

and eventually arrive at the RMISE bound (5.7) for the pdf estimator. \square

Remark 5.5. In Theorem 5.4, we assume the minimal smoothness required to apply the QMC theory from Theorem 2.1 (see also [24]), i.e., $\boldsymbol{\nu} = \mathbf{1}$, which in most cases is sufficient to also handle the interpolation error. For the special case $d = 1$, the interpolation component of the error converges linearly for $\widehat{F}_{N,M}$, i.e., $\sigma = 1$ in (5.6), but does not converge for $\widehat{f}_{N,M}$, i.e., $\sigma = 0$ in (5.7). Under stricter smoothness assumptions, one can of course prove higher convergence rates for the interpolation error when $d = 1$. However, in this case, after preintegration one only needs to

approximate a one-dimensional integral, for which it is more appropriate to use a one-dimensional quadrature rule, e.g., Gauss–Hermite, instead of QMC. Such a rule would additionally exploit any higher smoothness assumptions to obtain higher-order convergence for the quadrature error.

Theorem 5.4 implies that we can take σ up to $d-1$ (or d for the cdf approximation) and obtain a very fast convergence rate in terms of M . However, as σ increases the constant increases significantly. Hence, in practice one should take a moderate value for σ , e.g., around 2–5.

To see how Theorem 5.4 applies in practice, let now N be a prime or a prime power, in which case $\phi_{\text{tot}}(N) \sim N$. Then to balance the QMC and interpolation error, Theorem 5.4 implies that we should take $M \sim N^{1/\sigma}$. The final result is that the estimators converge at a rate arbitrarily close to $1/N$. It is summarized in the following corollary.

COROLLARY 5.6. *Let $d \geq 2$ and suppose that the conditions in Theorem 5.4 hold. Let N be a prime power and choose $M \sim N^{1/\sigma}$ for a moderate $\sigma \leq d-1$. Then, for $\epsilon > 0$ there exist constants $C_{F,\epsilon}$, $C_{f,\epsilon} < \infty$ such that the error of the cdf and pdf estimators satisfies*

$$\begin{aligned}\sqrt{\mathbb{E}_{\Delta} [\|F - \widehat{F}_{N,M}\|_{L^2}^2]} &\leq C_{F,\epsilon} N^{-1+\epsilon}, \\ \sqrt{\mathbb{E}_{\Delta} [\|f - \widehat{f}_{N,M}\|_{L^2}^2]} &\leq C_{f,\epsilon} N^{-1+\epsilon}.\end{aligned}$$

6. Numerical results. To test the method, we consider approximating the cdf and pdf of a random variable $X \in \mathbb{R}$ given by a sum of $d+1$ log-normals,

$$(6.1) \quad X = \sum_{i=0}^d \exp(W_i) = \sum_{i=0}^d \exp(\mathbf{A}_i \mathbf{Y}) =: \phi(\mathbf{Y}),$$

where $\mathbf{W} = (W_i)_{i=0}^d$ is a $(d+1)$ -dimensional multivariate normal vector with mean $\mathbf{0}$ and covariance Σ . In the second equality, we have factorized the covariance matrix as $\Sigma = \mathbf{A}\mathbf{A}^\top$ and made the change of variables $\mathbf{Y} = \mathbf{A}^{-1}\mathbf{W}$, so that ϕ is a function of the $(d+1)$ -dimensional standard normal vector $\mathbf{Y} = (Y_i)_{i=0}^d$. In (6.1), \mathbf{A}_i denotes the i th row of the matrix factor \mathbf{A} . We use the *principal components* or *PCA* factorization, which is based on the eigendecomposition of Σ with the eigenvalues ordered in nonincreasing value. Clearly, X fits the setting of this paper with $\rho_i(y_i) = e^{-y_i^2/2}/\sqrt{2\pi}$.

We test the method for two covariance matrices. The first example is in $d+1 = 32$ dimensions with covariance matrix $\Sigma^{(1)}$ and the second example takes $d+1 = 64$ and a covariance matrix $\Sigma^{(2)}$ with entries that are decaying in value:

$$\Sigma_{i,j}^{(1)} = \begin{cases} 1 & \text{for } i = j, \\ \frac{1}{2} & \text{for } i \neq j, \end{cases} \quad \text{and} \quad \Sigma_{i,j}^{(2)} = \frac{1}{\max(i,j)}.$$

It can easily be verified that Assumptions 1 and 2, with the weight functions $\psi_i(y) = e^{-\delta y^2/2}$ for $\delta \in (0, \frac{1}{2})$, are satisfied for both covariance matrices. For this choice of ψ_i , Table 3 in [21] indicates that (2.5) holds with $\omega = 2(1-\delta)$, giving a QMC convergence rate of $1-\delta$. Choosing the weight parameters $\{\gamma_\eta\}$ and performing a full error analysis that is explicit in the dependence on the dimension is left for future work.

For the QMC approximations we use embedded lattice rules given by the CBC construction from [7] and which are available at [20]. Although there is no accompanying theory, these rules have been shown to work well in practice. For $\Sigma^{(1)}$, we use the generating vectors `lattice-38005-1024-1048576.5000`, constructed using equal product weights $\gamma_i = 0.05$, and for $\Sigma^{(2)}$, we use `lattice-39102-1024-1048576.3600`, constructed using decaying product weights $\gamma_i = 1/i^2$.

The final estimate is the average over $R = 32$ random shifts, and we estimate the RMSE of this average by the sample standard error over the random shifts. For a fair comparison, each MC approximation uses $R \times N$ points in total and the RMSE is estimated by the sample standard error over all MC realizations. For each τ_n , the value of $\xi(t, \tau_n)$ is computed using Newton's method with a tolerance of 10^{-10} . All computations were run on the computational cluster Katana [28] at UNSW Sydney.

Convergence results for the cdf and pdf at the point $t = 60$ for both covariance matrices are given in Figures 1 and 2 for $N = 2^{10}, 2^{11}, \dots, 2^{20}$, where we plot the

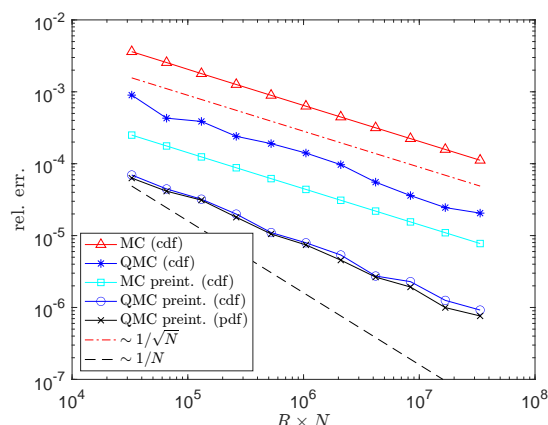


FIG. 1. Relative RMSE convergence in N for MC and QMC, with and without preintegration, for $F(60) = \mathbb{P}[X \leq 60]$, and also QMC with preintegration for $f(60)$, for $\Sigma^{(1)}$.

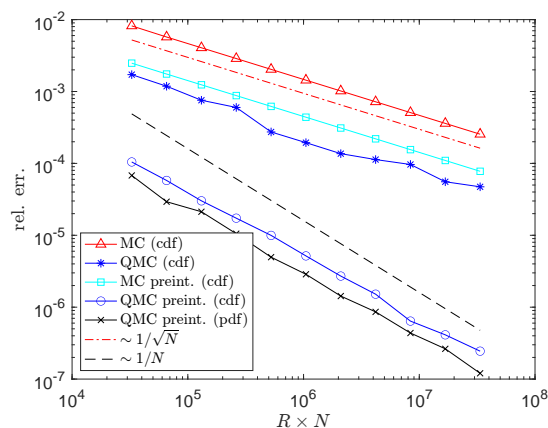


FIG. 2. Relative RMSE convergence in N for MC and QMC, with and without preintegration, for $F(60) = \mathbb{P}[X \leq 60]$, and also QMC with preintegration for $f(60)$, for $\Sigma^{(2)}$.

TABLE 1

CPU times for QMC approximations with a single random shift for $\Sigma^{(1)}$.

N	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}	2^{19}	2^{20}
QMC (cdf)	0.05	0.11	0.23	0.46	0.95	1.92	3.59	7.17
QMC preint. (cdf)	0.13	0.25	0.50	1.01	2.03	4.11	8.09	16.12
QMC preint. (pdf)	0.09	0.18	0.34	0.72	1.36	2.89	5.41	10.88
Increase factor (cdf)	2.6	2.3	2.2	2.2	2.1	2.1	2.3	2.2

TABLE 2

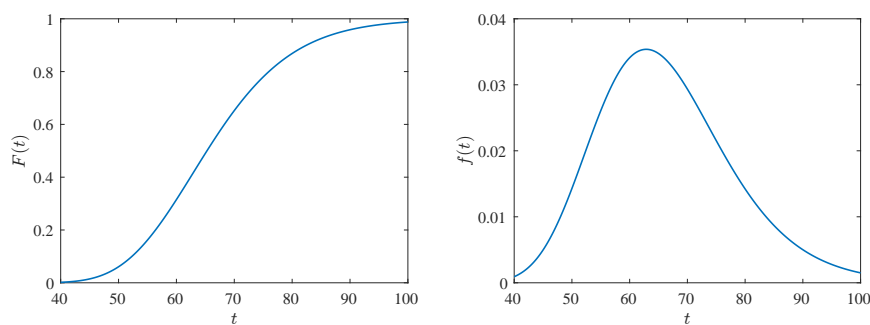
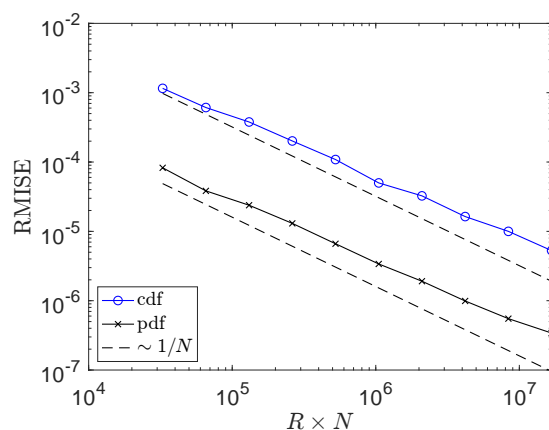
CPU times for QMC approximations with a single random shift for $\Sigma^{(2)}$.

N	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}	2^{18}	2^{19}	2^{20}
QMC (cdf)	0.08	0.17	0.32	0.67	1.28	2.60	5.14	10.22
QMC preint. (cdf)	0.15	0.31	0.63	1.31	2.49	5.03	10.04	19.83
QMC preint. (pdf)	0.12	0.23	0.49	0.92	1.92	3.71	7.49	14.68
Increase factor (cdf)	1.9	1.8	2.0	2.0	1.9	1.9	2.0	1.9

approximate relative RMSE (the RMSE divided by the estimated value). We see clearly that preintegration drastically improves the empirical results for QMC, especially for the matrix $\Sigma^{(2)}$ which has decaying eigenvalues. The results for $\Sigma^{(1)}$ are similar to those presented in [6]. A possible explanation for the better results observed for $\Sigma^{(2)}$, compared to $\Sigma^{(1)}$, is that although formally the dimension is larger, this problem may have a lower effective dimension. Since the eigenvalues of $\Sigma^{(2)}$ are decaying, this suggests that the variables are also decaying in importance, whereas the largest eigenvalue of $\Sigma^{(1)}$ is simple and the remaining $d = 31$ eigenvalues are equal, suggesting that after preintegration the remaining variables are all equally important.

Tables 1 and 2 give the CPU times for the QMC and QMC with preintegration approximations for the cdf at $t = 60$, as well as the QMC with preintegration approximation of the pdf at $t = 60$, for $\Sigma^{(1)}$ and $\Sigma^{(2)}$, respectively. The timing tests were run on a single processor for $N = 2^{13}, 2^{14}, \dots, 2^{20}$ with a single random shift. As expected, the CPU time increases linearly with N . In the last rows, we give the increase factor of the CPU time for QMC with preintegration compared to plain QMC for the cdf. Since this factor is around 2 in all cases, it is much less compared to the error reduction observed in Figures 1 and 2, which ranges from 10 to over 100. This demonstrates that preintegration is well worth the slight increase in cost.

In Figure 3, we plot the QMC with preintegration estimators for both the cdf F (left) and the pdf f (right) on the interval $[40, 100]$. Figure 4 then plots the convergence of the RMISE. The approximations of the cdf and pdf use $N = 2^{20}$ QMC points averaged over $R = 32$ random shifts and degree $M = 42$ Lagrange interpolation based on Chebyshev points. For the RMISE we use $N = 2^{10}, 2^{11}, \dots, 2^{19}$ with interpolation degree $M = \lceil N^{1/4} \rceil + 10$ and $R = 32$ random shifts. In this way, the number of interpolation points is coupled to the number of QMC points as in Corollary 5.6, with the choice $\sigma = 4$. To estimate the RMISE of $\hat{F}_{M,N}$ and $\hat{f}_{M,N}$, we first approximate the L^2 error by comparing each approximation to an approximation with much higher precision (i.e., $M = 42$, $N = 2^{20}$, $R = 32$) treated as the true cdf or pdf. Then we approximate the mean with respect to Δ by averaging the L^2 error over the random shifts. This captures both the QMC and interpolation contributions to the RMISE. As expected from Corollary 5.6, we observe almost $1/N$

FIG. 3. Approximate cdf (left) and approximate pdf (right) for $\Sigma^{(2)}$.FIG. 4. Convergence of RMISE in N for $\Sigma^{(2)}$.

convergence for the RMISE. This demonstrates that the method is an effective practical strategy.

Appendix A. Illustrative examples.

A.1. Some functions in \mathcal{H}_d^1 are not in BVHK after transformation. Mapping a weighted integral of an unbounded function g on \mathbb{R}^d to the d -dimensional unit cube as in (2.3) clearly results in an unbounded integrand over $[0, 1]^d$, since the closure of the range of g is unchanged by the transformation. This precludes membership of BVHK for the transformed integrand. For example, letting $d = 1$ and $g(y) = y^2$, and taking ρ to be the standard normal Gaussian density, results in an unweighted integral over $[0, 1]$ of an unbounded function, namely $\Phi^{-1}(u)^2$. An unbounded function of a single variable does not have bounded variation. The argument applies equally to any nonconstant polynomial on \mathbb{R}^d and to any unbounded function g that is integrable after multiplication by ρ .

On the other hand, choosing the weight function $\psi(y) = e^{-|y|}$, it is easily seen that $g(y) = y^2$, or any polynomial, belongs to \mathcal{H}_1^1 . Furthermore, for this pair of ρ and ψ , Theorem 2.1 holds with ω arbitrarily close to 2, and hence we get nearly first-order convergence.

Thus the setting used in the present paper allows a wider class of problems than assuming that the transformed function is in BVHK.

A.2. Smoothness fails for preintegrated functions if the random variables Y_i have compact support. The following example, defined on \mathbb{R}^4 , conforms with the definitions in the introduction for $d = 3$, except that in this case the underlying independent random variables Y_i have support $[0, 1]$ and have uniform distribution, corresponding to $\rho_i(y) = 1$ for $y \in [0, 1]$ and $\rho_i(y) = 0$ elsewhere. We take

$$\phi(\mathbf{y}) := \phi(y_0, y_1, y_2, y_3) = y_0 - \frac{y_1}{3} - \frac{y_2}{3} - \frac{y_3}{3} + \frac{1}{2}, \quad \mathbf{y} \in [0, 1]^4.$$

Preintegration with respect to y_0 of the indicator function composed with ϕ yields

$$\begin{aligned} \int_0^1 \text{ind}(\phi(\mathbf{y})) \, dy_0 &= \int_0^1 \text{ind}\left(y_0 - \frac{y_1}{3} - \frac{y_2}{3} - \frac{y_3}{3} + \frac{1}{2}\right) \, dy_0 \\ (A.1) \quad &= \int_{(\frac{y_1}{3} + \frac{y_2}{3} + \frac{y_3}{3} - \frac{1}{2})_+}^1 1 \, dy_0 = 1 - \left(\frac{y_1}{3} + \frac{y_2}{3} + \frac{y_3}{3} - \frac{1}{2}\right)_+, \end{aligned}$$

which is continuous on $[0, 1]^3$, but has a discontinuous gradient across the plane

$$(A.2) \quad y_1 + y_2 + y_3 = \frac{3}{2}.$$

The preintegrated function (A.1) is not in the class BVHK and hence does not have the smoothness assumed for QMC. To see this, for n a positive even integer, consider all the cubes of edge length $1/n$ with one vertex at $(\frac{i_1}{n}, \frac{i_2}{n}, \frac{i_3}{n})$ and the diagonally opposite vertex at $(\frac{i_1+1}{n}, \frac{i_2+1}{n}, \frac{i_3+1}{n})$, with $i_1, i_2, i_3 \in \{0, 1, \dots, n-1\}$ and such that $i_1 + i_2 + i_3 = \frac{3}{2}n - 2$. For each such cube, the vertex $(\frac{i_1+1}{n}, \frac{i_2+1}{n}, \frac{i_3+1}{n})$ satisfies

$$\frac{y_1}{3} + \frac{y_2}{3} + \frac{y_3}{3} - \frac{1}{2} = \frac{1}{3n} > 0,$$

and hence $(\frac{y_1}{3} + \frac{y_2}{3} + \frac{y_3}{3} - \frac{1}{2})_+$ has the value $\frac{1}{3n}$ at that vertex, while having the value 0 at the other seven vertices, since they lie on or below the plane (A.2). Thus each small cube contributes $\frac{1}{3n}$ to the variation in the sense of Vitali [25]. Moreover, because the plane (A.2) is a two-dimensional manifold with a nontrivial intersection with the unit cube, it is easy to see that there are of exact order n^2 such cubes in the unit cube. Thus the total variation in the sense of Vitali has a lower bound of exact order n . Letting $n \rightarrow \infty$, it follows that the variation in the sense of Vitali, and hence also of Hardy and Krause, is infinite. Similar results also hold in higher dimensions.

The paper [14] shows that for the case of the cube, as in this example, preintegration needs to be performed repeatedly if the mixed derivative smoothness is to be improved, with each successive preintegration yielding at most one additional order of smoothness. This implies that, in general, for the unit cube one must perform preintegration with respect to at least $d/2$ different variables to obtain first-order mixed derivative smoothness.

A.3. The monotonicity condition is necessary. Consider the bivariate function $g(y_1, y_2) := \sqrt[m]{(y_1 - y_2)_+}$, for $m \geq 2$, on $[0, 1]^2$. This is a simplified model (simplified in being restricted to the unit square instead of \mathbb{R}^d) of the typical singularity shown in [10] to arise when monotonicity with respect to the preintegration variable fails. A lower bound on the variation in the sense of Vitali, and hence of Hardy and Krause, can be obtained by taking a uniform partition of edge length $1/n$ of the unit square, and then considering the contribution to the variation in the sense of Vitali from only

those squares of size $1/n$ that are bisected by the main diagonal. From the values of g at the four vertices, each such square contributes $|0 - 0 + 0 - \sqrt[n]{1/n}| = \sqrt[n]{1/n}$. Since there are n such squares, a lower bound on the variation of g is $n^{1-1/m}$, which for all $m \geq 2$ is unbounded as $n \rightarrow \infty$. Thus g is not in the class BVHK.

A.4. Implications for analysis of QMC after preintegration. The results in [14] and the example in section A.2 show that preintegration is not guaranteed to be an effective method on compact domains, e.g., the unit cube. Hence, in this work we assume that the random variables Y_i , $i = 0, 1, \dots, d$, have support on the whole real line, resulting in an integration problem defined on \mathbb{R}^{d+1} , and we then use the theory from [18], which proves that on \mathbb{R}^{d+1} one step of preintegration is sufficient (under certain conditions). Furthermore, it is well-known that for problems on unbounded domains, the strategy of mapping back to the unit cube and using the Koksma–Hlawka inequality has the drawback that it cannot handle unbounded integrands, because they are not in BVHK (see also section A.1). Hence, in this paper we use the setting introduced in [24] (specifically an equivalent space), which can handle unbounded integrands by an appropriate choice of the weight functions $\{\psi_i\}$. The BVHK setting may allow for certain functions that do not have square-integrable mixed derivatives, as assumed in our setting.

In summary, preintegration is most effective for problems on unbounded domains, for which the results in [24] provide the appropriate setting to perform QMC error analysis. On the other hand, section A.3 and [10] show that if the monotonicity condition fails, then there may exist a square-root singularity, which can be handled by neither BVHK nor our setting.

Appendix B. Technical results.

LEMMA B.1. *Let $d \geq 1$, $\boldsymbol{\nu} \in \mathbb{N}_0^d$, and $[a, b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumption 1, and recall the definitions of U_t , ξ , and V in (3.2), (3.3), and (3.5), respectively. For any $q \in \mathbb{N}_0$ and $\boldsymbol{\eta} \leq \boldsymbol{\nu}$ satisfying $|\boldsymbol{\eta}| + q \leq |\boldsymbol{\nu}| + 1$, we consider functions $h_{q,\boldsymbol{\eta}} : V \rightarrow \mathbb{R}$ of the form (3.10). Then for any $i \in \{1 : d\}$, we can write*

$$D^i h_{q,\boldsymbol{\eta}}(t, \mathbf{y}) = \sum_{k=1}^{K_{q,\boldsymbol{\eta}}} h_{q,\boldsymbol{\eta}+\mathbf{e}_i}^{[k]}(t, \mathbf{y}) \quad \text{with} \quad K_{q,\boldsymbol{\eta}} \leq 8|\boldsymbol{\eta}| + 6q - 3,$$

where each function $h_{q,\boldsymbol{\eta}+\mathbf{e}_i}^{[k]}$ is of the form (3.10) with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta} + \mathbf{e}_i$.

Proof. For any $i \in \{1 : d\}$ and $h_{q,\boldsymbol{\eta}}(t, \mathbf{y}) = h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t, \mathbf{y})$ from (3.10) we have

$$D^i h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t, \mathbf{y}) = D^i \left(\underbrace{\frac{(-1)^r}{[D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})]^{r+q}}}_{=: T_1(t, \mathbf{y})} \underbrace{\rho_0^{(\beta)}(\xi(t, \mathbf{y}))}_{=: T_2(t, \mathbf{y})} \underbrace{\prod_{\ell=1}^r D^{\boldsymbol{\alpha}_\ell} \phi(\xi(t, \mathbf{y}), \mathbf{y})}_{=: T_3(t, \mathbf{y})} \right).$$

Using the chain rule and substituting $D^i \xi(t, \mathbf{y}) = -D^i \phi(\xi(t, \mathbf{y}), \mathbf{y}) / D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})$ (see (3.6)), and then simplifying our notation by suppressing the dependence on t and \mathbf{y} , we obtain

$$\begin{aligned} D^i T_1(t, \mathbf{y}) &= \frac{-(r+q)(-1)^r [D^i D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y}) + D^0 D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y}) D^i \xi(t, \mathbf{y})]}{[D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})]^{r+q+1}} \\ &= \frac{(r+q)(-1)^{r+1} D^{\mathbf{e}_0+\mathbf{e}_i} \phi(\xi)}{[D^0 \phi(\xi)]^{r+1+q}} + \frac{(r+q)(-1)^{r+2} D^{2\mathbf{e}_0} \phi(\xi) D^{\mathbf{e}_i} \phi(\xi)}{[D^0 \phi(\xi)]^{r+2+q}}, \end{aligned}$$

$$\begin{aligned}
D^i T_2(t, \mathbf{y}) &= \rho_0^{(\beta+1)}(\xi(t, \mathbf{y})) D^i \xi(t, \mathbf{y}) = -\frac{\rho_0^{(\beta+1)}(\xi) D^{e_i} \phi(\xi)}{D^0 \phi(\xi)}, \\
D^i T_3(t, \mathbf{y}) &= \sum_{m=1}^r [D^i D^{\alpha_m} \phi(\xi(t, \mathbf{y}), \mathbf{y}) + D^0 D^{\alpha_m} \phi(\xi(t, \mathbf{y}), \mathbf{y}) D^i \xi(t, \mathbf{y})] \prod_{\substack{\ell=1 \\ \ell \neq m}}^r D^{\alpha_\ell} \phi(\xi(t, \mathbf{y}), \mathbf{y}) \\
&= \sum_{m=1}^r \left[D^{\alpha_m + e_i} \phi(\xi) - \frac{D^{\alpha_m + e_0} \phi(\xi) D^{e_i} \phi(\xi)}{D^0 \phi(\xi)} \right] \prod_{\substack{\ell=1 \\ \ell \neq m}}^r D^{\alpha_\ell} \phi(\xi).
\end{aligned}$$

Using the product rule, we arrive at

$$\begin{aligned}
D^i h_{q, \eta, (r, \alpha, \beta)} &= (D^i T_1) T_2 T_3 + T_1 (D^i T_2) T_3 + T_1 T_2 (D^i T_3) \\
&= (S_{1a} + S_{1b}) + S_2 + \sum_{m=1}^r (S_{3a, m} + S_{3b, m}),
\end{aligned}$$

where

$$\begin{aligned}
S_{1a} &:= (r+q) h_{q, \eta + e_i, (r+1, \tilde{\alpha}, \beta)} & \text{with } \tilde{\alpha}_\ell &:= \begin{cases} \alpha_\ell & \text{for } \ell = 1, \dots, r, \\ e_i + e_0 & \text{for } \ell = r+1, \end{cases} \\
S_{1b} &:= (r+q) h_{q, \eta + e_i, (r+2, \tilde{\alpha}, \beta)} & \text{with } \tilde{\alpha}_\ell &:= \begin{cases} \alpha_\ell & \text{for } \ell = 1, \dots, r, \\ 2e_0 & \text{for } \ell = r+1, \\ e_i & \text{for } \ell = r+2, \end{cases} \\
S_2 &:= h_{q, \eta + e_i, (r+1, \tilde{\alpha}, \beta+1)} & \text{with } \tilde{\alpha}_\ell &:= \begin{cases} \alpha_\ell & \text{for } \ell = 1, \dots, r, \\ e_i & \text{for } \ell = r+1, \end{cases} \\
S_{3a, m} &:= h_{q, \eta + e_i, (r, \tilde{\alpha}, \beta)} & \text{with } \tilde{\alpha}_\ell &:= \begin{cases} \alpha_\ell & \text{for } \ell = 1, \dots, r, \ell \neq m, \\ \alpha_m + e_i & \text{for } \ell = m, \end{cases} \\
S_{3b, m} &:= h_{q, \eta + e_i, (r+1, \tilde{\alpha}, \beta)} & \text{with } \tilde{\alpha}_\ell &:= \begin{cases} \alpha_\ell & \text{for } \ell = 1, \dots, r, \ell \neq m, \\ \alpha_m + e_0 & \text{for } \ell = m, \\ e_i & \text{for } \ell = r+1. \end{cases}
\end{aligned}$$

Observe that all the $h_{q, \eta + e_i, [\dots]}$ functions above are of the form (3.10) with η replaced by $\eta + e_i$, and the conditions in (3.10) are satisfied by an inductive argument. For example, in S_{1b} , we gained two factors $D^{2e_0} \phi(\xi)$ and $D^{e_i} \phi(\xi)$ to join the product over ℓ , increasing the upper limit of the product from r to $r+2$, which is consistent with the increase in the exponent of $D^0 \phi(\xi)$ from $r+q$ to $r+2+q$. Furthermore, $r+2 \leq 2|\eta| + q - 1 + 2 = 2|\eta + e_i| + q - 1$, as required. Moreover, with $\tilde{\alpha}_{r+1} = 2e_0$ and $\tilde{\alpha}_{r+2} = e_i$, we have the updated sum $\beta e_0 + \sum_{\ell=1}^{r+2} \tilde{\alpha}_\ell = (r+q-1, \eta) + 2e_0 + e_i = (r+2+q-1, \eta + e_i)$, as required. The result for the other terms above can be justified in the same way. These $h_{q, \eta + e_i, [\dots]}$ functions are all different so there is no cancellation.

Treating the multiple $(r+q)h_{q, \eta + e_i, [\dots]}$ in S_{1a} as $r+q$ occurrences of the same function and doing this analogously for S_{1b} , we conclude that $D^i h_{q, \eta, (r, \alpha, \beta)}$ can be written as a sum of $K_{q, \eta}$ functions of the form (3.10) with η replaced by $\eta + e_i$, where

$$\begin{aligned}
K_{q, \eta} &= (r+q) + (r+q) + 1 + \sum_{m=1}^r (1+1) = 4r + 2q + 1 \\
&\leq 4(2|\eta| + q - 1) + 2q + 1 = 8|\eta| + 6q - 3.
\end{aligned}$$

This completes the proof. \square

LEMMA B.2. Let $d \geq 1$, $\boldsymbol{\nu} \in \mathbb{N}_0^d$, and $[a, b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumption 1, and recall the definitions of U_t , ξ and V in (3.2), (3.3), and (3.5), respectively. For any $q \in \mathbb{N}_0$ and $\boldsymbol{\eta} \leq \boldsymbol{\nu}$ satisfying $|\boldsymbol{\eta}| + q \leq |\boldsymbol{\nu}| + 1$, we consider functions $h_{q,\boldsymbol{\eta}}: V \rightarrow \mathbb{R}$ of the form (3.10). Then we can write

$$\frac{\partial}{\partial t} h_{q,\boldsymbol{\eta}}(t, \mathbf{y}) = \sum_{k=1}^{K_{q,\boldsymbol{\eta}}} h_{q+1,\boldsymbol{\eta}}^{[k]}(t, \mathbf{y}), \quad \text{with} \quad K_{q,\boldsymbol{\eta}} \leq 4|\boldsymbol{\eta}| + 3q - 1,$$

where each function $h_{q+1,\boldsymbol{\eta}}^{[k]}$ is of the form (3.10) with q replaced by $q + 1$.

Proof. For $h_{q,\boldsymbol{\eta}}(t, \mathbf{y}) = h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t, \mathbf{y})$ from (3.10) we have

$$\frac{\partial}{\partial t} h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t, \mathbf{y}) = \frac{\partial}{\partial t} \left(\underbrace{\frac{(-1)^r}{[D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})]^{r+q}}}_{=: T_1(t, \mathbf{y})} \underbrace{\rho_0^{(\beta)}(\xi(t, \mathbf{y}))}_{=: T_2(t, \mathbf{y})} \underbrace{\prod_{\ell=1}^r D^{\boldsymbol{\alpha}_\ell} \phi(\xi(t, \mathbf{y}), \mathbf{y})}_{=: T_3(t, \mathbf{y})} \right).$$

Using the chain rule and substituting $\frac{\partial}{\partial t} \xi(t, \mathbf{y}) = 1/D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})$ (see (3.7)), and then simplifying our notation by suppressing the dependence on t and \mathbf{y} , we obtain

$$\begin{aligned} \frac{\partial}{\partial t} T_1(t, \mathbf{y}) &= \frac{-(r+q)(-1)^r D^0 D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y}) \frac{\partial}{\partial t} \xi(t, \mathbf{y})}{[D^0 \phi(\xi(t, \mathbf{y}), \mathbf{y})]^{r+q+1}} = \frac{(r+q)(-1)^{r+1} D^{2\mathbf{e}_0} \phi(\xi)}{[D^0 \phi(\xi)]^{(r+1)+(q+1)}}, \\ \frac{\partial}{\partial t} T_2(t, \mathbf{y}) &= \rho_0^{(\beta+1)}(\xi(t, \mathbf{y})) \frac{\partial}{\partial t} \xi(t, \mathbf{y}) = \frac{\rho_0^{(\beta+1)}(\xi)}{D^0 \phi(\xi)}, \\ \frac{\partial}{\partial t} T_3(t, \mathbf{y}) &= \sum_{m=1}^r D^0 D^{\boldsymbol{\alpha}_m} \phi(\xi(t, \mathbf{y}), \mathbf{y}) \frac{\partial}{\partial t} \xi(t, \mathbf{y}) \prod_{\substack{\ell=1 \\ \ell \neq m}}^r D^{\boldsymbol{\alpha}_\ell} \phi(\xi(t, \mathbf{y}), \mathbf{y}) \\ &= \sum_{m=1}^r \frac{D^{\boldsymbol{\alpha}_m + \mathbf{e}_0} \phi(\xi)}{D^0 \phi(\xi)} \prod_{\substack{\ell=1 \\ \ell \neq m}}^r D^{\boldsymbol{\alpha}_\ell} \phi(\xi). \end{aligned}$$

Using the product rule, we arrive at

$$\frac{\partial}{\partial t} h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)} = \frac{\partial T_1}{\partial t} T_2 T_3 + T_1 \frac{\partial T_2}{\partial t} T_3 + T_1 T_2 \frac{\partial T_3}{\partial t} = S_1 + S_2 + \sum_{m=1}^r S_{3,m},$$

where now

$$\begin{aligned} S_1 &:= (r+q) h_{q+1,\boldsymbol{\eta},(r+1,\tilde{\boldsymbol{\alpha}},\beta)} & \text{with} \quad \tilde{\boldsymbol{\alpha}}_\ell &:= \begin{cases} \boldsymbol{\alpha}_\ell & \text{for } \ell = 1, \dots, r, \\ 2\mathbf{e}_0 & \text{for } \ell = r+1, \end{cases} \\ S_2 &:= h_{q+1,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta+1)}, \\ S_{3,m} &:= h_{q+1,\boldsymbol{\eta},(r,\tilde{\boldsymbol{\alpha}},\beta)} & \text{with} \quad \tilde{\boldsymbol{\alpha}}_\ell &:= \begin{cases} \boldsymbol{\alpha}_\ell & \text{for } \ell = 1, \dots, r, \ell \neq m, \\ \boldsymbol{\alpha}_m + \mathbf{e}_0 & \text{for } \ell = m. \end{cases} \end{aligned}$$

Again, all of the $h_{q+1,\boldsymbol{\eta},[\dots]}$ functions above are of the form (3.10) with q replaced by $q + 1$, and the conditions in (3.10) are satisfied by an inductive argument with justification similar to the arguments in the proof of Lemma B.1. These $h_{q+1,\boldsymbol{\eta},[\dots]}$ functions are all different so there is no cancellation.

Treating the multiple $(r + q)h_{q+1,\boldsymbol{\eta},[\dots]}$ in S_1 as $r + q$ occurrences of the same function, we conclude that $\frac{\partial}{\partial t}h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}$ can be written as a sum of $K_{q,\boldsymbol{\eta}}$ functions of the form (3.10) with q replaced by $q + 1$, where now

$$K_{q,\boldsymbol{\eta}} = (r + q) + 1 + \sum_{m=1}^r 1 = 2r + q + 1 \leq 2(2|\boldsymbol{\eta}| + q - 1) + q + 1 = 4|\boldsymbol{\eta}| + 3q - 1.$$

This completes the proof. \square

REFERENCES

- [1] N. ACHTSIS, R. COOLS, AND D. NUYENS, *Conditional sampling for barrier option pricing under the LT method*, SIAM J. Financial Math., 4 (2013), pp. 327–352.
- [2] N. ACHTSIS, R. COOLS, AND D. NUYENS, *Conditional sampling for barrier option pricing under the Heston model*, in Monte Carlo and Quasi-Monte Carlo Methods 2012, J. Dick, F. Y. Kuo, G. W. Peters, and I. H. Sloan, eds., Springer-Verlag, Berlin, 2013, pp. 253–269.
- [3] S. ASMUSSEN, *Conditional Monte Carlo for sums, with applications to insurance and finance*, Ann. Actuar. Sci., 12 (2018).
- [4] C. BAYER, C. BEN HAMMOUDA, AND R. TEMPONE, *Numerical smoothing with hierarchical adaptive sparse grids and quasi-Monte Carlo methods for efficient option pricing*, Quant. Finance, (2022), pp. 1–19.
- [5] A. BEN ABDELLAH, P. L'ECUYER, A. B. OWEN, AND F. PUCHHAMMER, *Density estimation by randomized quasi-Monte Carlo*, SIAM/ASA J. Uncertain. Quantif., 9 (2021), pp. 280–301.
- [6] Z. I. BOTEV, R. SALOMONE, AND D. MACKINLAY, *Fast and accurate computation of the distribution of sums of dependent log-normals*, Ann. Oper. Res., 280 (2019), pp. 19–46.
- [7] R. COOLS, F. Y. KUO, AND D. NUYENS, *Constructing embedded lattices rules for multivariate integration*, SIAM J. Sci. Comput., 28 (2006), pp. 2162–2188.
- [8] J. DICK, F. Y. KUO, AND I. H. SLOAN, *High dimensional integration – The quasi-Monte Carlo way*, Acta Numer., 22 (2013), pp. 133–288.
- [9] A. D. GILBERT, F. Y. KUO, AND I. H. SLOAN, *Equivalence between Sobolev spaces of first order dominating mixed smoothness and unanchored ANOVA spaces on \mathbb{R}^d* , Math. Comp., 91 (2022), pp. 1837–1869.
- [10] A. D. GILBERT, F. Y. KUO, AND I. H. SLOAN, *Preintegration is not smoothing when monotonicity fails*, in Advances in Modeling and Simulation, Z. Botev, et al., eds., Springer, New York, 2022, pp. 169–191.
- [11] M. B. GILES, T. NAGAPETIAN, AND K. RITTER, *Multilevel Monte Carlo approximation of distribution functions and densities*, SIAM/ASA J. Uncertain. Quantif., 3 (2013), pp. 267–295.
- [12] P. GLASSERMAN, *Monte Carlo Methods in Financial Engineering*, Springer, Berlin, 2003.
- [13] P. GLASSERMAN AND J. STAUM, *Conditioning on one-step survival for barrier option simulations*, Oper. Res., 49 (2001), pp. 923–937.
- [14] M. GRIEBEL, F. Y. KUO, AND I. H. SLOAN, *The smoothing effect of the ANOVA decomposition*, J. Complexity, 26 (2010), pp. 523–551.
- [15] M. GRIEBEL, F. Y. KUO, AND I. H. SLOAN, *The smoothing effect of integration in \mathbb{R}^d and the ANOVA decomposition*, Math. Comp., 82 (2013), pp. 383–400.
- [16] M. GRIEBEL, F. Y. KUO, AND I. H. SLOAN, *Note on “The smoothing effect of integration in \mathbb{R}^d and the ANOVA decomposition”*, Math. Comp., 86 (2017), pp. 1847–1854.
- [17] M. GRIEBEL, F. Y. KUO, AND I. H. SLOAN, *The ANOVA decomposition of a non-smooth function of infinitely many variables can have every term smooth*, Math. Comp., 86 (2017), pp. 1855–1876.
- [18] A. GRIEWANK, F. Y. KUO, H. LEÖVEY, AND I. H. SLOAN, *High dimensional integration of kinks and jumps — Smoothing by preintegration*, J. Comput. Appl. Math., 344 (2018), pp. 259–274.
- [19] M. HOLTZ, *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance*, Springer-Verlag, Berlin, 2011.
- [20] F. Y. KUO, *Lattice Rule Generating Vectors*, <https://web.maths.unsw.edu.au/~fkuo/lattice/index.html>.
- [21] F. Y. KUO, I. H. SLOAN, G. W. WASILKOWSKI, AND B. J. WATERHOUSE, *Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands*, J. Complexity, 26 (2010), pp. 135–160.

- [22] P. L'ECUYER AND C. LEMIEUX, *Variance reduction via lattice rules*, Management Sci., 46 (2000), pp. 1214–1235.
- [23] P. L'ECUYER, F. PUCHHAMMER, AND A. BEN ABDELLAH, *Monte Carlo and quasi-Monte Carlo density estimation via conditioning*, INFORMS J. Comput., 34 (2022), pp. 1729–1748.
- [24] J. A. NICHOLS AND F. Y. KUO, *Fast CBC construction of randomly shifted lattice rules achieving $O(n^{-1+\delta})$ convergence for unbounded integrands over \mathbb{R}^d in weighted spaces with POD weights*, J. Complexity, 30 (2014), pp. 444–468.
- [25] H. NIEDERREITER, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [26] Y. PENG, M. FU, J. HU, P. L'ECUYER, AND B. TUFFIN, *Variance Reduction for Generalized Likelihood Ratio Method by Conditional Monte Carlo and Randomized Quasi-Monte Carlo*, hal-03196379, 2021.
- [27] L. N. TREFETHEN, *Approximation Theory and Practice*, SIAM, Philadelphia, 2020.
- [28] UNSW Sydney, *Katana*, 2010, <https://doi.org/10.26190/669x-a286>.
- [29] C. WENG, X. WANG, AND Z. HE, *Efficient computation of option prices and greeks by quasi-Monte Carlo method with smoothing and dimension reduction*, SIAM J. Sci. Comput., 39 (2017), pp. B298–B322.