When we are given a fixed computational budget, there is a trade-off between large $N$ (low bias) and large $p$ (good estimator: low variance).

The quality of an estimator should be quantified using the Mean Square Error that combines bias and variance:

$$
\text{MSE}(\hat{\theta}) = E\{(\hat{\theta} - \theta)^2\} = E\{(\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta)^2\} = 0 \\
= E\{(\hat{\theta} - E\hat{\theta})^2\} + 2E\{(\hat{\theta} - E\hat{\theta})(\hat{\theta} - E\hat{\theta})\} \\
+ E\{(\hat{\theta} - E\hat{\theta})^2\} = (\hat{\theta} - E\hat{\theta})^2 + E\{(\hat{\theta} - E\hat{\theta})^2\} = \text{Bias}^2 + \text{Variance}.
$$

Generating Random Variables.

Computers provide algorithms that generate pseudo random numbers uniformly distributed on $[0,1]$. We assume here that we can sample from the uniform distribution on $[0,1]$ and we would to sample from other distributions.

Method 1: Inverse Transform Method.

We want to sample from a distribution on $\mathbb{R}$. If we know its cumulative distribution function $F : F(x) = P\{X \leq x\}$, $F$ is a non-decreasing function. If in addition it is continuous and strictly increasing, it has an inverse $F^{-1}$.

Claim: If $U \sim \text{Uniform}[0,1]$, then $F^{-1}(U) \sim X$.

Indeed $P\{F^{-1}(U) \leq x\} = P\{U \leq F(x)\} = F(x)$.

If $F$ is not strictly increasing nor continuous, define the generalized inverse

$$F^{-1}(x) = \inf \{ y : F(y) \geq x \}.$$

Then $\{ u : F^{-1}(u) \leq x \} = \{ u : u \leq F(x) \}$ and $F^{-1}(U) \sim X$.

Example: $X \sim \text{Exp}(\theta)$

$$F(x) = P\{X \leq x\} = 1 - e^{-\theta x}$$

$$F^{-1}(u) = \frac{1}{\theta} \log(1-u)$$
Method 2: Acceptance/Rejection method (Von Neumann)

Suppose we want to sample from a distribution whose density is \( f \)
and we know how to sample from another distribution whose density is \( g \)
and moreover there is a \( c \) such that \( \forall x \in \mathbb{R} \ f(x) \leq c g(x) \)
(in such a case, \( c \geq 1 \))

Here is the algorithm:

1. Sample from \( g : X \)
2. Sample from the uniform distribution \( U \) independently

\[
\text{Step 2: if } \quad U \leq \frac{f(x)}{cg(x)} \quad \text{return } x \\
\text{else go to step 1.}
\]

Claim: the output of the algorithm is a random variable with density \( f \).

Proof: the algorithm produces a random variable \( Y \) with law \( X \mid U \leq \frac{f(x)}{cg(x)} \)
where \( X \) and \( U \) are as in step 1. We want to show that \( Y \)
has a law with density \( f \), in other words that for any bounded function \( h \):

\[
E h(Y) = \int_{\mathbb{R}} h(y) f(y) \, dy.
\]

\[
E h(Y) = E \left[ h \left( X \mid U \leq \frac{f(x)}{cg(x)} \right) \right] = \frac{E \left[ h(X) 1 \{ U \leq \frac{f(x)}{cg(x)} \} \right]}{P \{ U \leq \frac{f(x)}{cg(x)} \}}
\]

\[
= \frac{E \{ h(X) \} E \left[ 1 \{ U \leq \frac{f(x)}{cg(x)} \} \mid X \right]}{E \left[ P \{ U \leq \frac{f(x)}{cg(x)} \mid X \} \right]} = \frac{E \{ h(X) f(x)/cg(x) \}}{E \{ f(x)/cg(x) \}}
\]

\[
= \frac{\int_{\mathbb{R}} h(x) \frac{f(x)}{cg(x)} \cdot g(x) \, dx}{\int_{\mathbb{R}} \frac{f(x)}{cg(x)} \cdot g(x) \, dx} = \int_{\mathbb{R}} h(x) f(x) \, dx.
\]

The probability of terminating at step 2 is

\[
P \{ U \leq \frac{f(x)}{cg(x)} \} = \frac{1}{c} \quad \text{(recall that } c \geq 1)\]

since \( U \) is independent, the expected number of steps to complete is \( c \).

We would like to have \( c \) as close to \( 1 \) as possible.

Example: sample from the Gaussian distribution using double exponential:

\[
\frac{1}{12 \pi} e^{-x^2/2} \leq e^{-|y|/2}.
\]
Recall that the density of the standard Gaussian distribution is \( \frac{e^{-x^2/2}}{\sqrt{\pi}} \).

Its cumulative distribution function is \( F(x) = \int_{-\infty}^{x} e^{-t^2/2} \, dt \).

If \( X \sim \text{Gen}(\mu, \sigma^2) \) then \( X \) has density \( \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \).

If \( Z \sim \text{Gen}(\sigma, 1) \) then \( \mu + \sigma Z \sim \text{Gen}(\mu, \sigma^2) \).

In \( \mathbb{R}^d \), recall that a random vector \( X \) has the Gaussian distribution if and only if any linear combination of its components has the Gaussian distribution on \( \mathbb{R} \).

Define \( \mu = \mathbb{E}(X) \) (\( d \)-dimensional vector)

\[ \Sigma_{ij} = \mathbb{E}((X_i - \mu_i)(X_j - \mu_j)) \]

\( \Sigma \) is a symmetric positive semi-definite matrix (\( d \times d \)).

If \( \Sigma \) is non-singular, \( X \) has density with respect to Lebesgue on \( \mathbb{R}^d \):

\[ \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right\} \]

If \( X \sim \text{Gen}(\mu, \Sigma) \) then \( X_i \sim \text{Gen}(\mu_i, \Sigma_{ii}) \) and \( \text{Cov}(X_i, X_j) = \Sigma_{ij} \).

**Moment Generating Function:** \( X \sim \text{Gen}(\mu, \Sigma) \)

\[ \mathbb{E} e^{t^T X} = e^{t^T \mu + \frac{1}{2} t^T \Sigma t} \quad t \in \mathbb{R}^d \]

**Linear Transformation Property:** \( X \sim \text{Gen}(\mu, \Sigma) \) \( A \) is a deterministic \( n \times d \) matrix

then \( AX \sim \text{Gen}(A\mu, A \Sigma A^T) \) \[ \text{proof:} \quad \mathbb{E} e^{t^T (AX)} = \mathbb{E} e^{(A^T t^T) X} = e^{(A^T \mu)^T t + \frac{1}{2} t^T (A^T \Sigma A) t} \]

**Conditioning Formula:** Let \( X \) be a \( d \)-dimensional Gaussian vector. Split it in \((X_{[d]}, X_{[q]})\)

when \( X_{[d]} \) is \( p \)-dimensional and \( X_{[q]} \) \( q \)-dimensional.

\[
\begin{pmatrix} X_{[d]} \\ X_{[q]} \end{pmatrix} \sim \text{Gen}\left( \begin{bmatrix} \mu_{[d]} \\ \mu_{[q]} \end{bmatrix}, \begin{bmatrix} \Sigma_{[d]} & \Sigma_{[d][q]} \\ \Sigma_{[q][d]} & \Sigma_{[q]} \end{bmatrix} \right)
\]
Generating Univariate Gaussian random variables.

It is enough to generate standard Gaussian random variables.

a) Inverse Transform Method. Not applicable because \( \Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} \, du \)

nor \( \Phi^{-1} \) have known explicit formulas. But these functions are so important that there exist very accurate polynomial approximations.

Physicists have numerical approximations for the error function \( \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} \, dt \)

and its inverse \( \Phi^{-1}(y) = \frac{1}{2} \left( \operatorname{erf}^{-1} \left( \frac{y}{2} \right) + 1 \right) \)

b) Acceptance - Rejection Method.

c) Box-Müller Transform: if \( U \) and \( V \) are independent \( \operatorname{Unif}[0,1] \) random variables, then \( \sqrt{-2 \log U} \cos 2\pi V \) and \( \sqrt{-2 \log U} \sin 2\pi V \) are independent \( \operatorname{Gau}(0,1) \) r.v.

\[
\operatorname{ef}(\sqrt{-2 \log U} \cos 2\pi V, \sqrt{-2 \log U} \sin 2\pi V) = \int_{[0,1]^2} \operatorname{f}(\sqrt{-2 \log u} \cos 2\pi v, \sqrt{-2 \log u} \sin 2\pi v) \, du \, dv
\]

Let \( x = \sqrt{-2 \log u} \cos 2\pi v \)

\( y = \sqrt{-2 \log u} \sin 2\pi v \)

\( x^2 + y^2 = -2 \log u \quad u = e^{-\frac{x^2+y^2}{2}} \)

\( 2\pi v = \arctg \frac{y}{x} \quad v = \frac{1}{2\pi} \arctg \frac{y}{x} \)

\( \frac{\partial u}{\partial x} = -x e^{-\frac{x^2+y^2}{2}} \quad \frac{\partial u}{\partial y} = -y e^{-\frac{x^2+y^2}{2}} \)

\( \frac{\partial u}{\partial x} = \frac{1}{2\pi} \left( -\frac{y}{x^2} \right) \quad \frac{1}{1+(y/x)^2} \quad \frac{\partial u}{\partial y} = \frac{1}{2\pi} \frac{1}{x} \quad \frac{1}{1+(y/x)^2} \)

\[
\int_{[0,1]^2} \operatorname{f}(\sqrt{-2 \log u} \cos 2\pi v, \sqrt{-2 \log u} \sin 2\pi v) \, du \, dv = \int_{\mathbb{R}^2} \operatorname{f}(xy) \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi} \, dx \, dy.
\]

Generating Multivariate Gaussian random variable.

Example: if \( (X_1, X_2) \sim \operatorname{Gau}((\mu_1, \mu_2), (\sigma_1^2, \sigma_2^2, \rho \sigma_1 \sigma_2)) \) if \( Z_1 \) and \( Z_2 \) are independent \( \operatorname{Gau}(0,1) \) r.v.

then

\[
\begin{pmatrix}
(\mu_1 + \sigma_1 \cdot Z_1 \\
(\mu_2 + \sigma_2 \cdot (\frac{Z_1 + \rho Z_2}{\sqrt{1-\rho^2}}))
\end{pmatrix}
\sim
\begin{pmatrix}
X_1 \\
X_2
\end{pmatrix}.
\]
In higher dimension we would like to run the same idea: if \( X \sim \text{Gn}(0, \Sigma) \) (independent \( \text{Gn}(0, I_d) \) random variables) then \( Y = \Lambda X \sim \text{Gn}(\mu, \Lambda \Lambda^T) \) so we must try to find a matrix \( \Lambda \) such that \( \Lambda \Lambda^T = \Sigma \).

There are many such decompositions.

a) One is particularly interesting: when \( \Lambda \) is lower triangular, \( \Lambda \Lambda^T = \Sigma \) is easier to compute because of the zero's.

It is called the Cholesky decomposition:

\[
\begin{bmatrix}
a_{11} & 0 & 0 & \cdots \\
a_{21} & a_{22} & 0 & \cdots \\
a_{31} & a_{32} & a_{33} & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_2 \\
a_3 \\
\vdots
\end{bmatrix}
= 
\begin{bmatrix}
\Sigma_{11} \\
\Sigma_{21} \\
\Sigma_{31} \\
\vdots
\end{bmatrix}
\]

By equating each term in the first column, we get:

\[ a_{11} = \Sigma_{11} \]
\[ a_{21} = \frac{1}{a_{11}} \Sigma_{21} \]
\[ a_{31} = \frac{1}{a_{11}} \Sigma_{31} \]
\[ \cdots \]
\[ a_{d1} = \frac{1}{a_{11}} \Sigma_{d1} \]

By equating each term in the second column, we get:

\[ a_{22} = \sqrt{\Sigma_{22} - a_{21}^2} \]
\[ a_{32} = \frac{1}{a_{22}} (\Sigma_{32} - a_{31} a_{21}) \]
\[ \cdots \]
\[ a_{d2} = \frac{1}{a_{22}} (\Sigma_{d2} - a_{d1} a_{21}) \]

More generally, \( \Sigma_{ij} = \sum_{k=1}^d a_{ik} a_{jk} \quad j \leq i \)

\[ a_{ii} = \sqrt{\Sigma_{ii} - \sum_{k=1}^{i-1} a_{ik}^2} \]
\[ a_{ij} = \frac{1}{a_{jj}} (\Sigma_{ij} - \sum_{k=1}^{j-1} a_{ik} a_{jk}) \quad j < i \]

with convention \( \sum_{k=1}^{0} = 0 \) and \( a_{jj} \neq 0 \).

If \( \Sigma \) is non-singular then \( a_{ii} \neq 0 \) for all \( i \), if not let \( a_{ij} = 0 \) if \( a_{ij} = 0 \) and that works.
Example: Simulation of the Gaussian random vector \((W_{t_1}, W_{t_2}, \ldots, W_{t_n})\) when \(W\) is Brownian motion and \(t_1 < \ldots < t_n\).

\[
\Sigma_{ij} = \inf(t_i, t_j) \quad \Sigma = \begin{bmatrix}
  t_1 & t_2 & t_3 & \cdots \\
  t_2 & t_2 & t_3 & \cdots \\
  t_3 & t_3 & t_3 & \cdots \\
  \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

The Cholesky decomposition is:

\[
A = \begin{bmatrix}
  \sqrt{t_1} & 0 & 0 & \cdots \\
  \sqrt{t_1} & \sqrt{t_2-t_1} & 0 & \cdots \\
  \sqrt{t_1} & \sqrt{t_2-t_1} & \sqrt{t_3-t_2} & \cdots \\
  \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

Which is simply the decomposition in terms of increments:

\((W_{t_1}, (W_{t_2} - W_{t_1}) + W_{t_1}, \ldots)\)

b) Another one is based on diagonalizing \(\Sigma\).

\[
\Sigma = P \Lambda P^T \quad P \text{ is orthogonal} \\
\Lambda \text{ is diagonal}
\]

\[
A = P \Lambda^{1/2} \quad (\text{where } \Lambda^{1/2} \text{ is the diagonal matrix with diagonal elements } \sqrt{\lambda_i})
\]

will be such that \(A A^T = \Sigma\).

\[
P = \begin{bmatrix}
  \vec{p}_1 & \cdots & \vec{p}_d
\end{bmatrix} \quad \vec{p}_i \text{ are the principal components (eigenvectors)}.
\]

\[
\mu + A Z = \vec{\mu} + \sum \sqrt{\lambda_i} \cdot \vec{p}_i + \cdots + \sqrt{\lambda_d} \cdot \vec{p}_d
\]

It has a nice statistical interpretation:

Example: \(d=2\) \(\vec{\mu} = 0\)

\[
\vec{p}_2 \times \vec{p}_1
\]

set when \(pdf > 5\%\).
Sampling from other diffusions.

Overstein-Uhlenbeck process.

The OU process is defined by the SDE:
\[ dX_t = -\alpha X_t \, dt + \sigma dW_t. \]

It has an explicit solution: Itô's:
\[ d(e^{\alpha t} X_t) = e^{\alpha t} \, \sigma dW_t. \] (\star)

\[ X_t = X_0 e^{-\alpha t} + e^{-\alpha t} \int_0^t e^{\alpha s} \, dW_s. \]

It is a Gaussian process mean function:
\[ E X_t = X_0 e^{-\alpha t} \]
and covariance function:
\[ \text{Cov}(X_s, X_t) = e^{-\alpha (s+t)} \int_0^{s+t} \sigma^2 e^{\alpha u} \, du \]
\[ = \frac{\sigma^2}{\alpha} e^{-\alpha (s+t)} (e^{\alpha t} - 1). \]

To draw a sample from \((X_{t_1}, X_{t_2}, \ldots, X_{t_n})\) one can use the general technique based on the Cholesky decomposition on the covariance matrix \(\text{Cov}(X_{t_i}, X_{t_j})\).

One can also do the following: from \((\star)\), we write:

\[ e^{\alpha t_{i+1}} X_{t_{i+1}} - e^{\alpha t_i} X_{t_i} = \int_{t_i}^{t_{i+1}} \sigma e^{\alpha s} \, dW_s. \]
\[ X_{t_{i+1}} = e^{-\alpha (t_{i+1} - t_i)} X_{t_i} + e^{-\alpha t_{i+1}} \int_{t_i}^{t_{i+1}} \sigma e^{\alpha s} \, dW_s. \]

Denote \(\xi_i = \int_{t_i}^{t_{i+1}} \sigma e^{\alpha s} \, dW_s \sim \text{N}(0, \frac{\sigma^2}{\alpha^2} (e^{2\alpha t_{i+1}} - e^{2\alpha t_i}))\)

and the \(\xi_i\)'s are independent of each other.

First draw \(\xi_0, \ldots, \xi_{n-1}\) are independent univariate Gaussians.

and then use \(X_{t_{i+1}} = e^{-\alpha (t_{i+1} - t_i)} X_{t_i} + e^{-\alpha t_{i+1}} \xi_i\), \(i = 1, \ldots, n\).

In Finance, OU are used in Variable Model (Interest rate): \(d\xi = a(b-\xi) \, dt + \sigma dW_t\)
[Note that \(d(b-\xi) = -a(b-\xi) \, dt - \sigma dW_t\)] or in stochastic volatility model:

\[
\begin{cases}
    \frac{dS_t}{S_t} = \mu \, dt + \sigma g(Y_t) \, dW_t \\
    dY_t = a(\bar{Y} - Y_t) \, dt + \nu \, dB_t
\end{cases}
\]