03 - Simulation - 03

Master in Foundations of Data Science
Bayesian Statistics and Probabilistic Programming
Fall 2018-2019

Josep Fortiana

Wednesday, October 10, 2018

Universitat de Barcelona

03 - Simulation - 03

Monte Carlo methods

Simple acceptance-rejection RNG

Acceptance-rejection with a candidate function

03 - Simulation - 03

Monte Carlo methods

Simple acceptance-rejection RNG

Acceptance-rejection with a candidate function

The name

Monte Carlo: collective name given to a class of algorithms to solve numerical problems by means of random numbers.

The name refers to gambling in the famous casino city in Monaco.

Metropolis, Nicholas (1987). *The beginning of Monte Carlo methods.* Los Alamos Science - Special Issue.

Monte Carlo methods in Physics

Enrico Fermi and *Stanislaw Ulam*, pioneering researchers, began introducing these methods to problems in Physics.

Fermi, in the 1930's, to perform computations on the physics of the recently discovered neutron.

Ulam, jointly with *John von Neumann* and *Nicholas Metropolis*, within the framework of the *Manhattan project* (II World War atomic bomb) at *Los Alamos*.

Stanislaw Ulam with the FERMIAC

A mechanical analog computer for Monte Carlo simulation of neutron diffusion (Enrico Fermi, 1947).

Monte Carlo methods for us

Nowadays, applications of Monte Carlo numerical simulation range over many, almost all, specialities in Engineering, in Physics, Computational Economics and Finance and, what is currently our main interest, in most computations in Bayesian Statistics.

Monte Carlo integration

Numerical definite integral of a real function on an interval $[a, b]$.

Throw uniform random points on a rectangle $[a, b] \times [c, d]$ containing all the function values.

Approximate the area of the region under the function graph with the proportion of points inside it it times the rectangle area.

Intuitively: why does it work?

With a uniform distribution on $\mathscr{R} = [a, b] \times [c, d]$, the probability of a region $\mathscr{A} \subset \mathscr{R}$ is:

$$\mathrm{P}(\mathscr{A}) = \frac{\mathrm{Area}(\mathscr{A})}{(b-a) \times (d-c)}.$$

The relative frequency of random points $\sim \mathrm{Unif}(\mathscr{R})$ falling inside $\mathscr{A}$ approximates $\mathrm{P}(\mathscr{A})$. From the LLN, this approximation improves with a larger sample size.

The Gaussian integral

Integral of the Gaussian *bell-shaped* function (proportional to the standard normal pdf)

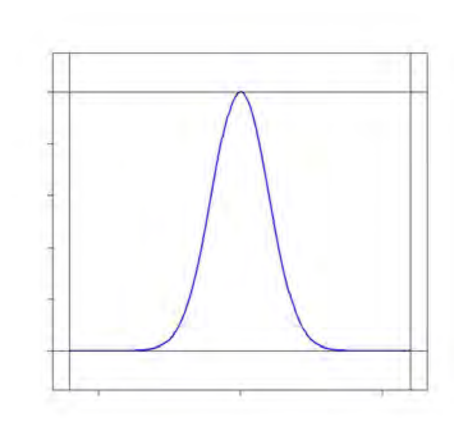
$$f(x) = \exp\left\{-\frac{1}{2}x^2\right\}, \quad -\infty < x < \infty.$$

We consider $f \approx 0$ outside of: $(a, b) = (-6, 6)$.

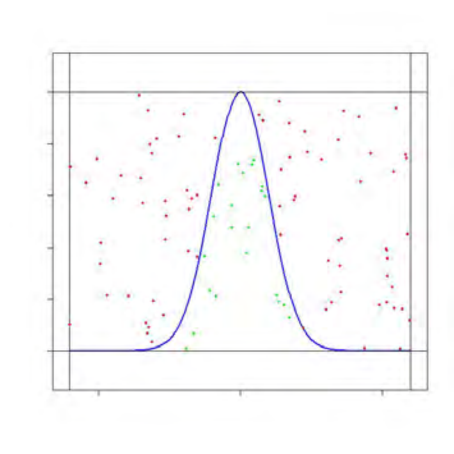
The Gaussian integral

The y interval: $(c, d) = (0, 1)$ contains all of f values.

These two intervals span our rectangle.

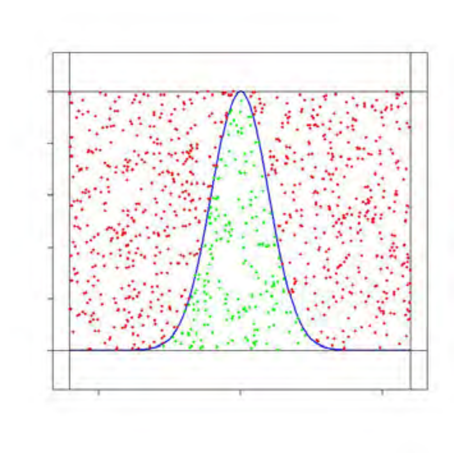


For $N = 100$, $I = 2.64$

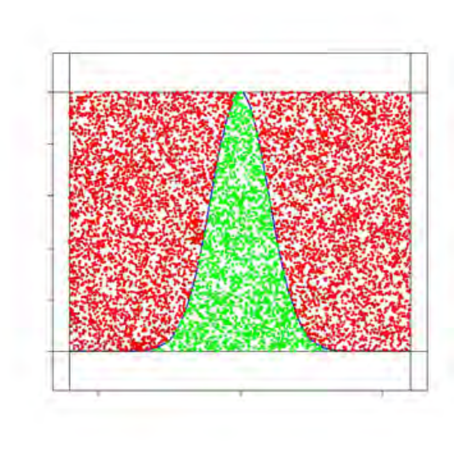


The well-known exact value is $I = \sqrt{2\,\pi} = 2.506628$.

For $N = 1000$, $I = 2.364$



For $N = 10000$, $I = 2.5716$



03 - Simulation - 03

Need for more methods of RNG

- The inverse cdf transformation method of RNG is valid only for one-dimensional r.v.
- With many r.v. either the cdf or its inverse or both of them must be numerically computed.

Example: Beta(α, β)

Take, for instance, Beta$(2, 2)$, whose pdf and cdf are:

$$f(x) = 6\,x\,(1 - x) = 6\,(x - x^2),$$

$$F(x) = \begin{cases} 0, & \text{if} \;\; x < 0, \\ x^2\,(3 - 2x), & \text{if} \;\; 0 \leq x < 1, \\ 1, & \text{if} \;\; 1 \leq x. \end{cases}$$

f attains its maximum value, 1.5, at $x = 0.5$.

Simple accept-reject procedure

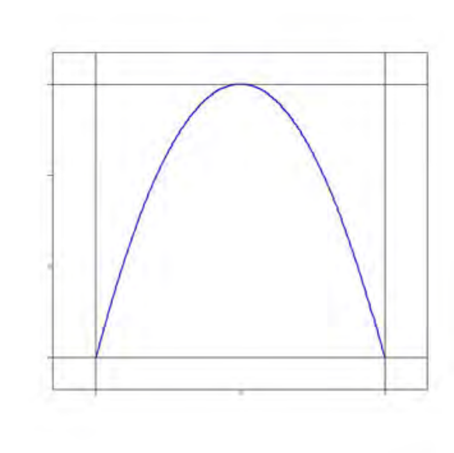
Throw uniform random points (u, v) on a rectangle $[a, b] \times [0, K]$ containing all the pdf values.

For $\mathrm{Beta}(2, 2)$, this is $[0, 1] \times [0, 1.5]$.

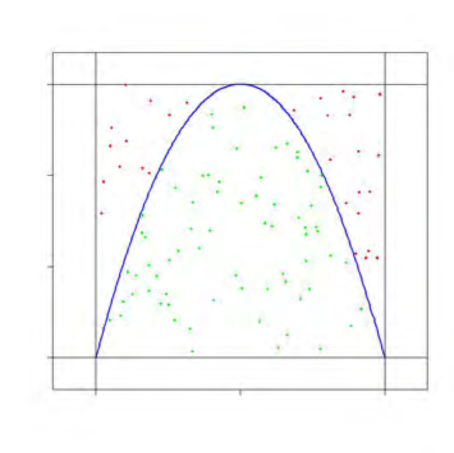
When a point (u, v) falls under the curve, i.e., $v \leq f(u)$, keep u in the list, otherwise throw it away.

Continue until the list is sufficiently long.

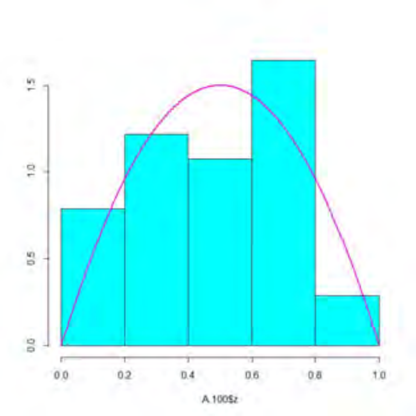
pdf of the Beta$(2, 2)$ distribution



Accept-reject with $N = 100$

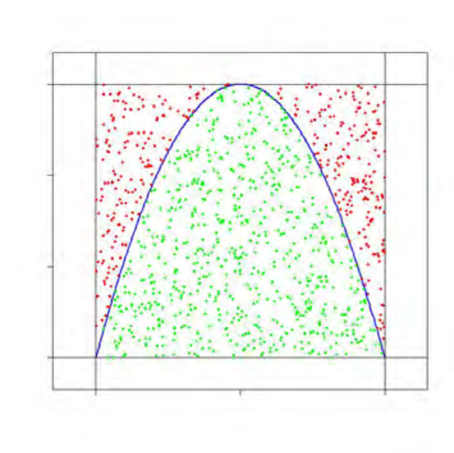


Accept-reject with $N = 100$. Histogram

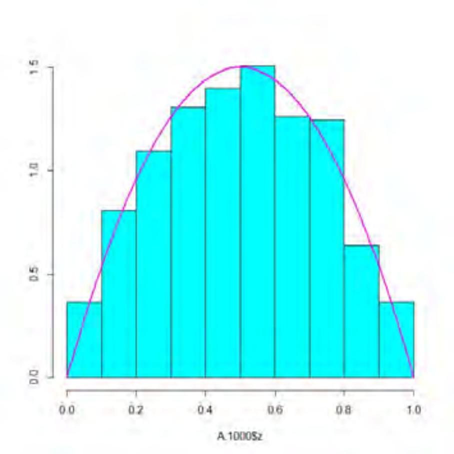


Acceptance rate: 0.7 ($\approx 1/1.5 = 0.6667$).

Accept-reject with $N = 1000$

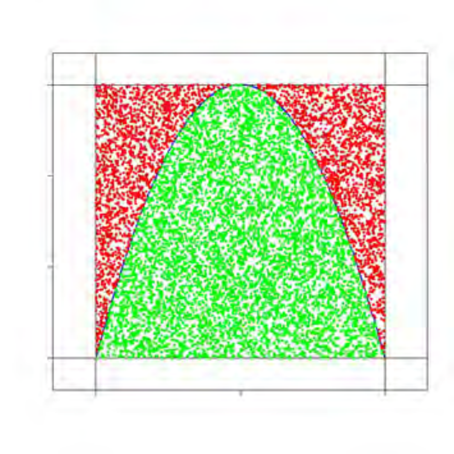


Accept-reject with $N = 1000$. Histogram

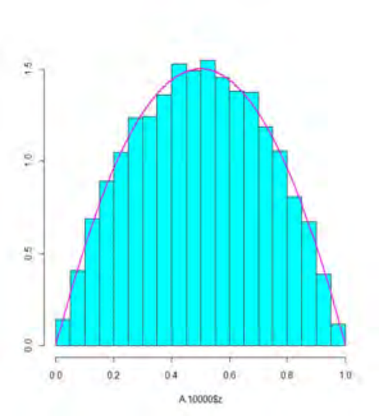


Acceptance rate: 0.657 ($\approx 1/1.5 = 0.6667$).

Accept-reject with $N = 10000$



Accept-reject with $N = 10000$. Histogram



Acceptance rate: 0.6715 ($\approx 1/1.5 = 0.6667$).

Intuitively, why does it work?

The number of accepted points (u, v) with u in:

$$(x, x + \Delta) \subset [a, b], \qquad \text{Δ a small value,}$$

is proportional to the area of the rectangle:

$$\approx \Delta \times f(x)$$

that is, $\mathrm{P}(x, x + \Delta)$ from the distribution with pdf f.

A formal proof

Straightforward computation of a conditional probability.

We are given a pdf f, whose cdf is F.

Throw uniform random points (u, v) on a rectangle $[a, b] \times [0, K]$ containing all the pdf values.

cdf of the accepted points

Take $U \sim \mathrm{Unif}(a, b)$, $V \sim \mathrm{Unif}(0, K)$. Their pdf's are:

$$f_U = \frac{1}{b-a}\ \mathbb{1}_{(a,b)}, \qquad f_V = \frac{1}{K}\ \mathbb{1}_{(0,K)}\,.$$

The accepted numbers r.v. is: $X \stackrel{\text{def}}{=\!=} [U | V \leq f(U)]$

We check the cdf of X is, indeed, F.

cdf of the accepted points

Given $x \in (a, b)$, the value at x of the wanted cdf is:

$$\mathrm{P}(U \leq x | V \leq f(U)) = \frac{\mathrm{P}(U \leq x,\ V \leq f(U))}{\mathrm{P}(V \leq f(U))}$$

$$= \frac{\displaystyle\int_a^x \frac{1}{b-a}\,du \cdot \int_0^{f(u)} \frac{1}{K}\,dv}{\displaystyle\int_a^b \frac{1}{b-a}\,du \cdot \int_0^{f(u)} \frac{1}{K}\,dv}$$

cdf of the accepted points

$$= \frac{\displaystyle\int_a^x \frac{1}{b-a}\,du \cdot \frac{f(u)}{K}}{\displaystyle\int_a^b \frac{1}{b-a}\,du \cdot \frac{f(u)}{K}} = \frac{\displaystyle\int_a^x f(u)\,du}{\displaystyle\int_a^b f(u)\,du}$$

$$= \frac{F(x)}{1} = F(x).$$

Probability of acceptance

It is the denominator in the expression above:

(before simplification)

$$\mathrm{P}(V \leq f(U)) = \frac{1}{(b-a)\,K}.$$

Drawbacks of this method

- Valid only for pdf's with a compact support.

- Has a large rate of rejection.

(large empty space above f)

03 - Simulation - 03

Monte Carlo methods

Simple acceptance-rejection RNG

Acceptance-rejection with a candidate function

A candidate function g, given the target pdf f

We seek (or design) another pdf g such that:

- f, g have equal support:

$$(f(x) > 0 \iff g(x) > 0)$$

- There is a constant M such that

$$f(x) \leq M \cdot g(x) \text{ for all } x$$

- We know how to generate RN $\sim g$

Accept-reject procedure

Generate a pair of random numbers (u, v),

- u from the distribution with pdf g,
- v from a $\mathrm{Unif}(0, 1)$.

$$\text{If } v \leq \frac{f(u)}{M \cdot g(u)}, \ \text{keep } u\text{, otherwise discard it.}$$

Continue until the list is sufficiently long.

Why does it work?

We are given a pdf f, whose cdf is F, and we have found g and M auch that:

$$f(x) \leq M \cdot g(x), \quad \text{for all } \; x.$$

We generate the pair of r.v. $U \sim g$ and $V \sim \mathsf{Unif}(0, 1)$.

Same conditional probability computation as above.

Why does it work?

We check the cdf of

$$X \stackrel{\text{def}}{=} \left[U \,\middle|\, V \leq \frac{f(U)}{M \cdot g(U)} \right]$$

is, indeed, F.

For $x \in \mathbb{R}$, we compute the value of this cdf at x.

cdf of the accepted points

$$\mathrm{P}\left(U \le x \,\middle|\, V \le \frac{f(U)}{M \cdot g(U)}\right) = \frac{\mathrm{P}\left(U \le x,\ V \le \frac{f(U)}{M \cdot g(U)}\right)}{\mathrm{P}\left(V \le \frac{f(U)}{M \cdot g(U)}\right)}$$

$$= \frac{\displaystyle\int_{-\infty}^{x} g(u)\, du \cdot \int_{0}^{f(u)/M\, g(u)} dv}{\displaystyle\int_{-\infty}^{\infty} g(u)\, du \cdot \int_{0}^{f(u)/M\, g(u)} dv}$$

cdf of the accepted points

$$= \frac{\displaystyle\int_{-\infty}^{x} g(u)\,du \cdot \frac{f(u)}{M\,g(u)}}{\displaystyle\int_{-\infty}^{\infty} g(u)\,du \cdot \frac{f(u)}{M\,g(u)}} = \frac{\displaystyle\int_{-\infty}^{x} du\,f(u)}{\displaystyle\int_{-\infty}^{\infty} du\,f(u)}$$

$$= \frac{F(x)}{1} = F(x).$$

Rate of acceptance

The expected rate of acceptance is:

$$\mathrm{P}\left(V \leq \frac{f(U)}{M \cdot g(U)}\right) = \frac{1}{M}.$$

Two important features of the accept-reject method

- Can be applied to multi-dimensional densities
- f can be known up to a multiplicative constant.