Lecture 3 - Random number generation

Computational Statistics and Applications

Vu Quoc Hoang (vqhoang@fit.hcmus.edu.vn)
Tran Thi Thao Nhi (thaonhitt2005@gmail.com)
FIT - HCMUS

February 5, 2023

Agenda

1. The beginning
2. Pseudo random number generators
3. Discrete distributions
4. The inverse transform method
5. Rejection sampling
6. Transformation of random variables
7. Special-purpose methods

Agenda

1. **The beginning**
2. Pseudo random number generators
3. Discrete distributions
4. The inverse transform method
5. Rejection sampling
6. Transformation of random variables
7. Special-purpose methods

The beginning

Problem. Given a coin and its probability of landing heads is p $(0 < p < 1)$. Without knowing the value of p, find an event whose probability is 50%.

Application. A football referee tosses a coin to determine which end of the field each team plays on during the game. How can we guarantee the fairness of the toss where the referee is unreliable (e.g. using an unfair coin)?!

The beginning (cont.)

Solution. Let's create an event whose probability is 50% by

- *Step 1.* Tossing a coin twice. Let m_1, m_2 be the side showing when it lands in the first and second toss, respectively.
- *Step 2.* Executing repeatedly Step 1 until $m_1 \neq m_2$.
- *Step 3.* Returning event "m_1 is heads".

(See more in Notebook)

The beginning (cont.)

```
def tung_dong_xu(p = 0.7): # unknown p
    return random.choices([0, 1], [1 - p, p])[0]
def tung_dong_xu_can_bang():
    while True:
        m1, m2 = tung_dong_xu(), tung_dong_xu()
        if m1 != m2:
            break
    return m1
def tan_xuat(e, N): return sum(e() for _ in range(N))/N

tan_xuat(tung_dong_xu, 10000)
# 0.7019
tan_xuat(tung_dong_xu_can_bang, 10000)
# 0.5017
```

Agenda

Pseudo random number generators

There are two fundamentally different classes of methods to generate random numbers:

- **True random numbers** are generated using some random physical phenomenon. Generating such numbers requires specialized hardware and can be expensive and slow.
- **Pseudo random numbers** are generated by computer programs. While these methods are normally fast and resource-effective, a challenge with this approach is that computer programs are inherently **deterministic** and therefore cannot produce truly random output.

Pseudo Random Number Generator - PRNG is an algorithm that outputs a sequence of numbers that can be used as a replacement for an **independent and identically distributed** (iid) sequence of true random numbers.

Linear congruential generator - LCG

LCG algorithm.

Input:

- $m > 1$ (modulus)
- $a \in \{1, 2, ..., m-1\}$ (multiplier)
- $c \in \{0, 1, ..., m-1\}$ (increment)
- $X_0 \in \{0, 1, ..., m-1\}$ (seed)

Output: a sequence $X_1, X_2, X_3, ...$ of pseudo random numbers.

for $n = 1, 2, 3, ...$ **do**
  $X_n \leftarrow (aX_{n-1} + c) \mod m$
  output X_n
end for

(See more in Notebook)

Linear congruential generator - Example

For parameter $m = 8, a = 5, c = 1$ and **seed** $X_0 = 0$, LCG gives

n	$5X_{n-1} + 1$	X_n
1	1	1
2	6	6
3	31	7
4	36	4
5	21	5
6	26	2
7	11	3
8	16	0
9	1	1
10	6	6

Linear congruential generator

- While the output of LCG "looks random", it is clear that the output has several properties which make it different from truly random sequences.
- The generated sequence is eventually **periodic**. In the example above, we get $X_8 = X_0, X_9 = X_1, X_{10} = X_2, \dots$
- Since X_n can take only m different values $\{0, 1, \dots, m-1\}$ the **period length** is no longer than m.
- In practice, a typical value of m is usually on the order of $m \approx 2^{32} \approx 4 \times 10^9$ and a, c are then chosen such that the generator actually achieves the maximally possible period length of m.

Linear congruential generator

Theorem. The LCG has period m if and only if the following three conditions are satisfied:

1. m and c are relatively prime,
2. $a - 1$ is divisible by every prime factor of m,
3. if m is a multiple of 4, then $a - 1$ is a multiple of 4.

In addition, the period length does not depend on the seed X_0.

Example: Let $m = 2^{32}, a = 1103515245, c = 12345$:

1. m has 2 as the only prime factor and c is odd, so m, c are relatively prime,
2. m has 2 as the only prime factor and a is even, thus $a - 1$ is divisible by every prime factor of m,
3. m is a multiple of 4 and $a - 1 = 1103515244 = 4 \times 275878811$ is also multiple of 4.

Therefore, the LCG with these parameters has period 2^{32} for every seed X_0 (independent of seed).

Quatily of pseudo random number generators

The criterion for the quality of the output of general PRNGs

- Period length of the output
- Distribution of samples
- Independence of samples

Distribution of samples

- The output of almost all PRNGs is constructed so that it can be used as a replacement for an iid sample of **uniformly distributed** random numbers. Since the output takes values in a finite set $S = \{0, 1, ..., m-1\}$, in the long run, for every set $A \subset S$, we should have

$$\frac{\#\{i | 1 \leq i \leq N, X_i \in A\}}{N} \approx \frac{\#A}{\#S},$$

  where $\#A$ stands for the number of elements in a finite set A.
- Uniformity of the output can be tested using statistical tests like the chi-squared test or the Kolmogorov–Smirnov test.
- Especially, statistical tests should be performed as **two-sided test** to detect when the output has "wrong distribution" (i.e. not every value appears with the same probability), or it could be "too regular".

Distribution of samples - Example

Assume that we have a PRNG with $m = 1024$ possible output values and that we perform a chi-squared test for the hypothesis H_0:

$$P(X_i \in \{64j, 64j+1, 64j+2, ..., 64j+63\}) = 1/16, j = 0, 1, ..., 15.$$

If we consider a sample $X_1, X_2, ..., X_N$, the observed numbers of samples in each block, and the expected count for block j are defined respectively by:

$$O_j = \#\{i | 64j \leq X_i \leq 64(j+1)\}, E_j = N/16, j = 0, 1, ..., 15.$$

For large sample size N, and under the hypothesis H_0, the test statistic of the corresponding chi-squared test is

$$Q = \sum_{j=0}^{15} \frac{(O_j - E_j)^2}{E_j}$$

where the value Q follows a χ^2-distribution with 15 degrees of freedom.

Distribution of samples - Example (cont.)

Some quantiles of the χ^2-distribution with 15 degrees of freedom

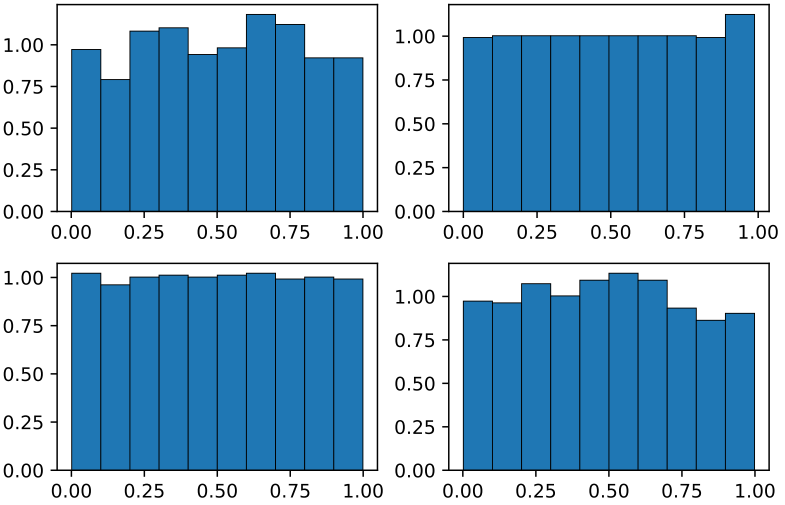
q	6.262	7.261	...	24.996	27.488
$P(Q \leq q)$	0.025	0.05	...	0.95	0.975

Consider some series with size $N = 10^6$:

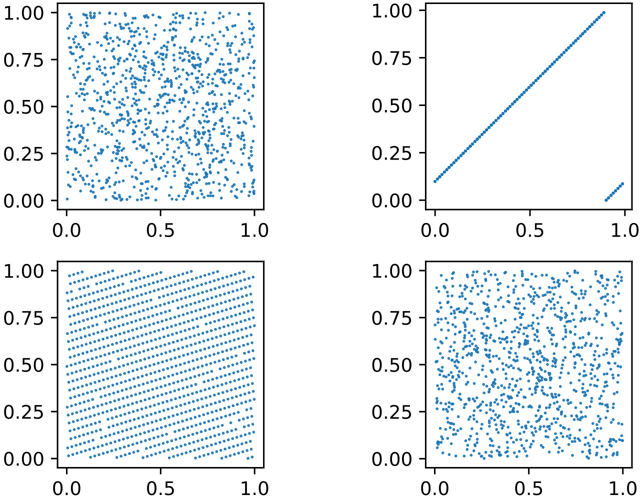
- "Too regular" series: $X_n = n \mod 1024$, we find $Q = 0.256$
- "Wrong distribution" series: $X_n = n \mod 1020$, we find $Q = 233.868$
- The series generated from the LCG $m = 1024, a = 493, c = 123, X_0 = 0$, we find $Q = 0.003$
- The series generated from Python function `numpy.random.randint`, we find $Q = 13.537$ (note, we will have different results each time executing)

(See more in Notebook)

Distribution of samples (cont.)



Independence of samples



Pseudo random number generators in practice

- It is advisable to use a well-established method for random number generation, typically a well-known software package or library, rather than implement your own PRNG.
- Because the seed is to initialise the state of the random number generator, it is useful to set the seed
  - a fixed value to get reproducible results or to ensure repeatability of published results.
  - some volatile quantity (like the current time) to get a different sequence of random numbers for different runs of the program.
- PRNGs often generate a integer sequence $(X_n)_{n\in\mathbb{N}}$, which is uniformly distributed on a finite set $\{0, 1, ..., m-1\}$. To generate a real sequence $(U_n)_{n\in\mathbb{N}}$ uniformly distributed on $(0, 1)$, or $\mathcal{U}(0, 1)$ whose uses to generate samples for other distribution, we can set

$$U_n = \frac{X_n + 1}{m + 1}.$$

Agenda

Discrete distributions

Proposition. Given $p \in [0,1]$ and $U \sim \mathcal{U}[0,1]$, event E is defined by

$$E = \{U \leq p\},$$

thus $P(E) = p$.

A random variable X has discrete uniform distribution on the set $\{0, 1, ..., n-1\}$, denoted by $X \sim \mathcal{U}\{0, 1, ..., n-1\}$, if

$$P(X = k) = \frac{1}{n}, \forall k \in \{0, 1, ..., n-1\}.$$

Proposition. Let $U \sim \mathcal{U}[0,1]$ and $n \in \mathbb{N}$, define a random variable X by

$$X = \lfloor nU \rfloor,$$

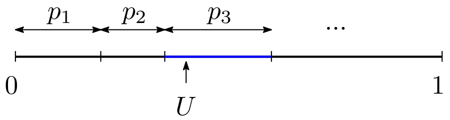
where $\lfloor . \rfloor$ denotes rounding down. Then, $X \sim \mathcal{U}\{0, 1, ..., n-1\}$.

Randomly generating discrete distribution

Proposition. Assume $A = \{a_i : i \in I\}$ where either $I = \{1, 2, ..., n\}$ for some $n \in \mathbb{N}$ or $I = \mathbb{N}$, and where $a_i \neq a_j$ whenever $i \neq j$. Given $(p_i)_{i \in I}$ with $p_i \geq 0, \forall i \in I$ and $\sum_{i \in I} p_i = 1$. Let $U \sim \mathcal{U}[0, 1]$, define a random variable K by

$$K = \min \left\{ k \in I : \sum_{i=1}^{k} p_i \geq U \right\}.$$

Then $X = a_K$ satisfies $P(X = a_k) = p_k, \forall k \in I$.
In conclusion, X is a random variable whose value lies in the subinterval A with probability (p_i).



Randomly generating discrete distribution - Example

Random variable X has a geometric distribution with parameter p $(0 \leq p \leq 1)$ if X takes value from $\mathbb{N} = \{1, 2, ...\}$ and the probability $P(X = i) = p^{i-1}(1-p), i \in \mathbb{N}$.

Suppose that $I = \mathbb{N}, a_i = i, p_i = p^{i-1}(1-p), i \in \mathbb{N}$, then

$$\sum_{i=1}^{k} p_i = (1-p)\sum_{i=1}^{k} p^{i-1} = (1-p)\frac{1-p^k}{1-p} = 1 - p^k.$$

Using the proposition above, $U \sim \mathcal{U}[0, 1]$, we can generate $X = a_K = K$ with

$$K = \min\left\{k \in I : \sum_{i=1}^{k} p_i \geq U\right\} = \min\left\{k \in I : 1 - p^k \geq U\right\}$$

$$= \min\left\{k \in I : k \geq \frac{\log(1-U)}{\log p}\right\} = \left\lceil \frac{\log(1-U)}{\log p} \right\rceil.$$

Agenda

Cumulative distribution function and Inverse

Cumulative distribution function - CDF of a random variable X on $\mathbb{R}$ is defined by

$$F(x) = F_X(x) = P(X \leq x), x \in \mathbb{R}.$$

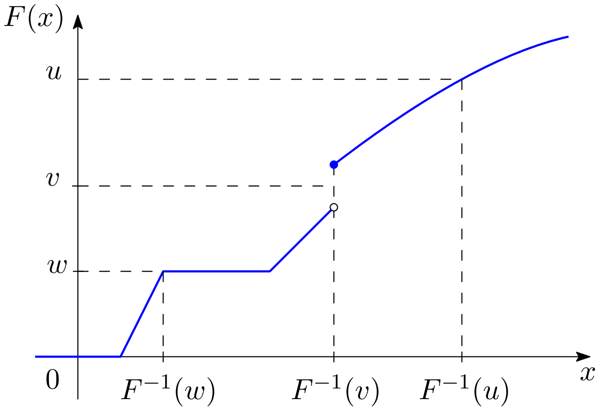
Then, it can be said that X has distribution F, denoted by $X \sim F$.

Let F be a CDF of random variable X, **inverse** of F is defined by

$$F^{-1}(u) = \inf\{x \in \mathbb{R} | F(x) \geq u\}, u \in (0, 1).$$

in the case where F is bijective, inverse F^{-1} is just the usual inverse of F, which means that $F^{-1}(u) = x$ if and only if $F(x) = u$.

Cumulative distribution function and Inverse



Inverse transform method

Inverse Transform Method - ITM
Input: inverse F^{-1} of a CDF F
Output: $X \sim F$

1: generate $U \sim \mathcal{U}[0,1]$
2: **return** $X = F^{-1}(U)$

Theorem. Let $F : \mathbb{R} \to [0,1]$ be a CDF with inverse F^{-1} and $U \sim \mathcal{U}[0,1]$, define random variable by

$$X = F^{-1}(U),$$

then $X \sim F$.

Inverse transform method - Example 1

The exponential distribution has Exp(λ) density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Using partial integration, we find CDF as

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} \lambda e^{-\lambda t} dt = -e^{-\lambda t}\Big|_{t=0}^{x} = 1 - e^{-\lambda x}, \forall x \geq 0.$$

Since this function is strictly monotonically inscreasing and continuous, F^{-1} is the usual inverse of F. We have $1 - e^{-\lambda x} = u \Leftrightarrow -\lambda x = \log(1-u) \Leftrightarrow x = -\log(1-u)/\lambda$ then $F^{-1}(u) = -\log(1-u)/\lambda$.
Thus, with $U \sim \mathcal{U}[0,1]$ then $X = -\log(1-U)/\lambda$ has Exp(λ) distribution.

Inverse transform method - Example 2

Rayleigh distribution with parameter $\sigma > 0$ has density

$$f(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We find the CDF

$$F(x) = \int_0^x \frac{t}{\sigma^2} e^{-t^2/2\sigma^2} dt = -e^{-t^2/2\sigma^2} \Big|_{t=0}^{x} = 1 - e^{-x^2/2\sigma^2}, \forall x \geq 0.$$

Solving the equation $u = F(x) = 1 - e^{-x^2/2\sigma^2}$ for x we have
$F^{-1}(u) = x = \sqrt{-2\sigma^2 \log(1-u)}$.
Thus, with $U \sim \mathcal{U}[0,1]$ then $X = \sqrt{-2\sigma^2 \log(1-U)}$ has Rayleigh distribution with parameter σ.

Inverse transform method - Example 3

Let random variable X have density

$$f(x) = \begin{cases} 3x^2 & \text{if } x \in [0,1] \\ 0 & \text{otherwise.} \end{cases}$$

Then, the CDF

$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & \text{if } x < 0 \\ x^3 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x. \end{cases}$$

Since F maps $(0,1)$ into $(0,1)$ bijectively, $F^{-1}(u) = u^{1/3}, \forall u \in (0,1)$.
Thus, if $U \sim \mathcal{U}[0,1]$, $X = U^{1/3}$ has the same distribution.

Inverse transform method - Example 4

Let X be discrete with $P(X = 0) = 0.6$ and $P(X = 1) = 0.4$. Then, the CDF

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.6 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x. \end{cases}$$

Using the definition of F^{-1}, we find

$$F^{-1}(u) = \begin{cases} 0 & \text{if } 0 < u \leq 0.6 \\ 1 & \text{if } 0.6 < u < 1. \end{cases}$$

Then, if $U \sim \mathcal{U}[0, 1]$,

$$X = \begin{cases} 0 & \text{if } U \leq 0.6 \\ 1 & \text{if } U > 0.6. \end{cases}$$

has the same distribution.

Inverse transform method

- The inverse transform method can always be applied when the inverse F^{-1} is easy to evaluate.
- For some distributions like the normal distribution this is not the case, and the inverse transform method cannot be applied directly.
- ITM can be applied (but may not be useful) for discrete distributions.
- The main restriction of the ITM is that distribution functions only exist in the one-dimensional case.
- For distribution on $\mathbb{R}^d$ $(d > 1)$, more sophisticated methods are required.

Agenda

Basic rejection sampling

Algorithm BRS (Basic Rejection Sampling)

Input:

- a probability sensity function g (the proposal density),
- a function p with values in the interval $[0, 1]$ (the acceptance probability).

Output: a sequence of iid random variables $X_{N_1}, X_{N_2}, X_{N_3}, \dots$ with density

$$f(x) = \frac{1}{Z} p(x) g(x) \text{ where } Z = \int p(x) g(x) dx.$$

for $n = 1, 2, 3, \dots$ **do**
  generate $X_n \sim g$ # the proposals
  generate $U_n \sim \mathcal{U}[0, 1]$
  if $U_n \leq p(X_n)$ **then**
    output X_n #X_n is accepted with the probability $p(X_n)$
  end if #else: X_n is rejected
end for

Basic rejection sampling

Proposition. For $k \in \mathbb{N}$, let X_{N_k} denote the k^{th} output of BRS algorithm, then the following statements hold:

1. The elements of the sequence $(X_{N_k})_{k\in\mathbb{N}}$ are iid with density function

$$f(x) = \frac{1}{Z}p(x)g(x) \text{ where } Z = \int p(x)g(x)dx.$$

2. Each proposal is accepted with probability Z and the number of proposals required to generate each X_{N_k} is geometrically distributed with mean $1/Z$.

Basic rejection sampling - Example

Let $X \sim \mathcal{U}[-1, 1]$ and accepted X with probability

$$p(X) = \sqrt{1 - X^2}.$$

Then, the accepted samples generated from BRS algorithm have density

$$f(x) = \frac{1}{Z} p(x) g(x) = \frac{1}{Z} \sqrt{1 - x^2} \frac{1}{2} \mathbb{I}_{[-1,1]}(x),$$

where

$$\mathbb{I}_{[-1,1]}(x) = \begin{cases} 1 & \text{if } x \in [-1, 1] \\ 0 & \text{otherwise,} \end{cases}$$

and

$$Z = \int \sqrt{1 - x^2} \frac{1}{2} \mathbb{I}_{[-1,1]}(x) dx = \frac{1}{2} \int_{-1}^{1} \sqrt{1 - x^2} dx = \frac{\pi}{4} \approx 0.7854.$$

Then, the density function of X is $f(x) = \frac{2}{\pi} \sqrt{1 - x^2} \mathbb{I}_{[-1,1]}(x)$. This distribution is known as **Wigner's semicircle distribution**.

Envelope rejection sampling

ERS algorithm (Envelope Rejection Sampling)

Input:

- a function f with values in $[0, \infty)$ (the non-normalised target density),
- a probability density function g (the proposal density),
- a constant $c > 0$ such that $f(x) \leq cg(x), \forall x$.

Output: a sequence of iid random variables $X_{N_1}, X_{N_2}, X_{N_3}, \dots$ with density function

$$\tilde{f}(x) = \frac{1}{Z_f} f(x) \quad \text{where} \quad Z_f = \int f(x)dx.$$

for $n = 1, 2, 3, \dots$ **do**
    generate $X_n \sim g$ # the proposals
    generate $U_n \sim \mathcal{U}[0, 1]$
    if $cg(X_n)U_n \leq f(X_n)$ **then**
        output X_n #X_n is accepted with probability $f(X_n)/(cg(X_n))$
    end if #else: X_n is rejected
end for

Envelope rejection sampling

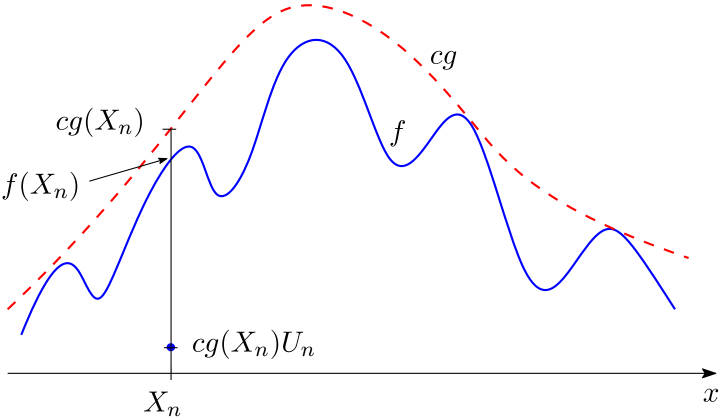
Proposition. For $k \in \mathbb{N}$, let X_{N_k} denote the k^{th} output value of ERS algorithm. Then the following statements hold:

1. The elements of the sequence $(X_{N_k})_{k\in\mathbb{N}}$ are iid with density function

$$\tilde{f}(x) = \frac{1}{Z_f} f(x) \;\text{ where }\; Z_f = \int f(x)dx.$$

2. Each proposal is accepted with probability Z_f/c and the number $M_k = N_k - N_{k-1}$ of proposals required to generate each X_{N_k} is geometrically distributed with mean $E(M_k) = c/Z_f$.

Envelope rejection sampling



Envelope rejection sampling - Example

We can use ERS algorithm to generate samples from the **half-normal distribution** with density function

$$f(x) = \begin{cases} \frac{2}{\sqrt{2\pi}} e^{-x^2/2} & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

If we assume that the proposals are exponential distribution $\mathrm{Exp}(\lambda)$, then the density of the proposals is

$$g(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

In order to apply ERS algorithm, we need to determine a constant $c > 0$ such that $f(x) \leq cg(x), \forall x \in \mathbb{R}$. For $x < 0$ we have $f(x) = g(x) = 0$. For $x \geq 0$, then

$$\frac{f(x)}{g(x)} = \frac{2}{\sqrt{2\pi}\lambda} e^{\left(-x^2/2+\lambda x\right)} \leq \sqrt{\frac{2}{\pi\lambda^2}} e^{\lambda^2/2} = c^*.$$

Envelope rejection sampling - Example (cont.)

Given our choice of g and $c = c^*$, then

$$cg(x)U \leq f(x) \Leftrightarrow \sqrt{\frac{2}{\pi\lambda^2}} e^{\lambda^2/2} \lambda e^{-\lambda x} U \leq \frac{2}{\sqrt{2\pi}} e^{-x^2/2}$$
$$\Leftrightarrow U \leq e^{-\frac{1}{2}(x-\lambda)^2}.$$

This leads to the following algorithm for generating samples from the half-normal distribution:

1: **for** $n = 1, 2, 3, \ldots$ **do**
2: generate $X_n \sim \text{Exp}(\lambda)$
3: generate $U_n \sim \mathcal{U}[0, 1]$
4: **if** $U_n \leq e^{-\frac{1}{2}(X_n-\lambda)^2}$ **then**
5: output X_n
6: **end if**
7: **end for**

Envelope rejection sampling - Example (cont.)

Since f is normalised, $Z_f = 1$. If $c = c^*$ and $\lambda = 1$, then the probability of accepted sample is given by

$$\frac{Z_f}{c} = \frac{1}{c^*} = \frac{1}{\sqrt{\frac{2}{\pi\lambda^2}}e^{\lambda^2/2}} = \sqrt{\frac{\pi}{2e}} \approx 76.02\%.$$

Since the normal distribution $\mathcal{N}(0,1)$ is symmetric, we can generate half-normal distributed values by

1: generate $Z \sim \mathcal{N}(0,1)$
2: output $X = |Z|$

This method has 100% of probability of accepted sample.

Another way to generate half-normal distributed values is as follows

1: generate $Z \sim \mathcal{N}(0,1)$
2: **if** $Z \geq 0$ **then**
3: output $X = Z$

This method has 50% of probability of accepted sample.

Conditional distributions

- Given random variable X and a fixed set A with $P(X \in A) > 0$, conditional distribution $P_{X|X\in A}$ of X if $X \in A$ is defined by

$$P_{X|X\in A}(B) = P(X \in B|X \in A) = \frac{P(X \in B, X \in A)}{P(X \in A)}, \forall B.$$

- Conditional distribution $P_{X|X\in A}$ corresponds to the remaining randomness in X when we already know that $X \in A$.
- Sampling from a conditional distribution can be easily done by the rejection sampling methods.

Rejection sampling for conditional distributions

RSCD algorithm. (Rejection Sampling for Conditional Distributions)
Input: a set A with $P(X \in A) > 0$
Randomness used: q sequence of iid copies $X_1, X_2, X_3, \dots$ of X (the proposals)
Output: a sequence of iid random variables $X_{N_1}, X_{N_2}, X_{N_3}, \dots$ with conditional distribution $P_{X|X \in A}$

for $n = 1, 2, 3, \dots$ **do**
    generate X_n
    if $X_n \in A$ **then**
        output X_n
    end if
end for

Rejection sampling for condition ditributions

Proposition. Let X be a random variable and a set A with $P(X \in A) > 0$. Furthermore, let X_{N_k} denote the k^{th} ($k \in \mathbb{N}$) output value of RSCD algorithm. Then the following statements hold:

1. The elements of the sequence $(X_{N_k})_{k\in\mathbb{N}}$ are iid which have the distribution satisfying

$$P(X_{N_k} \in B) = P_{X|X\in A}(B), \forall B.$$

2. Each accepted proposal has $P(X \in A)$, and the number $M_k = N_k - N_{k-1}$ of proposals required to generate each X_{N_k} is geometrically distributed with mean $E(M_k) = 1/P(X \in A)$.

Rejection sampling for conditional distributions - Example 1

We can use RSCD algorithm to generate samples $X \sim \mathcal{N}(0,1)$ conditioned on $X \geq a$ by executing repeatedly the following two steps until enough samples are output:

1. sinh $X \sim \mathcal{N}(0,1)$,
2. **if** $X \geq a$ **then** output X.

The efficiency of this method depends on the value of a. The following table shows the average number $X \sim \mathcal{N}(0,1)$ of samples required to generate one output sample $X \geq a$ (rounded to the nearest integer)

a	1	2	3	4	5	6
$E(N_a)$	6	44	741	31 574	3 488 556	1 013 594 635

Rejection sampling for conditional distributions - Example 2

We can use ERS algorithm to generate samples from the conditional distribution of $X \sim \mathcal{N}(0,1)$ conditioned on $X \geq a > 0$ (more efficiently than RSCD algorithm). The density function of the conditional distribution is

$$\tilde{f}(x) = \frac{1}{Z} e^{-x^2/2} \mathbb{I}_{[a,\infty)}(x) = \frac{1}{Z} f(x).$$

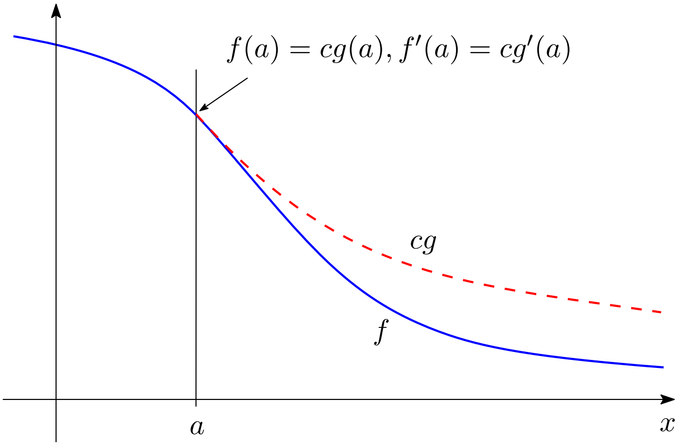
We can use proposals of the form $X = \tilde{X} + a$ where $\tilde{X} \sim \text{Exp}(\lambda)$ and density function

$$g(x) = \lambda e^{-\lambda(x-a)} \mathbb{I}_{[a,\infty)}(x).$$

Then, we find a constant $c > 0$ such that $f(x) \leq cg(x), \forall x \geq a$ and choose λ such that c is being kept to a minimum . The following graph indicates how we should choose parameters (to maximise the efficiency of this method)

$$\lambda = a \;\text{ and }\; c = \frac{e^{-a^2/2}}{a}.$$

Rejection sampling for conditional distributions - Example 2 (cont.)



Rejection sampling for conditional distributions - Example 2 (cont.)

Thus, we can apply ERS algorithm to generate samples $X \sim \mathcal{N}(0,1)$ conditioned on $X \geq a$ by executing repeatedly the following steps until enough samples:

1. generate $\tilde{X} \sim \text{Exp}(a)$,
2. generate $U \sim \mathcal{U}[0,1]$,
3. let $X = \tilde{X} + a$,
4. **if** $U \leq e^{-(X-a)^2/2}$ **then** output X.

We have

$$E(N_a) = \frac{c}{\int f(x)dx} = \frac{\exp(-a^2/2)/a}{\int_a^\infty \exp(-x^2/2)dx} = \frac{\exp(-a^2/2)}{a\sqrt{2\pi}(1-\Phi(a))}.$$

a	1	2	3	4	5	6
$E(N_a)$	1.53	1.19	1.09	1.06	1.04	1.03

Geometric interpretation

Given $A \subset \mathbb{R}^d$, we write $|A|$ for the $d-$dimensional "volume" of set A. For example,

1. The cube $Q = [a, b]^3 \subset \mathbb{R}^3$ has volume $|Q| = (b - a)^3$,
2. The unit circle $C = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 1\}$ has area (2-dimensional volume) $|C| = \pi$,
3. A line segment $L = [a, b] \subset \mathbb{R}$ has length (1-dimensional volume) $|L| = (b - a)$.

For a set $A \subset \mathbb{R}^d$, the volume of A can be found by integration

$$|A| = \int_{\mathbb{R}^d} \mathbb{I}_A(x) dx = \int ... \int \mathbb{I}_A(x_1, ..., x_d) dx_1 ... dx_d.$$

A random variable with values in $\mathbb{R}^d$ is uniformly distributed on a set $A \subset \mathbb{R}^d$ with $0 < |A| < \infty$, denoted by $X \sim \mathcal{U}(A)$, if

$$P(X \in B) = \frac{|A \cap B|}{|A|}, \forall B \subset \mathbb{R}^d.$$

Geometric interpretation

Proposition G1. Let $A \subset \mathbb{R}^d$ be a set with volume $0 < |A| < \infty$. Then the uniform distribution $\mathcal{U}(A)$ on set A has probability density function

$$f(x) = \frac{\mathbb{I}_A(x)}{|A|}, x \in \mathbb{R}^d.$$

Proposition G2. Given $X \sim \mathcal{U}(A)$ and B with $|A \cap B| > 0$, then the conditional distribution of X conditioned on $X \in B$ is uniform distributed on $A \cap B$, which means $P_{X|X \in B} \sim \mathcal{U}(A \cap B)$.

Geometric interpretation - Example

Let $(X_n), (Y_n) \sim \mathcal{U}[-1,1]$ be iid. Then, the pairs (X_n, Y_n) are uniformly distributed on the square $A = [-1,1] \times [-1,1]$. Now let $(Z_k)_{k \in \mathbb{N}}$ be the subsequence of all pairs (X_{n_k}, Y_{n_k}) satisfying

$$X_n^2 + Y_n^2 \leq 1$$

Then $(Z_k)_{k \in \mathbb{N}}$ is an iid sequence, uniformly distributed on the unit circle $B = \{x \in \mathbb{R}^2 : |x| \leq 1\}$.
The probability to accept each sample (X_n, Y_n) is given by

$$p = P((X_n, Y_n) \in B) = \frac{|A \cap B|}{|A|} = \frac{|B|}{|A|} = \frac{\pi}{4} \approx 78.5\%$$

and the number of proposals required to generate one sample is, on average, $1/p \approx 1.27$.

Geometric interpretation

Proposition G3. Let $f : \mathbb{R}^d \to [0, \infty)$ be a probability density function, and let

$$A = \{(x, y) \in \mathbb{R}^d \times [0, \infty) : 0 \leq y < f(x)\} \subset \mathbb{R}^{d+1},$$

Then $|A| = 1$ and the following two statements are equivalent:

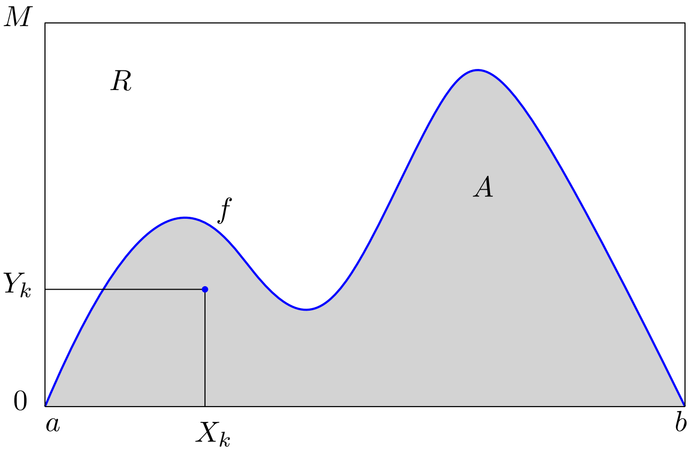
1. (X, Y) is uniformly distributed on A.
2. X is distributed with densify function f on $\mathbb{R}^d$ and $Y = Uf(X)$ where $U \sim \mathcal{U}[0, 1]$, independently of X.

Geometric interpretation

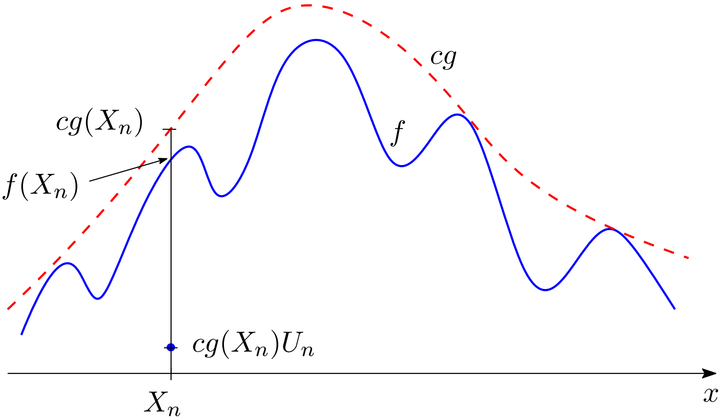
An easy application of Proposition G3 is to convert a uniform distribution of a subset of $\mathbb{R}^2$ to a distribution on $\mathbb{R}$ with a given density function $f : [a, b] \to \mathbb{R}$. For simplicity, we assume that the support of f is on a bounded interval $[a, b]$ and $f(x) \leq M, \forall x \in [a, b]$, we can generate samples from the distribution with density f as follows:

1. Let (X_k, Y_k) are be iid, uniformly distributed on the rectangle $R = [a, b] \times [0, M]$.
2. Consider the set $A = \{(x, y) \in \mathbb{R}^2 : y \leq f(x)\}$ and let $N = \min\{k \in \mathbb{N} : (X_k, Y_k) \in A\}$. By proposition G2, (X_N, Y_N) is uniformly distributed on A.
3. By proposition G3, the value X_N is distributed with density function f.

Geometric interpretation



Geometric interpretation - Illustration of the rejection sampling method



Agenda

Transformation of random variables

TRV Theorem. (Transformation of Random Variables) Let $A, B \subset \mathbb{R}^d$ be open sets, $\varphi : A \to B$ be bijective and differentiable with continuous partial derivatives, and let X be a random variable with valuein A. Furthermore let $g : B \to [0, \infty)$ be a probability density function and define $f : \mathbb{R}^d \to [0, \infty)$ by

$$f(x) = \begin{cases} g(\varphi(x)).|\det D\varphi(x)| & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Then f is a probability density function and the random variable X has density function f if and only if $\varphi(X)$ has density function g.

The matrix $D\varphi(x)$ used in the theorem is the **Jacobian matrix** of φ at x, which is the $d \times d$ matrix consisting of the partial derivatives of φ

$$D\varphi(x)_{ij} = \frac{\partial \varphi_i}{\partial x_j}(x), i, j = 1, 2, ..., d.$$

In case of 1×1 Jacobian matrix ($d = 1$), we have $|\det D\varphi(x)| = |\varphi'(x)|$.

Transformation of random variables - Example 1

Assume that we want to sample Y from the distribution with density function

$$g(y) = \frac{3}{2}\sqrt{y}\mathbb{I}_{[0,1]}(y).$$

We can "cancel the square root" from the definition of g by choosing $\varphi(x) = x^2$. Then, we can apply TRV theorem with $A = B = (0,1)$ and $|\det D\varphi(x)| = |\varphi'(x)| = |2x|$, we get

$$f(x) = g(\varphi(x)).|\det D\varphi(x)| = \frac{3}{2}x.2x = 3x^2, x \in [0,1].$$

We already know how to generate samples from the example of the preceding section. Thus, there are 3 steps to generate Y including:

1. generate $U \sim \mathcal{U}[0,1]$,
2. let $X = U^{1/3}$,
3. output $Y = \varphi(X) = X^2 = U^{2/3}$.

Transformation of random variables - Example 2

Assume that we want to sample (X, Y) from the two-dimensional standard normal distribution, that is from the distribution with density function

$$g(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, (x, y) \in \mathbb{R}^2.$$

We can simplify g using polar coordinates. The corresponding transformation φ is given by

$$(x, y) = \varphi(r, \theta) = (r \cos\theta, r \sin\theta), r > 0, \theta \in (0, 2\pi).$$

where φ is bijective from the open set $A = (0, \infty) \times (0, 2\pi)$ to a open set $B = \varphi(A) = \mathbb{R}^2 \backslash \{(x, y) : x \geq 0, y = 0\}$ in $\mathbb{R}^2$. Note that, if (X, Y) has the two-dimensional standard normal distribution, then

$$P\left((X, Y) \in \{(x, y) | x \geq 0, y = 0\}\right) = 0$$

Thus $P\left((X, Y) \in B\right) = 1$.

Transformation of random variables - Example 2 (cont.)

The Jacobian matrix of φ on (r, θ) is given by

$$D\varphi(r,\theta) = \begin{pmatrix} \frac{\partial \varphi_1}{\partial r} & \frac{\partial \varphi_1}{\partial \theta} \\ \frac{\partial \varphi_2}{\partial r} & \frac{\partial \varphi_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

thus $|\det D\varphi(r,\theta)| = |r\cos^2\theta + r\sin^2\theta| = r$.
Using TRV theorem, we have reduced the problem of sampling (X, Y) from a two-dimensional normal distribution to the problem of sampling (R, θ) from the density function

$$f(r,\theta) = g(\varphi(r,\theta)).|\det D\varphi(r,\theta)| = \frac{1}{2\pi} e^{-r^2/2} r$$

on $(0, \infty) \times (0, 2\pi)$.

Transformation of random variables - Example 2 (cont.)

The density function $f(r,\theta)$ can be rewritten as

$$f(r,\theta) = \frac{1}{2\pi} e^{-r^2/2} r = \left(\frac{1}{2\pi}\right)\left(re^{-r^2/2}\right) = f_1(\theta) f_2(r).$$

Then, (R,Θ) can be generated from $\Theta \sim \mathcal{U}(0,2\pi)$ independent of R with the density function f_2. From the example of the preceding section, we can sample from a two-dimensional normal distribution as follows:

1. generate $\Theta \sim \mathcal{U}(0,2\pi)$,
2. generate $U \sim \mathcal{U}[0,1]$ and let $R = \sqrt{-2\log U}$,
3. output $(X,Y) = \varphi(R,\Theta) = (R\cos\Theta, R\sin\Theta)$.

This method is known as the **Box-Muller transform**.

Ratio-of-uniforms method

RUM Theorem. (Ratio-of-Uniforms Method) Let $f : \mathbb{R}^d \to [0, \infty)$ be such that $Z = \int_{\mathbb{R}^d} f(x)dx < \infty$ and let X be the random vector which is uniformly distributed on the set

$$A = \left\{ (x_0, x_1, ..., x_d) : x_0 > 0, \frac{x_0^{d+1}}{d+1} < f\left(\frac{x_1}{x_0}, ..., \frac{x_d}{x_0}\right) \right\} \subset [0, \infty) \times \mathbb{R}^d.$$

Then the random vector

$$Y = \left(\frac{X_1}{X_0}, ..., \frac{X_d}{X_0}\right)$$

has density function $\frac{1}{Z} f$ on $\mathbb{R}^d$.

Ratio-of-uniforms method - Example

The Cauchy distribution has density function

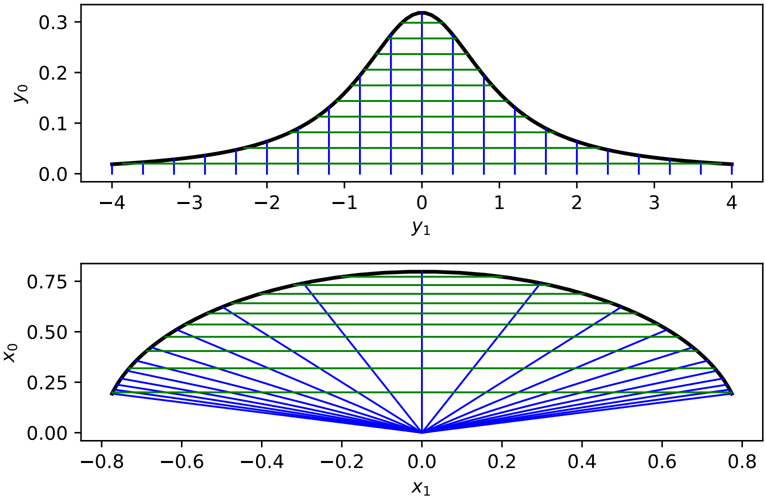
$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Suppose that

$$\begin{aligned} A &= \left\{ (x_0, x_1) : x_0 > 0, \frac{x_0^2}{2} < f\left(\frac{x_1}{x_0}\right) = \frac{1}{\pi\left(1 + \left(\frac{x_1}{x_0}\right)^2\right)} \right\} \\ &= \left\{ (x_0, x_1) : x_0 > 0, x_0^2 + x_1^2 \leq \frac{2}{\pi} \right\}. \end{aligned}$$

Using RUM theorem, if (X_0, X_1) is uniformly distributed on A, $Y = X_1/X_0$ has Cauchy distribution.

Ratio-of-uniforms method - Example (cont.)



Agenda

Special-purpose methods

- The methods discussed in the preceding sections are general purposes that can be used for a wide range of distributions on different applications.
- There are many specialised methods to generate samples from specific distributions. These specialised methods are often faster than the generic ones but quite complex also.
- Reference: **Chapter 4.** J. S. Dagpunar. *Simulation and Monte Carlo - With applications in finance and MCMC*. John Wiley & Sons, 2007.

References

Chapter 1. Jochen Voss. *An Introduction to Statistical Computing - A Simulation-based Approach*. John Wiley & Sons, 2014.

Chapter 2-4. J. S. Dagpunar. *Simulation and Monte Carlo - With applications in finance and MCMC*. John Wiley & Sons, 2007.