Quasi-Monte Carlo and Monte Carlo Methods and their Application in Finance

G. Larcher and G. Leobacher

Department of Financial Mathematics
University Linz Altenberger Strasse 69
A-4040 Linz Austria
e-mail: Gerhard.Larcher@jku.at,
Gunther.Leobacher@jku.at

Abstract

We give an introduction to and a survey on the use of Quasi-Monte Carlo and of Monte Carlo methods especially in option pricing and in risk management. We concentrate on new techniques from the Quasi-Monte Carlo theory.

1 Introduction

It is the aim of this paper to give a survey on recent developments in the application of Monte Carlo- (MC), but especially of quasi-Monte Carlo- (QMC) methods in financial mathematics. It moreover should be an advertising for extended use of quasi-Monte Carlo-methods at various suitable occasions within financial mathematics.

This advertising is motivated for example by graphs like the following, illustrating (without going into any details at this place) the behaviour of a Monte Carlo approximation (solid line) and of a quasi-Monte Carlo approximation (dotted line) to the fair price of a Mortgage Backed Security (figure 1 from Tezuka [70]).

Although in various papers on QMC-methods and in various further experiments like the one above the authors of these papers have shown the sometimes stunning superiority of the QMC-method over the MC-method, there still seems to exist a certain reserve, if not exaggerated caution or even an open objection to the use of QMC-methods by practitioners as well as by academic financial mathematicians.

The main reasons for this reserve seem to be twofold (and maybe even threefold): First there still is a broad gap between practice and theory of QMC-methods. This gap also is mirrored by a lack of suitable literature on the practice of QMC-methods for practitioners. There have been great advances in the theory of QMC-methods within the last two decades, especially by the work of for example Sobol, Faure, Niederreiter and many others. There also exists excellent theoretical literature on the topic. However these great theoretical results have not been translated for practitioners in a satisfactory way. Of course there have been some attempts (especially by J. Traub and his group at Columbia University) to make the theory accessible to practitioners, but these until now are only singular attempts.

The second reason maybe is more subtle: Although QMC-methods in many cases of applications show an overwhelming better performance than other methods and especially than MC-methods, there are other principally possible cases for an application where it is not indicated to use QMC-methods, where it maybe is even a dangerous mistake to use QMC, whereas it is

*Research supported by the Austrian research funds (FWF) project S8308
quite all right to use plain MC-methods. So maybe this slight uncertainty about the applicability of QMC-methods discourages further possible users of QMC-methods.

And we finally must not withhold a third possible argument against a carefree use of QMC: these methods in many concrete applications work very well, however in many cases until now it is not fully possible to theoretically explain this superiority. We will give some more details about this situation in section 2.

So in this survey paper we will try to do two things:

For the field of application of QMC-methods to different problems in financial mathematics we want to narrow the gap between the theory of QMC and possible use of QMC on the one hand, and to reduce the uncertainty about the applicability of QMC in financial mathematics a little on the other hand.

This paper addresses to two kinds of readers: to readers with some background in financial mathematics who are interested in a possible use of QMC-methods, and also to specialists in the theory of QMC who are looking for new fields of applications for their methods. So probably a specialist in QMC-methods will miss some details on this topic, whereas on the other hand a specialist in financial mathematics will not learn many news about financial mathematics.

For example, we will essentially be concerned with "crude" MC- and QMC-methods and we reduce only to short hints to the literature concerning further important topics, like variance reduction, variation reduction or ANOVA decomposition techniques.

The paper is organised as follows:

In Section 2 we summarise the necessary basic general facts and information about MC- and QMC-methods, and we report on (not empirical but theoretically well-founded) facts concerning the comparison between the quality of performance of MC- versus QMC-methods.

In Section 3 we try to give some concrete valuable hints for a practitioner on how and when to use QMC-methods on a first stage.

In Sections 4 and 5 we first provide a short introduction into the general princi-
amples of option pricing and then give several convincing examples for the use of QMC in this field.

In Section 6 we give a short introduction into the very actual field of measuring risks of several kinds in an adequate and efficient way by MC- and QMC-methods.

2 Basic facts on Monte Carlo- and Quasi-Monte Carlo-methods

Many simulation problems can be reduced to the task of approximately calculating an expected value in the sense of probability theory. This further means to (at least approximately) calculate a sometimes high-dimensional and, especially in applications in finance, a sometimes even very high-dimensional integral, say

$$\int_{\mathbb{R}^s} f(x) dx.$$

There are several well-working techniques to transform these integrals (again at least approximately) to integrals over a bounded region, say, over the $s$-dimensional unit-cube, so that the task reduces to the numerical evaluation of integrals of the form

$$\int_{[0,1]^s} f(x) dx,$$

where $f$ belongs to a certain class $F$ of integrands in question. This class $F$, which changes from case to case, in some sense represents the information we have about the integrand in the special application.

A global obvious approach is to consider this integral as the average value of $f$ on $[0,1]^s$, and to approximate it by a discrete average value

$$\frac{1}{N} \sum_{k=1}^{N} f(x_k)$$

over a sample of suitable chosen points $x_1, \ldots, x_N$ in $[0,1]^s$.

This approach is the basis for both "crude" Monte Carlo as well as "crude" quasi-Monte Carlo methods. By labelling these approaches by "crude" we want to hint that there of course exist refinements (weighted means, variance reduction methods, variation reduction methods, Brownian bridges, ...) and adaptive versions of these basic "crude" MC- and QMC-method.

So what are the suitable point sets $x_1, \ldots, x_N$ for this task? MC-theory and QMC-theory answer this question in different ways:

If one wants to apply MC-methods, then he should choose $x_1, \ldots, x_N$ in $[0,1]^s$ "as randomly distributed as possible". What does this mean? We would like to say: "Do not ask a random number specialist." The reason for this is: There are several philosophies on how measuring the "quality of randomness" of a given finite point sequence (an early exhaustive discussion can be found in Knuth [33], but see also for example Niederreiter [46], or Hellekalek [25] for a more recent discussion). Just informal and roughly speaking it should be a point sequence with "average properties". For example: it should be well distributed over the unit cube, but not extra-ordinarily well distributed. Or: it should have no, or only imperceptible correlations between consecutive points, and so on. We will just remark here that there are several tests for the quality of the randomness of a finite point sequence, and that various reliable pseudo-random number generators can be found for generating point sets which pass most of these tests.
A typical picture of a pseudo-random point set in the plane is shown in figure 2.

Figure 2: a 2-dim. random point set

A user of a QMC-method uses point sets which are as uniformly distributed in \([0, 1]^s\) as possible. What does this mean? It is much easier to give a concrete answer to this question:

There are two main (essentially equivalent) measures for the "quality of uniform distribution" of a point set \(x_1, \ldots, x_N\) in \([0, 1]^s\).

The two measures are the star-discrepancy \(D^*_N\) and the \(L_2\)-discrepancy \(L^{(2)}_N\) of the point set. They are defined as follows:

For a subinterval \(B\) of \([0, 1]^s\), let \(A_N(B)\) denote the number of points \(x_i\) in \(B\). For a uniformly distributed point set this quantity \(A_N(B)\) should be approximately equal to \(N\) times the volume \(\lambda(B)\) of \(B\). A discrepancy measures the deviation of the two quantities.

The star discrepancy measures the maximal deviation

\[
D^*_N = \sup_B \left| \frac{A_N(B)}{N} - \lambda(B) \right|
\]

and the \(L_2\)-discrepancy measures the average quadratic deviation

\[
L^{(2)}_N = \left( \int_{[0,1]^s} \left( \frac{A_N(\prod_{i=1}^s [0, b_i])}{N} - b_1 \cdot b_2 \cdots b_s \right)^2 db_1 db_2 \cdots db_s \right)^{1/2}.
\]

A point set will be called well-distributed if \(D^*_N\) respectively \(L^{(2)}_N\) is small.

What are the reasons, or the philosophy for MC- respectively QMC-users to prefer their special kinds of point sets? Their choice depends on their method to estimate the integration error:
The method to manage the integration error when using a MC-method is based on the following result:

Consider the class $F$ of square integrable functions on $[0, 1]^s$, i.e., $F = L_2([0, 1]^s)$. For $f \in F$ we consider the average of the quadratic integration error when using the basic approach, i.e.

$$\left( \int_{[0,1]^s} \cdots \int_{[0,1]^s} \left( \frac{1}{N} \sum_{k=1}^{N} f(x_k) - \int_{[0,1]^s} f(x)dx \right)^2 \right)^{1/2} dx_1 \cdots dx_s .$$

This quantity easily can be calculated and it turns out to equal

$$\frac{1}{\sqrt{N}} \cdot \sigma(f)$$

where $\sigma(f)$ denotes the standard deviation of $f$, that is

$$\sigma(f) = \left( \int_{[0,1]^s} f^2(x)dx - \left( \int_{[0,1]^s} f(x)dx \right)^2 \right)^{1/2},$$

a constant not depending on the number $N$ of sample points.

The most remarkable fact in this result is, that this average quadratic integration error does not explicitly depend on the dimension $s$ of the integration problem. This means that the speed of convergence as a function of the number $N$ of sample points does not depend on the dimension $s$. Moreover it is easily checked that the integration errors

$$\frac{1}{N} \sum_{k=1}^{N} f(x_k) - \int_{[0,1]^s} f(x)dx$$

are normally distributed around 0 with standard deviation $\frac{1}{\sqrt{N}} \cdot \sigma(f)$.

So the strategy of a MC-player is the following: he performs several trials to calculate the approximate integral value with changing (non-correlated) random point sets $x_1^{(i)}, \ldots, x_N^{(i)}$. When he finally takes the average over the obtained results then with a large probability this final value will be placed near to the true value, say within a tolerance of $\frac{1}{\sqrt{N}} \cdot \sigma(f)$, which, we emphasise it again, does not depend on the dimension $s$. So MC-methods use a probabilistic, (in the above sense) dimension-independent error estimate.

This is in contrast to classical product rules for high-dimensional numerical integration. Just for illustration we give a simple example of such a classical product rule:

To calculate

$$\int_{[0,1]^s} f(x)dx = \int_{0}^{1} \cdots \int_{0}^{1} f(x^{(1)}, \ldots, x^{(s)})dx^{(1)} \cdots dx^{(s)}$$

just approximate each single integral by a one-dimensional integration rule, say in a naive version by the average of $f$ over equidistant points

$$\frac{1}{2m}, \frac{3}{2m}, \ldots, \frac{2m-1}{2m}.$$ 

This means that the whole integral is approximated by

$$\frac{1}{m^s} \sum_{i_1, \ldots, i_s=1}^{m} f \left( \frac{2i_1 - 1}{2m}, \ldots, \frac{2i_s - 1}{2m} \right).$$
Since for every single dimension $i$, for a large class $F$ of functions $f$ (for example for the class $F = W^{(1,1,...,1)}([0,1]^s)$, the Sobolev space of functions which are once differentiable with respect to each variable) there can be given integration error estimates of the form $c_i m$ with a constant $c_i$ depending only on $f$, we altogether obtain an integration error estimate also of the form

$$\left| \int_{[0,1]^s} f(x) dx - \frac{1}{m^s} \sum_{i_1,...,i_s=1}^m f \left( \frac{2i_1 - 1}{2m}, ..., \frac{2i_s - 1}{2m} \right) \right| \leq \frac{c}{m} = \frac{c}{N^{1/s}}$$

with a constant $c$ depending only on $f$.

Hence the integral over $f$ is approximated by the average value of $f$ over the points of a regular lattice in $[0,1)^s$ with $N = m^s$ points.

Figure 3: Regular lattice in $[0,1)^2$

So here we have a concrete deterministic error estimate. However the speed of convergence in the number $N$ of sample points is very poor for higher dimensions $s$. In any case the bound is worse, at least asymptotically, than the MC-bound for $s \geq 3$, but: it is a deterministic bound!

Now QMC-methods try to combine the two possible merits, namely the deterministic bound, and an error estimate which essentially is independent on the dimension $s$.

A first essential step into this direction was provided by the proof of the now so-called inequality of Koksma and Hlawka [26] (see also [36]) in 1960. This inequality states the following:

For a function $f : [0,1]^s \rightarrow \mathbb{R}$ denote by $V(f)$ the variation of $f$ (in the sense of Hardy and Krause). Now we consider the class $F$ of functions $f$ with $V(f) < \infty$. Let $x_1, ..., x_N$ be a set of points in $[0,1]^s$ with star discrepancy $D_N^*$. Then

$$\left| \frac{1}{N} \sum_{k=1}^N f(x_k) - \int_{[0,1]^s} f(x) dx \right| \leq V(f) \cdot D_N^*.$$

So here we find a deterministic bound valid for arbitrary point sets $x_1, ..., x_N$ in $[0,1]^s$. 
But now the decisive question is: what about the quantity $D^*_N$, how small can it be made by a suitable choice of points $x_1, \ldots, x_N$?

First of all, as we might guess from the above considerations on classical product rules, the naive choice of a uniform lattice

$$\left(\frac{2i_1 + 1}{2m}, \ldots, \frac{2i_s + 1}{2m}\right) ; i_j = 0, \ldots, m - 1, \ j = 1, \ldots, s$$

with $N = m^s$, is not a prudent choice. Consider for example the rectangle $B = \prod_{i=1}^{s} [0, \frac{2m-1}{2m}]$ then $A_N(B) = N$, hence

$$D^*_N \geq \frac{A_N(B)}{N} - \lambda(B) = 1 - (1 - \frac{1}{2m})^s \geq \frac{1}{2m} = \frac{1}{2N^{1/s}}.$$

So we do not get a better bound than the strongly dimension-dependent bound for the product rule.

It was for the first time around 1960 that independently by Korobov [35] and by Hlawka [27] the existence of point sets with an essentially smaller order of convergence for $D^*_N$ was shown. They introduced the now so-called concept of good lattice points. The basic result in this context was given by Niederreiter [43]:

For every integer $N$ and any dimension $s$ there exist integers $a_1, \ldots, a_s$ such that the point set

$$x = \left(\left\{ k \cdot \frac{a_1}{N} \right\}, \ldots, \left\{ k \cdot \frac{a_s}{N} \right\} : k = 1, \ldots, N \right)$$

($\{x\}$ denotes the fractional part $x - [x]$ of $x$) has star discrepancy

$$D^*_N \leq c_s \cdot \frac{(\log N)^s}{N}$$

($c_s$ is a constant depending only on the dimension $s$, indeed $c_s$ grows large as $s$ becomes large, which will be pointed out again later on).

Let us give some remarks on this result on "good lattice point sets":

• Until today for all dimensions $s \geq 3$ this still is a pure existence result. But of course for values of $N$ and $s$ reasonable for applications there exist tabulated values for good choices $(a_1, \ldots, a_s)$ (see for example [19]).

• It is not known how far the estimate

$$D^*_N \leq c_s \cdot \frac{(\log N)^s}{N}$$

is best possible. In any case it is not too far from the truth since there exists a general lower estimate for the discrepancy of point sets in $[0, 1]^s$:

for any dimension $s$ there is a constant $c_s > 0$ such that for any point set $x_1, \ldots, x_N$ in $[0, 1]^s$

$$D^*_N \geq c_s \cdot \frac{(\log N)^{(s-1)/2}}{N}$$

holds (see [36]).

It is a famous conjecture in the theory of uniform distribution that even

$$D^*_N \geq c_s \cdot \frac{(\log N)^{s-1}}{N}$$

for any dimension $s$.
always holds.

Inserting the estimate for $D_N^s$ into the Koksma-Hlawka inequality provides an integration error

$$c_s \cdot V(f) \cdot \frac{(\log N)^s}{N},$$

which still depends on the dimension $s$, but concerning the global speed of convergence in $N$ the dependence of the $\frac{(\log N)^s}{N}$ estimate is much weaker than for the $\frac{1}{\sqrt{N}}$ estimate, and in fact, for given $s$, $\frac{(\log N)^s}{N}$ converges faster to zero than the probabilistic $\frac{1}{\sqrt{N}}$ bound of the MC method. But these are only theoretical aspects. We return to a discussion on practical aspects later on.

- Good lattice point sets are of special interest for the numerical integration of periodic and smooth functions:

  for $\alpha$ in $\mathbb{N}$ let $F_\alpha$ be the class of functions $f : \mathbb{R}^s \rightarrow \mathbb{R}$, periodic with period one in each variable, for which all partial derivatives $\frac{d^{a_1} f}{d^{a_1} x_1 \cdots d^{a_s} f}$ exist and are bounded.

  Let $(a_1, \ldots, a_s)$ be a good lattice point in the above sense, then for the corresponding good lattice point set $x_1, \ldots, x_N$ one has

$$\left| \frac{1}{N} \sum_{k=1}^{N} f(x_k) - \int_{[0,1]^s} f(x) dx \right| \leq c(s, f) \cdot \frac{(\log N)^{\alpha s}}{N^\alpha}$$

for all $f \in F_\alpha$. That means: good lattice points "recognise" smoothness of an integrand.

The first concrete (explicitly given) point sets with low discrepancy (let us informally use the notion "low-discrepancy point set" for $N$-element point sets with discrepancy around $\frac{(\log N)^{\alpha s}}{N}$) were provided by Halton [23] in 1960. These points are generated as follows (we give the slightly improved Hammersley variant):

  for given dimension $s$ choose different primes $p_1, \ldots, p_{s-1}$. For a prime $p$ and integer $n$ with base $p$ representation

$$n = a_0 p^r + \cdots + a_1 p + a_0$$

let

$$\phi_p(n) := \frac{a_0}{p} + \frac{a_1}{p^2} + \cdots + \frac{a_r}{p^{r+1}}.$$

Then a Hammersley-Halton point set is given by

$$x_n = \left( \frac{n}{N}, \phi_{p_1}(n), \ldots, \phi_{p_{s-1}}(n) \right); \quad n = 0, \ldots, N - 1.$$

This concrete point set (see for example [28]) has discrepancy

$$D_N^s \leq c_s \cdot \frac{(\log N)^{s-1}}{N}.$$

So, what is needed more: using such a concrete and easy to implement Hammersley-Halton point set for the numerical integration of a multivariate function $f$ and using the Koksma-Hlawka inequality for estimating the integration error provides a deterministic bound of an order which tends to zero (in the number $N$ of sample points) much faster than the purely probabilistic MC bound ?

We do want more, at least in two different aspects:
• for concrete integration purposes not only the order of convergence in $N$ plays a crucial role, but also the dimension-dependent constant $c_s$ in the discrepancy estimates. So what about these constants?

• Is it possible to get completely rid of the dimension $s$ in the order of convergence?

We will discuss these two topics in the remaining part of this section.

Remember that for good lattice point sets, respectively for Hammersley-Halton point sets we had discrepancy estimates of the form

\[ D_N^* \leq c_s(1) \cdot \frac{(\log N)^s}{N} \]  

(respectively

\[ D_N^* \leq c_s(2) \cdot \frac{(\log N)^{s-1}}{N} \])

However, in the currently known discrepancy estimates the occurring constants $c_s(i)$ are very large, namely $c_s(1) \approx 2^s$ and $c_s(2) \approx s^s$ (if $p_i$ in the definition of the point set denotes the $i$-th prime).

That means: for large dimensions $s$ these estimates are of no practical use!

And this fact heavily motivated the search for alternative point sets whose discrepancy, like the one of the former point sets, also is of (almost) minimal order in $N$, but which moreover have an essentially smaller constant $c_s$ in the discrepancy estimate. This search led to the utmost important concept of $(t, m, s)$-nets.

A first singular example of a $(t, m, s)$-net already was given by Sobol [67] in 1967. Then in 1982 essentially the same example, but in a much more general form and with explicit discrepancy estimate was given by Faure [18]. The detailed study of a general theory of $(t, m, s)$-nets then was initiated and was in great parts carried out by Niederreiter and by his collaborators since 1987 (see e.g.:[44],[45],[50], [38],[39]).

We do not give the full details concerning $(t, m, s)$-nets, we just give the facts which are necessary and of interest for a potential user. Especially we restrict in the following to the special subclass of digital $(t, m, s)$-nets in a prime base $b$ (over $\mathbb{Z}_b$).

In principle, this is a point set of $b^m$ points in $[0,1)^s$ with a certain quality parameter $t$ with $0 \leq t \leq m$, which gives information about the quality of the distribution of the point set. $t = 0$ means optimal quality.

A digital $(t, m, s)$-net over $\mathbb{Z}_b$ is constructed in the following way:

choose $s$ $m \times m$-matrices $C_1, \ldots, C_s$ with entries from the residue class field $\mathbb{Z}_b$ ($b$ is a prime !) modulo $b$. To generate the point set $x_0, \ldots, x_{N-1}$; $N = b^m$, the $i$-th coordinate $x_n(i)$ of $x_n$; $0 \leq n < N$ is generated as follows:

represent $n$ in base $b$

\[ n = a_{m-1}b^{m-1} + \cdots + a_1b + a_0. \]

multiply

\[ C_i \cdot \begin{pmatrix} a_0 \\ \vdots \\ a_{m-1} \end{pmatrix} =: \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \]

and set

\[ x_n(i) := \frac{y_1}{b} + \frac{y_2}{b^2} + \cdots + \frac{y_m}{b^m} \in [0,1). \]

This is the general digital construction method.
But what about the parameter $t$? It of course depends on the choice of the matrices $C_1, \ldots, C_s$ and it is determined as follows:

let the integer $d$ be maximal such that for all non-negative integers $d_1, \ldots, d_s \geq 0$ with $d_1 + \cdots + d_s = d$ we have:

the system of the

first $d_1$ rows of $C_1$ together with the
first $d_2$ rows of $C_2$ together with the
\[ \ldots \]
first $d_s$ rows of $C_s$

is linearly independent over $\mathbb{Z}_b$, then the quality parameter $t$ of the corresponding digital net satisfies

\[ t = m - d. \]

Note that we have $m$-dimensional row vectors, so in the extreme case $d$ is equal to $m$.

As a first simple example consider

$$
C_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix},
$$

in $\mathbb{Z}_b$. This, as is easily checked, provides a $(0, m, 2)$-net in base $b$ and is just a two-dimensional Hammersley-Halton point set with $N = b^m$ elements.

So one has to search for matrices over $\mathbb{Z}_b$ with far-reaching independence relations. And this search is a big field for the application of methods from various branches like algebra, combinatorics, coding theory, number theory, algebraic geometry, \ldots.

What are the fundamental results?

First, what about the star-discrepancy $D^*_N$ of a digital $(t, m, s)$-net in base $b$? we have

\[ D^*_N \leq c_s \cdot b^t \cdot \frac{(\log N)^{s-1}}{N}. \]

That is:

- we have minimal order of convergence in $N$
- we should like to have $t = 0$
- but what about $c_s$?

It is the great surprise that an estimate of the form

\[ c_s \approx \frac{1}{s^s} \]

is possible (however with an restriction: for $N \geq N_0(s, b)$. We will return to this later).
But what about the existence of digital nets with small $t$? First a negative result:

The optimal parameter $t = 0$ is, for given dimension $s$, possible only if the base $b$ is large enough. In detail: a $(0,m,s) - \text{net over } \mathbb{Z}_b$ ($b$ prime) exists if and only if $b \geq s - 1$. Such point sets for the first time were given by Faure in 1982 and we will introduce them to the reader and possible user in the next section.

So what should one want more now? We have (for these point sets)

$$D_N^* \leq c_s \cdot b^{0} \cdot \frac{(\log N)^{s-1}}{N},$$

i.e., probably best possible order of convergence in $N$, a constant tending to zero with growing dimension, and all this for explicitly given point sets.

At this place we must return to a small remark given above: the stated discrepancy estimate holds for $N \geq N_0(s,b)$, and this border $N_0(s,b)$ is strongly growing in the base $b$. And indeed it turned out that for applications in higher dimensions (say $s \geq 10$), smaller bases are much more favourable even if $t$ then is not zero any more!

And so in recent years much work was concentrated on the search for digital $(t,m,s)$-nets in base 2 with $t$ as small as possible. (See for example [46], [50], [63].) Again in Section 3 we will give a concrete example of interest for users.

The best theoretical result in this direction is the following deep result of Niederreiter and Xing [50] based on methods from algebraic geometry:

they give for every $s$ and $m$ the concrete construction of digital $(t,m,s)$-nets over $\mathbb{Z}_2$ with $t \leq 3(s-1)$. However until now there still exist difficulties in implementing these point sets for concrete use (see for example [60]).

So these on the whole seem to be good news, and, in fact, QMC-methods work, they even work very well in industrial and in financial applications!

But remember that we promised to discuss also the second question, if it were possible to get completely rid of the dimension $s$ in the discrepancy estimates. And now there will follow the bad news, and it will turn out that these are really bad news, since we will have to confess, that we in fact do not know why QMC-methods work so well, and especially why they in many cases work so much better than MC-methods.

So to this end let us compare once more the probabilistic MC-integration bound with the bound obtained from the Koksma-Hlawka inequality when using Niederreiter-Xing nets with optimal performance:

<table>
<thead>
<tr>
<th>MC</th>
<th>QMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma \frac{1}{\sqrt{N}}$</td>
<td>$\frac{1}{s} \cdot 2^{3(s-1)} \cdot \frac{(\log N)^{(s-1)}}{N}$</td>
</tr>
</tbody>
</table>

probabilistic deterministic

So the deterministic (though, of course, asymptotic) bound should be favourable in any aspect, why should we beyond try to even get completely rid of $s$?

The answer is: consider a realistic scenario, say the integration of a function $f$ in dimension $s \geq 20$, and assume you use, say, $\alpha^s$ points with some $\alpha > 1$. Then to obtain by Hlawka’s inequality a non-trivial bound, we need that

$$\frac{1}{s^s} \cdot 2^{3(s-1)} \cdot \frac{s^{s-1} \cdot (\log \alpha)^{s-1}}{\alpha^s} \ll 1,$$

but this means roughly $\alpha \gg 26$, and so we need at least $26^s$ sample points. Of course this is not possible, even for $s \geq 10$, say.
So the QMC-bound is brilliant for large \(N\), but in no way for a realistic choice of \(N\) if the dimension \(s\) is not really small! Hence the "shocking" consequence: the Koksma-Hlawka bound together with the known discrepancy estimates is - in spite of all the mentioned deep results - useless for integration in high dimensions. (Concerning these considerations see also [65].)

So does this mean that QMC-integration in higher dimensions is useless? No, it does not mean this, since - as will for example be illustrated in Sections 5 and 7 - it works!

However, researchers in the field must confess that they do not know why QMC methods work in (high-dimensional) practice. Much research is done now into this direction to give a satisfactory explanation (see for example [66], [24]). The attempts to give an explanation go into two main directions:

- Probably the known theoretical bounds for the star discrepancy of the point sets in use could be improved considerably for "small", i.e. realistic point sets in the sense indicated above. An essential hint into this direction was given quite recently in a pioneering work by Heinrich et al. [24], in which it is shown that "small" point sets with "small" discrepancy must exist. It was shown there, that:

  in any dimension \(s\) for every \(\epsilon > 0\) there are \(N\) points \(x_1, \ldots, x_N\) in \([0,1]^s\) with

\[
N \leq \frac{c \cdot s}{\epsilon^2}
\]

\((c\) an absolute constant) and with

\[
D^*_N \leq \epsilon.
\]

However until now this result is in no way constructive.

- A second possible explanation is, that the Koksma-Hlawka inequality is a too weak and too insensible tool to measure the integration error in high dimensions. This suspicion already is supported by the fact that the variation \(V(f)\) satisfies \(V(f) = \infty\) for many classes of rather simple functions (e.g. for the characteristic function of a triangle or of a sphere). However there is also another important aspect:

The bound \(V(f) \cdot D^*_N\) is absolutely insensible for imbalances of the integrand \(f\), i.e., it does not take into account that possibly certain directions of the integrand are of more or less importance. An extreme example would be an integrand of the form given in figure 4 which nominally is a function in two variables but which in fact only depends on the first coordinate.

So for integrating such a function by low-discrepancy point sets the second coordinate of the point sets and especially the distribution of the point set regarding the second coordinate does not play any role. So it may be that a point set with large two-dimensional discrepancy \(D^*_N\) integrates the function \(f\) very well, since it is merely very well distributed in the first coordinate. But this effect is not expressed in the Koksma-Hlawka bound.

Now especially in problems in financial mathematics the involved integrands frequently are of nominally high dimension, but with a strongly decreasing importance of the dimensions (in the above sense).

Now different authors (see especially [66]) have given dimension-weighted versions of a Koksma-Hlawka inequality, thereby enlightening a bit the problem why QMC-methods work so well in high dimensions. We cannot go into further details here.
Figure 4: a 2-dimensional function which depends only on the first coordinate

We conclude this section by calling attention to a further fundamental result in the field of QMC-methods, which also points out the importance of the $L_2$-discrepancy $L_N^{(2)}$ of a point set and by calling attention to a further version of QMC-methods.

The result (for the class $F$ of continuous functions $f$ on $[0, 1]^s$) was given by Wozniakowski [74] in 1991:

Let a fixed point set $x_1, \ldots, x_N$ be given in $[0, 1]^s$ and take now the average over an integration error

$$\left| \frac{1}{N} \sum_{k=1}^{N} f(x_k) - \int_{[0, 1]^s} f(x) \, dx \right|,$$

the average taken over all continuous functions $f$ with respect to the Wiener sheet measure. Then this average error is essentially the $L_2$-discrepancy of the point set $1 - x_1, \ldots, 1 - x_N$, where $1 = (1, \ldots, 1) \in \mathbb{R}^s$.

Concerning the $L_2$-discrepancy by a quite recent result of Chen and Skriganov [11], for every dimension $s$ and every $N$ concrete point sets (certain digital $(t, m, s)$-nets) $x_1, \ldots, x_N$ in $[0, 1]^s$ with minimal order of $L_2$-discrepancy, namely

$$L_N^{(2)} = c_s \cdot \frac{(\log N)^{(s-1)/2}}{N}$$

are known.

Finally we mention a so-called hybrid method developed by Owen [55], where it is tried to apply a combination of MC and QMC-methods. Here randomly chosen transformations are applied to a given net, thereby on the one hand maintaining the net properties and on the other hand bringing random, and thus statistical, properties into play.

So it is possible to prove for fixed, moderately smooth integrands $f$ (indeed the class $F$ of functions $f : [0, 1]^s \to \mathbb{R}$ which are two times differentiable in each coordinate is considered) expected integration errors of order

$$\frac{(\log N)^{(s-1)/2}}{N^{3/2}}.$$

For an application in financial mathematics see for example [54].
3 Concrete hints for the use of QMC

In this section we do not give new information for specialists in QMC-methods, we just give some maybe useful hints for beginners in QMC-simulation, to facilitate the first steps into high-dimensional QMC-methods.

At first we give two concrete, easy to implement digital \((t, m, s)\)-nets of high quality which provide for every reasonable dimension \(s\) and for every reasonable number \(N\) of points a point set \(x_1, \ldots, x_N\) in \([0,1]^s\) suitable for use in QMC-methods.

The first class of nets are the digital Faure \((0, m, s)\)-nets over \(Z_b\) with some \(b \geq s - 1\) (here we also repeat in short the general digital net method):

To produce a low-discrepancy point set in \([0,1]^s\) first choose a prime \(b \geq s - 1\) as small as possible and choose \(m\) for generating \(b^m\) points. Then take the matrices \(C_0, \ldots, C_{s-1}\) as follows:

\[
C_i = \begin{pmatrix}
(0) & 0 & \cdots & 0 & 0 \\
(0) & (1) & 0 & \cdots & 0 \\
(0) & (1)^2 & (2) & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(0) & (m-1) & (m-1)^2 & \cdots & (m-1) \\
\end{pmatrix}
\]

in \(Z_b\) for \(i = 0, \ldots, s - 2\) and

\[
C_{s-1} = \begin{pmatrix}
0 & 0 & \cdots & 0 & 1 \\
0 & 0 & \cdots & 1 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 1 & \cdots & 0 & 0 \\
1 & 0 & \cdots & 0 & 0 \\
\end{pmatrix}
\]

Then, as already explained in Section 2, to generate \(x_0, \ldots, x_{b^m-1}\) in \([0,1]^s\), the \(i\)th coordinate \(x_n^{(i)}\) of \(x_n\) is generated by:

represent \(n\) in base \(b\)

\[
n = n_{m-1}b^{m-1} + \cdots + n_0,
\]

let

\[
C_i \cdot \begin{pmatrix}
n_0 \\
\vdots \\
n_{m-1}
\end{pmatrix} = : \begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_m
\end{pmatrix}
\]

and set

\[
x_n^{(i)} = \frac{y_1}{b} + \frac{y_2}{b^2} + \cdots + \frac{y_m}{b^m}.
\]

If one needs \(N\) points with \(b^{m-1} < N < b^m\) (for large \(b\) the gap between two powers of \(b\) may be too large) then just taking the first \(N\) points \(x_1, \ldots, x_N\) is dangerous since this in general gives badly distributed point sets.

A good solution then is the following strategy:

choose \(b\) prime, \(b \geq s\), and choose the \(m \times m\)-matrices \(C_0, \ldots, C_{s-1}\) in the same way as \(C_0, \ldots, C_{s-2}\) above. So do not use the matrix

\[
\begin{pmatrix}
0 & 0 & \cdots & 0 & 1 \\
0 & 0 & \cdots & 1 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 1 & \cdots & 0 & 0 \\
1 & 0 & \cdots & 0 & 0
\end{pmatrix}
\]
Then any digital point set \( x_1, \ldots, x_N \) produced with these matrices for any \( N \) with \( b^{m-1} \leq N < b^m \) shows desirable distribution properties. Several authors prefer this second method anyway and they suggest moreover to use subsequences of the above point set of the form \( x_l, \ldots, x_{N+l-1} \), with, as a rule of thumb, \( l \) about 100,000.

As already mentioned earlier, the above point sets are not useful when \( s \) is too large, since then the base \( b \) has to be large with the effect that the desirable distribution properties of the net emerge only for \( m \) large enough, i.e., for eventually unrealistic large \( N = b^m \).

Then one should prefer the following type of digital \((t, m, s)\)-nets in base 2 which is a special case of a larger class suggested by Niederreiter [46] and in a slightly modified form by Tezuka [69] (see also [6] for further hints concerning the implementation of these point sets):

To construct again \( s \times m \)-matrices \( C_1, \ldots, C_s \) now over \( \mathbb{Z}_2 \) choose \( s \) different primitive polynomials \( p_1, \ldots, p_s \in \mathbb{Z}_2[x] \) with degree as small as possible. For the sake of the reader we list the "first" primitive in Table 1. such polynomials

| Table 1: List of primitive polynomials over \( \mathbb{Z}_2 \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 0               | 2               | 3               | 7               | 11              | 13              | 19              | 25              | 37              | 41              |
| 47              | 55              | 59              | 61              | 67              | 91              | 97              | 103             | 109             | 115             |
| 131             | 137             | 143             | 145             | 157             | 167             | 171             | 185             | 191             | 193             |
| 203             | 211             | 213             | 229             | 239             | 241             | 247             | 253             | 285             |
| ...             |                 |                 |                 |                 |                 |                 |                 |                 |                 |

Here for example 91 means, 91 = (1011011)\(_2\), hence the polynomial is \( x^6 + x^4 + x^3 + x + 1 \).

Let \( e_i := \text{degree} (p_i) \). Then develop the rational functions

\[
\frac{z^k \mod e_i}{(p_i(z))^j} = \sum_{r=0}^{\infty} a^{(i)}(j, k, r) \cdot \frac{1}{z^r}
\]

for \( k = 1, \ldots, m; \ j = 1, \ldots \left\lfloor \frac{m-1}{e_i} \right\rfloor ; \ i = 1, \ldots, s \) and \( r \) is needed only for \( r = 1, \ldots, m \).

This for example can be done with the help of a symbolic computing software package like Mathematica.

Then the \( i \)-th matrix

\[
C_i := \left(c_{kr}^{(i)}\right); \ k, r = 1, \ldots, m; \ i = 1, \ldots, s
\]

is defined by

\[
c_{kr}^{(i)} = a^{(i)} \left( \left\lfloor \frac{k-1}{e_i} \right\rfloor + 1, k, r \right).
\]

If \( x_0, \ldots, x_{2^m-1} \) are generated then this gives a digital \((t, m, s)\)-net in base 2 with

\[
t = \sum_{i=1}^{s} e_i - s.
\]

(See [69].)

Here it is well possible to take any subsequence of consecutive elements \( x_l, \ldots, x_{N+l-1} \).
In many applications it is necessary to transform the uniformly distributed point sets to another point set with another given distribution. An excellent reference for this task is the book of Devroye [13]. In the subsequent concrete examples and in section 6 we give further hints in this context.

In any case QMC-methods can be applied if it is the clear goal to approximate a certain integral. (So most expected values can be expressed as integrals.) If the goal of the simulation is not expressed as an integral then sometimes it is still possible to apply QMC, but one should be careful about whether undesirable correlations between the elements of the point set are able to falsify the expected result.

One can also try, in analogy to the variance reduction techniques of MC-methods, to reduce the variation of the integrand \( f \), which is of relevance for the Koksma-Hlawka inequality. It has turned out by experiments that for example the antithetic method in some applications does give significant improvement also when applied for QMC-methods [1]. For more details concerning variance reduction see for example [20]. Concerning further investigations about the role of nonlinearity and complexity for the performance of QMC-methods we refer to [56] and to articles collected in [48], [47], [49], [17] and [52].

4 MC- and QMC-methods in option pricing

The pricing of financial derivatives is of basic importance in practice because of at least two reasons:

- Financial institutions need to know the theoretical price when engineering "new" financial products, or when marking to market some non-liquid assets, because the market price of the product has never been observed before, or because the assets have not been traded lately in the actual financial market.

- They conduct a benchmark for comparing the theoretical price with the actual price. If the difference is not negligible, it implies an arbitrage profit opportunity or mispricing.

It is well-known that the first basic breakthrough in this field was made by Black and Scholes [3] who successfully developed an elegant mathematical theory for option pricing based on stock prices following a geometric Brownian motion. They indeed found an explicit closed form solution for pricing plain vanilla European options.

However, since then, more and more complex derivatives, such as path-dependent European or American exotic options, have been developed on the basis of underlying complicated - but more realistic - evolution models for stock prices (like the hyperbolic models of Eberlein and Keller [16]) or on the basis of underlying complicated stochastic differential equations such as HJM (Heath-Jarrow-Morton) multi factor models. For the most of the pricing problem this complexity makes it almost impossible to them analytically. So we have to resort to numerical methods in order to obtain solutions. Numerical techniques for solving such problems fall into three categories:

- finite-difference methods, (binomial and trinomial) lattice methods (see [29] or [73]), and MC- respectively QMC-methods.

In this section we just give the necessary basic facts on derivative security pricing, and especially on option pricing, and on the approach to it via MC- and QMC-simulation methods. In the next section then we report on some selected concrete pricing examples handled by MC- and by QMC-simulation.
The first to apply MC-simulations to finance was Boyle [4]. See also [14].

The use of low-discrepancy sequences for finance problems has started around 1992. Paskov and Traub [59] used low-discrepancy sequences for pricing Collateralised Mortgage Obligations issued by Goldman-Sachs, and reported that these sequences performed very well relative to simple Monte Carlo methods, as well as to antithetic Monte Carlo methods. Further published work on QMC-methods in finance can be found for example in papers of Joy, Boyle and Tan [31], of Caflisch and Morokoff [10], Ninomiya and Tezuka [51], by Acworth, Broadie and Glasserman [1], by Ökten [53] and by Papageorgiou and Traub [58]. See also the papers contained in [14].

We recall, that an option on a stock (or on a currency or on an interest rate, ...) is a contract between two parties, one in a long position (the holder), one in a short position (the writer), which is valid over a time interval $[0, T]$, and which guarantees the holder a certain non-negative payoff $f$, paid by the writer, which is a function of the values of the stock $S$ in the time interval $[0, T]$. This payoff will be paid out at time $T$ if the option is of European type. If the option is of American type, then the payoff has to be paid (once) by the writer at any time $t$ dependent on the free choice of the holder. The payoff $f = f(t, S)$ then also depends on $t \in [0, 1]$ and for every $t$ on the values of the stock in the interval $[0, t]$ only. If the payoff of a European type option only depends on the values of the stock in time $T$ then we speak of a plain vanilla option. Otherwise we will speak of exotic options. Sometimes options do not rely only on one underlying asset but on several such primary securities. We will speak then of multi-asset (or basket-) options. We give examples of the most popular types of options in the above classes:

- a European (plain vanilla) call option has the payoff function
  $$f(S_T) = \max(S_T - K, 0).$$
  Here $S_T$ denotes the stock price at time $T$ and $K$ is a pre-arranged constant, the so-called exercise price.

- a (plain vanilla) put option pays
  $$f(S_T) = \max(K - S_T, 0).$$

- An American call (put) option pays
  $$f(S) = \max(S_t - K, 0)$$
  (respectively
  $$f(S) = \max(K - S_t, 0),$$
  if the holder decides to realize the payoff in time $t \in [0, T]$. (Here $S_t$ denotes the value of the stock $S$ in time $t$.)

- examples for exotic options (of European type all) are Asian options whose payoffs imply certain averages over the stock price during the life-time of the option. An example is
  $$f(S) = \max\left(\int_0^T S_t dt - K, 0\right).$$

- Further examples of exotic options are so-called ”barrier options”, for example ”down and out calls” with a payoff function
  $$f(S) = \max(S(T) - K_1, 0) \cdot 1_{\{\min_{0 \leq t \leq T} S(t) > K_2\}}.$$
Here $1_B$ is 1 if $B$ holds and 0 otherwise and $K_1, K_2$ again are pre-arranged constants. (See for example [34] or [37] for further examples of exotic options.)

- Let $S^{(1)}, \ldots, S^{(\tau)}$ be $\tau$ different stocks, then a simple example of a multi-asset option on $S^{(1)}, \ldots, S^{(\tau)}$ pays
  \[ f(S^{(1)}, \ldots, S^{(\tau)}) = \max\left( \frac{1}{\tau} \sum_{i=1}^{\tau} S^{(i)}_T - K, 0 \right). \]

The option pricing problem asks for the "fair price" or the "set of fair prices" of such options in time $t = 0$, i.e. for those prices of the contract which neither to the holder nor to the writer give the possibility to make risk-free gains against the other party.

To obtain exact fair prices (and not only estimates) for an option, it is necessary to assume a certain model for the evolution of the underlying stock price.

The most popular and an until now quite successful assumption is, that the stochastic process $S_t = (S^t)_{t \in [0,T]}$, describing the stock price behaviour follows a geometric Brownian motion, i.e. that $S_t$ follows a stochastic differential equation
\[ dS_t = S_t(\mu \cdot dt + \sigma \cdot dW_t) \]
where $\mu$ and $\sigma$, the trend and the volatility of $S$ are pre-estimated constants and \{ $W_t, 0 \leq t \leq T$ \} is a Wiener process. The solution of this stochastic differential equation is given by
\[ S_t = S_0 \cdot \exp\left( (\mu - \frac{\sigma^2}{2})t + \sigma W_t \right), \]
that means
\[ S_t = S_0 \cdot \exp\left( (\mu - \frac{\sigma^2}{2})t + \sigma \sqrt{t} \cdot z \right), \]
where $S_0$ is the stock price at time 0 and $z$ is a standard normal random variable.

In recent years it turned out that other models (say for example equipped with a stochastic volatility $\sigma(t)$ instead of $\sigma$, or for example "hyperbolic models" which are able to model the characteristic "fat tails" of stock prices more realistic, ...) have to be taken into account. Further, as already mentioned, the security underlying an option do not necessarily have to be stocks, but can for example be interest rates or currencies where the geometric Brownian motion is not at all an adequate model.

(For a discussion of interest rate models see for example [29] or [37].)

It is the fundamental theorem of modern financial mathematics (see for example [37], [12]) and simultaneously it is the basic tool that MC- and QMC-methods can be brought into play, that the following holds:

No matter which model is assumed for the underlying asset $S = (S_t)_{t \in [0,T]}$: if there exists a unique measure $P$ such that the process $\tilde{S} := (e^{-rt} \cdot S_t)_{t \in [0,T]}$ is a martingale under $P$, (here $r$ is the assumed constant risk free interest rate for the time-interval $[0,T]$), then there exists a unique fair price for any European option over $S$ with non-negative payoff function $f$ and this price is given by
\[ E_P(e^{-rT} \cdot f(S)) = \int_{\Omega} e^{-rT} \cdot f(S) dP(S) \]
where $\Omega$ is the probability space underlying $S$. (We say that we are in a complete market.)
If there exists more than one martingale measure \( P \) (incomplete market) then there is an interval of possible fair prices of the option in question. This interval is given by
\[
\{ E_P(e^{-rT} \cdot f(S)) | P \text{ martingale measure for } S \}.
\]
In some of these (incomplete) cases there is a natural choice of a specific one of the possible martingale measures (for example entropy-minimising or variance-minimising martingale measures or the martingale measure determined via the Es scheher transform, see [21] and [22]) to obtain again a unique reasonable option price.

These expected values in most cases cannot be evaluated explicitly and so they may be attacked by MC- or QMC-methods. We will give some concrete examples for this in the following section.

If the option is of American type, then the price is given by
\[
\max_{\Delta \leq T} E_P[e^{-r\Delta} \cdot f(\Delta, S)],
\]
where the maximum here is not extended over simple spot times \( \Delta \) but over stopping times \( \Delta \) in the sense of the theory of stochastic processes. We cannot go into these details concerning the valuation of American options by MC- and QMC-methods here. We just mention that until quite recently it was believed that MC- and QMC-simulations can be used only for European-style derivatives. But new developments show that to some extent pricing of American options by simulation is also well possible. The paper [8] gives a good survey on the current state of the art and on relevant further literature.

5 Some concrete examples of derivative pricing with MC and QMC

Here we report on five concrete pricing examples, all treated with both MC- and QMC-methods and we compare these methods. The first three examples already can be found in the literature. Partly the fourth and the fifth example are quite recent experiments carried out by the authors of this article and by their collaborators. In fact the fifth example still is work in progress and we just present preliminary remarks here.

For the sake of uniformity in the following we use a sometimes unusual notation. With \( \tau \) we always denote the dimension of the QMC-simulation problem and with \( N \) the number of sample paths.

a) Simulating the Black-Scholes model

This example is rather simple and not really a pricing example, but it is simply understood, enlightens the method, is the basis for the following pricing examples, and shows the power of QMC-methods quite directly. The example was given by Tezuka in [70]:

He considers a discrete version of the Black-Scholes model
\[
\log S_t - \log S_{t-1} = \mu \cdot dt + \sigma \cdot dW; \ t = 1, \ldots, \tau
\]
here \( \mu = r \) is chosen (see also Example b)), where \( r \) is a fixed risk free return, \( \sigma \) a fixed volatility, \( dt = \frac{1}{365} \) (one day), \( dW \) is a normal random variable with mean zero and variance \( dt \).
For a given current price $S_0$ of the stock, the price $S_T$ at time $T = \tau \cdot dt$ is easily derived by

$$S_T(z_1, \ldots, z_\tau) = S_0 \cdot e^{(r - \frac{\sigma^2}{2})\tau dt + \sigma \sqrt{\tau} \sum_{i=1}^{\tau} z_i},$$

where $z_i; i = 1, 2, \ldots, \tau$ are independent random variables from the standard normal distribution.

The expected value of $S_T$ hence is

$$E(S_T) = S_0 \cdot e^{(r - \frac{\sigma^2}{2})\tau + \frac{\sigma^2}{2} \tau}.$$

It is tried to approximate this expected value by MC- and by QMC-simulation. The approximation is carried out for the concrete values $S_0 = 100$, $r = 0.05$, $\sigma = 0.3$ and $\tau = 100$.

A single simulation experiment consists in constructing $N$ sample paths, i.e. by $N$ times generating $\tau = 100$ pseudo-random numbers when applying MC, and by generating $N$ points, well distributed in $[0, 1]^{100}$ when applying QMC.

Pseudo-random numbers were generated by the generator ComTaus (see [71]) and a generalised version of the digital Faure $(0, m, s)$-net, introduced in Section 3 was used as QMC point set.

To transform the uniformly distributed pseudo-random (PR) and quasi-random (QR) point sets to normally distributed point sets an algorithm of Moro [42] is used.

For any of the $N$ strings of (transformed) PR-numbers, respectively for any of the $N$ QR-points the value

$$S_T(z_1, \ldots, z_\tau)$$

is calculated and the average value over the $N$ results is taken. Figure 5 shows the speed of convergence of the approximations (MC and QMC) to the exact value of 101.379289...

![Figure 5: convergence of MC (continuous) and QMC (dotted) towards the price of a share in the Black-Scholes model](image)

b) Pricing discrete down-and-out call barrier options
This example is carried out by Acworth, Broadie and Glasserman in [1]. The payoff function $f$ is given by (see section 4)

$$f(S) = \max(S_T - K, 0) \cdot 1_{\{\min_{1 \leq i \leq \tau} S_i > H\}}.$$ 

Here $K$ and $H$ are constants and $t_1, \ldots, t_\tau$ are points in $[0, T]$, given by $t_i = \frac{i}{\tau} T$; $i = 0, 1, \ldots, \tau$.

For $S$ again a discrete geometric Brownian motion is assumed. This means

$$S_{t_i} = S_0 \cdot e^{\left(\mu - \frac{\sigma^2}{2}\right) t_i + \sigma \sqrt{t_i} \sum_{i=1}^{t_i} z_i},$$

where $\mu$ is a fixed trend, $\sigma$ is a fixed volatility and $z_i : i = 1, \ldots, \tau$ are independent standard normally distributed random variables.

So here we have to calculate

$$e^{-rT} \cdot E_Q(f(S))$$

with respect to the equivalent martingale measure $Q$, which here simply means that for simulation purposes we have to replace $\mu$ in the stock price model by $r$.

Then like in example a) the path for the stock price is simulated, once by a string of $\tau$ PR-numbers (RAN1 generator [61]), once by a $\tau$-dimensional QR-point set (here a generalised version of the digital $(t, m, s)$-nets in base b, introduced in Section 3 ("GF points" in the table) is used).

This is carried out $N$ times and the averages over the resulting values for

$$e^{-rT} \cdot f(S)$$

are calculated as approximations to the true values.

The experiment was carried out by Acworth, Broadie and Glasserman for every choice for $\tau$ and $N$ for altogether 250 different choices for the parameters $r, \sigma, H, K$. The given values are the average relative errors for these 250 instances. (In the original paper the authors also carried out tests for MC with antithetic variates, and for QMC with using a Brownian Bridge construction. For details see [1].) The "exact" values used for comparing the results are provided by exhaustive numerics (see [9]). Examples for results are given in Table 2.

c) Pricing discrete Asian options

This example is carried out by Joy, Boyle and Tan in [31]. The payoff function in this example is given by

$$f(S) = \max \left( \frac{1}{\tau} \sum_{i=1}^{\tau} S_i - K, 0 \right),$$

$$t_i = i \cdot \frac{1}{52} \ (i\text{-th week}) \ \text{for} \ i = 1, \ldots, \tau. \ K \text{ is a constant strike price.}$$

Again the geometric Brownian motion is assumed to model the stock price (with $\mu = r$).

We give here the results for MC and QMC (GF-point sets, see example b)) for two choices for $\tau$:

in both examples we have $S_0 = 100$, $r = 0.09$, $\sigma = 0.5$, $K = 100$. In the first example we have $\tau = 5$, in the second example $\tau = 52$ (i.e. $T$ is 5 weeks, respectively one year). $N$ again is the number of samples over which the average is taken.
Table 2: Barrier Options, relative errors

<table>
<thead>
<tr>
<th></th>
<th>MC</th>
<th>GFauRe</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau = 10$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N =$ 1,250</td>
<td>5.09</td>
<td>1.32</td>
</tr>
<tr>
<td>5,000</td>
<td>2.28</td>
<td>0.62</td>
</tr>
<tr>
<td>20,000</td>
<td>1.29</td>
<td>0.54</td>
</tr>
<tr>
<td>80,000</td>
<td>0.73</td>
<td>0.51</td>
</tr>
<tr>
<td>$\tau = 50$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N =$ 1,250</td>
<td>4.73</td>
<td>2.72</td>
</tr>
<tr>
<td>5,000</td>
<td>2.59</td>
<td>0.98</td>
</tr>
<tr>
<td>20,000</td>
<td>1.19</td>
<td>0.40</td>
</tr>
<tr>
<td>80,000</td>
<td>0.61</td>
<td>0.17</td>
</tr>
<tr>
<td>$\tau = 100$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N =$ 1,250</td>
<td>4.61</td>
<td>3.25</td>
</tr>
<tr>
<td>5,000</td>
<td>2.57</td>
<td>1.74</td>
</tr>
<tr>
<td>20,000</td>
<td>1.32</td>
<td>0.65</td>
</tr>
<tr>
<td>80,000</td>
<td>0.65</td>
<td>0.41</td>
</tr>
</tbody>
</table>

**d) Pricing multi-asset options**

Here we combine results obtained in [1] and in [72]. The option is a geometric mean European option, i.e. we consider $\tau$ stocks $S^1, \ldots, S^\tau$ and an option with payoff

$$f(S^1, \ldots, S^\tau) = \max \left( \left( \prod_{i=1}^{\tau} S_T \right)^{1/\tau} - K, 0 \right).$$

For such options an exact formula for the option price is known [5], no matter if the underlying assets are correlated or not. Now the number of assets determines the dimension $\tau$ of the problem.

For simulating $S^1_T, \ldots, S^\tau_T$ we generate $\tau$ PR-numbers or a $\tau$-dimensional QR-point. The PR-numbers, respectively the coordinates of the QR-point are transformed (by using the algorithm of Moro [42], respectively by using the Box-Muller method [13]) to standard normal random variates. We write the result as a vector

$$\begin{pmatrix} z_1 \\ \vdots \\ z_\tau \end{pmatrix}$$

If $S^1, \ldots, S^\tau$ are assumed to be correlated via a covariance matrix $\Sigma$, then we construct the Cholesky decomposition (see [13])

$$A \cdot A' = \Sigma$$

of $\Sigma$ and set

$$\begin{pmatrix} w^1 \\ \vdots \\ w^\tau \end{pmatrix} = A \cdot \begin{pmatrix} z_1 \\ \vdots \\ z_\tau \end{pmatrix}$$

Finally

$$S^i_T = S^i_0 \cdot e^{\left( r - \frac{\sigma^2}{2} \right) T + \sigma \sqrt{T} \cdot w^i}$$
Table 3: Asian Option, 5 Weeks

for \( i = 1, \ldots, \tau \).

For the generation of PR-numbers in [72] the Unix implemented drand48() function was used. As QR-point sets the \((t, m, s)\)-nets in base 2, introduced in Section 3 were used.

We give two examples for the test results: the first one is given in [72]. Here the following parameters were used: \( \tau = 4, S_0^i = 42, \sigma^i = 0.2 \) for \( i = 1, 2, 3, 4 \), \( r = 0.1 \), \( K = 41 \), \( T = 0.5 \) and for the correlation coefficients

\[
\rho_{12} = 0.5, \quad \rho_{13} = 0.4, \quad \rho_{14} = 0.1, \quad \rho_{23} = 0.7, \quad \rho_{24} = 0.1, \quad \rho_{34} = 0.2.
\]

Figure 6 shows the resulting relative errors

![Figure 6: convergence of MC (continuous) and QMC (dotted) towards the value of a multi-asset option](image)

The second example is given in [1]. Here uncorrelated assets and the parameters \( \tau = 50, S_0^i = 100, T = 1 \), were used. The given relative errors (Table
5) are the average relative errors over experiments with 250 different choices for the parameters \( r, K, \sigma \).

<table>
<thead>
<tr>
<th>( \tau = 52 )</th>
<th>Crude</th>
<th>Quasi</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10.206</td>
<td>17.043</td>
</tr>
<tr>
<td>200</td>
<td>15.405</td>
<td>14.219</td>
</tr>
<tr>
<td>400</td>
<td>11.818</td>
<td>13.327</td>
</tr>
<tr>
<td>800</td>
<td>12.593</td>
<td>14.991</td>
</tr>
<tr>
<td>1,600</td>
<td>13.206</td>
<td>12.928</td>
</tr>
<tr>
<td>3,200</td>
<td>12.934</td>
<td>12.801</td>
</tr>
<tr>
<td>4,800</td>
<td>12.965</td>
<td>12.673</td>
</tr>
<tr>
<td>9,600</td>
<td>12.684</td>
<td>12.804</td>
</tr>
<tr>
<td>19,200</td>
<td>12.754</td>
<td>12.993</td>
</tr>
<tr>
<td>38,400</td>
<td>13.033</td>
<td>12.949</td>
</tr>
<tr>
<td>76,800</td>
<td>13.039</td>
<td>12.948</td>
</tr>
<tr>
<td>153,600</td>
<td>12.948</td>
<td>12.962</td>
</tr>
<tr>
<td>307,200</td>
<td>12.914</td>
<td>12.963</td>
</tr>
<tr>
<td>614,400</td>
<td>12.922</td>
<td>12.965</td>
</tr>
<tr>
<td>1,228,800</td>
<td>12.885</td>
<td>12.963</td>
</tr>
<tr>
<td>2,457,600</td>
<td>12.906</td>
<td>12.962</td>
</tr>
</tbody>
</table>

Table 4: Asian Option, 52 weeks

<table>
<thead>
<tr>
<th>( \tau = 50 )</th>
<th>MC</th>
<th>Sobol</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N = )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,250</td>
<td>3.44</td>
<td>3.55</td>
</tr>
<tr>
<td>5,000</td>
<td>1.38</td>
<td>0.50</td>
</tr>
<tr>
<td>20,000</td>
<td>0.80</td>
<td>0.18</td>
</tr>
<tr>
<td>80,000</td>
<td>0.43</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 5: Multi-asset options, relative errors

e) Pricing Asian options over a hyperbolic model

We do not have the place to go into the details of hyperbolic models in this paper. These models were introduced in [16] (see also [32]). It turned out that these models in some aspects provide a better instrument to model stock prices. (See for example figure 7)

However technically it is much more difficult to work with this hyperbolic model (which is based on Levy-processes) than with the usual Brownian motion. Moreover this model is not complete, i.e. option prices are not unique.

One approach to the pricing of options in this model is via Esscher transforms (see [22]) and MC- or QMC-simulation. This and a comparison with the values obtained over the geometric Brownian motion is carried out in [40] (see also [41]), a work which is still in progress.
In recent years the quantification of risk exposure has become one of the major tasks for financial institutions, almost equally important as the pricing of derivatives. Though the mathematical modelling of markets is essentially the same, the requirements are quite different. While when pricing derivatives we want to find the fair price of one contract as accurately as possible, we are content with a rather rough measure of risk, but for a portfolio consisting of hundreds of contracts. But the main difference (at least in this paper) is that the measurement of risk is statistical work, that is we use the physical measure rather than some equivalent martingale measure.

In the following we give a brief introduction to the most important risk measure, that is Value at Risk, and on how to attack it by MC and QMC methods.

Suppose we hold a portfolio consisting of several contracts such as bonds, shares and derivatives (We are focusing on so called market risk. See Jorion [30] for different types of risks). Now, at time 0, the worth \(a\) of the portfolio is known (or can be accurately estimated), whereas the worth at a future date \(T\) is uncertain. This uncertainty we express by modelling this future worth by a random variable \(X\). Now the discounted profit (or loss) of the portfolio over the horizon \([0, T]\) is the random variable \(X/(1 + r) - a\), where \(r\) is the risk-less return in the period \([0, T]\). Most often in practice \(T\) equals 1 to 10 days, such that we will take \(r = 0\). However, the generalisation to nonzero interest rates is straightforward. The distribution of \(X - a\) is called the profit/loss-distribution (P/L-distribution) of the portfolio.

The most widely used risk measure is Value at Risk (VaR). By the VaR over the target horizon \([0, T]\) at confidence level \(c\) we mean the worst loss over the target horizon which occurs with probability \(c\). With other words: With a probability \(c\) our loss over the target horizon is less or equal to VaR. Typical values for \(c\) are 95\% or 99\%. In mathematical terms this means that VaR is essentially the negative of the \((1 - c)\)-quantile of the P/L-distribution.

So we define VaR to be the number

\[
\text{VaR}_c(X) := -\inf\{x : P[X-a \leq x] > \alpha\}
\]

where \(\alpha = 1 - c\). Figure 8 shows the graphical interpretation for VaR for the case that \(X\) has a continuous density.

The book of Jorion [30] can be seen as the standard reference concerning VaR.

Despite of its popularity, VaR has at least two deficiencies: The first is that it gives us our worst losses if things are not going “bad”, but it does not tell us “how
bad” things can be if they are. This problem can be solved by considering the so called Tailed Conditional Expectation, TCE, at level alpha:

\[ \text{TCE}_\alpha(X) := -E[X|X < -\text{VaR}_\alpha]. \]

Here \( E[X|A] \) denotes the conditional expectation of \( X \) given the event \( A \), i.e.

\[ E[X|A] = \frac{1}{P[A]} \int_A XdP. \]

The second point is that VaR is not coherent, as defined in [2], since VaR fails to have the sub-additivity property, which essentially states that the risk of a portfolio should be smaller than the sum of the risks of the sub-portfolios. This problem can be overcome by the so called Worst Conditional Expectation, also defined in [2] by

\[ \text{WCE}_\alpha(X) := -\inf\{E[X|A] : P[A] < \alpha\}. \]

The three risk measures have in common that they can be obtained from the profit/loss-distribution of our portfolio. So in either case it seems to be a good thing to try to estimate the P/L-distribution and then to read off the risk measure best meeting one’s flavour (actually there is hardly a way around estimating the P/L-distribution when calculating VaR).

If, for instance, one assumes that the market factors (such as stock returns or interest rates), that influence our portfolio have joint normal distribution and that our portfolio depends linearly on each factor, then the P/L-distribution is itself a normal distribution. Then it “only” remains to estimate the parameters of the distributions, i.e. the corresponding means, variances and covariances from historical data and to compute the parameters, mean and variance, of the P/L-distribution. However we do not want to rely on these assumptions, since they almost never hold in reality. As already mentioned before, stock returns for example are not normally distributed, they admit much heavier tails than the normal distribution, which means that extreme events are more likely in practice. But these extreme events are exactly what we are interested in and the losses due to which we want to control in risk management. Moreover many contracts do not depend linearly on the underlying market factors. A brief glance at the Black-Scholes formula reveals that the value of an European call option is not linear with respect to the underlying stock price, therefore the returns of an option cannot be normally distributed.
even if the returns of the underlying stock are. But derivatives are one particular
generator of large losses (or gains) due to their leverage effect.

Since analytical computation of the P/L-distribution without the assumptions
of normality and linearity is virtually impossible, we are at the mercy of simulation
methods.

The estimation of the P/L-distribution by simulation methods is basically done
in four steps:

1. mark the whole portfolio to market, i.e. evaluate its current worth \( a \);
2. generate \( N \) possible future scenarios of market evolution;
3. evaluate the net worth of the portfolio at time \( T \) for each scenario and subtract
   \( a \). This results in \( N \) simulated values of the profit/loss \( x_1, \ldots, x_N \);
4. construct the empirical distribution function by defining
   \[
   \hat{F}(x) := \frac{k}{N},
   \]
   where \( k \) is the number of the \( x_i \) that are less than \( x \).

This simple recipe is the standard (Quasi-) Monte Carlo integration scheme. However it should be emphasised that steps 2 and 3 can be quite delicate and often
require the use of appropriate simulation by themselves, for example if the portfolio
contains exotic options.

The generation of scenarios

A scenario is one realization of the possible states of market factors. Market
factors can be stock prices, interest rates, (implied) volatilities and so on.

Basically there are two ways of generating the scenarios

- historical simulation
- parametric simulation

In historical simulation we take our scenarios from the past evolution of the
market. We illustrate this by an example: Suppose that we hold a portfolio of
\( M \) shares and we want to estimate its P/L distribution over the target horizon
of one day. Suppose further that we have recorded the daily price movement
of each share over the last \( N \) days. Then we can take the joint daily (relative)
movements of the prices as \( N \) possible future scenarios.

Historical simulation has the advantage that we are not obliged to make any
assumptions about the distribution of market factors. Provided that the dis-
tribution of the changes of market factors is stationary, i.e. does not change
over time, then we can expect quite accurate results if \( N \) is made large. But
here we already see the disadvantage of this method: The larger \( N \) is taken,
the further we have to go back in history and the more likely it is that the
assumption of stationarity is corrupted. If, for example, we want to take
\( N = 1000 \) scenarios of daily movements then this corresponds to records over
about 4 years. But this is rather a long time in todays financial markets
and a number of financial instruments does not even exist for such a long
period. Another weakness of historical simulation is that one cannot test for
the sensitivity of VaR to market parameters.

In parametric simulation we try to fit certain probability distributions, like
normal-, parametric distributions (see [62]) or hyperbolic distribution to the
historical data (see [15], [7]).
If we assume, for example, that the market factors are jointly normally distributed we can use the Cholesky decomposition technique described earlier when we considered multi asset options. Still the contracts in our portfolio are not required to depend linearly on the market factors.

It is in this case that Papageorgiou and Paskov [57] have found that deterministic simulation can be superior to the classical Monte Carlo approach. They considered two examples: The first portfolio consisted of 34 at-the-money equity and currency European call options, which gives a 34-dimensional problem. The second example considered collateralised mortgage obligations from a pool of 30 year mortgages with monthly payments, a 360-dimensional problem. They find that in these examples the estimation error of the classical methods is about ten times higher than that of the low discrepancy method.

If we want to allow more general distributions the problem becomes more subtle. It is not difficult to achieve that the market factors have the required distribution, besides estimation problems of the parameters. The main problem is to create sample points with a predefined correlation structure. One method, that solves this problem for Monte Carlo methods quite well is so called latin hypercube sampling or more accurate Stein’s Algorithm. We will describe the basic technique. For a more thorough discussion see [64] and [68].

The basic idea is to generate a sample with prescribed correlation structure and arbitrary distribution, to transform it to the unit cube without changing the correlation structure too much and to apply the inverse of the cumulative distribution function to each coordinate. By the way we get a stratified sample, i.e. one that has no random accumulation points. We show what we mean by stratification for the most simple case of a 1-dimensional sample in figure 9. For the upper picture we take uniformly distributed points $e_i; i = 1, \ldots, N$, for the lower picture we take stratified points $\frac{i + 0.5}{N}; i = 1, \ldots, N$. Both times the points on the vertical axes have uniform distribution, after applying the inverse cumulative distribution function $F^{-1}$ we get the desired sample.

In higher dimensions we still divide each coordinate in $N$ cells, but we do not choose a random point in every sub-cube, instead we let a sample with desired correlation structure determine the proper sub-cubes.

Now we give Stein’s algorithm:

1. Generate a sample of $N \tau$-dimensional scenarios with prescribed correlation structure, for example by using the Cholesky decomposition technique;
2. arrange these samples in a $N \times \tau$-matrix, $Y = (Y_{ij}); i = 1, \ldots, N; j = 1, \ldots, \tau$, in which the element $Y_{ij}$ denotes the $j^{th}$ coordinate of the $i^{th}$ scenario;
3. define the auxiliary matrix $Z = (Z_{ij}),$ the columns of which are those of $Y$ sorted by increasing value;
4. define another matrix, $R = (R_{ij}),$ in which the the element $R_{ij}$ denotes the rank of the element $Y_{ij}$ within the $j^{th}$ column of $Y,$ i.e. the row-index of the element $Y_{ij}$ in $Z$;
5. define the matrix $X$ by

$$X_{ij} := F_k^{-1} \left( \frac{R_{ij} + e_{ij} - 1}{N} \right),$$

where the $e_{ij}$ are independent uniformly distributed (pseudo) random variables and $F_k$ is the inverse cumulative distribution function of the
$k^{th}$ market factor. Figure 10 visualises Stein’s Algorithm. We abbreviate the numbers $R_{ij} + e_{ij}^{-1}$ to $U_{ij}$.

We know of no implementation of Value at Risk using non-normal distributions and deterministic, i.e. Quasi-Monte Carlo simulation.

If we want to allow for more complex portfolios like dynamic portfolios or such that contain path-dependent derivatives, we have to take into account intermediate values, such that a particular scenario consists of an entire path. The generalisation to paths is straightforward and runs exactly parallel to the preceding sections.

**Evaluating the net worth**

We distinguish between two fundamental ways of evaluating the portfolio in a special scenario:

- full valuation
- parametric valuation

Full valuation means marking to market the whole portfolio for each single scenario. Quite often it may be the case that the portfolio consists of several hundred contracts, some of which may require separate simulation for their valuation. It is clear that the computational cost of such a procedure can easily become infeasible.

In contrast parametric valuation means replacing the profit/loss function (PLF) of the portfolio by a suitable approximation. This can be done by Taylor approximation or by interpolation. Since we are interested in big relative changes, we can expect that Taylor approximation will not be the proper
Final: Apply the inverse cumulative distribution function according to a sample having the desired correlation.

Choose a stratification according to a sample having the desired correlation.

Choose random points in the filled subcubes.

Figure 10: illustration of Stein’s Algorithm, $\tau = 2$, $N = 10$

choice, since it is unnecessarily accurate for small movements and not at all accurate for larger ones. See the discussion in [64].

The basic idea is to compute a profit/loss approximation function for sensible chosen sub-portfolios of the portfolio by performing the full valuation only for relatively few scenarios and interpolating in between. One possible choice for these scenarios is a regular grid. But the number of the scenarios then becomes very large when there are many market factors. Shaw [64] proposed artificial intelligence techniques to find efficient interpolating points. But maybe this point will be an application for low discrepancy point sets in the future.

Now the global profit/loss approximation function is obtained by adding up the profit/loss approximation function of the sub portfolios.

To summarise: the standard method for the estimation of Value at Risk (respectively the P/L-distribution) is Monte Carlo simulation. Of course one can always replace random points by low-discrepancy points to obtain a Quasi-Monte Carlo method, but this can be quite unsatisfactory. For example a stratification technique like included in Stein’s algorithm is unnecessary for low discrepancy sequences, since they are already stratified. We feel that many of the virtues of low discrepancy sequences can be of use in the field of risk management, yet it seems
that they have remained undiscovered yet.

Acknowledgements

We thank the two referees for many valuable hints.

References


