COMP0043 Numerical Methods for Finance

2 Probability distributions

Guido Germano

Department of Computer Science, University College London

9 October 2024

Random numbers

In computer programming, random number generation always starts from standard uniform random variables $U(0,1)$. In C++ and Matlab they are produced by `rand()`, in Python by `random()`; see their documentation on

1. https://cplusplus.com/reference/cstdlib/rand
2. https://uk.mathworks.com/help/matlab/ref/rand.html
3. https://docs.python.org/2/library/random.html

We postpone how such uniform random generators work and first we focus on how random variables with other probability distributions are obtained from standard uniforms by

1. transformation
2. inversion
3. rejection.

As examples, we will see the following distributions:

1. exponential
2. normal
3. lognormal
4. noncentral chi-squared.

Transformation with a strictly monotonic one-dimensional function

For a strictly monotonic and thus bijective function $Y = g(X)$ of a random variable, how does the probability density function (PDF) of X, $f_X(x)$, transform into the PDF of Y, $f_Y(y)$? We start from the equality for the two probability distribution functions,

$$P_Y(dy) = P_X(dx) \tag{1}$$

from which we obtain the equality for the two cumulative distribution functions (CDFs),

$$|dF_Y(y)| = |dF_X(x)| \tag{2}$$

and the equality for the PDFs,

$$f_Y(y)|dy| = f_X(x)|dx|. \tag{3}$$

Dividing the latter by $|dy|$ gives

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|, \tag{4}$$

and substituting $x = g^{-1}(y)$,

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\left|\frac{dg^{-1}(y)}{dy}\right|. \tag{5}$$

An important special case of the transformation method is when $X \sim U(0,1)$, i.e. X is a standard uniform random variable, and $y = g(x) = F_Y^{-1}(x)$, i.e. the function of X is the inverse CDF (ICDF) of Y, also called the quantile function. Then, starting from Eq. (5),

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\left|\frac{dg^{-1}(y)}{dy}\right|, \tag{6}$$

and inserting

$$f_X(x) = \mathbf{1}_{[0,1]}(x) \tag{7}$$

$$x = g^{-1}(y) = F_Y(y) \in [0,1], \tag{8}$$

we obtain

$$f_Y(y) = \underbrace{\mathbf{1}_{[0,1]}(F_Y(y))}_{1}\left|\frac{dF_Y(y)}{dy}\right| = \frac{dF_Y(y)}{dy}, \tag{9}$$

i.e. $y = g(x) = F_Y^{-1}(x)$ transforms a standard uniform r.v. $X \sim U(0,1)$ into the r.v. Y.

Thus a standard uniform random variable $U(0,1)$ can easily be mapped to another random variable if its quantile function or ICDF is available in analytic form. However, this is not always possible in 1D; many distributions, including the normal, lack a closed-form CDF. A textbook example where this works is the exponential distribution.

Example: exponential distribution from standard uniform

The PDF of the exponential distribution is

$$f_Y(y) = \mathbf{1}_{[0,\infty)}(y)\,\tfrac{1}{\tau}\,e^{-\frac{y}{\tau}}. \tag{10}$$

The CDF is obtained integrating the PDF (in the following we assume $y \geqslant 0$),

$$F_Y(y) = \int_{-\infty}^{y} f_Y(y')dy' = \int_0^y e^{-\frac{y'}{\tau}}\,d\,\frac{y'}{\tau} = \int_y^0 e^{-\frac{y'}{\tau}}\,d\left(-\frac{y'}{\tau}\right) = e^u\Big|_{-\frac{y}{\tau}}^{0} = 1 - e^{-\frac{y}{\tau}}. \tag{11}$$

The quantile function or ICDF $y = F_Y^{-1}(x)$ is obtained inverting the CDF,

$$x = F_Y(y) = 1 - e^{-\frac{y}{\tau}} \tag{12}$$

$$e^{-\frac{y}{\tau}} = 1 - x \tag{13}$$

$$-\frac{y}{\tau} = \log(1-x) \tag{14}$$

$$y = g(x) = F_Y^{-1}(x) = -\tau\log(1-x). \tag{15}$$

Because $X \sim U(0,1)$ and thus $1 - X \sim X$, the function $g(x)$ can be simplified to

$$y = g(x) = -\tau\log x. \tag{16}$$

Example: lognormal distribution from normal

Consider $Y = g(X) = \exp X$ where $X \sim N(\mu, \sigma^2)$ is a normal random variable with PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}. \tag{17}$$

$y = g(x) = e^x$ is strictly increasing and thus invertible, $x = g^{-1}(y) = \log y$. From Eq. (5),

$$f_Y(y) = f_X\big(g^{-1}(y)\big)\left|\frac{dg^{-1}(y)}{dy}\right| = f_X(\log y)\frac{d\log y}{dy} = f_X(\log y)\frac{1}{y} \tag{18}$$

we obtain

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log y-\mu)^2}{2\sigma^2}\right), \qquad y \in \mathbb{R}_{>0}. \tag{19}$$

Geometric Brownian motion has a lognormal distribution. GBM was introduced by Osborne (1959) as a model for spot prices, popularised by Samuelson (1965) and used by Black, Scholes and Merton (1973) for their option pricing model.

Change of a multidimensional random variable

For the multidimensional case with random vectors $\mathbf{X} : \Omega \to \mathbb{R}^n$, $\mathbf{Y} : \Omega \to \mathbb{R}^m$ and the bijective function $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{Y} = \mathbf{g}(\mathbf{X})$, $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$, Correspondingly to Eqs. (1) and (3),

$$P_{\mathbf{Y}}(d\mathbf{y}) = P_{\mathbf{X}}(d\mathbf{x}) \tag{20}$$

$$f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}. \tag{21}$$

The differential volume element $d\mathbf{x} := dx_1 dx_2 \dots dx_n$ is positive and thus does not require an absolute value. With $\mathbf{y} = \mathbf{g}(\mathbf{x})$ and $\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y})$, the volume elements transform as

$$d\mathbf{y} = |\mathbf{J}_{\mathbf{g}}(\mathbf{x})|d\mathbf{x}, \qquad d\mathbf{x} = |\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{y})|d\mathbf{y}, \tag{22}$$

where $\mathbf{J}_{\mathbf{g}}(\mathbf{x})$ is the Jacobian matrix of $\mathbf{g}(\mathbf{x})$ and $|\mathbf{J}_{\mathbf{g}}(\mathbf{x})| = |\det \mathbf{J}_{\mathbf{g}}(\mathbf{x})|$ is the Jacobian of $\mathbf{g}(\mathbf{x})$. Inserting $d\mathbf{x} = |\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{y})|d\mathbf{y}$ and $\mathbf{x} = \mathbf{g}^{-1}(\mathbf{y})$ into Eq. (21) and simplifying $d\mathbf{y}$ yields

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{g}^{-1}(\mathbf{y}))|\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{y})| \tag{23}$$

that corresponds to Eq. (5) for $n = 1$.

Jacobian matrix

The Jacobian matrix of a function $\mathbf{f}(\mathbf{x}) : D \subset \mathbb{R}^n \to \mathbb{R}^m$ is

$$\mathbf{J}_\mathbf{f}(\mathbf{x}) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{pmatrix} \boldsymbol{\nabla}^\mathrm{T} f_1(\mathbf{x}) \\ \boldsymbol{\nabla}^\mathrm{T} f_2(\mathbf{x}) \\ \vdots \\ \boldsymbol{\nabla}^\mathrm{T} f_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \tag{24}$$

with elements $J_{ij} = \partial f_i / \partial x_j$; it provides the first-order approximation of the function,

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x_0}) + \mathbf{J}_\mathbf{f}(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}) + o(\|\mathbf{x} - \mathbf{x_0}\|). \tag{25}$$

The gradient $\boldsymbol{\nabla} f(\mathbf{x}) = \partial f / \partial \mathbf{x}$ of a function $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ gives the first-order approximation

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \boldsymbol{\nabla}^\mathrm{T} f(\mathbf{x})(\mathbf{x} - \mathbf{x_0}) + o(\|\mathbf{x} - \mathbf{x_0}\|). \tag{26}$$

The derivative $f'(x) = df/dx$ of a function $f(x) : \mathbb{R} \to \mathbb{R}$ gives the first-order approximation

$$f(x) = f(x_0) + f'(x)(x - x_0) + o(x - x_0). \tag{27}$$

Jacobian matrix, Jacobian, and inverse function theorem

The Jacobian matrix of a differentiable function $\mathbf{f}(\mathbf{x}) : D \subset \mathbb{R}^n \to \mathbb{R}^m$ can also be defined

$$\mathbf{J}_\mathbf{f}(\mathbf{x}) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \frac{\partial \mathbf{f}^\mathsf{T}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix} \begin{pmatrix} f_1 & f_2 & \dots & f_m \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \tag{28}$$

with elements $J_{ij} = \partial f_j / \partial x_i$; then the first-order approximation of the function is

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x_0}) + \mathbf{J}^\mathsf{T}(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}) + o(\|\mathbf{x} - \mathbf{x_0}\|). \tag{29}$$

This less common definition is more coherent with the gradient, which is a column vector.
The Jacobian $|\mathbf{J}| = |\det \mathbf{J}|$, that is short for the absolute value of the Jacobian determinant, has the same value with both definitions because for any matrix $\mathbf{A}$, $\det \mathbf{A} = \det \mathbf{A}^\mathsf{T}$.

For a differentiable and invertible function $\mathbf{f}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^n$, if $\mathbf{J}_\mathbf{f}(\mathbf{x})$ is nonsingular in $\mathbf{x}$,

$$\mathbf{J}_{\mathbf{f}^{-1}}(\mathbf{y}) = \mathbf{J}_\mathbf{f}^{-1}(\mathbf{x}) = \mathbf{J}_\mathbf{f}^{-1}(\mathbf{f}^{-1}(\mathbf{y})). \tag{30}$$

This generalises the inverse function theorem, which for $n = 1$ and $x = f^{-1}(y)$ states

$$(f^{-1}(y))' = (f'(x))^{-1} \quad \text{or} \quad \frac{df^{-1}(y)}{dy} = \left(\frac{df(x)}{dx}\right)^{-1} \quad \text{or} \quad \frac{dx}{dy} = \frac{1}{dy/dx}. \tag{31}$$

Example: Box-Muller method for standard normal r.v.

The inversion method works well only for the few target distributions with a closed-form ICDF or quantile function, e.g. the exponential distribution, Eq. (10). However, random variables from almost all distributions can be obtained transforming *two* (sometimes three) standard uniform random variables. A textbook example is the method by Box and Muller (1958) for standard normal random variables, actually due to Paley and Wiener (1934); see also Press et al. (2007), Section 7.3.4, p. 364. The functions $g(X_1, X_2)$ and $g^{-1}(Y_1, Y_2)$ are

$$\begin{cases} Y_1 = \sqrt{-2\log X_1}\cos 2\pi X_2 \\ Y_2 = \sqrt{-2\log X_1}\sin 2\pi X_2 \end{cases} \Leftrightarrow \begin{cases} X_1 = \exp\left(-\frac{1}{2}(Y_1^2 + Y_2^2)\right) \\ X_2 = \frac{1}{2\pi}\arctan\frac{Y_2}{Y_1}. \end{cases} \tag{32}$$

$$|\mathbf{J}_{\mathbf{g}^{-1}}| = \left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} -y_1 e^{-\frac{y_1^2+y_2^2}{2}} & \frac{-y_2}{2\pi(y_1^2+y_2^2)} \\ -y_2 e^{-\frac{y_1^2+y_2^2}{2}} & \frac{y_1}{2\pi(y_1^2+y_2^2)} \end{vmatrix} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y_2^2}{2}}. \tag{33}$$

Thus we have $Y_1, Y_2 \sim N(0,1) \Leftrightarrow X_1, X_2 \sim U(0,1)$, i.e. $(N_1, N_2) = g(U_1, U_2)$. For further reading, see Press et al. (2007), Section 7.3.8 Ratio-of-uniforms method, pp. 367–368, and Luc Devroye, Non-Uniform Random Variate Generation, Springer, New York, 1986.

Chi-squared distribution

The χ^2 **distribution** with k degrees of freedom is the distribution of the sum of the squares of k independent standard normal random variables $N_i(0,1) = Z_i$,

$$\chi^2(k) = \sum_{i=1}^{k} N_i^2(0,1) = \sum_{i=1}^{k} Z_i^2, \tag{34}$$

sometimes also indicated with $Q(k)$. It is a gamma distribution with $\alpha = \frac{k}{2}$ and $\beta = \frac{1}{2}$ and thus an extension of the exponential distribution,

$$f_{\chi_k^2}(x) = \frac{1_{(0,+\infty)}(x)}{2^{k/2}\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \qquad E(\chi_k^2) = k, \qquad \mathrm{Var}(\chi_k^2) = 2k. \tag{35}$$

The **noncentral** χ^2 **distribution** is a generalisation where each N_i has expectation μ_i,

$$\chi'^2(k,\lambda) = \sum_{i=1}^{k} N_i^2(\mu_i, 1) = \sum_{i=1}^{k} (Z_i + \mu_i)^2. \tag{36}$$

Besides the degrees of freedom k, the $\chi'^2_{k,\lambda}$ distribution has the noncentrality parameter

$$\lambda = \sum_{i=1}^{k} \mu_i^2. \tag{37}$$

Noncentral chi-squared distribution

The PDF of the noncentral chi-squared distribution is

$$f_{\chi'^2_{k,\lambda}}(x) = \frac{1}{2} e^{-\frac{x+\lambda}{2}} \left(\frac{x}{\lambda}\right)^{\frac{k}{4}-\frac{1}{2}} I_{\frac{k}{2}-1}\left(\sqrt{\lambda x}\right), \tag{38}$$

where $I_\alpha(y)$ is the modified Bessel function of the first kind, i.e. one of the two linearly independent solutions of the modified Bessel differential equation of order $\alpha \in \mathbb{C}$,

$$y^2 \frac{d^2 f}{dy^2} + y \frac{df}{dy} - (y^2 + \alpha^2) f = 0, \qquad y \in \mathbb{R}. \tag{39}$$

This is a special case of the Bessel differential equation of order $\alpha \in \mathbb{C}$,

$$z^2 \frac{d^2 f}{dz^2} + z \frac{df}{dz} + (z^2 - \alpha^2) f = 0, \qquad z \in \mathbb{C}, \tag{40}$$

when $\operatorname{Re} z = 0$ and $\operatorname{Im} z = y$, i.e. $z = iy$. One of the two linearly independent solutions of the Bessel differential equation is the Bessel function of the first kind $J_\alpha(z)$. The modified Bessel function of the first kind can be obtained from the Bessel function of the first kind by

$$I_\alpha(y) = i^{-\alpha} J_\alpha(iy). \tag{41}$$

The most important cases are when α is integer or half-integer.

Bessel functions

The Bessel functions have series representations, e.g.

$$I_\alpha(y) = \sum_{n=0}^{\infty} \frac{1}{n!\Gamma(n+\alpha+1)} \left(\frac{y}{2}\right)^{2n+\alpha} \tag{42}$$

$$J_\alpha(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+\alpha+1)} \left(\frac{z}{2}\right)^{2n+\alpha} \tag{43}$$

and others based on the general hypergeometric function or on Laguerre polynomials. The Bessel functions also have integral representations, the Bessel integrals, e.g.

$$I_\alpha(y) = \frac{1}{\pi}\int_0^\pi e^{y\cos\theta}\cos(\alpha\theta)d\theta - \frac{\sin\alpha\pi}{\pi}\int_0^\infty e^{-y\cosh x-\alpha x}dx \tag{44}$$

$$J_\alpha(z) = \frac{1}{\pi}\int_0^\pi \cos(\alpha\theta - z\sin\theta)d\theta - \frac{\sin\alpha\pi}{\pi}\int_0^\infty e^{-z\sinh x-\alpha x}dx \tag{45}$$

for $\operatorname{Re} z > 0$. If $\alpha \in \mathbb{Z}$, the 2nd integral on the RHS is zero. The PDF of the noncentral χ^2 distribution can be approximated with the inverse FFT of the characteristic function

$$\varphi_{\chi'^2_{k,\lambda}}(\xi) = \frac{e^{\frac{i\lambda\xi}{1-2i\xi}}}{(1-2i\xi)^{k/2}}. \tag{46}$$

From the characteristic function it is easy to compute $E(\chi'^2_{k,\lambda}) = k+\lambda$, $\operatorname{Var}(\chi'^2_{k,\lambda}) = 2k+4\lambda$.

Distributions overview

Distribution	PDF	CDF	ICDF	CF	Process
Exponential	✓	✓	✓	✓	KJD, CTRW
Normal	✓	✗	✗	✓	ABM, OU, BB, MJD, VG, CTRW
Lognormal	✓	✗	✗	✓	GBM
Noncentral χ^2	✗	✗	✗	✓	FSR

✓: analytic; ✗: non-analytic; CF: characteristic function; ABM: arithmetic Brownian motion; GBM: geometric Brownian motion; OU: Ornstein-Uhlenbeck; BB: Brownian bridge; FSR: Feller square-root; MJD: Merton jump-diffusion; KJD: Kou jump-diffusion; VG: variance gamma; CTRW: continuous-time random walk.

ABM, OU, GBM and FSR are the most important drift-diffusion processes in finance: ABM models log prices, GBM models spot stock prices, OU and FSR model mean-reverting quantities like short interest rates, exchange rates and volatilities.

The normal distribution is used also in the Monte Carlo simulation of GBM, and the FSR process. KJD actually uses the double-sided exponential or Laplace distribution.

For all Lévy processes other than ABM, e.g. MJD, KJD, VG, and the most relevant stochastic volatility processes, e.g. Heston and its extensions, only the CF, i.e. the Fourier transform of the PDF, is analytic, i.e. available in closed form; this is why Fourier transform methods are important in finance.