Approximating distribution functions and densities using quasi-Monte Carlo methods after smoothing by preintegration

Alexander D. Gilbert[1] Frances Y. Kuo[1] Ian H. Sloan[1]

December 2021

Abstract

The cumulative distribution or probability density of a random variable, which is itself a function of a high number of independent real-valued random variables, can be formulated as high-dimensional integrals of an indicator or a Dirac δ function, respectively. To approximate the distribution or density at a point, we carry out preintegration with respect to one suitably chosen variable, then apply a 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. We provide rigorous regularity and error analysis for the preintegrated function to show that our estimators achieve nearly first order convergence. Numerical results support the theory.

1 Introduction

Let X be a continuous real-valued random variable that cannot be simulated directly, and denote the cumulative distribution function (cdf) and probability density function (pdf) of X by F_X and f_X, respectively, which are also not known *a priori*. However, suppose that X is a function of $d+1$ independent random variables $Y_0, Y_1, \ldots, Y_d \in \mathbb{R}$ that can be simulated:

$$X = \phi(Y_0, Y_1, \ldots, Y_d)\,, \tag{1}$$

for some computable function $\phi : \mathbb{R}^{d+1} \to \mathbb{R}$. Suppose also that the density and distribution functions of each Y_i are known, and are denoted by ρ_i and Φ_i, respectively. Realisations 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 $Y_0, \ldots, Y_d$ 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 $Y_0, \ldots, Y_d$ are zero-mean normal random variables that correspond to the random parameters in the coefficient.

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

1. Approximate F_X and f_X at finitely many points $\{t_m\}_{m=0}^M \subset [a, b]$ using quasi-Monte Carlo (QMC) with preintegration (see Sections 2.2 and 3, respectively).

2. Reconstruct F_X and f_X on $[a, b]$ by interpolating the approximations at the points $\{t_m\}_{m=0}^M$.

[1]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

The key point to using QMC is that the cdf and pdf can each be written as expected values (i.e., high-dimensional integrals) with respect to $\boldsymbol{y}$: for $t \in [a,b]$

$$F_X(t) = \mathbb{E}\big[\mathrm{ind}(t - X)\big] = \int_{\mathbb{R}^{d+1}} \mathrm{ind}\big(t - \phi(y_0,\ldots,y_d)\big)\left(\prod_{i=0}^{d} \rho_i(y_i)\right)\mathrm{d}y_0\cdots\mathrm{d}y_d\,, \tag{2}$$

$$f_X(t) = \mathbb{E}\big[\delta(t - X)\big] = \int_{\mathbb{R}^{d+1}} \delta\big(t - \phi(y_0,\ldots,y_d)\big)\left(\prod_{i=0}^{d} \rho_i(y_i)\right)\mathrm{d}y_0\cdots\mathrm{d}y_d\,, \tag{3}$$

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

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

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

$$\begin{aligned} &\delta(z) = 0 \quad \text{for all } z \neq 0, \quad \text{and} \\ &\int_{-\infty}^{\infty} g(z)\,\delta(z)\,\mathrm{d}z = g(0) \quad \text{for all sufficiently smooth functions } g\,. \end{aligned} \tag{4}$$

The integral in (3) can be interpreted as the result of a differentiation, in the distributional sense, of (2) (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. Also, to ease the notation for the paper we simply write $F := F_X$ and $f := f_X$.

Previously, QMC theory has not been able to tackle such integrals (2) and (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 (3) is not even a function, but rather a distribution that is 0 everywhere except when $\phi(y_0,\ldots,y_d) = t$. However, recent work [16] 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 simple distributions involving a δ function, as well as extending the theory for the indicator function.

Smoothing by preintegration is a specific case of *conditional sampling* from the statistics and computational finance literature [2, 3, 4, 11, 12, 17, 24], where a given function is "smoothed" by the operation of integration (conditioning). We follow a specific method of preintegration within a numerical integration method, whereby a single variable is first integrated out to smooth the integrand, thus allowing a cubature rule to be applied in the remaining dimensions. The smoothing by preintegration step was recently analysed in [16] which extends the earlier work [13, 14, 15]. See also [9].

Despite the widespread use of QMC for high-dimensional integrals (i.e., expected values), their use in density estimation has been limited. The recent paper [1] used QMC to construct kernel density estimators and histograms, but they are only suitable for low dimensions since the need to balance the kernel/histogram bandwidth with the number of QMC points means the overall convergence rate deteriorates very rapidly with dimension. Another recent paper [20] studied density estimation using conditioning, essentially the same as the preintegration method in this paper, but with a key theoretical difference — here we provide a full regularity analysis whereas [20] just assumed that the integrand *after* conditioning was smooth enough for QMC. The paper [10] replaced the non-smooth functions in (2) and (3) by smooth approximations and then applied multilevel Monte Carlo methods. The paper [5] introduced smooth cdf and pdf estimators for the specific case of a sum of dependent lognormals, with promising numerical results using QMC but without supporting theory.

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

$$g(y_0, \ldots, y_d) = \mathrm{ind}\big(\phi(y_0, \ldots, y_d)\big)\,,$$

where the inner function $\phi : \mathbb{R}^{d+1} \to \mathbb{R}$ 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. To give some intuition, one of the key assumptions is that ϕ is strictly increasing in y_0 and tends to ∞ as $y_0 \to \infty$ (cf. Assumption A1 below). Then, to perform the preintegration step, we integrate with respect to this special variable y_0 to give a *preintegrated* function

$$P_0 g(y_1, \ldots, y_d) := \int_{-\infty}^{\infty} g(y_0, \ldots, y_d)\, \rho_0(y_0)\, \mathrm{d}y_0\,, \tag{5}$$

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

The key point from [16] is that if we fix $(y_1, \ldots, y_d) \in \mathbb{R}^d$ and treat $\phi(\cdot, y_1, \ldots, y_d)$ 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 because $\phi(y_0, \ldots, y_d) \geq 0$ for all $y_0 \in \mathbb{R}$. 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, which, due to the smoothness of $P_0 g$, will now converge at a faster rate, e.g., close to $\mathcal{O}(1/N)$ for QMC methods.

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 $\boldsymbol{y} = (y_1, \ldots, y_d) \in \mathbb{R}^d$ and all of the $d+1$ variables collectively by $\boldsymbol{y}_{0:d} \coloneqq (y_0, \ldots, y_d) \in \mathbb{R}^{d+1}$ or $(y_0, \boldsymbol{y})$. Similarly, we denote the products of univariate functions $(\rho_i)_{i=0}^d$, by

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

Let $\mathbb{N}_0 \coloneqq \{0, 1, 2, \ldots\}$ and $\mathbb{N} \coloneqq \{1, 2, \ldots\}$ denote the set of natural numbers with and without zero, respectively. Let $\{0 : d\} \coloneqq \{0, 1, \ldots, d\}$ and define $\{1 : d\}$ analogously. Let $\boldsymbol{\nu} = (\nu_0, \nu_1, \ldots, \nu_d) \in \mathbb{N}_0^{d+1}$ be a multi-index, and let $|\boldsymbol{\nu}| \coloneqq \sum_{i=0}^d \nu_i$ denote its *order* and $\mathrm{supp}(\boldsymbol{\nu}) \coloneqq \{j \in \{0 : d\} : \nu_j > 0\}$ 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 $i = 0, 1, \ldots, d$, and addition is defined by $\boldsymbol{\eta} + \boldsymbol{\nu} = (\eta_i + \nu_i)_{i=0}^d$. For $\boldsymbol{y}_{0:d} \in \mathbb{R}^{d+1}$ and $\boldsymbol{\nu} \in \mathbb{N}_0^{d+1}$, we denote the *active* variables by $\boldsymbol{y}_{\boldsymbol{\nu}} \coloneqq (y_i : \nu_i > 0)_{i=0}^d$ and the *inactive* variables by $\boldsymbol{y}_{-\boldsymbol{\nu}} \coloneqq (y_i : \nu_i = 0)_{i=0}^d$. Analogously to the notation $(y_0, \boldsymbol{y})$, we denote the $(d+1)$-dimensional concatenation of $\nu_0 \in \mathbb{N}_0$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_d) \in \mathbb{N}_0^d$ by $(\nu_0, \boldsymbol{\nu}) = (\nu_0, \nu_1, \nu_2, \ldots, \nu_d) \in \mathbb{N}_0^{d+1}$.

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 $i = 0, 1, \ldots, d$ and multi-index $\boldsymbol{\nu} \in \mathbb{N}_0^{d+1}$, let

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

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

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

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 $\boldsymbol{\nu}$:

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

We impose no restrictions on the behaviour of their elements at infinity.

To provide a function space setting for ϕ in (1), we introduce a class of Sobolev spaces of dominating mixed smoothness on $\mathbb{R}^{d+1}$, where the behaviour of derivatives at $\pm\infty$ is controlled by functions different from the densities ρ_i. To this end, for $i = 0, 1, \ldots, d$, let $\psi_i : \mathbb{R} \to \mathbb{R}$ 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 $\boldsymbol{\gamma} := \{\gamma_{\mathfrak{u}} > 0 : \mathfrak{u} \subseteq \{0 : d\}\}$ be a collection of positive real numbers called *weight parameters*; they model the relative importance of different collections of variables, i.e., $\gamma_{\mathfrak{u}}$ describes the importance of the collection of variables $(y_i : i \in \mathfrak{u})$. We set $\gamma_\emptyset := 1$.

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

$$\|g\|_{\mathcal{H}_{d+1}^{\boldsymbol{\nu}}}^2 := \sum_{\boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{1}{\gamma_{\boldsymbol{\eta}}} \int_{\mathbb{R}^{d+1}} |D^{\boldsymbol{\eta}} g(\boldsymbol{y}_{0:d})|^2\, \boldsymbol{\psi}_{\boldsymbol{\eta}}(\boldsymbol{y}_{\boldsymbol{\eta}})\, \boldsymbol{\rho}_{-\boldsymbol{\eta}}(\boldsymbol{y}_{-\boldsymbol{\eta}})\, \mathrm{d}\boldsymbol{y}_{0:d}$$

is finite, where we introduced the shorthand notations

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

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 $\mathcal{H}_d^{\boldsymbol{\nu}}$ with variables indexed from 1 to d.

An important property of the Sobolev space of *first-order* dominating mixed smoothness, i.e., $\mathcal{H}_d^{\boldsymbol{1}}$ with $\boldsymbol{1} := (1, 1, \ldots, 1)$, is that it is equivalent to the (unanchored) ANOVA space introduced in [21] over the unbounded domain $\mathbb{R}^d$. Explicitly, it was shown recently in [8] that if the weight functions ψ_i satisfy

$$\int_{-\infty}^{\infty} \frac{\Phi_i(z)(1-\Phi_i(z))}{\psi_i(z)}\,\mathrm{d}z < \infty, \quad \text{for all } i = 1, 2, \ldots, d, \tag{6}$$

then $\mathcal{H}_d^{\mathbf{1}}$ and the ANOVA space from [21] are equivalent. This equivalence is crucial for our analysis, because it immediately shows that the bounds on the QMC error from [21] also hold in $\mathcal{H}_d^{\mathbf{1}}$ (see Theorem 1 below). Since the preintegration theory in [16] is formulated in $\mathcal{H}_d^{\boldsymbol{\nu}}$, 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 $\mathcal{H}_d^{\boldsymbol{\nu}}$, and have no need to introduce the ANOVA space from [21]. We assume (6) holds throughout.

2.2 Quasi-Monte Carlo methods

In the classic case of the unit cube, an N-point QMC approximation (see e.g., [7]) for a function $g : [0,1]^d \to \mathbb{R}$ is given by

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

where here the cubature points $\{\boldsymbol{q}_n\}_{n=0}^{N-1}$ 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 randomised QMC methods called *randomly shifted rank-1 lattice rules*, for which the QMC points are given by

$$\boldsymbol{q}_n = \left\{\frac{n\boldsymbol{z}}{N} + \boldsymbol{\Delta}\right\} \quad \text{for } n = 0, 1, \ldots, N-1. \tag{7}$$

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

The benefits of randomly shifting the pointset are threefold: (i) the resulting approximation is unbiased; (ii) we can average over a small number of i.i.d. random shifts then use the sample variance to estimate the mean-square error; and (iii) randomly shifted lattice rules with good $\boldsymbol{z}$ can be constructed efficiently (see below) to achieve nearly $\mathcal{O}(N^{-1})$ convergence, depending on g and the function space setting.

To approximate an integral over an unbounded domain one must map the pointset $\{\boldsymbol{q}_n\}_{n=0}^{N-1}$ from the unit cube to $\mathbb{R}^d$. In the case of an integral with respect to a product of densities, each with a given pdf, as we have in (2), we can perform this mapping by applying the inverse cdf componentwise (also called the *Rosenblatt transform*). An N-point QMC approximation for a function $g : \mathbb{R}^d \to \mathbb{R}$ is then given by

$$Q_{d,N}(g) \coloneqq \frac{1}{N}\sum_{n=0}^{N-1} g(\boldsymbol{\Phi}^{-1}(\boldsymbol{q}_n)) \approx \int_{[0,1]^d} g(\boldsymbol{\Phi}^{-1}(\boldsymbol{u}))\,\mathrm{d}\boldsymbol{u} = \int_{\mathbb{R}^d} g(\boldsymbol{y})\,\boldsymbol{\rho}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,, \tag{8}$$

where $\boldsymbol{\Phi}^{-1}$ denotes 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

$$\boldsymbol{\tau}_n = \boldsymbol{\Phi}^{-1}(\boldsymbol{q}_n) \in \mathbb{R}^d \quad \text{for } n = 0, 1, \ldots, N-1. \tag{9}$$

It was proved in [21] that good generating vectors $\boldsymbol{z}$ for the approximation (8) can be constructed using a component-by-component (CBC) algorithm to achieve almost the optimal convergence rate in a certain *first-order ANOVA space* (which as we have discussed is equivalent to $\mathcal{H}_d^{\mathbf{1}}$). Below we restate the error bound from [21], but now in terms of $\mathcal{H}_d^{\mathbf{1}}$ rather than the equivalent ANOVA space used in [21].

Theorem 1 *Suppose* (6) *holds. Let* $\omega \in (1, 2]$ *and* $c < \infty$ *be such that*

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

Let $N \in \mathbb{N}$ *and suppose that* $\boldsymbol{z}$ *is a generating vector constructed using the CBC algorithm from [21]. Then for* $g \in \mathcal{H}_d^{\mathbf{1}}$ *the root-mean-square error (with the expectation taken with respect to the random shift* $\boldsymbol{\Delta}$*) of the randomly shifted lattice rule approximation* (8) *corresponding to* $\boldsymbol{z}$ *satisfies*

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[\left|\int_{\mathbb{R}^d} g(\boldsymbol{y})\boldsymbol{\rho}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - Q_{d,N}(g)\right|^2\right]} \le C_{d,\boldsymbol{\gamma},\lambda}\,[\phi_{\mathrm{tot}}(N)]^{-1/(2\lambda)}\,\|g\|_{\mathcal{H}_d^{\mathbf{1}}} \tag{10}$$

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

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

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 [21]. Theorem 8 from [21] gives the error bound (10) instead for $g \in \mathcal{W}_d$ and with the $\mathcal{W}_d$-norm on the right. Since [8, Theorem 13] shows that $\mathcal{W}_d$ and $\mathcal{H}_d^{\mathbf{1}}$ are equivalent under assumption (6) on the weight functions, with $\|g\|_{\mathcal{W}_d} \le \|g\|_{\mathcal{H}_d^{\mathbf{1}}}$, the result is proved. □

Observe that the convergence rate $N^{-1/(2\lambda)}$ is governed by the parameter $\omega \in (1, 2]$ 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, even if (6) holds. We refer to Table 3 in [19], which gives values of $r_2 := \omega/2$ for several common pairs of ρ_i and ψ_i.

2.3 Lagrange interpolation in one dimension

There are two steps to our 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* $L_M : C[a, b] \to V_M$ is given by

$$L_M g := \sum_{m=0}^{M} g(t_m)\,\chi_{M,m}, \qquad \chi_{M,m}(t) := \prod_{\substack{\ell=0\\ \ell\neq m}}^{M} \frac{t - t_\ell}{t_m - t_\ell}\,. \tag{11}$$

We now state the classical error bounds for Lagrange interpolation based on Chebyshev points from, e.g., [22]. 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 $\|g\|_{W^{\sigma,\infty}} := \max_{q=0,1,\ldots,\sigma} \|g^{(q)}\|_{L^\infty}$. 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

$$\|g - L_M g\|_{L^\infty} \leq \frac{4\,\|g^{(\sigma+1)}\|_{L^1}}{\pi\sigma(M-\sigma)^\sigma}\,. \tag{12}$$

The original result [22, 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. For notational convenience we take y_0 to be this special variable. In this section we formalise the preintegration step for indicator functions by following [16], 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 (3).

First we make the following assumptions about the function ϕ in (1).

Assumption A1 For $d \geq 1$ and $\boldsymbol{\nu} \in \mathbb{N}_0^d$, let $\phi : \mathbb{R}^{d+1} \to \mathbb{R}$ satisfy

1. $D^0\phi(y_0, \boldsymbol{y}) > 0$ for all $(y_0, \boldsymbol{y}) \in \mathbb{R}^{d+1}$; and
2. for each $\boldsymbol{y} \in \mathbb{R}^d$, $\phi(y_0, \boldsymbol{y}) \to \infty$ as $y_0 \to \infty$; 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 [13, 15, 16] whether the *monotonicity* assumption (Assumption A1.1) is necessary. Recently it was proved in [9] that this is indeed necessary: if it fails then there remains a singularity after preintegration.

3.1 Smoothing by preintegration for indicator functions

Motivated by the cdf (2), for $t \in [a,b]$ we define a discontinuous function $g_t : \mathbb{R}^{d+1} \to \mathbb{R}$ by

$$g_t(y_0, \boldsymbol{y}) := \mathrm{ind}\big(t - \phi(y_0, \boldsymbol{y})\big)\,, \qquad y_0 \in \mathbb{R}\,,\ \boldsymbol{y} \in \mathbb{R}^d\,, \tag{13}$$

where $\phi : \mathbb{R}^{d+1} \to \mathbb{R}$ satisfies Assumption A1. Note that [16] considered functions of the more general form $g(y_0, \boldsymbol{y}) = \theta(y_0, \boldsymbol{y})\,\mathrm{ind}(\phi(y_0, \boldsymbol{y}))$, 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 (2), the restricted form in (13) is sufficient. Also, note that we consider here shifted indicator functions $\mathrm{ind}(t - \cdot)$ instead of $\mathrm{ind}(\cdot)$ as in [16]. This results only in minor changes in the presentation, and does not fundamentally affect any of the theory.

Since in Assumption A1 we assume that $\phi(\cdot, \boldsymbol{y})$ is strictly increasing with respect to y_0 for fixed $\boldsymbol{y} \in \mathbb{R}^d$, and also tends to ∞ as $y_0 \to \infty$, there are only two possible scenarios:

either the discontinuity of $g_t(\cdot, \boldsymbol{y})$ for fixed $t \in [a, b]$ and $\boldsymbol{y} \in \mathbb{R}^d$ occurs at a unique point in dimension 0, or the discontinuity does not occur at all because $\phi(y_0, \boldsymbol{y}) > t$ 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 $\boldsymbol{y} \in \mathbb{R}^d$ for which the discontinuity occurs by

$$U_t := \left\{\boldsymbol{y} \in \mathbb{R}^d : \phi(y_0, \boldsymbol{y}) = t \text{ for some } y_0 \in \mathbb{R}\right\}, \tag{14}$$

which (unlike in [16]) now depends on the point t. We have the following equivalence

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

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

$$\phi(\xi(t, \boldsymbol{y}), \boldsymbol{y}) = t. \tag{15}$$

Because $\phi(\cdot, \boldsymbol{y})$ is increasing, it follows from (15) that for $\boldsymbol{y} \in U_t$,

$$\phi(y_0, \boldsymbol{y}) < t \iff y_0 < \xi(t, \boldsymbol{y}).$$

The following Implicit Function Theorem adapted from [13] shows that ξ is a well-defined function of both t and $\boldsymbol{y}$, and implies that ξ "inherits" the smoothness of ϕ.

Theorem 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 A1, and define*

$$V := \left\{(t, \boldsymbol{y}) \in (a, b) \times \mathbb{R}^d : \phi(y_0, \boldsymbol{y}) = t \textit{ for some } y_0 \in \mathbb{R}^d\right\} \subset [a, b] \times \mathbb{R}^d. \tag{16}$$

If V is not empty, then there exists a unique function $\xi \in C^{(\nu_0, \boldsymbol{\nu})}(\overline{V})$ satisfying (15) for all $(t, \boldsymbol{y}) \in \overline{V}$. Furthermore, for $(t, \boldsymbol{y}) \in V$ the first-order derivatives are given by

$$D^i\xi(t, \boldsymbol{y}) = \frac{\partial}{\partial y_i}\xi(t, \boldsymbol{y}) = -\frac{D^i\phi(\xi(t, \boldsymbol{y}), \boldsymbol{y})}{D^0\phi(\xi(t, \boldsymbol{y}), \boldsymbol{y})} \quad \textit{for all } i = 1, \ldots, d, \quad \textit{and} \tag{17}$$

$$\frac{\partial}{\partial t}\xi(t, \boldsymbol{y}) = \frac{1}{D^0\phi(\xi(t, \boldsymbol{y}), \boldsymbol{y})}. \tag{18}$$

Proof. The result $\xi \in C^{(\nu_0, \boldsymbol{\nu})}(\overline{V})$ follows by applying [13, Theorem 2.3] to the function $\phi(y_0, \boldsymbol{y}) - t$, with $j = 1$ along with the variables labelled as $x_{i+1} = y_i$ for $i = 0, 1, \ldots, d$ and $x_{d+2} = t$. Since the proof of [13, Theorem 2.3] is based on a local argument about an arbitrary point $\boldsymbol{x}$, 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 $C^{(\nu_0, \boldsymbol{\nu})}(\overline{V})$. Finally, differentiating (15) with respect to y_i leads to the formula (17). Similarly, differentiating (15) with respect to t implies (18). □

With P_0 the preintegration operator defined by (5), we now apply preintegration to the function g_t defined by (13) for $t \in [a, b]$, obtaining

$$P_0 g_t(\boldsymbol{y}) = \int_{-\infty}^{\infty} \operatorname{ind}\big(t - \phi(y_0, \boldsymbol{y})\big)\, \rho_0(y_0)\, \mathrm{d}y_0,$$

which for $\boldsymbol{y} \in U_t$ can now be written

$$P_0 g_t(\boldsymbol{y}) = \int_{-\infty}^{\xi(t, \boldsymbol{y})} \rho_0(y_0)\, \mathrm{d}y_0 = \Phi_0(\xi(t, \boldsymbol{y})), \tag{19}$$

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

The main result from [16, Theorem 3] showed that if ϕ satisfies Assumption A1, along with some extra technical conditions in Assumption A2 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

$$D^i[P_0 g_t(\boldsymbol{y})] = D^i[\Phi_0(\xi(t,\boldsymbol{y}))] = \rho_0(\xi(t,\boldsymbol{y}))\, D^i \xi(t,\boldsymbol{y}) = -\frac{\rho_0(\xi(t,\boldsymbol{y}))\, D^i\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})}{D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})}, \quad (20)$$

where we used (17). This motivates the general form of functions in Assumption A2 below. Our assumption is formulated differently from [16] 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 below to extend [16, Theorem 3] by providing an explicit bound on the norm. We use $\boldsymbol{e}_i$ to denote a multi-index whose ith component is 1 and all other components are 0.

Assumption A2 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 A1. Recall the definitions of U_t, ξ and V in (14), (15) and (16), respectively. Given $q \in \mathbb{N}_0$ and $\boldsymbol{\eta} \leq \boldsymbol{\nu}$ satisfying $|\boldsymbol{\eta}| + q \leq |\boldsymbol{\nu}| + 1$, we consider functions $h_{q,\boldsymbol{\eta}} : \overline{V} \to \mathbb{R}$ of the form

$$\begin{cases} h_{q,\boldsymbol{\eta}}(t,\boldsymbol{y}) = h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t,\boldsymbol{y}) := \dfrac{(-1)^r \rho_0^{(\beta)}(\xi(t,\boldsymbol{y}))\, \prod_{\ell=1}^r D^{\boldsymbol{\alpha}_\ell}\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})}{[D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})]^{r+q}}, \\ \text{with } r \in \mathbb{N}_0,\ \boldsymbol{\alpha} = (\boldsymbol{\alpha}_\ell)_{\ell=1}^r,\ \boldsymbol{\alpha}_\ell \in \mathbb{N}_0^{d+1} \setminus \{\boldsymbol{e}_0, \boldsymbol{0}\},\ \beta \in \mathbb{N}_0 \text{ satisfying} \\ r \leq 2|\boldsymbol{\eta}| + q - 1,\ \alpha_{\ell,0} \leq |\boldsymbol{\eta}| + q,\ \beta \leq |\boldsymbol{\eta}| + q - 1,\ \beta \boldsymbol{e}_0 + \displaystyle\sum_{\ell=1}^r \boldsymbol{\alpha}_\ell = (r+q-1, \boldsymbol{\eta}). \end{cases} \quad (21)$$

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

$$\lim_{\boldsymbol{y} \to \partial U_t} h_{q,\boldsymbol{\eta}}(t,\boldsymbol{y}) = 0 \quad \text{for all } t \in [a,b], \quad (22)$$

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

$$\sup_{t \in [a,b]} \int_{U_t} |h_{q,\boldsymbol{\eta}}(t,\boldsymbol{y})|^2\, \boldsymbol{\psi}_{\boldsymbol{\eta}}(\boldsymbol{y}_{\boldsymbol{\eta}})\, \boldsymbol{\rho}_{-\boldsymbol{\eta}}(\boldsymbol{y}_{-\boldsymbol{\eta}})\, \mathrm{d}\boldsymbol{y} \leq B_{q,\boldsymbol{\eta}} < \infty. \quad (23)$$

It is worthwhile to briefly discuss Assumption A2. 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 $\boldsymbol{y}$. The conditions (22) and (23) then ensure that the derivatives are well-behaved enough for the preintegrated function to be sufficiently smooth. It follows from Assumption A1 and Theorem 2 that each function $h_{q,\boldsymbol{\eta}}$ of the form (21) is continuous on $\overline{V}$.

Having introduced preintegration and our key assumptions, we now state the main smoothing by preintegration theorem for functions of the form (13). It is a refined version of [16, Theorem 3].

Theorem 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 Assumption A1 and Assumption A2 for $q = 0$ and all $\boldsymbol{0} \neq \boldsymbol{\eta} \leq \boldsymbol{\nu}$. Then for $t \in [a,b]$ the function*

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

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

$$\sup_{t\in[a,b]} \|P_0 g_t\|_{\mathcal{H}_d^{\boldsymbol{\nu}}} \le \left(1 + \sum_{\mathbf{0}\neq\boldsymbol{\eta}\le\boldsymbol{\nu}} \frac{\left(8^{|\boldsymbol{\eta}|-1}(|\boldsymbol{\eta}|-1)!\right)^2 B_{0,\boldsymbol{\eta}}}{\gamma_{\boldsymbol{\eta}}}\right)^{1/2} < \infty\,. \tag{24}$$

Proof. From (19) the preintegrated function can be written as

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

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 $\mathbf{0}\neq\boldsymbol{\eta}\le\boldsymbol{\nu}$, we first prove by induction on $|\boldsymbol{\eta}|\ge 1$ that the $\boldsymbol{\eta}$th derivative of $P_0 g_t$ is given by

$$D^{\boldsymbol{\eta}}[P_0 g_t(\boldsymbol{y})] = \begin{cases} \displaystyle\sum_{j=1}^{J_{0,\boldsymbol{\eta}}} h_{0,\boldsymbol{\eta}}^{[j]}(t,\boldsymbol{y}) & \text{if } \boldsymbol{y}\in U_t\,, \quad \text{with} \quad J_{0,\boldsymbol{\eta}} \le 8^{|\boldsymbol{\eta}|-1}(|\boldsymbol{\eta}|-1)!\,,\\ 0 & \text{if } \boldsymbol{y}\in\mathbb{R}^d\setminus U_t\,, \end{cases} \tag{25}$$

where each function $h_{0,\boldsymbol{\eta}}^{[j]}$ is of the form (21) with $q=0$.

For the base case $\boldsymbol{\eta}=\boldsymbol{e}_i$ with any $i\in\{1:d\}$, we take $r=1$, $\boldsymbol{\alpha}_1=\boldsymbol{e}_i$, $\beta=0$ and $J_{0,\boldsymbol{e}_i}=1$ to recover the single function (20). Suppose next that (25) holds for some $\boldsymbol{\eta}\in\mathbb{N}_0^d$ with $|\boldsymbol{\eta}|\ge 1$ and consider any $i\in\{1:d\}$ and $\boldsymbol{y}\in U_t$. We have

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

In the second equality we used Lemma 9 in the Appendix with $q=0$, which states that each function $D^i h_{0,\boldsymbol{\eta}}^{[j]}$ can be written as a sum of $K_{0,\boldsymbol{\eta}}\le 8|\boldsymbol{\eta}|-3$ functions of the form (21) with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta}+\boldsymbol{e}_i$. We enumerated these functions with the notation $h_{0,\boldsymbol{\eta}+\boldsymbol{e}_i}^{[j,k]}$ and then relabeled all functions for different combinations of indices j and k with the notation $h_{0,\boldsymbol{\eta}+\boldsymbol{e}_i}^{[j']}$. The total number of functions satisfies

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

as required. This completes the induction proof for (25).

Since every function $h_{0,\boldsymbol{\eta}}^{[j]}(t,\cdot)$ in (25) is continuous on U_t, it follows by induction 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 (22) implies that $D^{\boldsymbol{\eta}}[P_0 g_t(\boldsymbol{y})]\to 0$ as $\boldsymbol{y}\to\partial U_t$. Hence, it follows by an adaptation of [16, 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

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

where we used the assumption (23) 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 A1 for Theorem 3 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 below.

3.2 Smoothing by preintegration for Dirac δ distributions

In this section we show that the same smoothing by preintegration methodology also works for distributions that are constructed by a Dirac δ function, which will allow us to also estimate the pdf as formulated in (3).

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

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

where $\delta(\cdot)$ is the Dirac δ function as characterised by (4) and $\phi : \mathbb{R}^{d+1} \to \mathbb{R}$ satisfies Assumption A1.

For $t \in [a, b]$, $\boldsymbol{y} \in U_t$ and assuming $U_t \neq \emptyset$, we have $t \in \phi(\mathbb{R}, \boldsymbol{y})$, i.e., t is in the image of $\mathbb{R}$ under $\phi(\cdot, \boldsymbol{y})$. Let $\xi(t, \boldsymbol{y})$ be the unique point of discontinuity in dimension 0 as in (15). Applying the preintegration operator (5) to the distribution g_t and using the change of variables $z = \phi(y_0, \boldsymbol{y})$ so that $y_0 = \xi(z, \boldsymbol{y})$, we obtain

$$\begin{aligned} P_0 g_t(\boldsymbol{y}) &= \int_{-\infty}^{\infty} \delta\big(t - \phi(y_0, \boldsymbol{y})\big)\, \rho_0(y_0)\, \mathrm{d}y_0 = \int_{\phi(\mathbb{R},\boldsymbol{y})} \delta(t - z)\, \rho_0(\xi(z, \boldsymbol{y}))\, \frac{\partial}{\partial z}\xi(z, \boldsymbol{y})\, \mathrm{d}z \\ &= \int_{\phi(\mathbb{R},\boldsymbol{y})} \delta(t - z)\, \rho_0(\xi(z, \boldsymbol{y}))\, \frac{1}{D^0\phi(\xi(z, \boldsymbol{y}), \boldsymbol{y})}\, \mathrm{d}z = \frac{\rho_0(\xi(t, \boldsymbol{y}))}{D^0\phi(\xi(t, \boldsymbol{y}), \boldsymbol{y})}, \end{aligned} \tag{26}$$

where we used (15), (18), and the definition of the $\delta(\cdot)$ function (4). For $\boldsymbol{y} \in \mathbb{R}^d \setminus U_t$ and so $t \notin \phi(\mathbb{R}, \boldsymbol{y})$, we have $\delta(t - z) = 0$ for *all* $z \in \phi(\mathbb{R}, \boldsymbol{y})$ and hence $P_0 g_t(\boldsymbol{y}) = 0$. As expected, (26) is the derivative of (19) with respect to t.

With a similar proof to Theorem 3, we now show that the preintegrated distribution $P_0 g_t$ is also smooth, with a different bound on its norm.

Theorem 4 *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 A1 and Assumption A2 for $q = 1$ and all $\boldsymbol{\eta} \leq \boldsymbol{\nu}$. Then for $t \in [a, b]$ the distribution*

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

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

$$\sup_{t \in [a,b]} \|P_0 g_t\|_{\mathcal{H}_d^{\boldsymbol{\nu}}} \leq \left(\sum_{\boldsymbol{\eta} \leq \boldsymbol{\nu}} \frac{\big(8^{|\boldsymbol{\eta}|} |\boldsymbol{\eta}|!\big)^2 B_{1,\boldsymbol{\eta}}}{\gamma_{\boldsymbol{\eta}}} \right)^{1/2} < \infty\,. \tag{27}$$

Proof. From (26) the preintegrated distribution can be written as

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

For any $\boldsymbol{\eta} \in \mathbb{N}_0^d$ with $\boldsymbol{\eta} \leq \boldsymbol{\nu}$, we first prove by induction on $|\boldsymbol{\eta}|$ that the $\boldsymbol{\eta}$th derivative of $P_0 g_t$ is given by

$$D^{\boldsymbol{\eta}}[P_0 g_t(\boldsymbol{y})] = \begin{cases} \displaystyle\sum_{j=1}^{J_{1,\boldsymbol{\eta}}} h_{1,\boldsymbol{\eta}}^{[j]}(t,\boldsymbol{y}) & \text{if } \boldsymbol{y} \in U_t\,, \quad \text{with} \quad J_{1,\boldsymbol{\eta}} \leq 8^{|\boldsymbol{\eta}|}|\boldsymbol{\eta}|!\,, \\ 0 & \text{if } \boldsymbol{y} \in \mathbb{R}^d \setminus U_t\,, \end{cases} \tag{28}$$

where each function $h_{1,\boldsymbol{\eta}}^{[j]}$ is now of the form (21) with $q = 1$.

For the base case $\boldsymbol{\eta} = \boldsymbol{0}$, we take $r = 0$ (and so there are no $\boldsymbol{\alpha}_\ell$ terms), $\beta = 0$, and $J_{1,\boldsymbol{0}} = 1$ to recover the single function $\rho_0(\xi(t,\boldsymbol{y}))/D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})$. Suppose now that (28) holds for some $\boldsymbol{\eta} \in \mathbb{N}_0^d$ and consider any $i \in \{1:d\}$ and $\boldsymbol{y} \in U_t$. We have

$$D^i D^{\boldsymbol{\eta}}[P_0 g_t(\boldsymbol{y})] = \sum_{j=1}^{J_{1,\boldsymbol{\eta}}} D^i h_{1,\boldsymbol{\eta}}^{[j]}(t,\boldsymbol{y}) = \sum_{j=1}^{J_{1,\boldsymbol{\eta}}} \sum_{k=1}^{K_{1,\boldsymbol{\eta}}} h_{1,\boldsymbol{\eta}+\boldsymbol{e}_i}^{[j,k]}(t,\boldsymbol{y}) = \sum_{j'=1}^{J_{1,\boldsymbol{\eta}+\boldsymbol{e}_i}} h_{1,\boldsymbol{\eta}+\boldsymbol{e}_i}^{[j']}(t,\boldsymbol{y}).$$

In the second equality we used Lemma 9 in the Appendix with $q = 1$, which states that each function $D^i h_{1,\boldsymbol{\eta}}^{[j]}$ can be written as a sum of $K_{1,\boldsymbol{\eta}} \leq 8|\boldsymbol{\eta}|+3$ functions of the form (21) with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta}+\boldsymbol{e}_i$. We enumerated these functions with the notation $h_{1,\boldsymbol{\eta}+\boldsymbol{e}_i}^{[j,k]}$ and then relabeled all functions for different combinations of indices j and k with the notation $h_{1,\boldsymbol{\eta}+\boldsymbol{e}_i}^{[j']}$. The total number of functions satisfies

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

as required. This completes the induction proof for (28).

Since every function $h_{1,\boldsymbol{\eta}}^{[j]}(t,\cdot)$ in (28) is continuous on U_t, it follows by induction 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 (22) implies that $D^{\boldsymbol{\eta}}[P_0 g_t(\boldsymbol{y})] \to 0$ as $\boldsymbol{y} \to \partial U_t$. Hence, it follows by an adaptation of [16, Lemma 9] that $P_0 g_t \in C^{\boldsymbol{\nu}}(\mathbb{R}^d)$.

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

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

where we used the assumption (23) with $q = 1$. This completes the proof. □

4 Distribution function and density estimators

In this section we develop our 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 (2) for the cdf as

$$F(t) = \int_{\mathbb{R}^d} P_0\big(\mathrm{ind}\big(t - \phi(\cdot,\boldsymbol{y})\big)\big)\, \boldsymbol{\rho}(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y} = \int_{U_t} \Phi_0(\xi(t,\boldsymbol{y}))\boldsymbol{\rho}(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}\,, \tag{29}$$

where in the last step we have substituted in the simplified formula (19) for a preintegrated indicator function. Similarly, using the representation (3) along with (26) we can write

$$f(t) = \int_{\mathbb{R}^d} P_0\big(\delta\big(t - \phi(\boldsymbol{y})\big)\big)\boldsymbol{\rho}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_{U_t} \frac{\rho_0(\xi(t,\boldsymbol{y}))}{D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})}\boldsymbol{\rho}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,. \tag{30}$$

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 (8) to the d-dimensional integrals (29) and (30), we obtain the approximations $\widehat{F}_N$ and $\widehat{f}_N$ as follows:

$$F(t) \approx \widehat{F}_N(t) \coloneqq Q_{d,N}\big(\Phi_0(\xi(t,\cdot))\big) = \frac{1}{N}\sum_{n=0}^{N-1}\Phi_0\big(\xi(t,\boldsymbol{\tau}_n)\big), \tag{31}$$

$$f(t) \approx \widehat{f}_N(t) \coloneqq Q_{d,N}\bigg(\frac{\rho_0(\xi(t,\cdot))}{D^0\phi(\xi(t,\cdot),\cdot)}\bigg) = \frac{1}{N}\sum_{n=0}^{N-1}\frac{\rho_0(\xi(t,\boldsymbol{\tau}_n))}{D^0\phi(\xi(t,\boldsymbol{\tau}_n),\boldsymbol{\tau}_n)}\,. \tag{32}$$

We are particularly interested in using randomly shifted lattice points (7) with (9), but the description in this section applies 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 $\widehat{F}_N$ and $\widehat{f}_N$ 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, which leads to *biased* estimators $\widetilde{F}_N$ and $\widetilde{f}_N$. We now detail how to construct these biased estimators efficiently.

First, consider approximating the cdf F at $t \in [a,b]$. For each QMC point $\boldsymbol{\tau}_n$ we must: (i) find the point of discontinuity $\xi(t,\boldsymbol{\tau}_n)$, 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

$$\phi_{0,n} \coloneqq \phi(\cdot,\boldsymbol{\tau}_n),$$

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

To find the point of discontinuity we use a numerical root-finding algorithm, e.g., Newton's method. Since $\phi \in C^{(\nu_0,\boldsymbol{\nu})}(\mathbb{R}^{d+1})$ with $\nu_0 = |\boldsymbol{\nu}| + 1$, we have $\phi_{0,n} \in C^{\nu_0}(\mathbb{R})$ for each $\boldsymbol{\tau}_n$. If $|\boldsymbol{\nu}| \geq 1$ 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 $\widetilde{\xi}$.

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 $\Phi_0(y_0) = \int_{-\infty}^{y_0}\rho_0(z)\,\mathrm{d}z$ by a quadrature rule. In both cases we denote the approximation of Φ_0 by $\widetilde{\Phi}_0$.

Approximating the pdf f at $t \in [a,b]$ is similar: (i) obtaining the point $\xi(t,\boldsymbol{\tau}_n)$ 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.

Algorithm 1 Pointwise cdf estimator

Given $t \in [a,b]$, $N \in \mathbb{N}$ and $\{\boldsymbol{\tau}_n\}_{n=0}^{N-1}$ a d-dimensional QMC pointset:

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

Algorithm 2 Pointwise pdf estimator

Given $t \in [a,b]$, $N \in \mathbb{N}$ and $\{\boldsymbol{\tau}_n\}_{n=0}^{N-1}$ a d-dimensional QMC pointset:

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

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

$$\widetilde{F}_N(t) := \frac{1}{N}\sum_{n=0}^{N-1} \widetilde{\Phi}_0\big(\widetilde{\xi}(t, \boldsymbol{\tau}_n)\big)\,, \tag{33}$$

$$\widetilde{f}_N(t) := \frac{1}{N}\sum_{n=0}^{N-1} \frac{\rho_0\big(\widetilde{\xi}(t, \boldsymbol{\tau}_n)\big)}{D^0\phi\big(\widetilde{\xi}(t, \boldsymbol{\tau}_n), \boldsymbol{\tau}_n\big)}\,. \tag{34}$$

Algorithms 1 and 2 give explicit implementations of (33) and (34).

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 $\mathrm{cost}(\widehat{F}_N(t)) = \mathcal{O}(N)$ and $\mathrm{cost}(\widehat{f}_N(t)) = \mathcal{O}(N)$. 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.

To analyse the cost of the pointwise approximations $\widetilde{F}_N(t)$ and $\widetilde{f}_N(t)$ in Algorithms 1 and 2, we assume that the number of evaluations of $\phi_{0,n}$ and $\phi'_{0,n}$ in the root-finding method to compute $\widetilde{\xi}(t, \boldsymbol{\tau}_n)$ 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 1, we also assume that $\mathrm{cost}(\rho_0) = \mathcal{O}(1)$, 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 $\boldsymbol{\tau}_n$ 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 $\phi'_{0,n}$:

$$\begin{cases} \mathrm{cost}(\phi) = \$(d), & \mathrm{cost}(\phi_{0,n}) = \$(1) \text{ with precomputed contribution of } \boldsymbol{\tau}_n, \\ \mathrm{cost}(D^0\phi) = \$(d), & \mathrm{cost}(\phi'_{0,n}) = \$(1) \text{ with precomputed contribution of } \boldsymbol{\tau}_n, \end{cases} \tag{35}$$

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

The cost of Algorithms 1 and 2 are then

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

For large d this will be much more efficient than a naive implementation without precomputed contribution of $\boldsymbol{\tau}_n$, which would have

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

4.3 Approximating the cdf and pdf

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 $\widehat{F}_N$ and $\widehat{f}_N$. We denote the approximations by

$$\widehat{F}_{N,M} \coloneqq L_M(\widehat{F}_N) = L_M\Big(Q_{d,N}\big(\Phi_0(\xi(\bullet,\cdot))\big)\Big), \tag{36}$$

$$\widehat{f}_{N,M} \coloneqq L_M(\widehat{f}_N) = L_M\Big(Q_{d,N}\Big(\frac{\rho_0(\xi(\bullet,\cdot))}{D^0\phi(\xi(\bullet,\cdot),\cdot)}\Big)\Big), \tag{37}$$

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 $\widetilde{\xi} \approx \xi$, and for the cdf also the one-dimensional integral by $\widetilde{\Phi}_0 \approx \Phi_0$. This leads to the biased estimators

$$\widetilde{F}_{N,M} \coloneqq L_M(\widetilde{F}_N) = L_M\Big(Q_{d,N}\Big(\widetilde{\Phi}_0\big(\widetilde{\xi}(\bullet,\cdot)\big)\Big)\Big), \tag{38}$$

$$\widetilde{f}_{N,M} \coloneqq L_M(\widetilde{f}_N) = L_M\left(Q_{d,N}\left(\frac{\rho_0\big(\widetilde{\xi}(\bullet,\cdot)\big)}{D^0\phi\big(\widetilde{\xi}(\bullet,\cdot),\cdot\big)}\right)\right). \tag{39}$$

Recall from the definition of the Lagrange interpolation operator (11) that to construct the estimators $\widetilde{F}_{N,M}$ and $\widetilde{f}_{N,M}$ we must compute the pointwise approximations $\widetilde{F}_N$ and $\widetilde{f}_N$ at all of the interpolation nodes $\{t_m\}_{m=0}^M$. One way to implement the estimator $\widetilde{F}_{N,M}$ as in (38) is to simply run Algorithm 1 for each t_m for $m = 0, 1, \ldots, M$, with $\mathrm{cost}(\widetilde{F}_{N,M}) = (M+1) \times \mathrm{cost}(\widetilde{F}_N)$. However, since we can use the same QMC rule for each interpolation node t_m, it is more efficient to instead *vectorise* Algorithm 1 and utilise precomputed contributions of each QMC point $\boldsymbol{\tau}_n$ so that we only have to deal with $M+1$ univariate functions. Similar arguments can also be made for the cdf estimator $\widetilde{f}_{N,M}$. Explicit algorithms detailing how to construct the estimators $\widetilde{F}_{N,M}$ and $\widetilde{f}_{N,M}$ are given in Algorithms 3 and 4.

Algorithm 3 cdf estimator

Given $M \in \mathbb{N}$, $\{t_m\}_{m=0}^{M} \subset [a, b]$, $N \in \mathbb{N}$ and $\{\boldsymbol{\tau}_n\}_{n=0}^{N-1}$ a d-dimensional QMC pointset:

Initialise: $\widetilde{F}_N(t_m) \leftarrow 0$ for each $m = 0, 1, \ldots, M$
for $n = 0, 1, \ldots, N-1$ **do**
  Precompute the contribution of $\boldsymbol{\tau}_n$ to $\phi_{0,n} = \phi(\cdot, \boldsymbol{\tau}_n)$ and $\phi'_{0,n} = D^0\phi(\cdot, \boldsymbol{\tau}_n)$
  for $m = 0, 1, \ldots, M$ **do**
    Compute the point of discontinuity $\widetilde{\xi}(t_m, \boldsymbol{\tau}_n)$
    Approximate the 1D integral $\widetilde{\Phi}_0\big(\widetilde{\xi}(t_m, \boldsymbol{\tau}_n)\big) \approx \int_{-\infty}^{\widetilde{\xi}(t_m, \boldsymbol{\tau}_n)} \rho_0(z)\,\mathrm{d}z$
    Sum: $\widetilde{F}_N(t_m) \leftarrow \widetilde{F}_N(t_m) + \widetilde{\Phi}_0\big(\widetilde{\xi}(t_m, \boldsymbol{\tau}_n)\big)$
  end for
end for
Average: $\widetilde{F}_N(t_m) \leftarrow \widetilde{F}_N(t_m)/N$ for each $m = 0, 1, \ldots, M$
Interpolate: $\widetilde{F}_{N,M} \leftarrow \sum_{m=0}^{M} \widetilde{F}_N(t_m)\,\chi_{M,m}$

Algorithm 4 pdf estimator

Given $M \in \mathbb{N}$, $\{t_m\}_{m=0}^{M} \subset [a, b]$, $N \in \mathbb{N}$ and $\{\boldsymbol{\tau}_n\}_{n=0}^{N-1}$ a d-dimensional QMC pointset:

Initialise: $\widetilde{f}_N(t_m) \leftarrow 0$ for each $m = 0, 1, \ldots, M$
for $n = 0, 1, \ldots, N-1$ **do**
  Precompute the contribution of $\boldsymbol{\tau}_n$ to $\phi_{0,n} = \phi(\cdot, \boldsymbol{\tau}_n)$ and $\phi'_{0,n} = D^0\phi(\cdot, \boldsymbol{\tau}_n)$
  for $m = 0, 1, \ldots, M$ **do**
    Compute the point of discontinuity $\widetilde{\xi}(t_m, \boldsymbol{\tau}_n)$
    Sum: $\widetilde{f}_N(t_m) \leftarrow \widetilde{f}_N(t_m) + \frac{\rho_0\big(\widetilde{\xi}(t_m, \boldsymbol{\tau}_n)\big)}{D^0\phi\big(\widetilde{\xi}(t_m, \boldsymbol{\tau}_n), \boldsymbol{\tau}_n\big)}$
  end for
end for
Average: $\widetilde{f}_N(t_m) \leftarrow \widetilde{f}_N(t_m)/N$ for each $m = 0, 1, \ldots, M$
Interpolate: $\widetilde{f}_{N,M} \leftarrow \sum_{m=0}^{M} \widetilde{f}_N(t_m)\,\chi_{M,m}$

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 (35) 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 3 and 4 are then

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

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 $\mathrm{cost}(\widetilde{f}_{N,M}) = \mathcal{O}(N\,M\,K_{\mathrm{root}}\,\$(d))$.

5 Error analysis

5.1 Regularity of F and f

In order to utilise the results on the error for interpolation on $[a,b]$ from Section 2.3 we need to know quantitatively how smooth 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). Since in Assumption A1 we assume that ϕ is $|\boldsymbol{\nu}|+1$ times differentiable with respect to variable y_0, we can expect a similar level of smoothness for F and f.

To see the dependence on t more explicitly, recall that the formulas (2) for cdf and (3) for the pdf can be formulated as d-dimensional integrals as in (29) and (30), 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 2 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 beyond those already required.

Theorem 5 *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 A1 and Assumption A2 for $\boldsymbol{\eta} = \mathbf{0}$ and all $q = 1, 2, \ldots, |\boldsymbol{\nu}|+1$. Assume additionally that $U_t = \mathbb{R}^d$ for all $t \in [a,b]$. Then $F \in W^{|\boldsymbol{\nu}|+1,\infty}[a,b]$ and $f \in W^{|\boldsymbol{\nu}|,\infty}[a,b]$, and for $q = 1, \ldots, |\boldsymbol{\nu}|+1$ 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 $F \in W^{|\boldsymbol{\nu}|+1,\infty}[a,b]$. Then since $f = F'$ the result for the pdf follows immediately.

Consider the derivative of order $q \leq |\boldsymbol{\nu}|+1$. First, differentiating (29) with respect to t, by the Leibniz rule we obtain

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

Recall that we have

$$\frac{\partial}{\partial t}\Phi_0(\xi(t,\boldsymbol{y})) = \frac{\rho_0(\xi(t,\boldsymbol{y}))}{D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})}. \tag{40}$$

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

$$\frac{\partial^q}{\partial t^q}\Phi_0(\xi(t,\boldsymbol{y})) = \sum_{j=1}^{J_{q,\mathbf{0}}} h_{q,\mathbf{0}}^{[j]}(t,\boldsymbol{y}), \quad \text{with} \quad J_{q,\mathbf{0}} \leq 3^{q-1}\,(q-1)!\,, \tag{41}$$

where each function $h_{q,\mathbf{0}}^{[j]}$ is of the form (21) with $\boldsymbol{\eta} = \mathbf{0}$. The base step $q = 1$ holds for the single function (40) with $r = 0$ (no $\boldsymbol{\alpha}$), $\beta = 0$, and $J_{1,\mathbf{0}} = 1$. Suppose next that (41) holds for some $q \geq 1$. Then we have

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

In the second equality we used Lemma 10 in the Appendix 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 (21), 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 $h^{[j']}_{q+1,\mathbf{0}}$. The total number of functions satisfies

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

as required. This completes the induction proof for (41).

Thus we have

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

and

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

where we used assumption (23) with $\boldsymbol{\eta} = \mathbf{0}$ and the Cauchy-Schwarz inequality. Hence $F \in W^{|\boldsymbol{\nu}|+1,\infty}[a,b]$ and also $f = F' \in W^{|\boldsymbol{\nu}|,\infty}[a,b]$. □

5.2 Error of cdf and pdf estimators

In this subsection we analyse the error of the unbiased estimators from Section 4. First, we prove bounds for the *root-mean-square error (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 $\mathbb{E}_{\boldsymbol{\Delta}}$, is taken with respect to the random shift $\boldsymbol{\Delta}$ in the lattice rule.

Theorem 6 (Pointwise RMSE) *Let $d \ge 1$, $\boldsymbol{\nu} = \mathbf{1} \in \mathbb{N}^d$ and $[a,b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumption A1 and Assumption A2 for all $\boldsymbol{\eta} \in \{0,1\}^d$ 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* (8)*, then, for $t \in [a,b]$, the estimators $\widehat{F}_N(t)$ and $\widehat{f}_N(t)$ as given in* (31) *and* (32) *satisfy, for all $\lambda \in (1/\omega, 1]$,*

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\big[\big|F(t) - \widehat{F}_N(t)\big|^2\big]} \le C_{F,\lambda}\,\phi_{\mathrm{tot}}(N)^{-1/(2\lambda)}\,, \tag{42}$$

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\big[\big|f(t) - \widehat{f}_N(t)\big|^2\big]} \le C_{f,\lambda}\,\phi_{\mathrm{tot}}(N)^{-1/(2\lambda)}\,, \tag{43}$$

where, with ω and c as in Theorem 1,

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

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

Proof. First for the cdf estimator, using (29) and the definition (31) of $\widehat{F}_N$ we can write the mean square error as

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

Then since ϕ and ρ_0 satisfy Assumption A1 and Assumption A2 for all $\boldsymbol{\eta} \in \{0,1\}^d$ with $q = 0$, we can apply Theorem 3 to show that the preintegrated function $\Phi_0(\xi(t, \cdot))$ belongs to $\mathcal{H}_d^{\mathbf{1}}$ and its norm is bounded by (24) with $\boldsymbol{\nu} = \mathbf{1}$. Substituting this norm bound into the CBC error bound (10) 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 (27) in Theorem 4 instead of Theorem 3. □

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

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

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

Theorem 7 *Let $d \geq 1$, $\boldsymbol{\nu} = \mathbf{1} \in \mathbb{N}^d$ and $[a, b] \subset \mathbb{R}$. Suppose ϕ and ρ_0 satisfy*

$$\begin{cases} \textit{Assumption A1; and} \\ \textit{Assumption A2 for } \boldsymbol{\eta} = \mathbf{0} \textit{ and all } q \leq d+1\textit{; and} \\ \textit{Assumption A2 for all } \boldsymbol{\eta} \in \{0,1\}^d \textit{ with } q = 0 \textit{ for the cdf case; and} \\ \textit{Assumption A2 for all } \boldsymbol{\eta} \in \{0,1\}^d \textit{ with } q = 1 \textit{ for the pdf case.} \end{cases}$$

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 (7) *and let L_M be Lagrange interpolation on $[a, b]$ based on Chebyshev points as in* (11)*. Then for $\sigma \in \mathbb{N}$ with $\sigma \leq d - 1$ and $M > \sigma$, the estimators $\widehat{F}_{N,M}$ and $\widehat{f}_{N,M}$ in* (36) *and* (37) *satisfy, for all $\lambda \in (1/\omega, 1]$,*

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[\|F - \widehat{F}_{N,M}\|_{L^2}^2\right]} \leq C_{F,\lambda,\sigma}\Big(\phi_{\mathrm{tot}}(N)^{-1/(2\lambda)} + M^{-\sigma}\Big), \tag{44}$$

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}\left[\|f - \widehat{f}_{N,M}\|_{L^2}^2\right]} \leq C_{f,\lambda,\sigma}\Big(\phi_{\mathrm{tot}}(N)^{-1/(2\lambda)} + M^{-\sigma}\Big), \tag{45}$$

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 6, $\omega \in (1, 2]$ as in Theorem 1, and

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

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

$$\mathbb{E}_{\boldsymbol{\Delta}}\left[\|F - \widehat{F}_{N,M}\|_{L^2}^2\right] \leq 2\,\mathbb{E}_{\boldsymbol{\Delta}}\left[\|F - \widehat{F}_N\|_{L^2}^2\right] + 2\,\mathbb{E}_{\boldsymbol{\Delta}}\left[\|\widehat{F}_N - L_M\widehat{F}_N\|_{L^2}^2\right]. \tag{46}$$

The first term in (46) can be bounded by the pointwise error as follows. By Fubini's Theorem we may swap the expected value with respect to $\boldsymbol{\Delta}$ and the integral over $[a, b]$ to obtain

$$\mathbb{E}_{\boldsymbol{\Delta}}\left[\|F - \widehat{F}_N\|_{L^2}^2\right] = \int_a^b \mathbb{E}_{\boldsymbol{\Delta}}\left[|F(t) - \widehat{F}_N(t)|^2\right]\mathrm{d}t \leq (b-a)\,C_{F,\lambda}^2\,\phi_{\mathrm{tot}}(N)^{-1/\lambda}, \tag{47}$$

where we have substituted in the bound (42).

For the second term in (46) we will use the Lagrange interpolation error bound (12) by first adapting the proof of Theorem 5 to show that $\widehat{F}_N \in W^{\sigma+1,\infty}[a,b]$ for all random shifts $\boldsymbol{\Delta}$. Differentiating (31) with respect to t then substituting in the formula (41) for $q = \sigma + 1$ gives

$$\widehat{F}_N^{(\sigma+1)}(t) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\partial^{\sigma+1}}{\partial t^{\sigma+1}} \Phi_0\big(\xi(t, \boldsymbol{\tau}_n^{\boldsymbol{\Delta}})\big) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{J_{\sigma+1,\mathbf{0}}} h_{\sigma+1,\mathbf{0}}^{[j]}(t, \boldsymbol{\tau}_n^{\boldsymbol{\Delta}}), \tag{48}$$

with $J_{\sigma+1,\mathbf{0}} \le 3^\sigma\, \sigma!$, where we emphasized the explicit dependence of each QMC point on the random shift with the superscript $\boldsymbol{\Delta}$. Hence, we can apply (12) to obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\Delta}}\big[\|\widehat{F}_N - L_M \widehat{F}_N\|_{L^2}^2\big] &\le (b-a)\, \mathbb{E}_{\boldsymbol{\Delta}}\big[\|\widehat{F}_N - L_M \widehat{F}_N\|_{L^\infty}^2\big] \\
&\le (b-a) \left(\frac{4}{\pi\sigma(M-\sigma)^\sigma}\right)^2 \mathbb{E}_{\boldsymbol{\Delta}}\big[\|\widehat{F}_N^{(\sigma+1)}\|_{L^1}^2\big] \\
&\le (b-a) \left(\frac{4\,(\sigma+1)^\sigma}{\pi\sigma}\right)^2 \mathbb{E}_{\boldsymbol{\Delta}}\big[\|\widehat{F}_N^{(\sigma+1)}\|_{L^1}^2\big]\, M^{-2\sigma}, \qquad (49)
\end{aligned}$$

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

To bound the expected value in (49), we use the formula (48), the Cauchy–Schwarz inequality for integral and sum, and Fubini's Theorem to obtain

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\Delta}}\big[\|\widehat{F}_N^{(\sigma+1)}\|_{L^1}^2\big] &\le \frac{(b-a)\, J_{\sigma+1,\mathbf{0}}}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{J_{\sigma+1,\mathbf{0}}} \int_a^b \int_{[0,1]^d} |h_{\sigma+1,\mathbf{0}}^{[j]}(t, \boldsymbol{\tau}_n^{\boldsymbol{\Delta}})|^2 \,\mathrm{d}\boldsymbol{\Delta}\,\mathrm{d}t \\
&= \frac{(b-a)\, J_{\sigma+1,\mathbf{0}}}{N} \sum_{n=0}^{N-1} \sum_{j=1}^{J_{\sigma+1,\mathbf{0}}} \int_a^b \int_{\mathbb{R}^d} |h_{\sigma+1,\mathbf{0}}^{[j]}(t, \boldsymbol{y})|^2\, \boldsymbol{\rho}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}t \\
&\le (b-a)^2\, J_{\sigma+1,\mathbf{0}}^2\, B_{\sigma+1,\mathbf{0}} \le [(b-a)\, 3^\sigma\, \sigma!]^2\, B_{\sigma+1,\mathbf{0}}, \qquad (50)
\end{aligned}$$

where we made a change of variables and then used the upper bound (23) as well as the bound $J_{\sigma+1,\mathbf{0}} \le 3^\sigma\, \sigma!$. Substituting (50) into (49), we conclude that

$$\mathbb{E}_{\boldsymbol{\Delta}}\big[\|\widehat{F}_N - L_M \widehat{F}_N\|_{L^2}^2\big] \le (b-a)\, C_{F,\sigma}^2\, M^{-2\sigma}, \tag{51}$$

with $C_{F,\sigma}$ as defined in the theorem. Substituting (47) and (51) into (46), we obtain the RMISE bound (44) 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 $\|\widehat{f}_N^{(\sigma+1)}\|_{L^1}$. Similar to the relationship $f = F'$, it follows from (40) that $\widehat{f}_N$ is the derivative with respect to t of $\widehat{F}_N$. Thus

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

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

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

and eventually arrive at the RMISE bound (45) for the pdf estimator. □

Theorem 7 implies that we can take σ up to $d-1$ 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 7 applies in practice let now N be a prime or a prime power, in which case $\phi_{\mathrm{tot}}(N) \sim N$. Then to balance the QMC and interpolation error Theorem 7 implies that we should take $M \sim N^{1/\sigma}$. The final result is that our estimators converge at a rate arbitrarily close to $1/N$. It is summarised in the following Corollary.

Corollary 8 *Suppose that the conditions in Theorem 7 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 satisfy*

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

6 Numerical results

To test our 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,

$$X = \sum_{i=0}^{d} \exp(W_i) = \sum_{i=0}^{d} \exp(\boldsymbol{A}_i \boldsymbol{Y}) =: \phi(\boldsymbol{Y}). \tag{52}$$

where $\boldsymbol{W} = (W_i)_{i=0}^d$ is a $(d+1)$-dimensional multivariate normal vector with mean $\mathbf{0}$ and covariance Σ. In the second equality, we have factorised the covariance matrix as $\Sigma = AA^\top$ and made the change of variables $\boldsymbol{Y} = A^{-1}\boldsymbol{W}$, so that ϕ is a function of the $(d+1)$-dimensional standard normal vector $\boldsymbol{Y} = (Y_i)_{i=0}^d$. In (52) $\boldsymbol{A}_i$ denotes the ith row of the matrix factor A. We use the *principal components* or *PCA* factorisation, 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 our 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:

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

It can be easily verified that Assumption A1 is satisfied for both covariance matrices.

For the QMC approximations we use “off-the-shelf” embedded lattice rules [6] based on the generating vectors “C” for $\Sigma^{(1)}$ and “A” for $\Sigma^{(2)}$ from [18]. The final estimate is the average over $R = 32$ random shifts, and we estimate the RMSE 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 realisations. All computations were run on the computational cluster Katana [23] at UNSW Sydney. Convergence results for the cdf and pdf at the point $t = 60$ for both covariance matrices are given in Figure 1 for $N = 2^{10}, 2^{11}, \ldots, 2^{20}$. 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 for $\Sigma^{(1)}$ are similar to those presented in [5].

In Figure 2, we plot the QMC with preintegration estimators for both the cdf F (left) and pdf f (middle) on the interval $[40, 100]$, along with the convergence of the RMISE (right). The approximations of the cdf and pdf use $N = 2^{20}$ QMC points averaged over

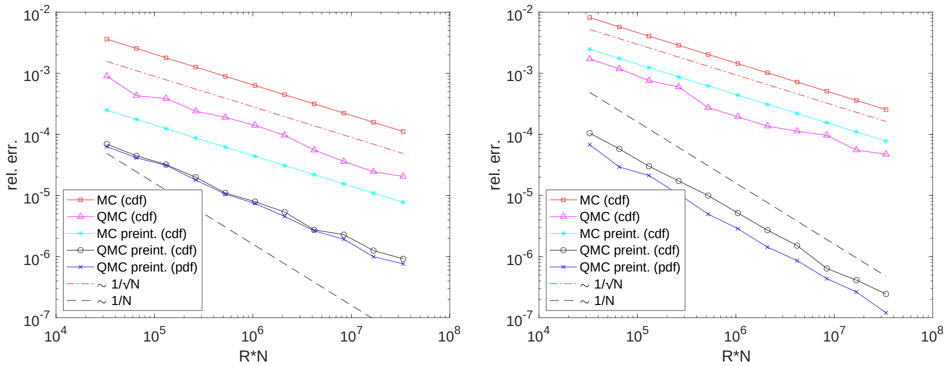

Figure 1: Relative RMSE convergence in N for MC, QMC, and QMC with preintegration for $F(60) = \mathbb{P}[X \leq 60]$, and also QMC with preintegration for $f(60)$, for $\Sigma^{(1)}$ (left) and $\Sigma^{(2)}$ (right).

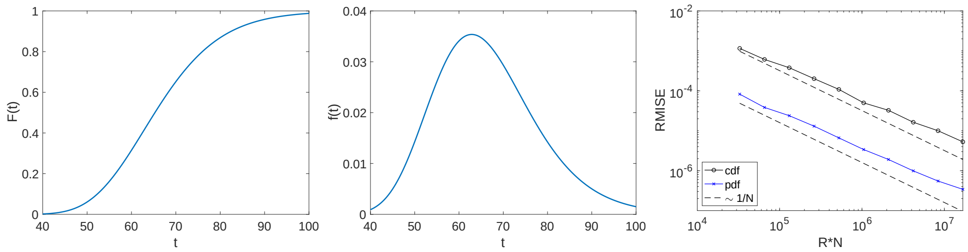

Figure 2: Approximate cdf, approximate pdf and RMISE convergence in N for $\Sigma^{(2)}$.

$R = 32$ random shifts and degree $M = 42$ Lagrange interpolation based on Chebyshev points. For the RMISE we use $N = 2^{10}, 2^{11}, \ldots, 2^{19}$ with interpolation degree $M = \lceil N^{1/4} \rceil + 10$ and $R = 32$ random shifts. To estimate the RMISE, 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 perform a MC approximation of the mean by averaging the L^2 error over the random shifts. As expected we observe almost $1/N$ convergence for the RMISE. This demonstrates that our method is an effective practical strategy.

References

[1] A. B. Abdellah, P. L'Ecuyer, A. B. Owen, F. Puchhammer, *Density estimation by randomized quasi-Monte Carlo*, SIAM/ASA J. Uncertaint. Quantif. **9**, 280–301 (2021).

[2] N. Achtsis, R. Cools, D. Nuyens, *Conditional sampling for barrier option pricing under the LT method*, SIAM J. Financial Math. **4** (2013), 327–352.

[3] N. Achtsis, R. Cools, D. Nuyens, *Conditional sampling for barrier option pricing under the Heston model*, in: J. Dick, F.Y. Kuo, G.W. Peters, I.H. Sloan (eds.), Monte Carlo and Quasi-Monte Carlo Methods 2012, pp. 253–269, Springer-Verlag, Berlin/Heidelberg (2013).

[4] C. Bayer, C. B. Hammouda, R. Tempone, *Numerical smoothing with hierarchical adaptive sparse grids and quasi-Monte Carlo methods for efficient option pricing*, arXiv:2111.01874.

[5] Z. I. Botev, R. Salomone, D. Mackinlay, *Fast and accurate computation of the distribution of sums of dependent log-normals*, Ann. Oper. Res. **280** (2019), 19–46.

[6] R. Cools, F. Y. Kuo, D. Nuyens. *Constructing embedded lattices rules for multivariate integration*, SIAM J. Sci. Comp., **28** (2006), 2162–2188.

[7] J. Dick, F. Y. Kuo, I. H. Sloan, *High dimensional integration – the quasi-Monte Carlo way*, Acta Numer. **22** (2013), 133–288.

[8] A. D. Gilbert, F. Y. Kuo, I. H. Sloan, *Equivalence between Sobolev spaces of first order dominating mixed smoothness and unanchored ANOVA spaces on* $\mathbb{R}^d$, Math. Comp., to appear.

[9] A. D. Gilbert, F. Y. Kuo, I. H. Sloan, *Preintegration is not smoothing when monotonicity fails*, arXiv:2112.11621, (2021).

[10] M. B. Giles, T. Nagapetyan, K. Ritter, *Multilevel Monte Carlo approximation of distribution functions and densities*, SIAM/ASA J. Uncertain. Quantif. **3** (2013), 267–295.

[11] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, Berlin (2003).

[12] P. Glasserman, J. Staum, *Conditioning on one-step survival for barrier option simulations*, Oper. Res. **49** (2001), 923–937.

[13] M. Griebel, F. Y. Kuo, I. H. Sloan, *The smoothing effect of integration in* $\mathbb{R}^d$ *and the ANOVA decomposition*, Math. Comp. **82** (2013), 383–400.

[14] M. Griebel, F. Y. Kuo, I. H. Sloan, *Note on "The smoothing effect of integration in* $\mathbb{R}^d$ *and the ANOVA decomposition"*, Math. Comp. **86** (2017), 1847–1854.

[15] M. Griebel, F. Y. Kuo, I. H. Sloan, *The ANOVA decomposition of a non-smooth function of infinitely many variables can have every term smooth*, Math. Comp. **86** (2017), 1855–1876.

[16] A. Griewank, F. Y. Kuo, H. Leövey, I. H. Sloan, *High dimensional integration of kinks and jumps — smoothing by preintegration*, J. Comput. App. Math. **344** (2018), 259–274.

[17] M. Holtz, *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance* (PhD thesis), Springer-Verlag, Berlin, 2011.

[18] F. Y. Kuo, `https://web.maths.unsw.edu.au/~fkuo/lattice/index.html`.

[19] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, B. J. Waterhouse, *Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands*, J. Complexity **26** (2010), 135–160.

[20] P. L'Ecuyer, F. Puchhammer, A. B. Abdellah, *Monte Carlo and quasi-Monte Carlo density estimation via conditioning*, arXiv:1906.04607.

[21] J. A. Nichols, 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), 444–468.

[22] L. N. Trefethen, *Approximation Theory and Practice*, SIAM, Philadelphia (2020).

[23] UNSW Sydney, *Katana* (2010), `https://doi.org/10.26190/669x-a286`.

[24] C. Weng, X. Wang, 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), B298–B322.

A Technical results

Lemma 9 *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 A1, and recall the definitions of U_t, ξ and V in* (14), (15) *and* (16), *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 $h_{q,\boldsymbol{\eta}} : V \to \mathbb{R}$ of the form* (21). *Then for any $i \in \{1 : d\}$ we can write*

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

where each function $h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i}^{[k]}$ is of the form (21) *with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta} + \boldsymbol{e}_i$.*

Proof. For any $i \in \{1 : d\}$ and $h_{q,\boldsymbol{\eta}}(t,\boldsymbol{y}) = h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t,\boldsymbol{y})$ from (21) we have

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

Using the chain rule and substituting $D^i\xi(t,\boldsymbol{y}) = -D^i\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})/D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})$ (see (17)), and then simplifying our notation by suppressing the dependence on t and $\boldsymbol{y}$, we obtain

$$\begin{aligned}
D^iT_1(t,\boldsymbol{y}) &= \frac{-(r+q)\,(-1)^r\big[D^iD^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y}) + D^0D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})\,D^i\xi(t,\boldsymbol{y})\big]}{[D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})]^{r+q+1}}\\
&= \frac{(r+q)\,(-1)^{r+1}\,D^{\boldsymbol{e}_0+\boldsymbol{e}_i}\phi(\xi)}{[D^0\phi(\xi)]^{r+1+q}} + \frac{(r+q)\,(-1)^{r+2}\,D^{2\boldsymbol{e}_0}\phi(\xi)\,D^{\boldsymbol{e}_i}\phi(\xi)}{[D^0\phi(\xi)]^{r+2+q}},\\
D^iT_2(t,\boldsymbol{y}) &= \rho_0^{(\beta+1)}(\xi(t,\boldsymbol{y}))\,D^i\xi(t,\boldsymbol{y}) = -\frac{\rho_0^{(\beta+1)}(\xi)\,D^{\boldsymbol{e}_i}\phi(\xi)}{D^0\phi(\xi)},\\
D^iT_3(t,\boldsymbol{y}) &= \sum_{m=1}^{r}\big[D^iD^{\boldsymbol{\alpha}_m}\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y}) + D^0D^{\boldsymbol{\alpha}_m}\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})D^i\xi(t,\boldsymbol{y})\big]\prod_{\substack{\ell=1\\ \ell\neq m}}^{r} D^{\boldsymbol{\alpha}_\ell}\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})\\
&= \sum_{m=1}^{r}\left[D^{\boldsymbol{\alpha}_m+\boldsymbol{e}_i}\phi(\xi) - \frac{D^{\boldsymbol{\alpha}_m+\boldsymbol{e}_0}\phi(\xi)\,D^{\boldsymbol{e}_i}\phi(\xi)}{D^0\phi(\xi)}\right]\prod_{\substack{\ell=1\\ \ell\neq m}}^{r} D^{\boldsymbol{\alpha}_\ell}\phi(\xi).
\end{aligned}$$

Using the product rule, we arrive at

$$\begin{aligned}
D^i h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)} &= (D^iT_1)\,T_2\,T_3 + T_1\,(D^iT_2)\,T_3 + T_1\,T_2\,(D^iT_3)\\
&= \big(S_{1a} + S_{1b}\big) + S_2 + \sum_{m=1}^{r}\big(S_{3a,m} + S_{3b,m}\big),
\end{aligned}$$

where

$$S_{1a} := (r+q)\, h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,(r+1,\widetilde{\boldsymbol{\alpha}},\beta)}, \qquad \text{with} \quad \widetilde{\boldsymbol{\alpha}}_\ell := \begin{cases} \boldsymbol{\alpha}_\ell & \text{for } \ell = 1,\dots,r, \\ \boldsymbol{e}_i + \boldsymbol{e}_0 & \text{for } \ell = r+1, \end{cases}$$

$$S_{1b} := (r+q)\, h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,(r+2,\widetilde{\boldsymbol{\alpha}},\beta)}, \qquad \text{with} \quad \widetilde{\boldsymbol{\alpha}}_\ell := \begin{cases} \boldsymbol{\alpha}_\ell & \text{for } \ell = 1,\dots,r, \\ 2\boldsymbol{e}_0 & \text{for } \ell = r+1, \\ \boldsymbol{e}_i & \text{for } \ell = r+2, \end{cases}$$

$$S_{2} := h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,(r+1,\widetilde{\boldsymbol{\alpha}},\beta+1)}, \qquad \text{with} \quad \widetilde{\boldsymbol{\alpha}}_\ell := \begin{cases} \boldsymbol{\alpha}_\ell & \text{for } \ell = 1,\dots,r, \\ \boldsymbol{e}_i & \text{for } \ell = r+1, \end{cases}$$

$$S_{3a,m} := h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,(r,\widetilde{\boldsymbol{\alpha}},\beta)}, \qquad \text{with} \quad \widetilde{\boldsymbol{\alpha}}_\ell := \begin{cases} \boldsymbol{\alpha}_\ell & \text{for } \ell = 1,\dots,r,\ \ell \neq m, \\ \boldsymbol{\alpha}_m + \boldsymbol{e}_i & \text{for } \ell = m, \end{cases}$$

$$S_{3b,m} := h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,(r+1,\widetilde{\boldsymbol{\alpha}},\beta)}, \qquad \text{with} \quad \widetilde{\boldsymbol{\alpha}}_\ell := \begin{cases} \boldsymbol{\alpha}_\ell & \text{for } \ell = 1,\dots,r,\ \ell \neq m, \\ \boldsymbol{\alpha}_m + \boldsymbol{e}_0 & \text{for } \ell = m, \\ \boldsymbol{e}_i & \text{for } \ell = r+1, \end{cases}$$

Observe that all the $h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,[\cdots]}$ functions above are of the form (21) with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta}+\boldsymbol{e}_i$, and the conditions in (21) are satisfied by an inductive argument. For example, in S_{1b}, we gained two factors $D^{2\boldsymbol{e}_0}\phi(\xi)$ and $D^{\boldsymbol{e}_i}\phi(\xi)$ 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, $r+2 \le 2|\boldsymbol{\eta}|+q-1+2 = 2|\boldsymbol{\eta}+\boldsymbol{e}_i|+q-1$ as required. Moreover, with $\widetilde{\boldsymbol{\alpha}}_{r+1} = 2\boldsymbol{e}_0$ and $\widetilde{\boldsymbol{\alpha}}_{r+2} = \boldsymbol{e}_i$, we have the updated sum $\beta\boldsymbol{e}_0 + \sum_{\ell=1}^{r+2}\widetilde{\boldsymbol{\alpha}}_\ell = (r+q-1,\boldsymbol{\eta}) + 2\boldsymbol{e}_0 + \boldsymbol{e}_i = (r+2+q-1,\boldsymbol{\eta}+\boldsymbol{e}_i)$ as required. Other terms above can be justified in the same way. These $h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,[\cdots]}$ functions are all different so there is no cancellation.

Treating the multiple $(r+q)h_{q,\boldsymbol{\eta}+\boldsymbol{e}_i,[\cdots]}$ in S_{1a} as $r+q$ occurrences of the same function and doing this analogously for S_{1b}, we conclude that $D^i h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}$ can be written as a sum of $K_{q,\boldsymbol{\eta}}$ functions of the form (21) with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\eta}+\boldsymbol{e}_i$, where

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

This completes the proof. □

Lemma 10 *Let $d \ge 1$, $\boldsymbol{\nu} \in \mathbb{N}_0^d$, and $[a,b] \subset \mathbb{R}$. Suppose that ϕ and ρ_0 satisfy Assumption A1, and recall the definitions of U_t, ξ and V in* (14), (15) *and* (16)*, respectively. For any $q \in \mathbb{N}_0$ and $\boldsymbol{\eta} \le \boldsymbol{\nu}$ satisfying $|\boldsymbol{\eta}| + q \le |\boldsymbol{\nu}| + 1$, we consider functions $h_{q,\boldsymbol{\eta}} : V \to \mathbb{R}$ of the form* (21)*. Then we can write*

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

where each function $h_{q+1,\boldsymbol{\eta}}^{[k]}$ is of the form (21) *with q replaced by $q+1$.*

Proof. For $h_{q,\boldsymbol{\eta}}(t,\boldsymbol{y}) = h_{q,\boldsymbol{\eta},(r,\boldsymbol{\alpha},\beta)}(t,\boldsymbol{y})$ from (21) we have

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

Using the chain rule and substituting $\frac{\partial}{\partial t}\xi(t,\boldsymbol{y}) = 1/D^0\phi(\xi(t,\boldsymbol{y}),\boldsymbol{y})$ (see (18)), and then simplifying our notation by suppressing the dependence on t and $\boldsymbol{y}$, we obtain

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

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

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

Observe that all the $h_{q+1,\boldsymbol{\eta},[\cdots]}$ functions above are of the form (21) with q replaced by $q+1$, and the conditions in (21) are satisfied by an inductive argument. For example, in S_1, we gained a factor $D^{2\boldsymbol{e}_0}\phi(\xi)$ to join the product over ℓ, increasing the upper limit of the product from r to $r+1$, while the exponent of $D^0\phi(\xi)$ in the denominator increased from $r+q$ to $(r+1)+(q+1)$ which is further justified by the increase of q to $q+1$. Moreover, with $\widetilde{\boldsymbol{\alpha}}_{r+1} = 2\boldsymbol{e}_0$ we have the updated sum $\beta\boldsymbol{e}_0 + \sum_{\ell=1}^{r+1}\widetilde{\boldsymbol{\alpha}}_\ell = (r+q-1,\boldsymbol{\eta}) + 2\boldsymbol{e}_0 = ((r+1)+(q+1)-1,\boldsymbol{\eta})$ as required. Other terms above can be justified in the same way. These $h_{q+1,\boldsymbol{\eta},[\cdots]}$ functions are all different so there is no cancellation.

Treating the multiple $(r+q)h_{q+1,\boldsymbol{\eta},[\cdots]}$ 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 (21) with q replaced by $q+1$, where

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

This completes the proof. □