

FORMAL GROUP LAWS AND NON-UNIFORM QUASI-RANDOM SEQUENCES

Marco Pollanen
Trent University, Peterborough, ON, K9J 7B8, Canada
marcopollanen@trentu.ca

Abstract

In recent years, Quasi-Monte Carlo (QMC) integration methods have been successfully used in place of Monte-Carlo methods in many applications. However, in practice, QMC integration is often applied to integrands on unbounded domains with non-uniform probability measures, integrals for which there is little theoretical validation. We introduce group-theoretic methods to generate some non-uniform deterministic Weyl-like sequences. We also introduce a new importance sampling technique, which can be used with these group-theoretic sequences or lattice rules to create QMC integration rules with a high asymptotic order of convergence.

AMS Subject Classification: 11K60, 65C05, 65C10, 65D32

Key Words and Phrases: Quasi-Monte Carlo integration; Non-uniform sequences; Importance sampling; Low-discrepancy sequences; Weyl sequences

1 Introduction

The Monte Carlo method is widely used to perform numerical integration too complicated to solve analytically. In the unit cube $I^d = [0, 1]^d$ the Monte Carlo approximation for the Lebesgue integral of f is

$$\int_{I^d} f(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{n=1}^N f(\mathbf{x}_n) \quad (1)$$

where $\{\mathbf{x}_n\}$ is an independent identically distributed sequence of points sampled from the uniform distribution in I^d . If f is an L_2 integrand, the error is $O(1/\sqrt{N})$. This error bound is statistical, and is therefore not guaranteed and valid for only truly random sequences (which in practice are impractical, if not impossible).

By replacing the random sequence $\{\mathbf{x}_n\}$ with a well-chosen deterministic one that converges to a uniform distribution faster than a random sequence, it is possible in many circumstances to achieve faster convergence with guaranteed error bounds (as the sequence is predetermined). This is the essence of the Quasi-Monte Carlo method (see [17]), which has recently been gaining acceptance as a substitute for the Monte Carlo method in such diverse fields as statistics, physics, computer graphics, and mathematical finance [18, 21, 12, 14].

The key Quasi-Monte Carlo error estimate is the Koksma-Hlawka inequality [17] which is usually written in the form

$$\left| \int_{I^d} f(\mathbf{x}) d\mathbf{x} - \frac{1}{N} \sum_{n=1}^N f(\mathbf{x}_n) \right| \leq D_N(\{\mathbf{x}_n\})V(f) \quad (2)$$

where $D_N(\{\mathbf{x}_n\})$ is the discrepancy of the first N terms of the sequence and $V(f)$ is the variation of the function.

The discrepancy is usually taken as the *star discrepancy* with the associated variation being the variation in the sense of Hardy and Krause.

Definition 1. Let $S = \{\mathbf{x}_n\}_{n=1}^N$ be a finite sequence in $[0, 1)^d$. The star discrepancy $D_N^*(S)$ is defined by

$$D_N^*(S) = D_N^*(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sup_J \left| \frac{1}{N} \sum_{n=1}^N \chi_J(\mathbf{x}_n) - \lambda(J) \right|,$$

where the supremum is over all subintervals of $[0, 1)^d$ of the form $J = \prod_{i=1}^d [0, u_i)$. Moreover, λ denotes the k -dimensional Lebesgue measure.

The integration error (2) of a given function thus depends only on the discrepancy of the sequence. The sequences with the lowest known discrepancy have $D_N^*(S) = O\left(\frac{\log^d N}{N}\right)$ for infinite sequences and $D_N^*(S) = O\left(\frac{\log^{d-1} N}{N}\right)$ for sequences with predetermined length N . The regularity of the integrand is not reflected in these estimates. For smooth periodic functions, better asymptotic estimates are possible through a different QMC approach known as *lattice rules* [22].

Thus far, QMC theory has largely been confined to integration with respect to the uniform distribution in the unit cube. The main problem is that, except possibly for functions whose discontinuities are parallel to the coordinate axes, discontinuous functions are not of bounded variation in the sense of Hardy and Krause. Thus, characteristic set functions as simple as triangles are not of bounded variation. Special methods must be developed even for QMC integration using the uniform distribution on common domains such as spheres and tetrahedra (see [7], [24]).

In the MC case, generating non-uniform random sequences is a difficult but well studied problem. Common techniques for MC integration with respect to non-uniform distributions include (see [4]) acceptance-rejection methods, importance sampling, and inverse CDF transformations. Limited work has been done on generating non-uniform quasi-random sequences. The acceptance-rejection method in general cannot be used with QMC integration as decision-making processes can introduce characteristic functions into our integrand. However, in [24], a smoothed QMC acceptance-rejection method was introduced using importance sampling for bounded domains.

In many applications, we need to integrate unbounded functions over a tailed probability distribution. For instance, this occurs in statistics, when finding moments of tailed distributions, or in finance, when pricing options. For example, when pricing a European option, we must find the discounted expectation of a payoff function, which has unbounded linear growth, with respect to the risk-neutral transition probability distribution, which has log-normal tails for geometric Brownian motion. For typical “out of the money” options encountered, the “tail performance” is important, as the value of the option is related to the upside (i.e., making extreme events in the tails significant).

For QMC, unbounded functions with respect to tailed distributions are problematic. For instance, computing the mean of any tailed probability distribution P is equivalent (after transformation) to computing the integral in the unit cube of an improper integral (with singularities at 0 and/or 1), as $\int_{-\infty}^{\infty} x dP(x) = \int_0^1 P^{-1}(u) du$. The integrand is unbounded and thus not of bounded variation. This is an example of how the inverse CDF method can fail with QMC methods. The MC method does not have the same shortcoming due to the law of large numbers.

QMC integration rules have been studied only in the case of bounded domains. However, recently there has been interest [5, 11] in the related problem of QMC methods for functions which are unbounded on the boundary of the unit cube. For unbounded domains, the following definition of a non-uniform deterministic sequence is a natural extension, and is similar to what we would expect for random sequences:

Definition 2. The sequence $\{\mathbf{x}_n\}_{n=1}^{\infty}$ of points in \mathbb{R}^d is *P-distributed* if and only if P is a cumulative distribution function satisfying

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \chi_{(-\infty, \mathbf{x}]}(\mathbf{x}_n) = P(\mathbf{x}) = \int_{\mathbb{R}^d} \chi_{(-\infty, \mathbf{x}]}(\mathbf{y}) dP(\mathbf{y})$$

for every $\mathbf{x} \in \mathbb{R}^d$. Furthermore, we call such a P the *distribution function* of the sequence $\{\mathbf{x}_n\}$.

The following definition, introduced in [7], is a natural extension of the concept of star-discrepancy to non-uniform distributions.

Definition 3. If P is a cumulative distribution, the sequence $\{\mathbf{x}_n\}_{n=1}^{\infty}$ of points in \mathbb{R}^d has *P-discrepancy*

$$D_P(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{1}{N} \sum_{n=1}^N \chi_{(-\infty, \mathbf{x}]}(\mathbf{x}_n) - P(\mathbf{x}) \right|.$$

Like star-discrepancy, this notion of discrepancy measures the extreme difference between the empirical distribution function of the sample and the actual CDF. In fact, $D_P(\mathbf{x}_1, \dots, \mathbf{x}_N)$ is the Kolmogorov-Smirnov statistic for goodness of fit.

2 Quasi-Random Sequences from Rational Group Laws

A well-known theorem [6] of Weyl gives a method by which the additive group on the torus can be used to generate infinite uniform sequences:

Theorem 4 (Weyl). *The sequence $\{n\alpha\} \bmod 1$ is uniformly distributed in $[0, 1)$ iff α is irrational.*

In this section we use a group-theoretic Weyl-like method to directly generate a sequence that converges to the Cauchy distribution.

Given $G = \mathbb{R} \cup \{\infty\}$, G can be made into a group under the operation

$$x \oplus y = \frac{x + y}{1 - xy}, \text{ for all } x, y \in G.$$

This is clear, as the tangent function satisfies the identity

$$\tan(x + y) = \frac{\tan(x) + \tan(y)}{1 - \tan(x)\tan(y)}.$$

Let us define the sequence

$$x_n = \frac{x_{n-1} + x_0}{1 - x_{n-1}x_0}.$$

This is equivalent to $x_n = \tan(n\alpha)$, where $\alpha = \arctan x_0$, and thus, provided α/π is irrational, by Weyl's theorem the sequence converges to the density

$$p(x) = \frac{1}{\pi} (\tan^{-1}(x))' = \frac{1}{\pi} \frac{1}{1 + x^2},$$

which is the density for the Cauchy distribution.

The irrationality of α/π follows from Corollary 3.12 of [15], in which it is proved that given a rational r , the only rational values of $\tan 2\pi r$ are $0, \pm 1$. Thus $\frac{\arctan(x)}{\pi}$ cannot be rational for rational $x \neq 0, \pm 1$. However, we can make a stronger claim:

Theorem 5. *If x is rational and $x \neq 0, \pm 1$, then $\frac{\arctan x}{\pi}$ is transcendental.*

To see this, we use the identity

$$\log(c + id) = \log \sqrt{c^2 + d^2} + i \arctan\left(\frac{c}{d}\right)$$

and rewrite it as

$$\frac{\arctan(\frac{c}{d})}{\pi} = \frac{\log(c/\sqrt{c^2 + d^2} + di/\sqrt{c^2 + d^2})}{\log(-1)}.$$

The right-hand side is a ratio of logarithms of algebraic numbers, and so we can apply the Gelfond-Schneider theorem [15]:

Theorem 6 (Gelfond-Schneider). *If α and γ are non-zero algebraic numbers, and if $\alpha \neq 1$, then $(\log \gamma)/(\log \alpha)$ is either rational or transcendental.*

We see that if $\frac{c}{d}$ is rational and not equal to $-1, 0$, or 1 , then as $\frac{\arctan(\frac{c}{d})}{\pi}$ is not rational, it must, in fact, be transcendental.

Thus, we have the following proposition:

Proposition 7. *Given a rational $x_0 \neq 0, \pm 1$, the recursion $x_{n+1} = \frac{x_n + x_0}{1 - x_n x_0}$ defines a sequence which converges to the standard Cauchy distribution.*

Although sequences generated in this manner are uniform, the quality of the sequence depends on the Diophantine properties of the underlying irrational.

We will examine the Diophantine properties of the multivariate case where we have a Cartesian product of Cauchy distributions. Consider the linear form of logarithms

$$\Lambda = \beta_0 \log \alpha_0 + \beta_1 \log \alpha_1 + \cdots + \beta_n \log \alpha_n$$

and let us assume that β_j 's are integers and also that

$$\alpha_0 = -1, \alpha_j = c_j/\sqrt{c_j^2 + d_j^2} + d_j i/\sqrt{c_j^2 + d_j^2}$$

where c_j and d_j are integers. Then, according to Baker, [1], if $\Lambda \neq 0$,

$$|\Lambda| > (\max_{j \geq 0} (4, |\beta_j|) \log A)^{-K \log A},$$

where $K > 0$ and A are independent of the β_j 's.

It follows that if $\Lambda \neq 0$, then

$$\left| \beta_0 + \sum_{j=1}^n \beta_j \frac{\arctan \frac{c_j}{d_j}}{\pi} \right| > \frac{K'}{(\max_{j \geq 0} (4, |\beta_j|))^k},$$

for some K' and k independent of the β_j 's.

Hence, there exists constants k' and K'' such that

$$\left| \beta_0 + \sum_{j=1}^n \beta_j \frac{\arctan \frac{c_j}{d_j}}{\pi} \right| > \frac{K''}{\left(\prod_{j=0}^n \max(1, |\beta_j|) \right)^{k'}}.$$

And so, finally we can conclude that if

$$\left\{ 1, \frac{\arctan \frac{c_1}{d_1}}{\pi}, \dots, \frac{\arctan \frac{c_n}{d_n}}{\pi} \right\}$$

is an independent set over the rationals, then there exist constants $\sigma, C(\sigma)$, such that

$$\min_{m \in \mathbb{Z}} \left| m - \sum_{i=1}^n \beta_i \frac{\arctan \frac{c_i}{d_i}}{\pi} \right| > \frac{C}{\left(\prod_{i=1}^n \max(1, |\beta_i|) \right)^\sigma}.$$

This follows as the absolute value of the summand on the left-hand side is bounded by $0.5n \max_{j \geq 1} |\beta_j|$. If we take η to be the minimum σ that satisfies the above inequality, then

$$\left[\frac{\arctan \frac{a_1}{b_1}}{\pi}, \dots, \frac{\arctan \frac{a_n}{b_n}}{\pi} \right]$$

is by definition (see [16]) a type- η vectors of irrationals. So we have shown:

Theorem 8. *If a_i and b_i are non-zero integers such that*

$$1, \frac{\arctan \frac{a_1}{b_1}}{\pi}, \dots, \frac{\arctan \frac{a_n}{b_n}}{\pi}$$

are independent over the rationals, then the vector

$$\left[\frac{\arctan \frac{a_1}{b_1}}{\pi}, \dots, \frac{\arctan \frac{a_n}{b_n}}{\pi} \right]$$

is of finite type.

It would be of interest to find all group laws $x \oplus y = R(x, y)$ defined by a rational function R . Unfortunately, by a theorem in [3], the rational group laws over \mathbb{Q} are of the form:

$$R(x, y) = \frac{x + y + cxy}{1 - dxy}, \quad c, d, \in \mathbb{Q}.$$

As will be shown in the next section, this converges to the density

$$\frac{K}{1 + cx + dx^2}, \text{ provided } d > c^2/4, \text{ where } K \text{ is a normalizing constant.}$$

As there are no other interesting groups defined by rational group laws, a natural extension is to look at formal group laws, where we replace $R(x, y)$ with a formal power series $F(x, y)$.

3 Formal Groups and Weyl-Sequences

Definition 9 (see [10]). A formal group law over a ring R is a power series $F(x, y) \in R[[x, y]]$ satisfying

- i. $F(x, 0) = F(0, x) = x$
- ii. $F(y, x) = F(x, y)$
- iii. $F(x, F(y, z)) = F(F(x, y), z)$

From this definition, the following lemma holds:

Lemma 10 (see [10]). *Let $F(x,y)$ be a formal group law over a ring R . Then there exists a power series $l(x) = -x + bx^2 + \dots$ with coefficients in R such that $F(x, l(x)) = 0$.*

From now on, we assume that R is a real number field. Now, suppose F is continuous. If $u < v$, then by property (i), $F(u, 0) < F(v, 0)$. If there exists s such that $F(u, s) \geq F(v, s)$, then by continuity there exists t such that $F(u, t) = F(v, t)$. It follows from properties (i) and (iii), and the above lemma that

$$u = F(F(u, t), l(t)) = F(F(v, t), l(t)) = v$$

Hence, F is monotone in its first variable, and can be similarly shown to be monotone in its second variable. It follows that the sequence $x_n = F(x_{n-1}, x_0)$, $n > 1$ must be either increasing or decreasing and so will not converge to a density.

Thus, instead of assuming that F is continuous, we will need to allow F to have infinite discontinuities. However, we will assume near $y = 0$ that F is monotone, i.e. that $\frac{\partial F}{\partial y}(x, 0) > 0$.

Theorem 11. *Given a formal group law $F(x, y)$ with on $\mathbb{R} \cup \{-\infty\}$ such that $F(x, y)$ is analytic (where is is finite) and that*

$$\frac{\partial F}{\partial y}(x, 0) \text{ is positive and analytic with } \omega = \int_{-\infty}^{\infty} \left(\frac{\partial F}{\partial y}(x, 0) \right)^{-1} dx < \infty$$

it follows that the sequence $x_n = F(x_{n-1}, x_0)$, $n > 1$ converges to a distribution with density

$$\left(\omega \frac{\partial F}{\partial y}(x, 0) \right)^{-1} \text{ iff } x_0 \text{ is a point of infinite order in the group.}$$

Proof. Using formal calculations, we will show the existence of a group isomorphism with the additive group.

Suppose $F(x, y) = \sum c_{ij} x^i y^j$, then, by property (i), $F(y, 0) = F(0, y) = y$, and so $\frac{\partial F}{\partial y}(0, 0) = 1$ and thus $\frac{\partial F}{\partial y}(x, 0)$ is invertible in $R[[x]]$.

Hence, we may define a bijection $\phi : \mathbb{R} \cup \{-\infty\} \rightarrow [0, 1)$ by

$$\phi(x) = \int_{-\infty}^x \left(\omega \frac{\partial F}{\partial y}(t, 0) \right)^{-1} dt.$$

Letting $g(x, y) = \phi(F(x, y)) - \phi(x) - \phi(y)$, we will show that $g(x, y) \equiv 0$.

$$\begin{aligned} \frac{\partial g}{\partial y}(x, y) &= \phi'(F(x, y)) \frac{\partial F}{\partial y}(x, y) - \phi'(y) \\ &= \left(\omega \frac{\partial F}{\partial y}(F(x, y), 0) \right)^{-1} \frac{\partial F}{\partial y}(x, y) - \left(\omega \frac{\partial F}{\partial y}(y, 0) \right)^{-1} \end{aligned}$$

But by property (iii), $F(F(x, y), z) = F(x, F(y, z))$, we have, differentiating with respect to z and evaluating at $z = 0$:

$$\frac{\partial F}{\partial y}(F(x, y), 0) = \frac{\partial F}{\partial y}(x, y) \frac{\partial F}{\partial y}(y, 0).$$

Thus $\frac{\partial g}{\partial y} \equiv 0$, and so $\phi(F(x, y)) = \phi(x) + \phi(y)$, and hence ϕ is in fact a group isomorphism from the formal group to the additive group.

Now, using the convergence of $\phi(x_0)$ and the fact ω is finite, we can establish that there is in fact an isomorphism with the torus $[0, 1)$. Since the Weyl-sequence $\{n\phi(x_0)\}$, $n > 1$ is uniformly distributed in $[0, 1)$ iff $\phi(x_0)$ is irrational, it follows that the sequence $x_n = F(x_{n-1}, x_0)$, $n > 1$ converges to $\phi'(x)$ iff x_0 is a point of infinite order. \square

The above theorem provides an explicit “logarithm” from a 1-dimensional formal group to a torus. In general, there is a logarithm that provides an isomorphism from a d -dimensional formal group to an additive group [8]. The quality of the sequences constructed in this fashion is dependent on the Diophantine properties in the torus of the logarithm of the initial seed value, with “low-type” initial seed vectors being highly desirable.

Several examples of type-1 vectors are known. For example [16], if $1, \alpha_1, \dots, \alpha_d$ are algebraic numbers independent over the rationals, then $(\alpha_1, \dots, \alpha_d)$ is a vector of type-1. Also, if r_1, \dots, r_d are distinct rationals, then $(e^{r_1}, \dots, e^{r_d})$ is a type-1 vector.

Jacobian groups of algebraic curves have formal group representations [8]. In particular, Jacobians of hyperelliptic curves are isomorphic to additive groups on tori, and have effective algorithms for computation [2]. We will present a sequence derived from elliptic curves in the next section.

4 Non-Uniform Quasi-Random Sequences and Elliptic Curves

In this section, we will provide examples of the generation of non-uniform deterministic sequences by using formal groups from elliptic curves.

Definition 12 (see [13]). The elliptic curve over a field K defined by

$$y^2 = x^3 + ax + b \tag{3}$$

where $a, b \in K$ with $4a^3 + 27b^2 \neq 0$, is the set of points $(x, y) \in K^2$ that satisfy equation (3) in addition to a “formal point at infinity” denoted \mathcal{O} .

Given two points $P = [x_1, y_1]$ and $Q = [x_2, y_2]$ on an elliptic curve $y^2 = x^3 + ax + b$ we may define addition by

$$P \oplus P = \left[\left(\frac{3x_1^2 + a}{2y_1} \right)^2 - 2x_1, \frac{(3x_1^2 + a)(x_1 - x_3)}{2y_1} - y_1 \right],$$

where x_3 denotes the x -coordinate of $P \oplus P$, and for $P \neq Q$,

$$P \oplus Q = \left[\left(\frac{y_2 - y_1}{x_2 - x_1} \right)^2 - x_1 - x_2, \frac{(y_2 - y_1)(x_1 - x_3)}{x_2 - x_1} - y_1 \right],$$

where x_3 again denotes the x -coordinate of $P \oplus Q$. Also, the point \mathcal{O} will be taken to be the additive identity.

Under this definition the elliptic curve becomes an additive group.

Definition 13. A lattice L is an additive subgroup of \mathbb{C} which is generated by two elements $\omega_1, \omega_2 \in \mathbb{C}$ that are linearly independent over \mathbb{R} .

Definition 14. The Weierstrass \wp -function relative to the lattice L is the function $\wp_L : \mathbb{C} \rightarrow \mathbb{C}$ given by:

$$\wp(z) = \frac{1}{z^2} + \sum_{\omega \in L \setminus \{0\}} \left(\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right).$$

Note that, although \wp depends on L , it is customary to omit it from the notation.

The map

$$z \rightarrow P = (1, \wp(z), \wp'(z))$$

into the projective plane induces an isomorphism between the elliptic curve $y^2 = 4x^3 - g_2x - g_3$ over the field \mathbb{C} , denoted by $E(\mathbb{C})$ and \mathbb{C}/L

$$\mathbb{C}/L \rightarrow E(\mathbb{C}) \subset \mathbb{P}^2(\mathbb{C})$$

where $\mathbb{P}^2(\mathbb{C})$ denotes the projective plane over \mathbb{C} .

Here the modular invariants $g_2(L)$ and $g_3(L)$ can be calculated by

$$g_2(L) = 60 \sum_{\omega \in L \setminus \{0\}} \frac{1}{\omega^4}, \quad g_3(L) = 140 \sum_{\omega \in L \setminus \{0\}} \frac{1}{\omega^6}.$$

Conversely, given any elliptic curve $y^2 = 4x^3 + ax + b$, there exists a lattice whose modular invariants satisfy $g_2 = -a$ and $g_3 = -b$.

The inverse of the above map is provided by the elliptic logarithm of the point $P \in E(\mathbb{C})$, which can be defined by the following elliptic integral

$$\text{ELog}(P) = \int_0^P \frac{dz}{\sqrt{z^3 + az + b}} \pmod{L}.$$

Let us work over the field \mathbb{R} and take $\omega = \omega_1$ to be real and ω_2 purely imaginary.

Any cubic equation has either one or three real roots. If $x^3 + ax + b = 0$ has one real root γ then we may write

$$\text{ELog}(x(P)) = \int_0^{x(P)} \frac{dx}{\sqrt{x^3 + ax + b}} \pmod{\omega}.$$

However, if $x^3 + ax + b = 0$ has three roots (say $\gamma_1 < \gamma_2 < \gamma_3$), then $E(\mathbb{R})$ has two components

$$\begin{aligned} E_0(\mathbb{R}) &= \{P \in E(\mathbb{R}) \mid x(P) > \gamma_3\} \\ E_C(\mathbb{R}) &= \{P \in E(\mathbb{R}) \mid \gamma_1 < x(P) < \gamma_2\}. \end{aligned}$$

Thus, we write

$$\text{ELog}(x(P)) = \begin{cases} \int_{\gamma_1}^{x(P)} \frac{dx}{\sqrt{x^3 + ax + b}}, & \text{if } x(P) \in E_C(\mathbb{R}) \\ \frac{\omega}{2} + \int_{\gamma_3}^{x(P)} \frac{dx}{\sqrt{x^3 + ax + b}}, & \text{if } x(P) \in E_0(\mathbb{R}) \end{cases}.$$

Now the map

$$x(P) \rightarrow \frac{1}{\omega} \text{ELog}(x(P))$$

induces an isomorphism between $E(\mathbb{R})$ and $[0, 1)$.

To use the above properties to generate a non-uniform sequence, let $\wp(z)$ denote the Weierstrass \wp -function relative to the lattice L and (3), and P be a point of infinite order on (3).

Now, using addition on the elliptic curve, we define a sequence of points by using the x -coordinates of nP , i.e., $x_n = (nP)_x$. As P is a point of infinite order, the sequence $u_n = \text{ELog}(nP) = n\text{ELog}(P) \pmod{\omega}$ defines a uniform sequence in $[0, \omega)$.

The Weierstrass \wp -function is the inverse of the elliptic logarithm, and thus $x_n = (nP)_x$ must converge to the density function (for those values of x in the domain):

$$p(x) = \frac{1}{\omega} (\wp^{-1}(x))' = \frac{1}{\omega} \frac{1}{\sqrt{x^3 + ax + b}}.$$

We can summarize this in the following proposition:

Proposition 15. *Given an Elliptic Curve $y^2 = x^3 + ax + b$ over \mathbb{Q} , and a point of infinite order P on the curve, then $x_n = (nP)_x$ defines a sequence with distribution proportional to $\frac{1}{\sqrt{x^3 + ax + b}}$.*

If the initial seed is a rational point of infinite order, then the sequence is a rational sequence whose Diophantine properties follow from the Baker-Feldman theorem [13]. In fact, if $1, \alpha_1, \dots, \alpha_d$ are independent over the rationals, where each α_i is the elliptic logarithm with respect to some rational, then using the Baker-Feldman theorem and a similar calculation to that in section 2, we see that $(\alpha_1, \dots, \alpha_d)$ is of finite type.

5 Integration with Respect to Smooth Distributions

Consider the problem of integrating a function f with respect to a distribution P . Often P is difficult to generate directly by transformation, and it is likely that we do not have a group law to generate it indirectly. In this case, we can use importance sampling, in which we try to find a distribution $G(\mathbf{x})$ that is similar to $P(\mathbf{x})$ by using the fact that

$$\int_{\mathbb{R}^d} f(\mathbf{x}) dP(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x}) p(\mathbf{x}) \frac{g(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x} = \int_{\mathbb{R}^d} f(\mathbf{x}) \frac{p(\mathbf{x})}{g(\mathbf{x})} dG(\mathbf{x}),$$

where $g(\mathbf{x})$ and $f(\mathbf{x})$ are the respective densities of the distributions.

Thus, if we can generate the distribution $G(\mathbf{x})$, we can perform the integration. The problem with importance sampling is that if $p(\mathbf{x}) \not\approx g(\mathbf{x})$, then, in general, the constant in the order of convergence becomes quite large (variance in the Monte Carlo case). This technique is used quite often with Monte Carlo method. However, if you do not know if $p(\mathbf{x}) \approx g(\mathbf{x})$, it is in general better not to use importance sampling [20]. As the Monte Carlo error is proportional to the standard deviation $\sigma(f)$, this is an issue for QMC methods as well. This follows from the Koksma-Hlawka inequality and the fact that discrepancy is bounded by 1:

$$\begin{aligned} \sigma(f) &\leq \sup(f) - \inf(f) \leq \left| \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} - \sup(f) \right| + \left| \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} - \inf(f) \right| \\ &\leq 1 \cdot V(f) + 1 \cdot V(f) \leq 2V(f). \end{aligned}$$

We will now show how importance sampling can be used with Weyl-like sequences to create QMC rules with high orders of convergence.

Letting

$$h(\mathbf{u}) = \frac{f(G^{-1}(\mathbf{u}))p(G^{-1}(\mathbf{u}))}{g(G^{-1}(\mathbf{u}))},$$

we may write

$$\int_{\mathbb{R}^d} f(\mathbf{x}) dP(\mathbf{x}) = \int_{[0,1]^d} \frac{f(G^{-1}(\mathbf{u}))p(G^{-1}(\mathbf{u}))}{g(G^{-1}(\mathbf{u}))} d\mathbf{u} = \int_{[0,1]^d} h(\mathbf{u}) d\mathbf{u}.$$

If g is thick-tailed enough in comparison with p , then $(f \cdot p/g)(\mathbf{x})$ will approach zero for large $|\mathbf{x}|$, and so h will be zero on the boundary of the unit cube. In fact, when g is sufficiently thick-tailed and f, g, p are sufficiently differentiable, h can be extended into a periodic function with high-order derivatives.

For clarity, let us consider the situation where as the importance sampling distribution we use a product of Cauchy distributions, i.e.,

$$g(\mathbf{x}) = \frac{1}{\pi^d} \prod_{i=1}^d \frac{1}{1+x_i^2}.$$

The inverse cumulative distribution function is given by:

$$G^{-1}(\mathbf{u}) = (\tan \pi(u_1 - 1/2), \tan \pi(u_2 - 1/2), \dots, \tan \pi(u_d - 1/2)).$$

Definition 16. We will say that f has smooth tails of order k if $|x_i^k \frac{\partial^j f}{\partial x_i^j}(\mathbf{x})| \rightarrow 0$ as $x_i \rightarrow \pm\infty$, and

$$\int_{\mathbb{R}^d} x_i^k f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

exists for $i = 1, 2, \dots, d$ and $j = 1, 2, \dots, k$.

This definition aims to avoid any pathological distributions whose tails approach zero in measure but not point-wise. This condition is reasonable, as by integration by parts we should expect:

$$\left| \int_{-\infty}^{\infty} f(\mathbf{x}) dx_i \right| = \left| \int_{-\infty}^{\infty} x_i \frac{\partial f}{\partial x_i}(\mathbf{x}) dx_i \right| = \dots = \left| \int_{-\infty}^{\infty} x_i^k \frac{\partial^k f}{\partial x_i^k}(\mathbf{x}) dx_i \right|.$$

We will formalize this idea with the following theorem:

Theorem 17. *Let the product $z(\mathbf{x}) = (f \cdot p)(\mathbf{x})$ be a k -times differentiable integrand on \mathbb{R}^d with smooth tails of order k . Then, using a product Cauchy distribution $g(\mathbf{x})$ as an importance sampling distribution, the integral of $z(\mathbf{x})$ is equivalent to an integral of a $(k-2)$ -times differentiable integrand in the unit cube such that all partial derivatives of order $k-2$ or less vanish on the boundary of the cube.*

Proof. The differentiability is clear everywhere except possibly on the boundary of the unit cube. However, since the transformed integrand is zero on this boundary, the only partial derivatives we need to check are those perpendicular to the coordinate axes.

Thus, as the case $u_i \rightarrow 1^-$ will hold analogously, so letting $y_i = \tan(\pi(u_i - 1/2))$, all we need to show is that

$$\lim_{u_i \rightarrow 0^+} \frac{\partial^j}{\partial u_i^j} \frac{f(y_1, y_2, \dots, y_d) p(y_1, y_2, \dots, y_d)}{g(y_1, y_2, \dots, y_d)} = 0 \quad (4)$$

for $j = 1, 2, \dots, k$ and $i = 1, 2, \dots, d$.

Accordingly we need to compute the limit as $u_i \rightarrow 0^+$ for $j = 1, 2, \dots, k$ of

$$\frac{\partial^j}{\partial u_i^j} z(y_1, y_2, \dots, \tan \pi(u_i - 1/2), \dots, y_d) (\sec \pi(u_i - 1/2))^2.$$

Upon taking a number of derivatives, the differentiant can be written in the form

$$\sum_{r=1}^q c_r \sec^{l_r} \pi(u_i - 1/2) \sin^{n_r} \pi(u_i - 1/2) \frac{\partial^{m_r}}{\partial x_i^{m_r}} z(y_1, \dots, \dots, y_d)$$

for some constants c_r .

Now performing one additional differentiation the above summand becomes

$$\begin{aligned} & \pi c_r \left(\sec^{l_r+2} \pi(u_i - 1/2) \sin^{n_r} \pi(u_i - 1/2) \frac{\partial^{m_r+1}}{\partial u_i^{m_r+1}} z(y_1, \dots, y_d) \right. \\ & + l_r \sec^{l_r+1} \pi(u_i - 1/2) \sin^{n_r+1} \pi(u_i - 1/2) \frac{\partial^{m_r}}{\partial u_i^{m_r}} z(y_1, \dots, y_d) \\ & \left. + n_r \sec^{l_r-1} \pi(u_i - 1/2) \sin^{n_r-1} \pi(u_i - 1/2) \frac{\partial^{m_r}}{\partial u_i^{m_r}} z(y_1, \dots, y_d) \right). \end{aligned} \quad (5)$$

The effect is that in the second term the order of the secant factor increases by one and in the third term it decreases by one.

If z has smooth tails of order at least m , then

$$\begin{aligned} & \lim_{u_i \rightarrow 0^+} \sec^m \pi(u_i - 1/2) \cdot \frac{\partial^m}{\partial u_i^m} z(y_1, \dots, \tan \pi(u_i - 1/2), \dots, y_d) \\ & = \lim_{u_i \rightarrow 0^+} \tan^m \pi(u_i - 1/2) \cdot \frac{\partial^m}{\partial u_i^m} z(y_1, \dots, \tan \pi(u_i - 1/2), \dots, y_d) \\ & = \lim_{x_i \rightarrow -\infty} x_i^m \frac{\partial^m}{\partial x_i^m} z(x_1, \dots, x_i, \dots, x_d) = 0. \end{aligned} \quad (6)$$

So in essence, although the secant factor increases by two in the first term of equation (5), it can effectively be thought of as increasing by at most one, since the other factor can be grouped with the increased derivative. Thus, if we start with a secant factor of order 2 (as in equation (4)), we see that the transformed integrand can be extended into a periodic $(k-2)$ -times differentiable function. \square

Increasing the regularity of the integrand does not have an effect on Koksma-Hlawka error bounds. However, for smooth periodic integrands, increasing the smoothness provides greatly improved asymptotic bounds when using Weyl-like sequences and Fourier-based estimates for a special class of smooth integrands which we will now define.

Definition 18 (See [17]). Let $\alpha > 1$ and $C > 0$ be real numbers. Then $E_\alpha^d(C)$ is defined to be the class of all continuous periodic functions f on \mathbb{R}^d with period interval $[0, 1]^d$ such that for all non-zero $\mathbf{h} = (h_1, \dots, h_d) \in \mathbb{Z}^d$

$$|\hat{f}(\mathbf{h})| \leq \frac{C}{(\bar{h}_1 \bar{h}_2 \cdots \bar{h}_d)^\alpha}$$

where $\bar{h}_i = \max(1, |h_i|)$ and $\hat{f}(\mathbf{h})$ are the Fourier coefficients of f .

A sufficient condition that $f \in E_\alpha^d(C)$ for an explicit value of C [25] is that $\alpha > 1$ is an integer and all partial derivatives

$$\frac{\partial^{m_1+\dots+m_d} f}{\partial x_1^{m_1} \cdots \partial x_d^{m_d}} \text{ with } 0 \leq m_i \leq \alpha \text{ for } 1 \leq i \leq d$$

exist and are continuous on \mathbb{R}^d .

From the definition of $E_\alpha^d(C)$, the following theorem, which is an easy extension of one found in [23], follows.

Theorem 19. Let $w^{(k)}(x) = \frac{(2k+1)!}{k!k!} x^k (1-x)^k$, where k is a positive integer. If $f \in E_{\eta k + \lambda}^d(C)$, $\lambda > 0$ and $\{\mathbf{x}_j\} = j(\beta_1, \dots, \beta_d) \bmod 1$ is a Weyl-sequence where β_i are

type- η irrationals such that $1, \beta_1, \dots, \beta_d$ are linearly independent over the rationals, then,

$$\left| \int_{I^d} f(\mathbf{x}) d\mathbf{x} - \frac{1}{N} \sum_{j=0}^{N-1} w^{(k)} \left(\frac{j}{N} \right) f(\mathbf{x}_j) \right| = O(N^{-k}).$$

Proof. Following the proof in [23] we have that the integration error is

$$\begin{aligned} & \left| \int_{I^d} f(\mathbf{x}) d\mathbf{x} - \frac{1}{N} \sum_{j=0}^{N-1} w^{(k)} \left(\frac{j}{N} \right) f(\mathbf{x}_j) \right| \\ & \leq \frac{2(2k+1)!}{N^k (2\pi)^k k!} \left(|\hat{f}(\mathbf{0})| \zeta(k) + (1 + \zeta(k)) \sum_{\mathbf{h} \neq \mathbf{0}} \frac{|\hat{f}(\mathbf{h})|}{(\min_{m \in \mathbb{Z}} |m - \sum_{i=1}^n h_i \beta_i|)^k} \right). \end{aligned}$$

Using the fact that $|\hat{f}(\mathbf{h})| \leq C \prod_i (\max(1, |h_i|))^{-k\eta - \lambda}$ for some constant C we have that the error is bounded by

$$\begin{aligned} & \frac{1}{N^k} \left(C_1 + C_2 \sum_{\mathbf{h} \neq \mathbf{0}} \frac{\prod_i (\max(1, |h_i|))^{-k\eta - \lambda}}{(\min_{m \in \mathbb{Z}} |m - \sum_{i=1}^n h_i \beta_i|)^k} \right) \\ & \leq \frac{1}{N^k} \left(C_1 + C_3 \sum_{\mathbf{h} \neq \mathbf{0}} \frac{\prod_i (\max(1, |h_i|))^{-\eta - \lambda/k}}{(\min_{m \in \mathbb{Z}} |m - \sum_{i=1}^n h_i \beta_i|)^k} \right) \end{aligned}$$

by equation (2), where C_1, C_2 and C_3 are constants. Sums of the form in the last expression were shown to be convergent in the proof of theorem 8.1 in [16]. Thus, the desired result follows. \square

Incorporating the above sufficient conditions on smoothness we have:

Theorem 20. Let $w^{(k)}(x) = \frac{(2k+1)!}{k!k!} x^k (1-x)^k$, where k is a positive integer. Suppose $(f \cdot p)(\mathbf{x})$ is such that all partial derivatives

$$\frac{\partial^{m_1 + \dots + m_d} (f \cdot p)}{\partial x_1^{m_1} \dots \partial x_d^{m_d}} \text{ with } 0 \leq m_i \leq \eta k + 1 \text{ for } 1 \leq i \leq d$$

exist and are continuous on \mathbb{R}^d , and that $(f \cdot p)(\mathbf{x})$ has smooth tails of order $\eta k + 3$. If $\{\mathbf{x}_j\}$ is distributed as a product of Cauchy distributions generated by a Weyl-sequence of type- η irrationals β_i such that $1, \beta_1, \dots, \beta_d$ are linearly independent over the rationals, then,

$$\left| \int_{\mathbb{R}^d} f(\mathbf{x}) dP(\mathbf{x}) - \frac{1}{N} \sum_{j=0}^{N-1} w^{(k)} \left(\frac{j}{N} \right) \pi^d (1 + \mathbf{x}_j^2) (f \cdot p)(\mathbf{x}_j) \right| = O(N^{-k}).$$

6 Empirical Results

In the first figure we plot the star-discrepancy between the Halton sequence [17], the group-theoretic sequence associated with the elliptic curve $y^2 = x^3 + 8$ and initial point $(1, 3)$, which has been transformed to the standard uniform distribution, as well as the transformed group-theoretic Cauchy sequence generated with initial starting value $1/2$. The discrepancies of all three sequences are very close, with the elliptic curve sequence being consistently the best.

Comparing the P -discrepancy, for instance, of the elliptic curve sequence and transformed Halton sequence would have the same result, as by its definition P -discrepancy is invariant under transformation of the sequences.

It is interesting to note that for both group-theoretic methods presented, if the initial starting values are rational, then so are the entire sequences. However, if a large number of terms are needed, it will ultimately be necessary to use floating point arithmetic. Both the Cauchy sequence and elliptic curve sequences can be generated with inverse CDF methods. However, while the Cauchy distribution can be generated easily from a standard uniform sequence u_i with the transformation $x_i = \tan(\pi(u_i - 1/2))$, the elliptic curve sequences will require relatively more demanding computations of Weierstrass functions.

To compare the inverse method to the group-theoretic method, a version of the Cauchy sequence generator was implemented in Java on a 3.4GHz Pentium IV computer running Linux using v1.5 of Sun Microsystems' Java environment. Using double precision arithmetic to generate one term of the Cauchy sequence using the group law took on average 19ns. This is considerably faster than using the inverse CDF transformation of a Weyl sequence, which took on average 254ns per term of the sequence. For comparison purposes, the Java internal random number generator took on average 137ns to generate a random number on $[0, 1]$.

Although the group-law algorithm for the Cauchy distribution involves floating point division, which is not usually numerically stable, it appears that the Diophantine properties of the transformed sequence (see equation (2)) force the sequence away from 0 and 1 (points of instability). This is illustrated in the second figure, where the absolute error of the group law with initial value $1/2$ is compared to the error using a transformed Weyl sequence. Using double precision arithmetic, after 100 million terms of the sequence the error for the group-law method starting with initial value $1/2$ is 4×10^{-13} , while transforming the corresponding Weyl sequence, i.e. $x_n = n(\arctan(1/2)/\pi + 1/2) \bmod 1$, yields a considerably larger error of 1×10^{-8} . For comparison purposes, the precision of double arithmetic in the Java implementation is 1×10^{-16} .

The third figure summarizes the results of calculating the moment $E[x_1x_2x_3]$ of a mixture of two trivariate (in variables x_1, x_2, x_3) standard Gaussian distributions with means $(0, 0, 0)$ and $(1, 1, 1)$. The Monte Carlo method, the QMC-method with the Halton sequence (with the standard prime bases), and rank-1 Korobov-type lattice rules with optimal coefficients [22] were used to evaluate the integrand without importance sampling. In this case, the Gaussians are generated by transformation and the mixture is obtained by adding a fourth variable and characteristic function to select the Gaussian. It is important to note that, for the QMC-method and the lattice rules method, this is equivalent to an inverted problem of integrating a function in the unit cube that is unbounded on the boundary, as discussed in section 1, and so has not been theoretically validated. However, by using a thick-tailed importance sampling distribution, the inverted problem becomes smooth and zero on the boundary of the cube. The above results were compared to Cauchy importance sampling using a transformed Halton sequence and two Fourier-based methods, rank-1 lattice rules and a product of group-theoretic sequences using theorem 20, with initial point $(1/3, 1/5, 1/7)$ and weight function $w^{(4)}(x)$, to take advantage of the smoothness. In this case the inverted problem is infinitely differentiable. The fact that the Fourier methods converge quickly, while importance sampling with the Halton sequence does not, demonstrates the effect of the smoothing.

The final figure summarizes the results of integrating the function $(x_1x_2 - 1/3)(x_3x_4 - 1/2)(x_5x_6 - 1)(x_7x_8 - 2)(x_9x_{10} - 3)$ with respect to the t -distribution given by the density

$$\frac{33}{200\pi^3(1 + (x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2)/20)^{13}}.$$

In this case, Cauchy importance sampling is applied to the MC method, the QMC method

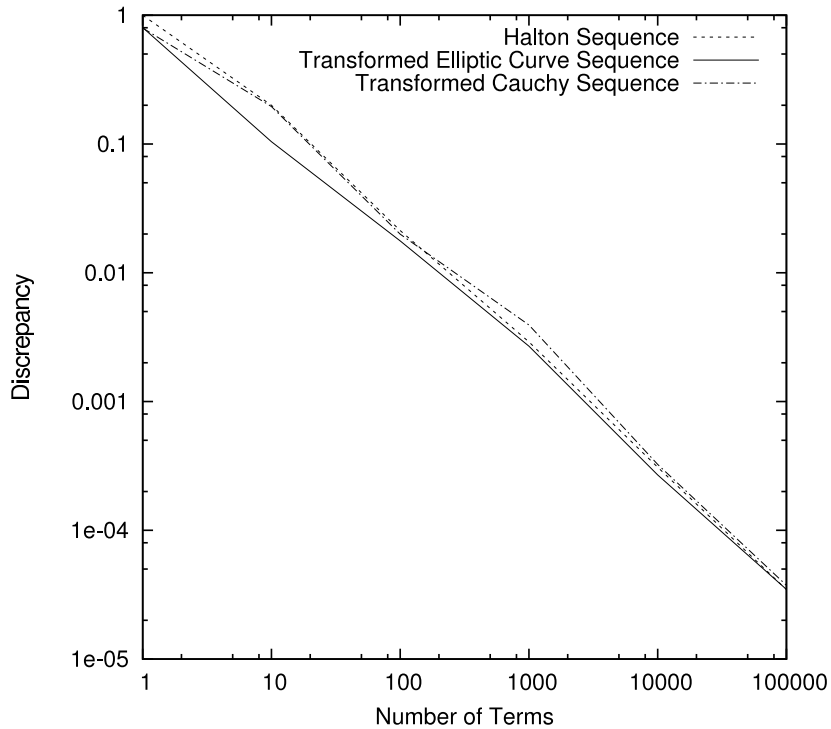


Figure 1: Star-Discrepancy of a Halton Sequence versus Transformed Group-Theoretic Elliptic Curve and Cauchy Sequences

with the Halton sequence, rank-1 lattice rules, as well as the Weyl-sequence with initial point $(\sqrt{2}, \sqrt{3}, \sqrt{5}, \sqrt{7}, \sqrt{11}, \sqrt{13}) \bmod 1$, which is of type-1. In this last case theorem 20 is applied with weight function $w^{(1)}(x)$. Once again the Fourier methods perform the best, but they have lost some of their advantage.

7 Conclusion

In this paper we have introduced a group-theoretic method to generate a few Weyl-like non-uniform quasi-random sequences. We have also introduced a thick-tailed quasi-random importance sampling technique that can be used for some problems involving distributions which we cannot generate directly. This importance sampling technique creates an equivalent smooth inverse problem in the cube that provides not only a theoretical validation for QMC methods involving unbounded integrands, but higher rates of convergence when used with Fourier-based techniques such as lattice rules or our weighted integration rule of theorem 20. While lattice rules are simple to evaluate and very effective, finding a good lattice point as an initial seed can be computationally intensive and is dependent on both dimension and number of required evaluations. In comparison, theorem 20 provides a very simple method to integrate smooth integrands of a moderate dimension (perhaps ≤ 6) with respect to smooth distributions.

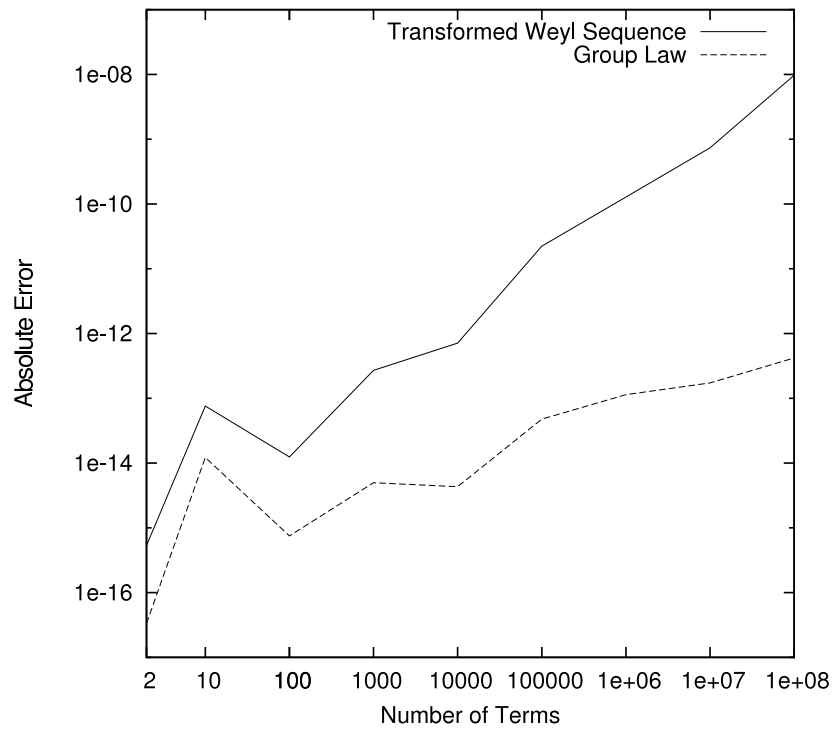


Figure 2: Comparison of Error Propagation for a Cauchy Sequence

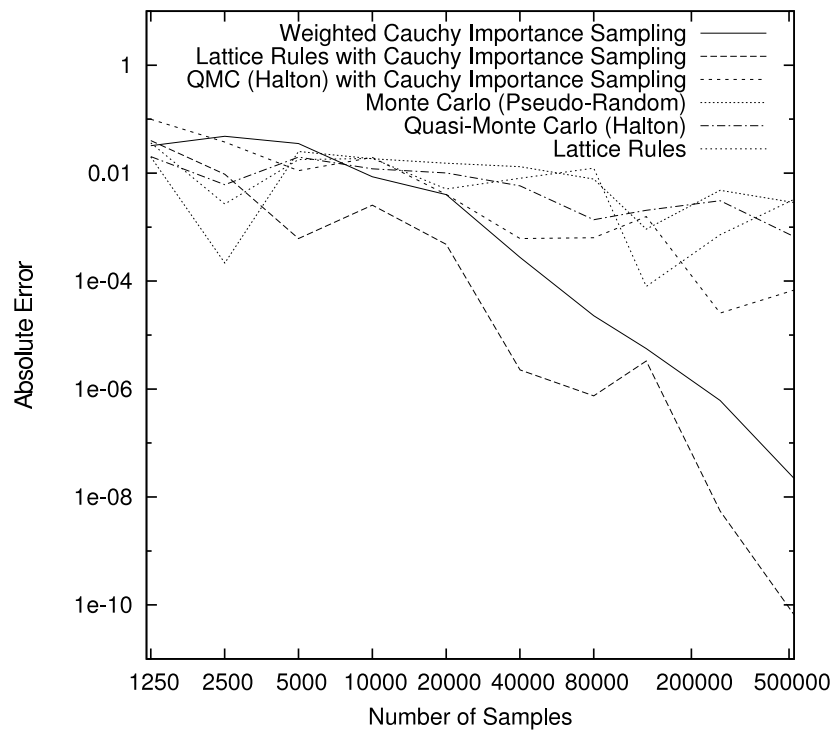


Figure 3: Integration with Respect to a Trivariate Mixture of Gaussians

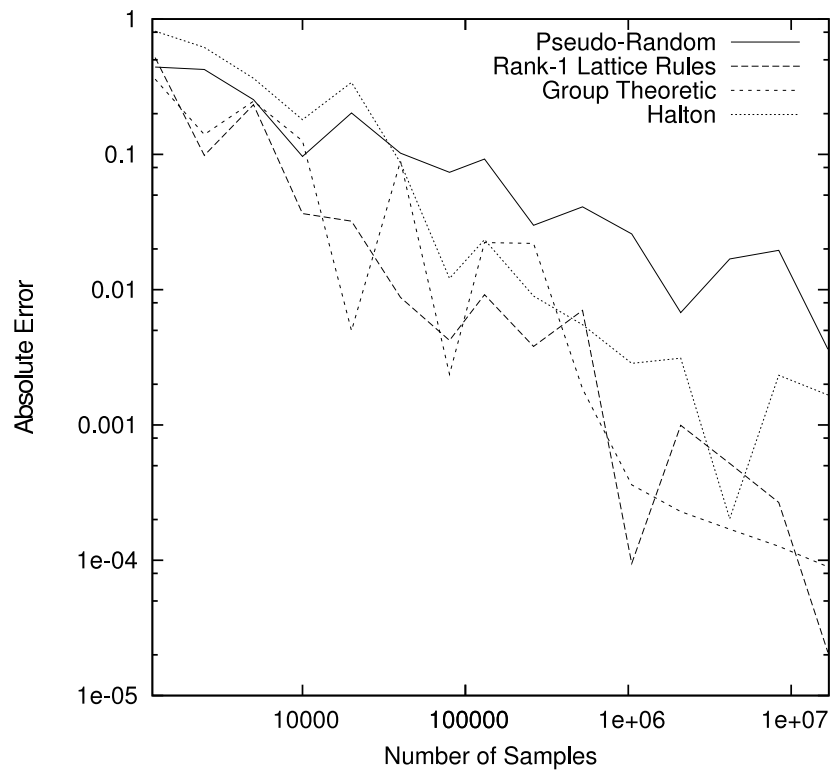


Figure 4: Integration with Respect to a Six-Variate t-Distribution Using Cauchy Importance Sampling

References

- [1] A. Baker, A Central Theorem in Transcendence Theory, in: C.F. Osgood (Ed.), *Diophantine Approximation and Its Applications*, Academic Press, New York (1973), 1–23.
- [2] D. G. Cantor, Computing in the Jacobian of a Hyper-Elliptic Curve, *Mathematics of Computation*, **48** (1987), 95–101.
- [3] R.F. Coleman, F.O. McGuinness, Rational Formal Group Laws, *Pacific Journal of Mathematics*, **147** (1991), 25–27.
- [4] L.P. Devroye, *Non-Uniform Random Variate Generation*, Springer, New York (1986).
- [5] E. de Doncker, Y. Guan, Error Bounds for the Integration of Singular Functions Using Equidistributed Sequences, *Journal of Complexity*, **19** (2003), 259–271.
- [6] M. Drmota, R.F. Tichy. *Sequences, Discrepancies and Applications*, Springer, Berlin (1997).
- [7] K.-T. Fang, Y. Wang, *Number-Theoretic Methods in Statistics*, Chapman and Hall, New York (1994).
- [8] M. N. Freije, The Formal Group of the Jacobian of an Algebraic Curve, *Pacific Journal of Mathematics*, **157** (1993), 241–255.
- [9] James E. Gentle, *Random Number Generation and Monte Carlo Methods*, Springer, New York (1998).
- [10] M. Hazewinkel, *Formal Groups and Applications*, Academic Press, New York (1978).
- [11] J. Hartinger, R. Kainhofer, R. Tichy, Quasi-Monte Carlo algorithms for unbounded, weighted integration problems, *Journal of Complexity* **20** (2004), 654–668.
- [12] A. Keller, Quasi-Monte Carlo Methods in Computer Graphics, in: O. Mahrenholtz, K. Marti, R. Mennicken (Eds.), *ICIAM / GAMM 95, Special Issue of ZAMM, Issue 3: Applied Stochastics and Optimization*, (1996), 109–112.
- [13] S. Lang, *Elliptic Curves: Diophantine Analysis*, Springer-Verlag, New York (1978).
- [14] C. Lemieux, P. L'Ecuyer, On the Use of Quasi-Monte Carlo Methods in Computational Finance, *Lecture Notes in Computer Science*, vol. **2073** (2001), Springer-Verlag, 607–616.
- [15] I. Niven, *Irrational Numbers*, MAA, Washington D.C (1963).
- [16] H. Niederreiter, Application of Diophantine Approximations to Numerical Integration, in: C.F. Osgood (Ed.), *Diophantine Approximation and Its Applications*, Academic Press, New York (1973), 129–199.
- [17] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, Society for Industrial and Applied Mathematics, Philadelphia (1992).
- [18] M. Ostland, B. Yu, Exploring Quasi-Monte Carlo for Marginal Density Approximation, *Statistics and Computing*, **7** (1997), 217–228.
- [19] T. Pillards, R. Cools, Transforming Low-Discrepancy Sequences from a Cube to a Simplex, *Journal of Computational and Applied Mathematics*, **174** (2005), 29–42.

- [20] J. E. H. Shaw, A Quasirandom Approach to Integration in Bayesian Statistics, *Annals of Statistics*, **16** (1998), 895–914.
- [21] B.V. Shuhman, Applications of Quasirandom Points for Simulation of Gamma Radiation Transfer, *Progress in Nuclear Energy*, **24** (1990), 89–95.
- [22] I. H. Sloan, S. Joe, *Lattice Methods for Multiple Integration*, Oxford University Press, Oxford (1994).
- [23] M. Sugihara, K. Murota, A Note on Haselgrove’s Method for Numerical Integration, *Mathematics of Computation*, **39** (1982), 549–554.
- [24] X. Wang, Improving the Rejection Sampling Method in Quasi-Monte Carlo Methods, *Journal of Computational and Applied Mathematics*, **114** (2000), 231–246.
- [25] S. K. Zaremba, Some Applications of Multidimensional Integration by Parts, *Ann. Polon. Math.*, **21** (1968), 85–96.