Course Notes on Computational Finance

2

Random Numbers

2. Computation of Random Numbers

Definition (sample from a distribution)

A sequence of numbers is called *sample* from a distribution function F, if the numbers are independent realizations of a random variable with distribution F.

Examples

If F is the uniform distribution on the interval $[0,1]$, then we call the samples from F *uniform deviates*. Notation: $\sim \mathcal{U}[0,1]$.

If F is the standard normal distribution, then we call the samples from F *standard normal deviates*. Notation: $\sim \mathcal{N}(0,1)$.

The basis of random number generation is to draw numbers $\sim \mathcal{U}[0,1]$.

2.1 Uniform Deviates

A. Linear Congruential Generators

Choose $a, b, M \in \mathbb{N}$, $a \neq 0$, $a, b < M$, and define for $N_0 \in \mathbb{N}$ (*"seed"*) a sequence of numbers by

Algorithm (linear congruential generator)

choose N_0 .
For $i = 1, 2, ...$ calculate
$N_i = (aN_{i-1} + b) \bmod M$

Define $U_i \in [0, 1)$ by

$$U_i = \frac{N_i}{M} .$$

The numbers U_i are used as uniform deviates.

Obvious **Properties**

(a) $N_i \in \{0, 1, \ldots, M-1\}$

(b) The sequence of N_i is periodic with a period $p \leq M$.
(*because* there are not $M+1$ distinct numbers N_i. Hence two out of $\{N_0, \ldots, N_M\}$ must be equal, $N_i = N_{i+p}$ with $p \leq M$. p-periodicity follows.)

Literature: [D. Knuth: The Art of Computer Programming, Volume 2]

The above numbers U_i are no real random numbers, but are deterministically defined and reproducible. We call such numbers *pseudo random*. In this chapter, we omit the modifier "pseudo" because it is clear from the context. The aim is to find parameters M, a, b such that the numbers U_i are good substitutes of real random numbers.

Example

$$M = 244944, \quad a = 1597, \quad b = 51749$$

Useful parameters a, b, M are in [Press et al.: Numerical Recipes].

Question: What are "good" random numbers?

A practical (and hypothetical) answer: The numbers should pass "all" tests.

First requirement: The period p must be large, hence M as large as possible. For example, in a binary computer with mantissa length l, one aims at $M \approx 2^l$. Suitable a, b can be derived with methods from number theory. [Knuth].

Second requirement: The numbers must be distributed as intended (density f, expectation μ, variance σ^2). Check this by **statistical tests** following these lines: First apply the algorithm to produce a large number of U_i-values. Then

(a) Calculate the mean $\hat{\mu}$ and the variance $\hat{s}^2$ of the sample. Check $\hat{\mu} \approx \mu$ and $\hat{s}^2 \approx \sigma^2$.

(b) Test for correlations of the U_i with previous U_{i-j}. For example, correlation could mean that small values of U are likely to be followed again by small values. In this case the generator would be of low quality.

(c) Estimate the density function $\hat{f}$ of the sample, and check for $\hat{f} \approx f$. A prototypical test is this: Divide the unit interval $[0, 1]$ into equidistant subintervals

$$k\Delta U \leq U < (k+1)\Delta U \,,$$

where ΔU denotes the length of the subintervals. (For other distributions choose an interval that contains all sample points U_i, and the subintervals will be defined accordingly.) When altogether j samples are calculated, let j_k be the number of samples that fall into the kth subinterval. The probability that the kth subinterval is hit is $\frac{j_k}{j}$. This should approximate

$$\int_{k\Delta U}^{(k+1)\Delta U} f(x)\,\mathrm{d}x \quad (f = 1 \text{ for the uniform distribution}) .$$

This integral is

$$\Delta U f(\bar{U}) ,$$

with $\bar{U}$ in the kth subinterval. Hence a good generator should satisfy

$$\Delta U \hat{f}(\bar{U}) = \frac{j_k}{j} \stackrel{!}{=} \Delta U f(\bar{U}) ,$$

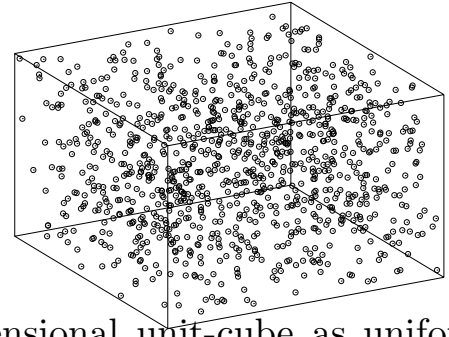
at least for small ΔU. Hence the empirical density on the kth subinterval is

$$\hat{f} = \frac{j_k}{j\Delta U} .$$

Third requirement:

The **lattice structure** should be OK.
To check this, arrange vectors out of m consecutive numbers:

$$(U_i, U_{i+1}, \ldots, U_{i+m-1})$$



For $U \sim \mathcal{U}[0,1]$, these points should fill the m-dimensional unit-cube as uniformly as possible. The sequences of points/vectors lie on $(m-1)$-dimensional hyperplanes. Trivial case: M parallel planes through $U = \frac{i}{M}, \quad i = 0, \ldots, M-1$ (any one of the m components).

A bad situation occurs when all points fall on only a *few* planes. Then the gaps between the planes without any points would be wide. This leads to analyze the lattice structure of the random points. The focus lies on the smallest number of planes, on which all points in $[0,1)^m$ "land."

Analysis for $m = 2$: In this planar case, the hyperplanes in (U_{i-1}, U_i)-space are straight lines $z_0 U_{i-1} + z_1 U_i = \lambda$, for parameters z_0, z_1, λ. From

$$\begin{aligned} N_i &= (aN_{i-1} + b) \bmod M \\ &= aN_{i-1} + b - kM \quad \text{for } \; kM \le aN_{i-1} + b < (k+1)M \end{aligned}$$

conclude for arbitrary numbers z_0, z_1

$$\begin{aligned} z_0 N_{i-1} + z_1 N_i &= z_0 N_{i-1} + z_1(aN_{i-1} + b - kM) \\ &= N_{i-1}(z_0 + az_1) + z_1 b - z_1 kM \\ &= M \underbrace{\left(N_{i-1}\frac{z_0 + az_1}{M} - z_1 k\right)}_{=:c=c(i)} + z_1 b \end{aligned}$$

Dividing by M leads to

$$z_0 U_{i-1} + z_1 U_i = c + z_1 b M^{-1}\,,$$

a straight line in the (U_{i-1}, U_i)-plane. For fixed z_0, z_1 this defines a family of parallel lines/"planes," parameterized by c.

Question: Is there a family of such lines (planes) defined by a pair (z_0, z_1), such that only few lines (planes) cut the unit-cube? (The minimal number of parallel hyperplanes holding all points is the worst case.)

For analyzing the number of planes, the cardinality of the c's matters. To find the worst case with a small set of c's, assume $z_1, z_0 \in \mathbb{Z}$ and $z_0 + az_1 \bmod M = 0$. Then $c \in \mathbb{Z}$, and

$$c = z_0 U_{i-1} + z_1 U_i - z_1 b M^{-1} \in \mathbb{Z}\,.$$

($z_1 b M^{-1}$ is a constant parallel shift not affecting the number of planes.) How many of such c's exist? For $0 \le U < 1$ obtain a range for the c's by a maximal set $I_c \subset \mathbb{Z}$, such that

$$c \in I_c \quad \Rightarrow \quad \text{the line touches or cuts the unit-cube .}$$

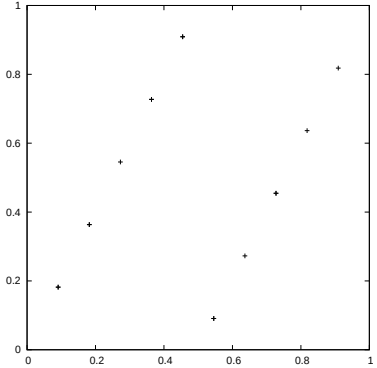
The cardinality of the set I_c gives a clue on the distance between the parallel lines (planes). It is unfavorable when the set consists of only a few elements.

Academic Example $N_i = 2N_{i-1} \mod 11$ (i.e. $a = 2,\ b = 0,\ M = 11$)

The pair $(z_0, z_1) = (-2, 1)$ solves $z_0 + a z_1 = 0 \mod M$.

Hence $-2U_{i-1} + U_i = c$.

$0 \le U < 1$ implies $-2 < c < 1$.
In view of $c \in \mathbb{Z}$, the only parameters are $c = -1$ and $c = 0$. For this choice of (z_0, z_1) all 10 points in $[0, 1)^2$ fall on only two straight lines.



(0 does not occur for $N_0 \ne 11k$, $k \in \mathbb{Z}$.)

(U_{i-1}, U_i)-plane

Example $N_i = (1229 N_{i-1} + 1) \bmod 2048$

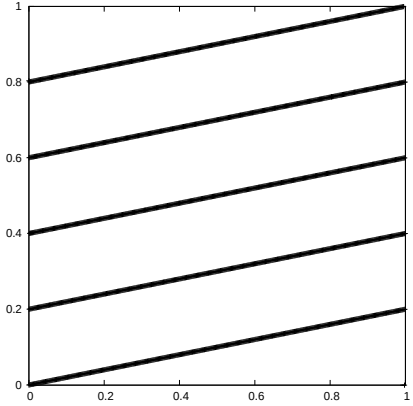
The condition $z_0 + a z_1 = 0 \bmod M$

$$\frac{z_0 + 1229 z_1}{2048} \in \mathbb{Z}$$

is satisfied by $z_0 = -1,\ z_1 = 5$, because

$$-1 + 1229 \cdot 5 = 6144 = 3 \cdot 2048\,.$$

$c = -U_{i-1} + 5U_i - \frac{5}{2048}$ implies $-1 - \frac{5}{2048} < c < 5 - \frac{5}{2048}$. Hence the c's consist of only six values, $c \in \{-1, 0, 1, 2, 3, 4\}$, and all points in $[0,1)^2$ fall on six straight lines. — The U_i-distance between two neighboring lines is $\frac{1}{z_1} = \frac{1}{5}$.



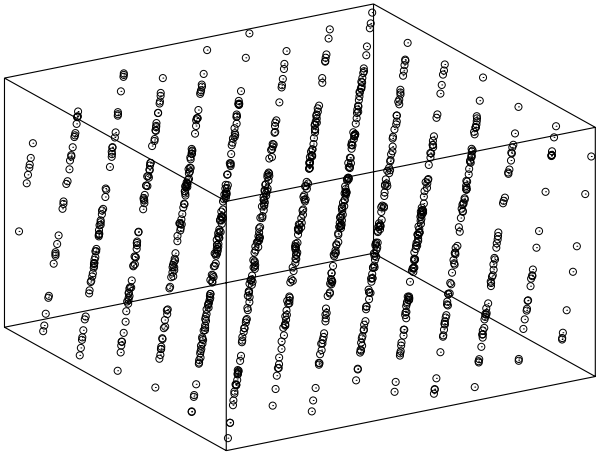
(U_{i-1}, U_i)-plane.
In this figure,
the discrete points
are not separated.
The sixth line
consists of one point.

The (U_{i-1}, U_i)-points of these two examples are obviously not equidistributed. The next example is fine for $m = 2$ but shows that the good distribution for $m = 2$ does not carry over to higher m.

Example (RANDU)

$N_i = aN_{i-1} \bmod M$, with $a = 2^{16} + 3$, $M = 2^{31}$

For $m = 2$ experiments show that the dots (U_{i-1}, U_i) are nicely equidistributed in the unit square. For $m = 3$ it turns out that the random points in the cube $[0, 1)^3$ fall on only 15 planes.



Analysis for larger m is analogous. ($\longrightarrow$ Topic 14)

B. Fibonacci Generators

There are other classes of random-number generators, for example, the **Fibonacci generators**. A prototype of such generators is defined by

$$N_{i+1} := N_{i-\nu} - N_{i-\mu} \bmod M$$

for suitable μ, ν (also with "+" or with more terms). Literature: [Knuth]

Example

$$U_i := U_{i-17} - U_{i-5},$$
$$\text{in case } U_i < 0 \text{ set } U_i := U_i + 1.0$$

This is a simple example with reasonable features, but there are correlations.

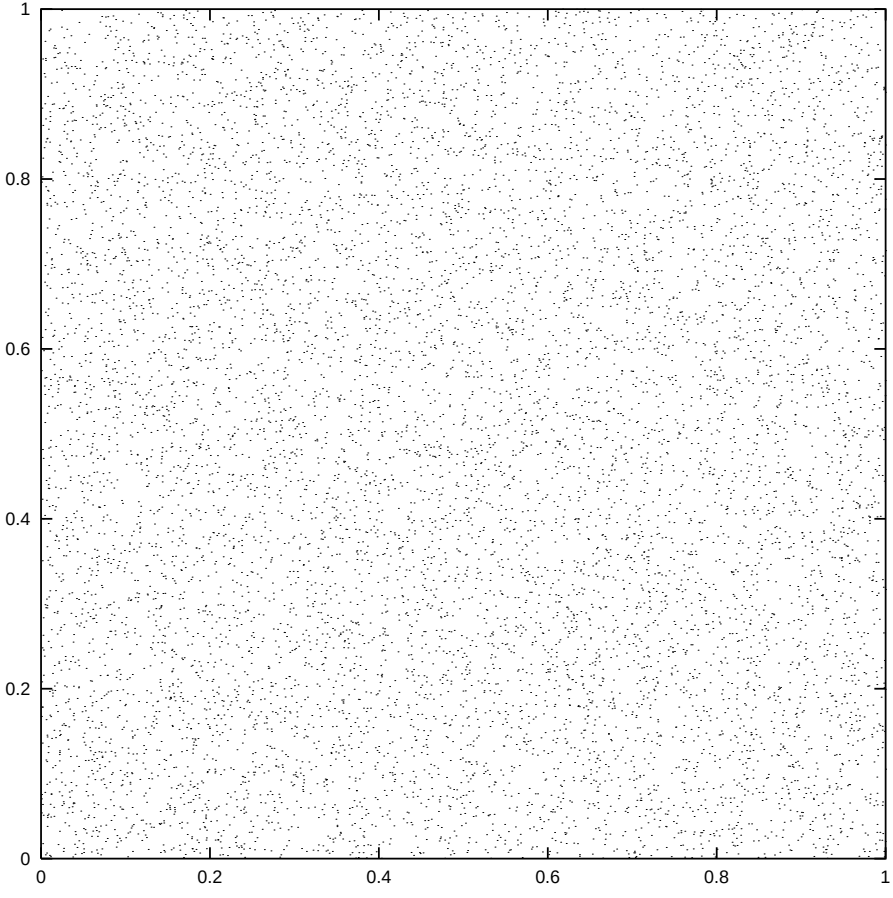
The algorithm has a "leg" with length 17. This requires an initial phase that provides 17 U-values to start the algorithm.

Algorithm (loop of a simple Fibonacci generator)

Repeat: $\zeta := U_i - U_j$
in case $\zeta < 0$, set $\zeta := \zeta + 1$
$U_i := \zeta$
$i := i - 1\,, \quad j := j - 1$
in case $i = 0$, set $i := 17$
in case $j = 0$, set $j := 17$

Initialization: Set $i = 17,\ j = 5$, and calculate $U_1, \ldots, U_{17}$ with a congruential generator with, for example, $M = 714025,\ a = 1366,\ b = 150889$.

A professional generator to calculate uniform random numbers is the "Mersenne Twister" by Matsumoto, Nishimura, in: ACM Transactions on Modelling and Computer Simulations **8** (1998), p.3-30. This generator has excellent features with a huge period, and enables equidistributed points even in high dimensions m.



10000 random numbers (U_{i-1}, U_i), calculated with a Fibonacci Generator

2.2 Random Numbers from Other Distributions

The generation of all kind of deviates is based on uniform deviates. For the calculation of random numbers from a given distribution we can apply several methods, namely, inversion, transformations, and rejection methods.

A. Inversion

Let $F(x) := \mathsf{P}(X \le x)$ be a distribution function, for a random variable X, and P is the corresponding probability.

Theorem (inversion)

Suppose $U \sim \mathcal{U}[0,1]$ and let F be a continuous strictly increasing distribution function. Then $X := F^{-1}(U)$ is a sample from F.

Proof:

$U \sim \mathcal{U}[0,1]$ means $\mathsf{P}(U \le \xi) = \xi$ for $0 \le \xi \le 1$. Hence

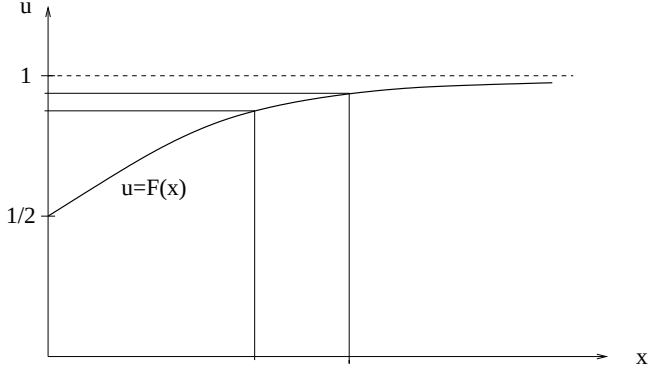
$$\mathsf{P}(F^{-1}(U) \le x) = \mathsf{P}(U \le F(x)) = F(x).$$

Application

Calculate $u \sim \mathcal{U}[0,1]$ and evaluate $F^{-1}(u)$. These numbers have the desired distribution. Mostly inversion is done numerically because F^{-1} in general is not known analytically.

There are two variants:

(a) $F(x) = u$ is a nonlinear equation for x, which can be solved iteratively with standard methods of numerical analysis (e.g. Newton method). For the normal distribution (Figure) the iteration requires tricky termination criteria, because for $u \approx 0,\ u \approx 1$ small perturbations in u lead to large perturbations in x.



(b) Construct an approximating function G such that $G(u) \approx F^{-1}(u)$. Then only $x = G(u)$ needs to be evaluated. The construction of G must observe the asymptotic behavior, which amounts to the poles of G. For the standard normal

distribution the symmetry w.r.t. $(x, u) = (0, \frac{1}{2})$ can be exploited and only the pole for $u = 1$ needs to be observed. This can be done with a rational function $G(u)$, with a denominator having a zero at $u = 1$.

B. Transformation

We begin with the scalar case: Let X be a random variable. What is the distribution of a transformed $h(X)$?

Theorem (scalar transformation)

Suppose X is a random variable with density function f and distribution function F. Further assume

$$h : S \to B$$

with $S, B \subseteq \mathbb{R}$, where S is the support of f, and let h be strictly monotonic.

(a) $Y := h(X)$ is random variable with distribution function

$$\begin{aligned} &F(h^{-1}(y)) \quad \text{for increasing } h \\ &1 - F(h^{-1}(y)) \quad \text{for decreasing } h \end{aligned}$$

(b) If h^{-1} is absolutely continuous, then for almost all y the density of $h(X)$ is

$$f(h^{-1}(y)) \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| .$$

Proof: (write also F^X for F)

(a) $F^Y(y) := \mathsf{P}(h(X) \le y) =$

(in case h is increasing:)

$= \mathsf{P}(X \le h^{-1}(y)) = F^X(h^{-1}(y))$

(in case h is decreasing:)

$= \mathsf{P}(X \ge h^{-1}(y)) = 1 - \mathsf{P}(X < h^{-1}(y)) = 1 - F^X(h^{-1}(y))$

(b) For absolutely continuous h^{-1} the density of $Y = h(X)$ is equal to the derivative of the distribution function almost everywhere. Evaluation of $\frac{\mathrm{d}F(h^{-1}(y))}{\mathrm{d}y}$ with the chain rule implies the assertion; distinguish between increasing and decreasing h.

Application

Start with $X \sim \mathcal{U}[0,1]$ and the density of the uniform distribution,

$$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

i.e. $S = [0,1]$. Random numbers Y with prescribed target density $g(y)$ are to be calculated. Hence we require a transformation h such that

$$f(h^{-1}(y)) \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| = 1 \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| \stackrel{!}{=} g(y)\,.$$

Then $h(X)$ is distributed as intended.

Example (exponential distribution)

The exponential distribution with parameter $\lambda > 0$ has the density

$$g(y) = \begin{cases} \lambda \mathrm{e}^{-\lambda y} & \text{for } y \ge 0 \\ 0 & \text{for } y < 0. \end{cases}$$

B consists of the non-negative real numbers. As transformation $[0,1] \to B$ we choose the monotonic decreasing function

$$y = h(x) := -\frac{1}{\lambda} \log x$$

with inverse $h^{-1}(y) = \mathrm{e}^{-\lambda y}$ for $y \geq 0$. Since

$$f(h^{-1}(y)) \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| = 1 \cdot \left| (-\lambda)\mathrm{e}^{-\lambda y} \right| = \lambda \mathrm{e}^{-\lambda y} \ = g(y)\,,$$

$h(X)$ is distributed exponentially as long as $X \sim \mathcal{U}[0,1]$.

Application

Calculate $U_1, U_2, \ldots \sim \mathcal{U}[0,1]$. Then

$$-\frac{1}{\lambda} \log(U_1), \quad -\frac{1}{\lambda} \log(U_2), \quad \ldots \text{ are distributed exponentially.}$$

(Hint: The distances between jump times of Poisson processes are distributed exponentially.)

Attempt with the normal distribution: Search for h such that

$$1 \cdot \left| \frac{\mathrm{d}h^{-1}(y)}{\mathrm{d}y} \right| = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} y^2 \right).$$

This is a differential equation for h^{-1} without analytic solution. In this situation the multidimensional version of the transformation helps.

Theorem (transformation in $\mathbb{R}^n$)

Suppose X is a random variable in $\mathbb{R}^n$ with density $f(x) > 0$ on the support S. Let the transformation $h : S \to B$, $S, B \subseteq \mathbb{R}^n$ be invertible and the inverse continuously differentiable on B. Then $Y := h(X)$ has the density

$$f(h^{-1}(y)) \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right|, \quad y \in B, \tag{2.7}$$

where $\frac{\partial(x_1,\ldots,x_n)}{\partial(y_1,\ldots,y_n)}$ denotes the determinant of the Jacobian matrix of $h^{-1}(y)$.

Proof: see Theorem 4.2 in [L. Devroye: Non-Uniform Random Variate Generation (1986)]

In Section 2.3 the two-dimensional version will be applied to calculate normal variates.

C. Acceptance-Rejection Method

This method is based on the following facts: Let f be a density function on $S \subset \mathbb{R}$ and $\mathcal{A}_f$ the area between the x-axis and the graph of f. Assume two random variables U and X independent of each other with $U \sim \mathcal{U}[0,1]$ and X distributed with density f. Then the points

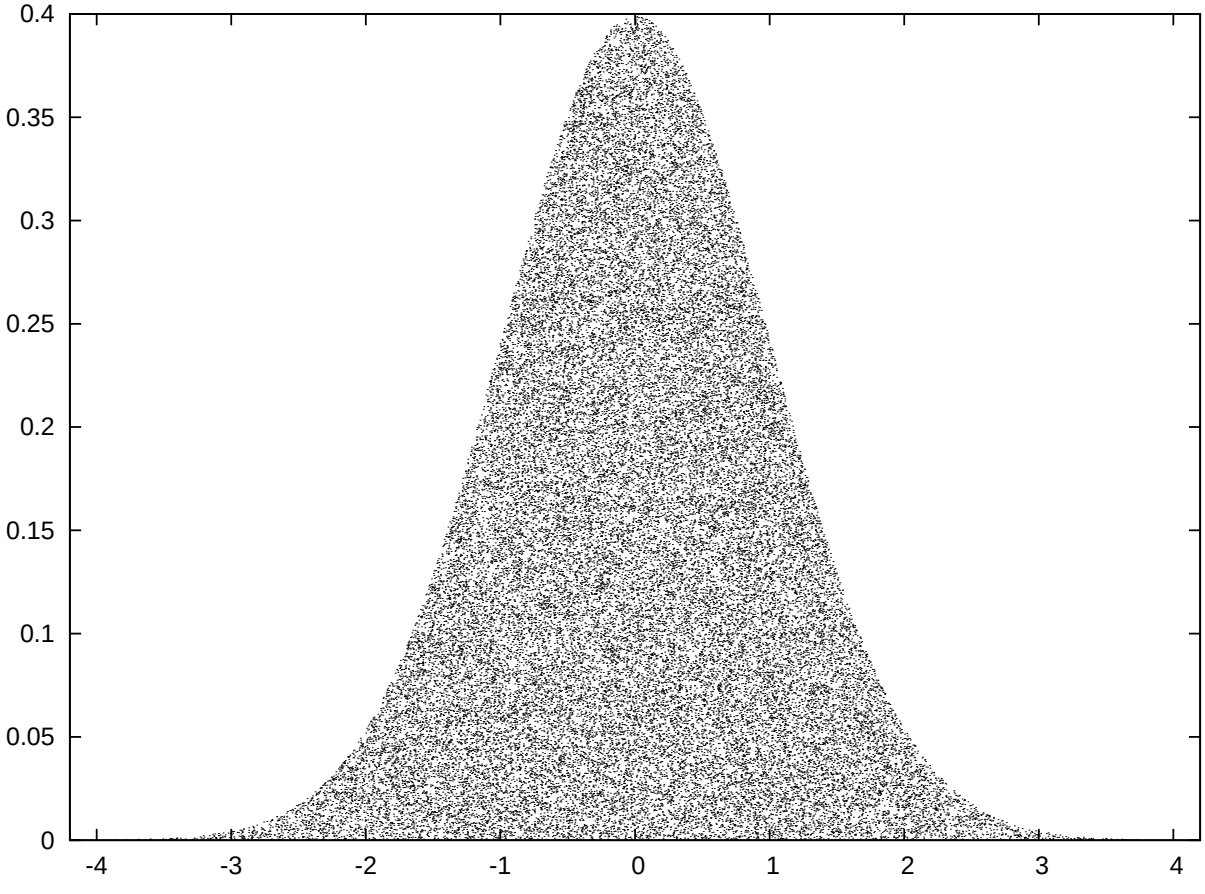
$$(x, y) := (X,\, U \cdot f(X))$$

are uniformly distributed on $\mathcal{A}_f$ (and vice versa). In the Figure this is illustrated for the normal distribution. If one cuts off a piece of the area $\mathcal{A}_f$, then the remaining points are still distributed uniformly. This is exploited by rejection methods.

Let g be another density on S, and assume for a constant $c \geq 1$

$$f(x) \leq c\, g(x) \quad \text{for all } x \in S\,.$$

The function cg is major to f, and the set $\mathcal{A}_f$ is subset of the area $\mathcal{A}_{cg}$ underneath the graph of cg. The rejection algorithm assumes that g-distributed x-samples can be calculated easily. Then the points $(x, ucg(x))$ are distributed uniformly on $\mathcal{A}_{cg}$. The aim is to calculate f-distributed random numbers. Cutting off the part of $\mathcal{A}_{cg}$ above $\mathcal{A}_f$ means to reject points with $ucg(x) > f(x)$. The x-coordinates of the remaining points with $ucg(x) \leq f(x)$ are accepted and are distributed as desired.



50000 points
$(X,\, U \cdot f(X))$
$X \sim \mathcal{N}(0,1)$
$U \sim \mathcal{U}[0,1]$
f density of X

Algorithm (rejection method)

Repeat:

$x :=$ random number distributed according to g,

$u :=$ random number $\sim \mathcal{U}[0,1]$ independent of x

until $u\, c\, g(x) \le f(x)$

return: x

Example (as exercise): Laplace-density $g(x) := \frac{1}{2}\exp(-|x|)$, f density of the standard normal distribution. What is c?*

* colored in Topic 3 of the *Topics for CF*

2.3 Normal Deviates

This section applies the transformation theorem in $\mathbb{R}^2$ to the calculation of normally distributed random numbers, and sketches the ziggurat method.

A. Method of Box and Muller

$S := [0,1]^2$, X distributed normally on S with density $f = 1$ on S. Transformation h:

$$\begin{cases} y_1 = \sqrt{-2\log x_1}\,\cos 2\pi x_2 =: h_1(x_1,x_2) \\ y_2 = \sqrt{-2\log x_1}\,\sin 2\pi x_2 =: h_2(x_1,x_2) \end{cases}$$

inverse h^{-1}:

$$\begin{cases} x_1 = \exp\left\{-\frac{1}{2}(y_1^2+y_2^2)\right\} \\ x_2 = \dfrac{1}{2\pi}\arctan\dfrac{y_2}{y_1} \end{cases}$$

For this transformation the determinant is

$$\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)} = \det\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} =$$

$$= -\frac{1}{2\pi}\exp\left\{-\tfrac{1}{2}(y_1^2+y_2^2)\right\}\,.$$

Its absolute value is the density of the two-dimensional normal distribution. Since

$$\left|\frac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\right| = \left[\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}y_1^2\right)\right]\cdot\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\tfrac{1}{2}y_2^2\right)\right],$$

the two-dimensional density is the product of the one-dimensional densities of the standard normal distribution. As a consequence, the two components y_1, y_2 of the vector Y are independent.

Application

When the two components x_1, x_2 are distributed $\sim \mathcal{U}[0,1]$, then the transformation provides two independent $y_1, y_2 \sim \mathcal{N}(0,1)$.

Algorithm (Box-Muller)

(1) generate $U_1 \sim \mathcal{U}[0,1]$ and $U_2 \sim \mathcal{U}[0,1]$.
(2) $\theta := 2\pi U_2, \quad \rho := \sqrt{-2\log U_1}$
(3) $Z_1 := \rho\cos\theta$ is $\sim \mathcal{N}(0,1)$
(same as $Z_2 := \rho\sin\theta$).

B. Variant of Marsaglia

Prepare the input x_1, x_2 for the Box–Muller transformation such that trigonometric functions are avoided. From $U \sim \mathcal{U}[0,1]$ obtain $V := 2U - 1 \sim \mathcal{U}[-1,1]$. Two such numbers V_1, V_2 define a point in $\mathbb{R}^2$. Define the disk

$$\mathcal{D} := \{(V_1, V_2) \ : \ V_1^2 + V_2^2 < 1\}.$$

Accept only those pairs (U_1, U_2) such that $(V_1, V_2) \in \mathcal{D}$. These accepted points are uniformly distributed on $\mathcal{D}$. Transformation to (radius)^2 and normalized angle:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} V_1^2 + V_2^2 \\ \frac{1}{2\pi} \arg(V_1, V_2) \end{pmatrix}.$$

These (x_1, x_2) are distributed uniformly in S ($\longrightarrow$ exercise) and serve as input for Box&Muller. The advantage is:

$$\cos(2\pi x_2) = \frac{V_1}{\sqrt{V_1^2 + V_2^2}}, \qquad \sin(2\pi x_2) = \frac{V_2}{\sqrt{V_1^2 + V_2^2}}$$

Algorithm (polar method)

(1) *Repeat:* generate $U_1, U_2 \sim \mathcal{U}[0,1]$;
$V_1 := 2U_1 - 1, \;\; V_2 := 2U_2 - 1$;
until $w := V_1^2 + V_2^2 < 1$.

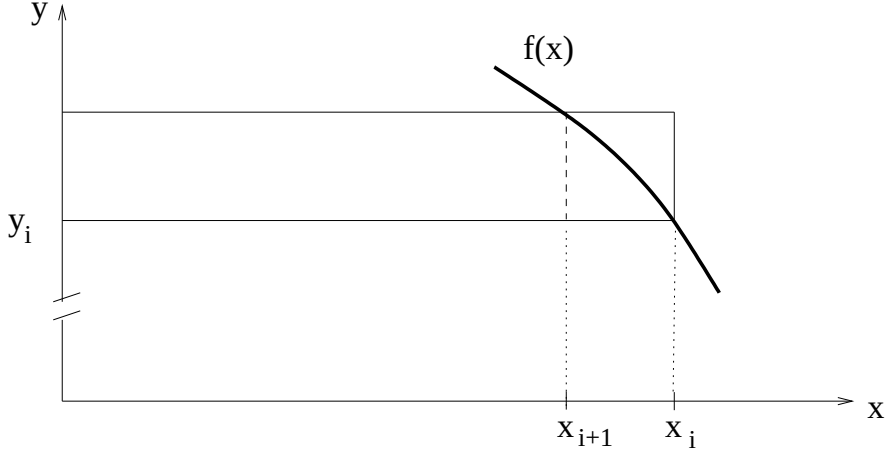
(2) $Z_1 := V_1 \sqrt{-2\log(w)/w}$ is $\sim \mathcal{N}(0,1)$
(as well as $Z_2 := V_2 \sqrt{-2\log(w)/w}$).

The probability of acceptance ($w < 1$) is the ratio of the areas $\frac{\pi}{4} \approx 0.785...$. That is, 21% of all draws (U_1, U_2) are rejected. But these costs are compensated by the saving of trigonometric functions, and Marsaglia's polar method is more efficient than standard Box&Muller.

C. Ziggurat Algorithm

A most efficient algorithm for the generation of normal deviates is the ziggurat algorithm, which is a acceptance-rejection method.

Essentially, g is a step function $\geq$ the Gaussian density f. Construction with N horizontal layers of $N-1$ rectangles *with the same area*, and one bottom segment with the same area, which is no rectangle but infinite because of the tail of f. The figure illustrates schematically a situation for $x \geq 0$, where the rectangle consists of two portions (for each i with $0 < i < N-1$), which make an extremely efficient test for acceptance possible (random choice of the layer i; uniformly distributed test point).



2.4 Correlated Normal Random Variates

The aim is the generation of a normal random vector $X = (X_1, \ldots, X_n)$ with prescribed

$$\mu = \mathsf{E}X = (\mathsf{E}X_1, \ldots, \mathsf{E}X_n),$$

covariance matrix with elements

$$\Sigma_{ij} = (\mathsf{Cov}X)_{ij} := \mathsf{E}\left((X_i - \mu_i)(X_j - \mu_j)\right); \quad \sigma_i^2 = \Sigma_{ii}$$

and correlations

$$\rho_{ij} := \frac{\Sigma_{ij}}{\sigma_i \sigma_j}\,.$$

For the following assume that Σ is symmetric and positive definite.

Recall: The density function $f(x_1, \ldots, x_n)$ of $\mathcal{N}(\mu, \Sigma)$ is

$$f(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det \Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^{tr}\, \Sigma^{-1}(x-\mu)\right\}.$$

Assume $Z \sim \mathcal{N}(0, I)$, where I is the unit matrix.

Apply a **linear transformation** $x = Az$, $A \in \mathbb{R}^{n\times n}$, where z is a realization of Z and A is nonsingular. Then

$$AZ \sim \mathcal{N}(0, AA^{tr})$$

This follows from the transformation theorem in Section 2.2B, with $X = h(Z) := AZ$. The density of X is

$$\begin{aligned} f(A^{-1}x)\,|\det(A^{-1})| &= \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(A^{-1}x)^{tr}(A^{-1}x)\right\} \frac{1}{|\det(A)|} \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{|\det(A)|} \exp\left\{-\frac{1}{2}x^{tr}(AA^{tr})^{-1}x\right\} \end{aligned}$$

for arbitrary nonsingular matrices A. In case AA^{tr} is a factorization of Σ, $\Sigma = AA^{tr}$, and hence $|\det A| = (\det \Sigma)^{1/2}$, we conclude:

$$AZ \sim \mathcal{N}(0, \Sigma)\,.$$

Translation with vector μ implies

$$\mu + AZ \sim \mathcal{N}(\mu, \Sigma)\,.$$

Example: Choose the Cholesky decomposition of Σ.
Alternative: decomposition out of a principal component analysis.

Algorithm (correlated normal deviates)

(1) Decompose Σ into $AA^{tr} = \Sigma$
(2) Draw $Z \sim \mathcal{N}(0, I)$ componentwise
with $Z_i \sim \mathcal{N}(0,1)$ for $i = 1, ..., n$, for example,
with Marsaglia's polar method
(3) $\mu + AZ$ is distributed $\sim \mathcal{N}(\mu, \Sigma)$

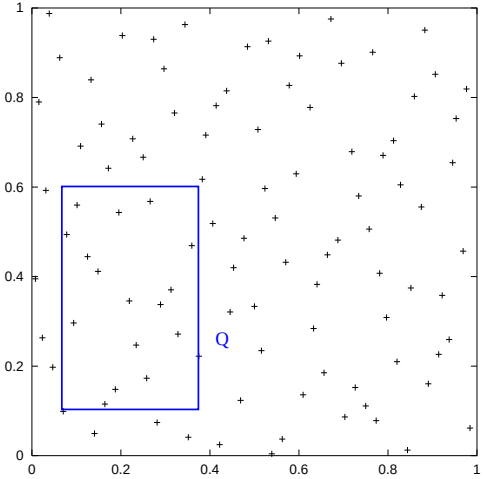
Example: If $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ is required, the solution is

$$\begin{pmatrix} \sigma_1 Z_1 \\ \sigma_2 \rho\, Z_1 + \sigma_2 \sqrt{1-\rho^2}\, Z_2 \end{pmatrix}.$$

2.5 Sequences of Numbers with Low Discrepancy

The aim is to construct points distributed similarly as random numbers, but avoid clustering or holes. In order to characterize **equidistributedness**, take any box (hyperrectangle) in $[0,1]^m$, $m \geq 1$. It would be desirable if for all Q

$$\frac{\#\text{ of the } x_i \in Q}{\#\text{ all points in } [0,1]^m} \approx \frac{\text{vol}(Q)}{\text{vol}([0,1]^m)}$$



For $m = 2$ the figure illustrates this idea in the unit square $[0,1]^2$.

Definition (discrepancy)

The discrepancy of a set $\{x_1, \ldots, x_N\}$ of N points $x_i \in [0,1]^m$ is

$$D_N := \sup_Q \left| \frac{\#\text{ of the } x_i \in Q}{N} - \text{ vol}(Q) \right| .$$

We wish to find sequences of points, whose discrepancy D_N for $N \to \infty$ tends to zero "quickly." To assess the decay we compare with the sequence

$$\frac{1}{\sqrt{N}} ,$$

which characterizes the probabilistic error of Monte Carlo methods. For true random points the discrepancy has a similar order of magnitude, namely,

$$\sqrt{\frac{\log \log N}{N}} .$$

Definition (sequence of low discrepancy)

A sequence of points $x_1, \ldots, x_N, \ldots \in [0,1]^m$ is called low-discrepancy sequence if there is a constant C_m such that for all N

$$D_N \leq C_m \frac{(\log N)^m}{N} \, .$$

Comment

The denominator in $\frac{1}{N}$ stands for relatively rapid decay of D_N with the number of points N, rapid as compared with the $\frac{1}{\sqrt{N}}$ of Monte Carlo.

But we have to observe the numerator $(\log N)^m$. Since $\log N$ grows only modestly, for low dimension m the decay of D_N is much faster than the decay of the probabilistic Monte Carlo error.

Table: different convergence rates to zero

N	$\frac{1}{\sqrt{N}}$	$\sqrt{\frac{\log\log N}{N}}$	$\frac{\log N}{N}$	$\frac{(\log N)^2}{N}$	$\frac{(\log N)^3}{N}$
10^1	.31622777	.28879620	.23025851	.53018981	1.22080716
10^2	.10000000	.12357911	.04605170	.21207592	.97664572
10^3	.03162278	.04396186	.00690776	.04771708	.32961793
10^4	.01000000	.01490076	.00092103	.00848304	.07813166
10^5	.00316228	.00494315	.00011513	.00132547	.01526009
10^6	.00100000	.00162043	.00001382	.00019087	.00263694
10^7	.00031623	.00052725	.00000161	.00002598	.00041874
10^8	.00010000	.00017069	.00000018	.00000339	.00006251
10^9	.00003162	.00005506	.00000002	.00000043	.00000890

Do sequences of low discrepancy exist?

Example: $(m = 1)$ **Van der Corput sequence**

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \frac{1}{16}, \ldots$$

Let us study its construction by means of the example $x_6 = \frac{3}{8}$. The binary representation of the index 6 is 110. This is radix-inverted: .011, which gives $\frac{3}{8}$.

Definition (radical-inverse function)

For $i = 1, 2, ...$ let

$$i = \sum_{k=0}^{j} d_k b^k$$

be the expansion in base b (integer ≥ 2), with $d_k \in \{0, 1, \ldots, b-1\}$. The radical-inverse function is defined by

$$\phi_b(i) := \sum_{k=0}^{j} d_k b^{-k-1} .$$

A one-dimensional example is the Van der Corput sequence: $x_i := \phi_2(i)$.

Definition (Halton sequence)

Let $p_1, \ldots, p_m$ be pairwise prime integers. The *Halton sequence* is defined as the sequence of vectors

$$x_i := (\phi_{p_1}(i), \ldots, \phi_{p_m}(i)) , \quad i = 1, 2, ...$$

The Halton sequence is of low discrepancy with $C_2 = 0.2602$ for $m = 2$ and easy to generate.

Other sequences of low discrepancy:

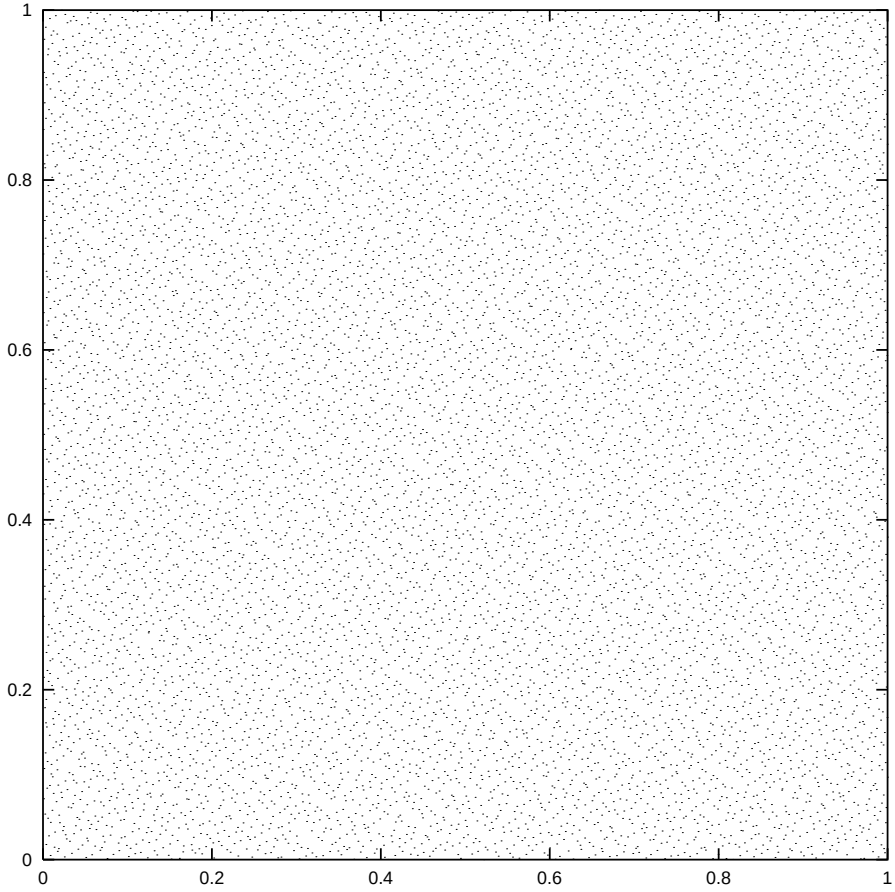
- Faure sequence
- Sobol sequence
- Niederreiter sequence
- Halton "leaped": For large m the Halton sequence suffers from correlation. This can be cured taking

$$x_i := (\phi_{p_1}(li), \ldots, \phi_{p_m}(li)), \quad i = 1, 2, ...$$

  for suitable prime l different from the p_k, for example, $l = 409$.

The deterministic sequences of low discrepancy are called **quasi-random numbers**. (They are not random!)

Literature on quasi-random numbers: [H. Niederreiter: Random Number Generation and Quasi-Monte Carlo Methods (1992)]



The figure shows the first 10000 Halton points with $m = 2$ and $p_1 = 2$, $p_2 = 3$.