Direct sampling methods

Random & pseudo-random numbers

There exist several ways to generate so-called "random" numbers according to known distributions

NB: computer programs do not generate truly random numbers

Rather **pseudo-random**, which seem random but are actually generated by a deterministic process (depending on a "**seed**" parameter).

Uniform sample generation

Linear congruential algorithm: sample pseudo-random numbers according to the Uniform distribution on $[0,1]$ (Lehmer, 1948)

1. Generate a sequence of integers y_n such as: $y_{n+1} = (ay_n + b) \text{ mod. } m$
2. $x_{n+1} = \frac{y_{n+1}}{m-1}$

choose a, b and m so that y_n has a long period & $(x_1, \dots, x_n)$ can be considered *iid*

with y_0 the "**seed**", i.e. the starting point

Remark: $0 \leq y_n \leq m-1 \Rightarrow$ in practice m very large (e.g. 2^{19937}, default in R which uses the Mersenne-Twister variation)

In the following, sampling pseudo-random numbers uniformly on $[0,1]$ will be considered reliable and used by the different sampling algorithms

Other usual distributions

Relying on **relationships between the different usual distributions** starting from $U_i \sim \mathscr{U}_{[0,1]}$

Other usual distributions

Relying on **relationships between the different usual distributions** starting from $U_i \sim \mathcal{U}_{[0,1]}$

Binomial $Bin(n,p)$:

$$Y_i = \mathbb{1}_{U_i \leq p} \sim \text{Bernoulli}(p)$$
$$X = \sum_{i=1}^{n} Y_i \sim Bin(n,p)$$

Other usual distributions

Relying on **relationships between the different usual distributions** starting from $U_i \sim \mathcal{U}_{[0,1]}$

Binomial $Bin(n,p)$:

$$Y_i = \mathbb{1}_{U_i \leq p} \sim \text{Bernoulli}(p)$$
$$X = \sum_{i=1}^{n} Y_i \sim Bin(n,p)$$

Normal $\mathcal{N}(0,1)$ (Box-Müller algorithm):

U_1 and U_2 are 2 independent uniform variables on $[0;1]$

$$Y_1 = \sqrt{-2\log U_1}\cos(2\pi U_2)$$
$$Y_2 = \sqrt{-2\log U_1}\sin(2\pi U_2)$$

$\Rightarrow$ Y_1 & Y_2 are independent random variables each following a $\mathcal{N}(0,1)$

Inverse transform sampling

Definition: For a function F defined on $\mathbb{R}$, its **generalized inverse** is defined as: $F^{-1}(u) = \inf\{x \text{ such that } F(x) > u\}$

Inverse transform sampling

Definition: For a function F defined on $\mathbb{R}$, its **generalized inverse** is defined as: $F^{-1}(u) = \inf\{x \text{ such that } F(x) > u\}$

Property: Let
- F be a cumulative probability distribution function
- U be a uniform random variable on $[0,1]$

Then $F^{-1}(U)$ defines a random variable whith cumulative probability distribution function F

If
1. one knows F, the cumulative probability distribution function from which to sample
2. one can invert F

$\Rightarrow$ then one can sample this distribution from a uniform sample on $[0,1]$

Inverse transform sampling: illustration

Example: sample from the Exponential distribution with parameter λ

Inverse transform sampling: illustration

Example: sample from the Exponential distribution with parameter λ

- density of the Exponential distribution: $f(x) = \lambda \exp(-\lambda x)$
- its cumulative probability distribution function (its integral): $F(x) = 1 - \exp(-\lambda x)$

Let $F(x) = u$

Then $x = \dots$

Inverse transform sampling: illustration

Example: sample from the Exponential distribution with parameter λ

- density of the Exponential distribution: $f(x) = \lambda \exp(-\lambda x)$
- its cumulative probability distribution function (its integral): $F(x) = 1 - \exp(-\lambda x)$

Let $F(x) = u$

Then $x = -\dfrac{1}{\lambda}\log(1-u)$

$\Rightarrow$ and if $U \sim U_{[0;1]}$, then $X = F^{-1}(U) = -\frac{1}{\lambda}\log(1-U) \sim E(\lambda)$.

Your turn !

Practical: exercise 3