Optimization and Simulation

Drawing from distributions

Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

Outline

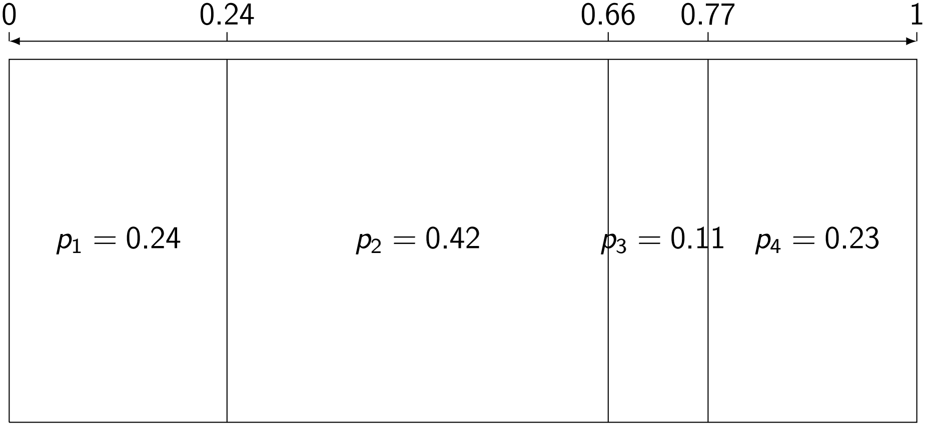
Discrete distributions

- Let X be a discrete r.v. with pmf:

$$P(X = x_i) = p_i, \; i = 1, \ldots, L$$

  where $\sum_{i=1}^{L} p_i = 1$.
- The support can be finite or infinite.
- We know how to draw from $U(0,1)$.
- How can we draw from X?

Inverse Transform Method: illustration



Discrete distributions

Inverse transform method (discrete)

1. Draw $r \sim U(0, 1)$.
2. Compute cumulative probabilities $F_k = \sum_{i=1}^{k} p_i$.
3. Return $X = x_k$ for the smallest k such that $r \leqslant F_k$.

Implementation remark

Precompute $(F_k)_k$ once and use a search (often binary search) for each draw.

Discrete inverse transform in Python (NumPy)

```
import numpy as np

x = np.array([1, 2, 3, 4])
p = np.array([0.24, 0.42, 0.11, 0.23])
F = np.cumsum(p)  # [0.24, 0.66, 0.77, 1.00]

rng = np.random.default_rng(2026)
r = rng.random(size=5)
k = np.searchsorted(F, r)     # smallest k with r <= F[k]
samples = x[k]

print(r)
print(samples)
[0.17893481 0.63991317 0.4672684  0.37050053 0.35491733]
[1 2 2 2 2]
```

Discrete distributions

Acceptance-rejection

- Attributed to von Neumann.
- We want to draw from X with pmf p_i.
- We know how to draw from Y with pmf q_i.

Define a constant $c \geqslant 1$ such that

$$\frac{p_i}{q_i} \leqslant c \; \forall i \text{ s.t. } p_i > 0.$$

Algorithm

1. Draw y from Y
2. Draw r from $U(0,1)$
3. If $r < \frac{p_y}{cq_y}$, return $x = y$ and stop. Otherwise, start again.

Acceptance-rejection: analysis

Probability to be accepted during a given iteration

$$\begin{aligned} P(Y=y,\text{accepted}) &= P(Y=y) \quad P(\text{accepted}|Y=y) \\ &= q_y \qquad\qquad p_y/cq_y \\ &= \frac{p_y}{c} \end{aligned}$$

Probability to be accepted

$$\begin{aligned} P(\text{accepted}) &= \sum_y P(\text{accepted}|Y=y)P(Y=y) \\ &= \sum_y \frac{p_y}{cq_y} q_y \\ &= 1/c. \end{aligned}$$

Acceptance-rejection: analysis

Probability to draw x at iteration n

$$P(X = x|n) = (1 - \tfrac{1}{c})^{n-1} \frac{p_x}{c}$$

Acceptance-rejection: analysis

$$\begin{aligned}
P(X = x) &= \sum_{n=1}^{+\infty} P(X = x|n) \\
&= \sum_{n=1}^{+\infty} \left(1 - \frac{1}{c}\right)^{n-1} \frac{p_x}{c} \\
&= c\frac{p_x}{c} \\
&= p_x.
\end{aligned}$$

Reminder: geometric series:

$$\sum_{n=0}^{+\infty} x^n = \frac{1}{1-x}$$

Acceptance-rejection: analysis

Remarks

- Average number of iterations: c
- The closer c is to 1, the closer the pmf of Y is to the pmf of X.

Accept–reject: acceptance rate in practice

```
import numpy as np

# Example: target p over {0,1,2,3}, proposal q uniform
p = np.array([0.24, 0.42, 0.11, 0.23])
q = np.ones_like(p) / len(p)
c = np.max(p / q)  # envelope constant

rng = np.random.default_rng(123)
R = 100000
accepted = 0
for _ in range(R):
    y = rng.integers(0, len(p))        # draw from q
    r = rng.random()
    if r < p[y] / (c * q[y]):
        accepted += 1
```

Accept–reject: acceptance rate in practice

```
print(f"c␣=␣{c:.3f}")
print(f"theory␣accept␣rate␣=␣{1/c:.3f}")
print(f"empirical␣accept␣rate␣=␣{accepted/R:.3f}")
c = 1.680
theory accept rate = 0.595
empirical accept rate = 0.594
```

Outline

Continuous distributions

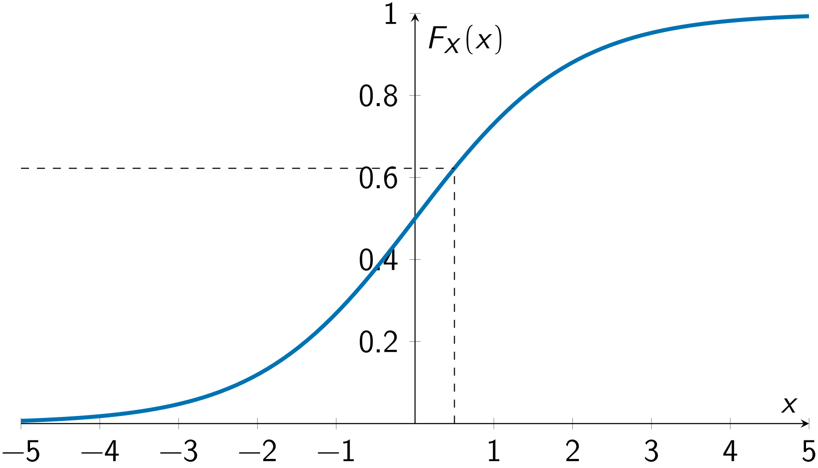
Inverse Transform Method

- Let X be a continuous r.v. with CDF $F_X(x)$
- Draw r from a uniform $U(0,1)$
- Generate $F_X^{-1}(r)$.

Motivation

- F_X is monotonically increasing
- It implies that $x_1 \leqslant x_2$ is equivalent to $F_X(x_1) \leqslant F_X(x_2)$.

Inverse Transform Method



Inverse Transform Method

More formally

- Denote $F_U(x) = x$ the CDF of the r.v. $U(0,1)$
- Let G be the distribution of the r.v. $F_X^{-1}(U)$

$$
\begin{aligned}
G(x) &= \Pr(F_X^{-1}(U) \leqslant x) \\
&= \Pr(F_X(F_X^{-1}(U)) \leqslant F_X(x)) \\
&= \Pr(U \leqslant F_X(x)) \\
&= F_U(F_X(x)) \\
&= F_X(x)
\end{aligned}
$$

Inverse Transform Method

Examples: let r be a draw from $U(0,1)$

Name	$F_X(\varepsilon)$	Draw
Exponential(b)	$1 - e^{-\varepsilon/b}$	$-b \ln r$
Logistic(μ,σ)	$1/(1 + \exp(-(\varepsilon - \mu)/\sigma))$	$\mu - \sigma \ln(\frac{1}{r} - 1)$
Power(n,σ)	$(\varepsilon/\sigma)^n$	$\sigma r^{1/n}$

Note

The CDF is not always available (e.g. normal distribution).

Continuous distributions

Rejection Method

- We want to draw from X with pdf f_X.
- We know how to draw from Y with pdf f_Y.

Define a constant c such that

$$\frac{f_X(\varepsilon)}{f_Y(\varepsilon)} \leqslant c \ \forall \varepsilon$$

Algorithm

1. Draw y from Y
2. Draw r from $U(0,1)$
3. If $r < \frac{f_X(y)}{cf_Y(y)}$, return $x = y$ and stop. Otherwise, start again.

Rejection Method: example

Draw from a normal distribution

- Let $\bar{X} \sim N(0,1)$ and $X = |\bar{X}|$
- Probability density function: $f_X(\varepsilon) = \frac{2}{\sqrt{2\pi}} e^{-\varepsilon^2/2}, \quad 0 < \varepsilon < +\infty$
- Consider an exponential r.v. with pdf $f_Y(\varepsilon) = e^{-\varepsilon}, \quad 0 < \varepsilon < +\infty$
- Then

$$\frac{f_X(\varepsilon)}{f_Y(\varepsilon)} = \frac{2}{\sqrt{2\pi}} e^{\varepsilon - \varepsilon^2/2}$$

- The ratio takes its maximum at $\varepsilon = 1$, therefore

$$\frac{f_X(\varepsilon)}{f_Y(\varepsilon)} \leqslant \frac{f_X(1)}{f_Y(1)} = \sqrt{2e/\pi} \approx 1.315.$$

- Rejection method, with $\frac{f_X(\varepsilon)}{cf_Y(\varepsilon)} = \frac{1}{\sqrt{e}} e^{\varepsilon - \varepsilon^2/2} = e^{\varepsilon - \frac{\varepsilon^2}{2} - \frac{1}{2}} = e^{-\frac{(\varepsilon-1)^2}{2}}$

Rejection Method: example

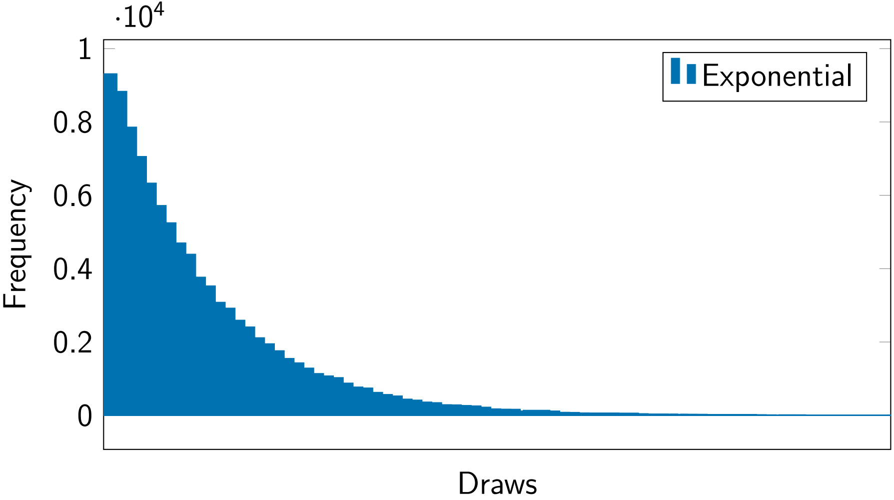
Algorithm: draw from a normal

1. Draw r from $U(0,1)$
2. Let $y = -\ln r$ (draw from the exponential)
3. Draw s from $U(0,1)$
4. If $s < e^{-\frac{(y-1)^2}{2}}$ return $x = y$ and go to step 5. Otherwise, go to step 1.
5. Draw t from $U(0,1)$.
6. If $t \leqslant 0.5$, return x. Otherwise, return $-x$.

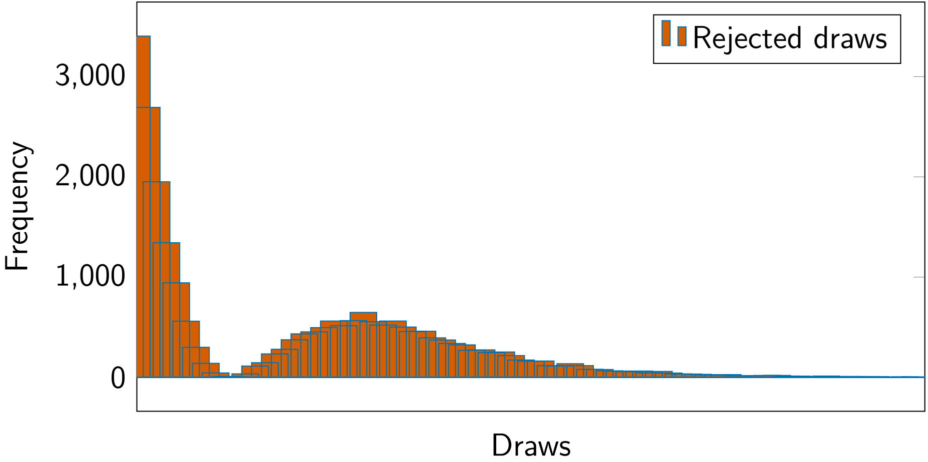
Note

This procedure can be improved. See [Ross, 2012] (Chapter 5).

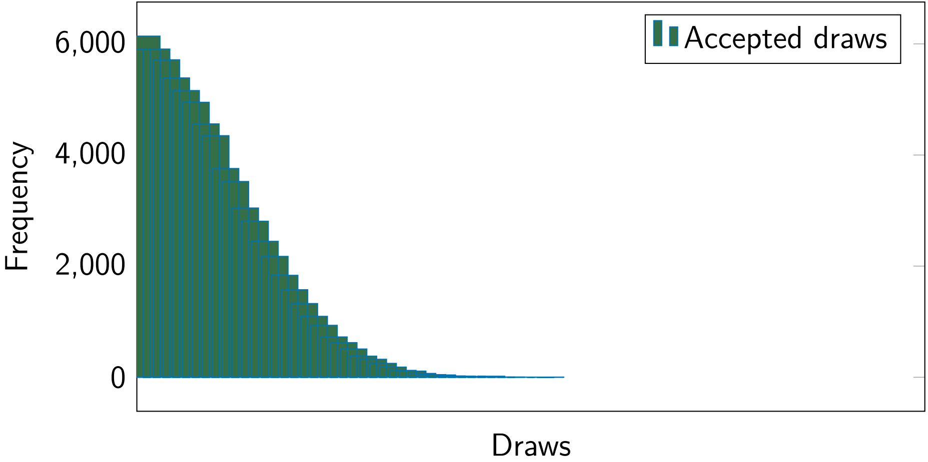
Draws from the exponential



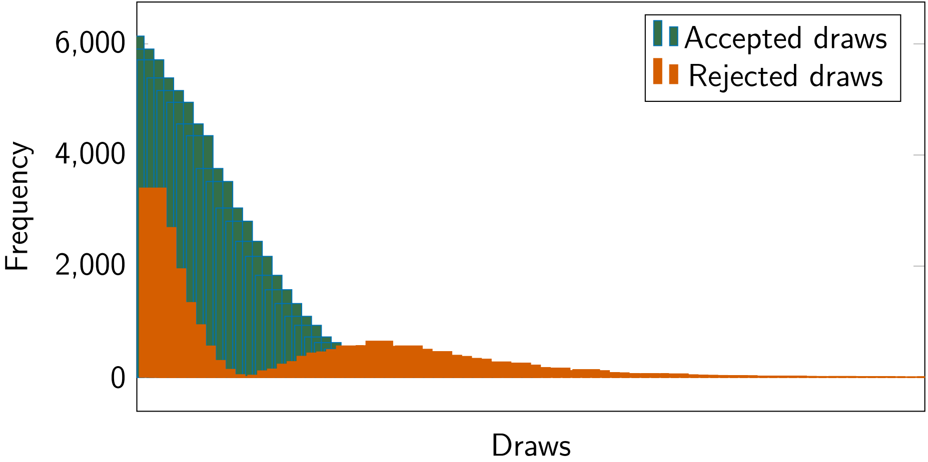
Rejected draws



Accepted draws



Rejected and accepted draws



Drawing from an unnormalized distribution

Rejection Method

- We want to draw from X with pdf

$$f_X = \frac{g_X}{K},$$

where

$$K = \int_{\varepsilon} g_X(\varepsilon) d\varepsilon$$

is difficult or impossible to calculate.

- Therefore, we know g_X but we don't know f_X.
- We know how to draw from Y with pdf f_Y.

Drawing from an unnormalized distribution

- Define a constant c_u such that
$$\frac{g_X(\varepsilon)}{f_Y(\varepsilon)} \leqslant c_u \ \forall \varepsilon.$$
Therefore,
$$\frac{f_X}{f_Y} = \frac{g_X}{K f_Y} \leqslant \frac{c_u}{K},$$
and the rejection method can be applied with $c = c_u/K$.
- Accept probability:
$$\frac{f_X}{c f_Y} = \frac{g_X}{K}\frac{K}{c_u}\frac{1}{f_Y} = \frac{g_X}{c_u f_Y},$$
and K does not play any role.

Drawing from the standard normal distribution

- Accept–reject is possible, but it is not the most efficient approach for $N(0, 1)$.
- In practice we use specialized methods (e.g. polar / Box–Muller; Ziggurat in many libraries).
- Polar method (no tuning, no envelope): see Appendix.

Transformations of standard normal

- If r is a draw from $N(0,1)$, then
$$s = br + a$$
is a draw from $N(a, b^2)$
- If r is a draw from $N(a, b^2)$, then
$$e^r$$
is a draw from a log normal $LN(a, b^2)$ with mean
$$e^{a+(b^2/2)}$$
and variance
$$e^{2a+b^2}(e^{b^2} - 1)$$

Multivariate normal

- If $r_1, \ldots, r_n$ are independent draws from $N(0, 1)$, and

$$r = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}$$

- then

$$s = a + Lr$$

  is a vector of draws from the n-variate normal $N(a, LL^T)$, where
  - L is lower triangular, and
  - LL^T is the Cholesky factorization of the variance-covariance matrix

Multivariate normal

Example:

$$L = \begin{pmatrix} \ell_{11} & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} \end{pmatrix}$$

$$\begin{aligned} s_1 &= \ell_{11} r_1 \\ s_2 &= \ell_{21} r_1 + \ell_{22} r_2 \\ s_3 &= \ell_{31} r_1 + \ell_{32} r_2 + \ell_{33} r_3 \end{aligned}$$

Outline

Transforming draws

Method

- Consider draws from the following distributions:
  - normal: $N(0,1)$ (draws denoted by ξ below)
  - uniform: $U(0,1)$ (draws denoted by r below)
- Draws R from other distributions are obtained from nonlinear transforms.

Lognormal(a,b)

$$f(x) = \frac{1}{xb\sqrt{2\pi}} \exp\left(\frac{-(\ln x - a)^2}{2b^2}\right) \qquad R = e^{a+b\xi}$$

Transforming draws

Cauchy(a,b)

$$f(x) = \left(\pi b \left(1 + \left(\frac{x-a}{b}\right)^2\right)\right)^{-1} \qquad R = a + b \tan\left(\pi(r - \frac{1}{2})\right)$$

$\chi^2(a)$ (a integer)

$$f(x) = \frac{x^{(a-2)/2} e^{-x/2}}{2^{a/2}\Gamma(a/2)} \qquad R = \sum_{j=1}^{a} \xi_j^2$$

Transforming draws

Exponential(a)

$$F(x) = 1 - e^{-x/a} \qquad R = -a \ln r$$

Extreme Value(a,b)

$$F(x) = 1 - \exp\left(-e^{-(x-a)/b}\right) \qquad R = a - b \ln(-\ln r)$$

Logistic(a,b)

$$F(x) = \left(1 + e^{-(x-a)/b}\right)^{-1} \qquad R = a + b \ln\left(\frac{r}{1-r}\right)$$

Transforming draws

Pareto(a,b)

$$F(x) = 1 - \left(\frac{a}{x}\right)^b \qquad R = a(1-r)^{-1/b}$$

Standard symmetrical triangular distribution

$$f(x) = \begin{cases} 4x & \text{if } 0 \leqslant x \leqslant 1/2 \\ 4(1-x) & \text{if } 1/2 \leqslant x \leqslant 1 \end{cases} \qquad R = \frac{r_1 + r_2}{2}$$

Transforming draws

Weibull(a,b)

$$F(x) = 1 - e^{-\left(\frac{x}{a}\right)^b} \quad R = a(-\ln r)^{1/b}$$

Erlang(a,b) (b integer)

$$f(x) = \frac{(x/a)^{b-1} e^{-x/a}}{a(b-1)!} \qquad R = -a \sum_{j=1}^{b} \ln r_i$$

Choosing a sampling method

Four common situations

- **Inverse transform:** exact; needs CDF and (numerical) inverse.
- **Accept–reject:** general; needs an envelope cf_Y close to f_X (accept rate $1/c$).
- **Specialized methods (e.g. normal):** fastest and most reliable for standard distributions.
- **MCMC:** for complex / unnormalized / high-dimensional targets when global envelopes are hard.

Outline

Monte-Carlo integration

Expectation

- X r.v. on $[a, b]$, $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R} \cup \{+\infty\}$
- Expectation of X:

$$\mathrm{E}[X] = \int_a^b x f_X(x) dx.$$

- If $g : \mathbb{R} \to \mathbb{R}$ is a function, then

$$\mathrm{E}[g(X)] = \int_a^b g(x) f_X(x) dx.$$

Monte-Carlo integration

Simulation

$$\mathrm{E}[g(X)] \approx \frac{1}{R}\sum_{r=1}^{R} g(x_r).$$

Approximating the integral

$$\int_a^b g(x) f_X(x) dx = \lim_{R\to\infty} \frac{1}{R}\sum_{r=1}^{R} g(x_r).$$

so that

$$\int_a^b g(x) f_X(x) dx \approx \frac{1}{R}\sum_{r=1}^{R} g(x_r).$$

Monte-Carlo integration

Calculating $I = \int_a^b g(x) f_X(x) dx$

- Consider X with pdf f_X.
- Convenient choice: $X \sim U[0,1]$, as $f_U(x) = 1$, $\forall x$.
- Generate R draws x_r, $r = 1, \ldots, R$ from X;
- Calculate
$$I \approx \widehat{I} = \frac{1}{R} \sum_{r=1}^{R} g(x_r).$$

Monte-Carlo integration

Approximation error

- Sample variance:

$$V_R = \frac{1}{R-1} \sum_{r=1}^{R} (g(x_r) - \widehat{I})^2.$$

- By simulation: as

$$\mathrm{Var}[g(X)] = \mathrm{E}[g(X)^2] - \mathrm{E}[g(x)]^2,$$

we have

$$V_R \approx \frac{1}{R} \sum_{r=1}^{R} g(x_r)^2 - \widehat{I}^2.$$

Monte-Carlo integration

Approximation error

95% confidence interval: $[\widehat{I} - 1.96 e_R \leqslant I \leqslant \widehat{I} + 1.96 e_R]$ where

$$e_R = \sqrt{\frac{V_R}{R}}.$$

Monte-Carlo integration

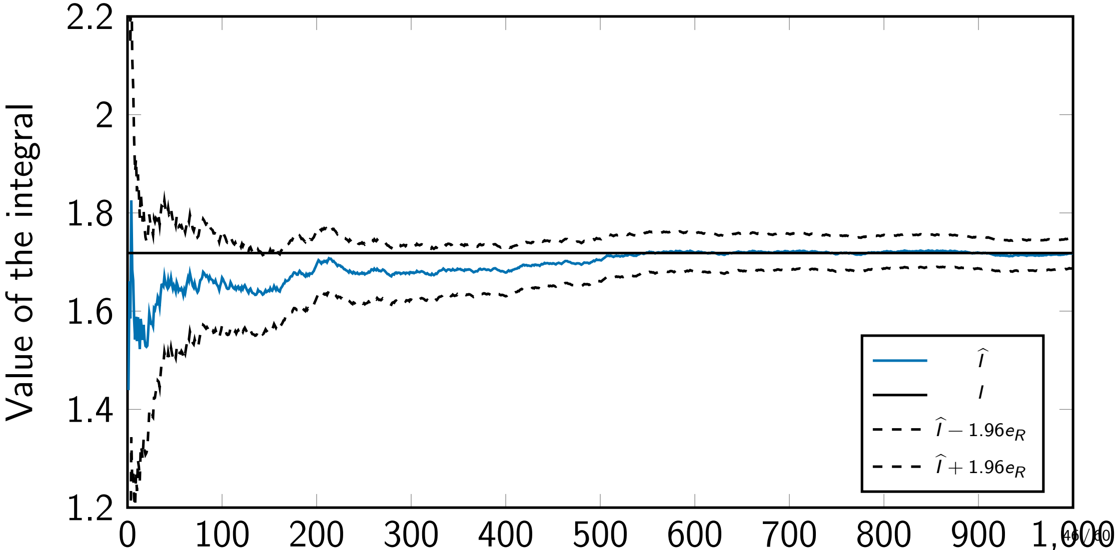
Example

$$\int_0^1 e^x dx = e - 1 = 1.7183$$

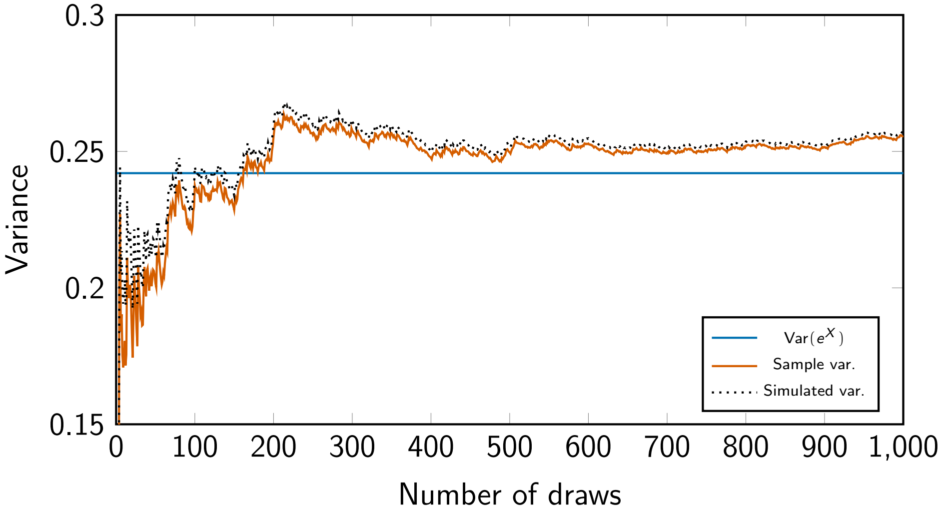
- Random variable X uniformly distributed $(f_X(\varepsilon) = 1)$
- $g(X) = e^X$
- $\text{Var}(e^X) = \frac{e^2-1}{2} - (e-1)^2 = 0.2420$

R	10	100	1000
$\widehat{I}$	1.8270	1.7707	1.7287
Sample variance	0.1607	0.2125	0.2385
Simulated variance	0.1742	0.2197	0.2398

Monte-Carlo integration



Monte-Carlo integration



Outline

Summary

- Draws from uniform distribution: available in any programming language
- Inverse transform method: requires the pmf or the CDF.
- Accept-reject: needs a "similar" r.v. easy to draw from.
- Transforming uniform and normal draws.
- First application: Monte-Carlo integration.

Outline

Uniform distribution: $X \sim U(a, b)$

pdf

$$f_X(x) = \begin{cases} 1/(b-a) & \text{if } a \leqslant x \leqslant b, \\ 0 & \text{otherwise.} \end{cases}$$

CDF

$$F_X(x) = \begin{cases} 0 & \text{if } x \leqslant a, \\ (x-a)/(b-a) & \text{if } a \leqslant x \leqslant b, \\ 1 & \text{if } x \geqslant b. \end{cases}$$

Mean, median

$(a+b)/2$

Variance

$(b-a)^2/12$

Normal distribution: $X \sim N(a, b)$

pdf

$$f_X(x) = \frac{1}{b\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2b^2}\right)$$

CDF

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Mean, median

a

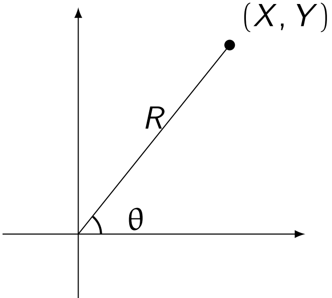
Variance

b^2

The polar method

Draw from a normal distribution

- Let $X \sim N(0,1)$ and $Y \sim N(0,1)$ independent
- pdf:
$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}.$$
- Let R and θ such that $R^2 = X^2 + Y^2$, and $\tan\theta = Y/X$.



The polar method

Change of variables (reminder)

- Let A be a multivariate r.v. distributed with pdf $f_A(a)$.
- Consider the change of variables $b = H(a)$ where H is bijective and differentiable
- Then $B = H(A)$ is distributed with pdf

$$f_B(b) = f_A(H^{-1}(b)) \left| \det \left(\frac{dH^{-1}(b)}{db} \right) \right| .$$

Here: $A = (X, Y)$, $B = (R^2, \theta) = (T, \theta)$

$$H^{-1}(b) = \begin{pmatrix} T^{\frac{1}{2}} \cos\theta \\ T^{\frac{1}{2}} \sin\theta \end{pmatrix} \quad \frac{dH^{-1}(b)}{db} = \begin{pmatrix} \frac{1}{2} T^{-\frac{1}{2}} \cos\theta & -T^{\frac{1}{2}} \sin\theta \\ \frac{1}{2} T^{-\frac{1}{2}} \sin\theta & T^{\frac{1}{2}} \cos\theta \end{pmatrix}$$

The polar method

$$H^{-1}(b) = \begin{pmatrix} T^{\frac{1}{2}} \cos\theta \\ T^{\frac{1}{2}} \sin\theta \end{pmatrix} \quad \frac{dH^{-1}(b)}{db} = \begin{pmatrix} \frac{1}{2} T^{-\frac{1}{2}} \cos\theta & -T^{\frac{1}{2}} \sin\theta \\ \frac{1}{2} T^{-\frac{1}{2}} \sin\theta & T^{\frac{1}{2}} \cos\theta \end{pmatrix}$$

Therefore,

$$\left| \det\left(\frac{dH^{-1}(b)}{db} \right) \right| = \frac{1}{2}.$$

and

$$f_B(T, \theta) = \frac{1}{2} \frac{1}{2\pi} e^{-T/2}, \quad 0 < T < +\infty, \quad 0 < \theta < 2\pi.$$

Product of

- an exponential with mean 2: $\frac{1}{2} e^{-T/2}$
- a uniform on $[0, 2\pi[$: $1/2\pi$

The polar method

Therefore

- R^2 and θ are independent
- R^2 is exponential with mean 2
- θ is uniform on $(0, 2\pi)$

Algorithm

1. Let r_1 and r_2 be draws from $U(0, 1)$.
2. Let $R^2 = -2 \ln r_1$ (draw from exponential of mean 2)
3. Let $\theta = 2\pi r_2$ (draw from $U(0, 2\pi)$)
4. Let
$$\begin{aligned} X &= R\cos\theta = \sqrt{-2\ln r_1}\cos(2\pi r_2) \\ Y &= R\sin\theta = \sqrt{-2\ln r_1}\sin(2\pi r_2) \end{aligned}$$

The polar method

Issue

Time consuming to compute sine and cosine

Solution

Generate directly the result of the sine and the cosine

- Draw a random point (s_1, s_2) in the circle of radius one centered at $(0, 0)$.
- How? Draw a random point in the square $[-1, 1] \times [-1, 1]$ and reject points outside the circle
- Let (R, θ) be the polar coordinates of this point.
- $R^2 \sim U(0, 1)$ and $\theta \sim U(0, 2\pi)$ are independent

$$
\begin{aligned}
R^2 &= s_1^2 + s_2^2 \\
\cos\theta &= s_1/R \\
\sin\theta &= s_2/R
\end{aligned}
$$

The polar method

Original transformation

$$
\begin{aligned}
X &= R\cos\theta = \sqrt{-2\ln r_1}\cos(2\pi r_2) \\
Y &= R\sin\theta = \sqrt{-2\ln r_1}\sin(2\pi r_2)
\end{aligned}
$$

Draw (s_1, s_2) in the circle

$$
\begin{aligned}
t &= s_1^2 + s_2^2 \\
X &= R\cos\theta = \sqrt{-2\ln t}\,\frac{s_1}{\sqrt{t}} = s_1\sqrt{\frac{-2\ln t}{t}} \\
Y &= R\sin\theta = \sqrt{-2\ln t}\,\frac{s_2}{\sqrt{t}} = s_2\sqrt{\frac{-2\ln t}{t}}
\end{aligned}
$$

The polar method

Algorithm

1. Let r_1 and r_2 be draws from $U(0, 1)$.
2. Define $s_1 = 2r_1 - 1$ and $s_2 = 2r_2 - 1$ (draws from $U(-1, 1)$).
3. Define $t = s_1^2 + s_2^2$.
4. If $t > 1$, reject the draws and go to step 1.
5. Return
$$x = s_1 \sqrt{\frac{-2 \ln t}{t}} \text{ and } y = s_2 \sqrt{\frac{-2 \ln t}{t}}.$$

Bibliography

Ross, S. (2012).
Simulation.
Academic Press, fifth edition edition.