Sampling via Rejection

Mihir Arjunwadkar

Centre for Modeling and Simulation
Savitribai Phule Pune University

When is inverse-CDF sampling useful?

- For the transformation method / inverse-CDF sampling to work, it should be possible to compute $F_X^{-1}(x)$
    - efficiently; and
    - in a numerically stable fashion.
- Hence, this method is (most) useful when a closed-form expression for the inverse CDF $F_X^{-1}(x)$ is available and is easy to compute.

What if it is not?

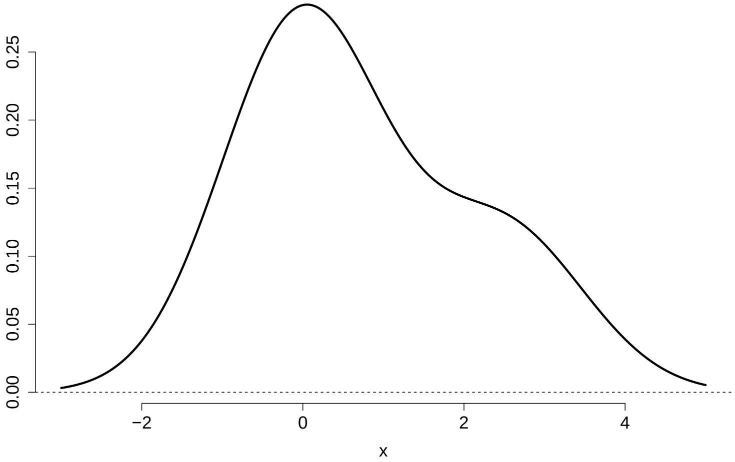
For example, when

- $F_X^{-1}(x)$ is difficult or expensive to compute;
- only the PDF is available / computable (i.e., $F_X^{-1}(x)$ is not).

We need a method for such situations.

What if it is not?

$f_X(x) = 0.7 \times N(0,1) + 0.3 \times N(2.5,1)$

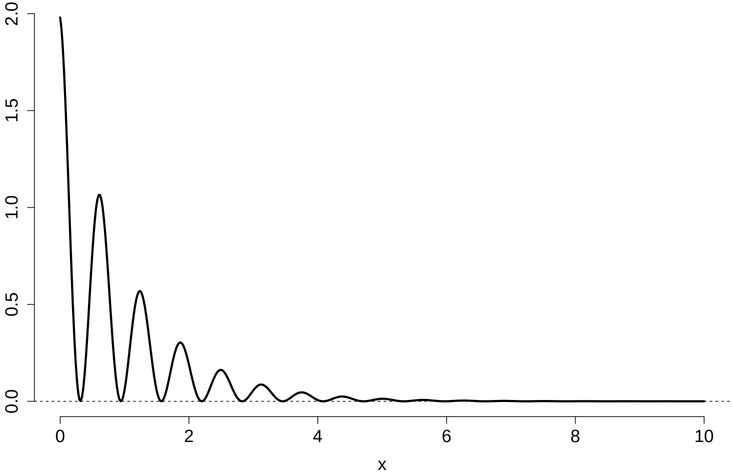


$F_X^{-1}(x)$: Not available in an easily computable form. Univariate transformation is NOT available; can't use the inverse-CDF method.

However, an efficient method for normal mixtures is available.

What if it is not?

$f_X(x) = \frac{101}{102} e^{-x}(1 + \cos(10x))$ for $x \geq 0$

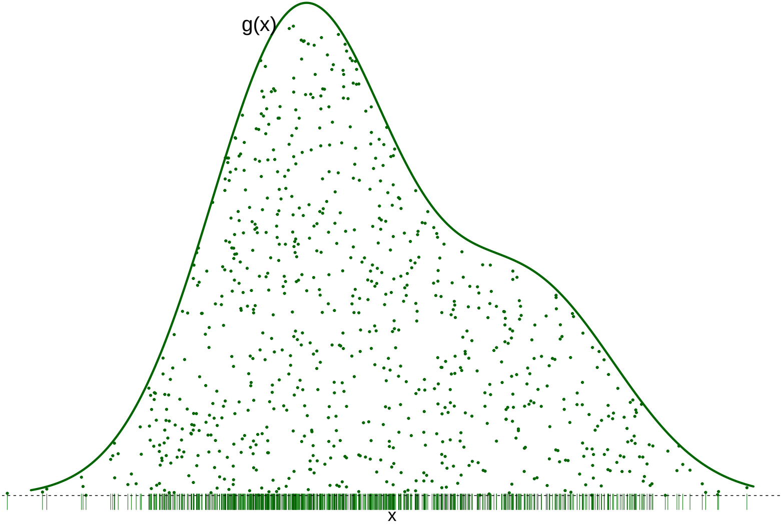


$F_X(x) = 1 - \frac{1}{102} e^{-x} \left(101 + \cos(10x) - 10 \sin(10x)\right)$

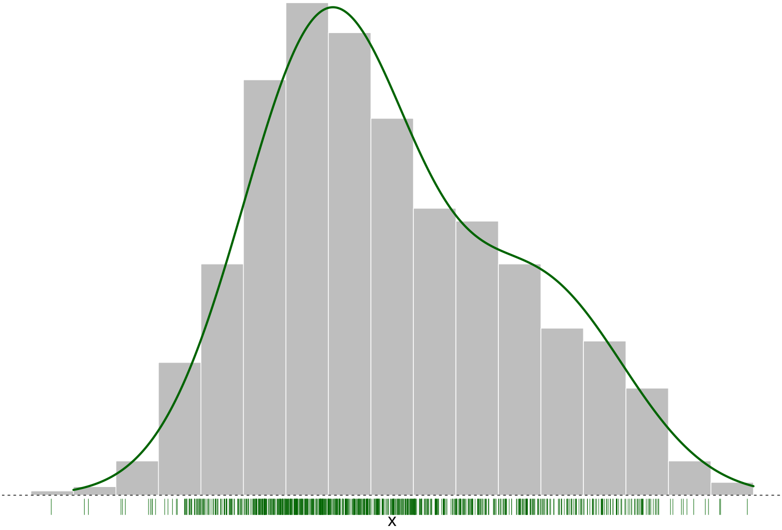
Difficult/impossible to invert analytically: Can't use the inverse-CDF method.

The key idea behind rejection sampling

If area under $g(x) \geq 0$ is sampled with uniform density ...



… then (marginal) PDF of X has the same shape as $g(x)$



... then (marginal) PDF of X has the same shape as $g(x)$

By assumption,

$$f_{XY}(x,y) = \begin{cases} c & \text{if } 0 \le y \le g(x) \\ 0 & \text{otherwise} \end{cases},$$

where the constant c is determined by the normalization condition

$$\int f_{XY}(x,y)dxdy = 1.$$

Therefore, the marginal PDF of X is

$$f_X(x) = \int f_{XY}(x,y)dy = c\int_0^{g(x)} dy = cg(x)$$

Hence the result: $f_X(x)$ has the same *shape* as $g(x)$.

Incidentally, $c = \frac{1}{\int g(x)dx}$.

A sampling prescription at a high level

Suppose f_X is the *target* or *desired* density;
i.e., the PDF which you wish to sample from.

High-level prescription

- Sample the area bounded by the x-axis and f_X with uniform density.
- Simply ignore the y coordinates of the sampled points.
- Take only the x coordinates.

A sampling prescription at a high level

But how does one sample the region $0 \leq y \leq f_X(x)$ with uniform density?

How to sample area between $g(x)$ & x-axis uniformly

Consider a function g such that $g(x) \geq 0 \; \forall x$ and $\int g(x)dx < \infty$.

Then $f_X(x) = \dfrac{g(x)}{\int g(t)dt}$ is a valid PDF.
Suppose a sampler/RNG is available for the density f_X.

Prescription

1. Sample a random x from f_X; i.e., $X \sim f_X$.
2. Sample a random y at this x as $Y|(X = x) \sim \text{Uniform}(0, g(x))$.

How to sample area between $g(x)$ & x-axis uniformly

Claim: (x, y) pairs generated by this prescription have a uniform density over the region between the curve $g(x)$ and the x-axis.

We have

$X \sim f_X$, where $f_X(x) = \dfrac{g(x)}{\int g(t)dt}$

$Y|(X = x) \sim \text{Uniform}(0, g(x)) \implies f_{Y|X}(y|x) = \begin{cases} \frac{1}{g(x)} & 0 \leq y \leq g(x) \\ 0 & \text{otherwise} \end{cases}$

Hence,

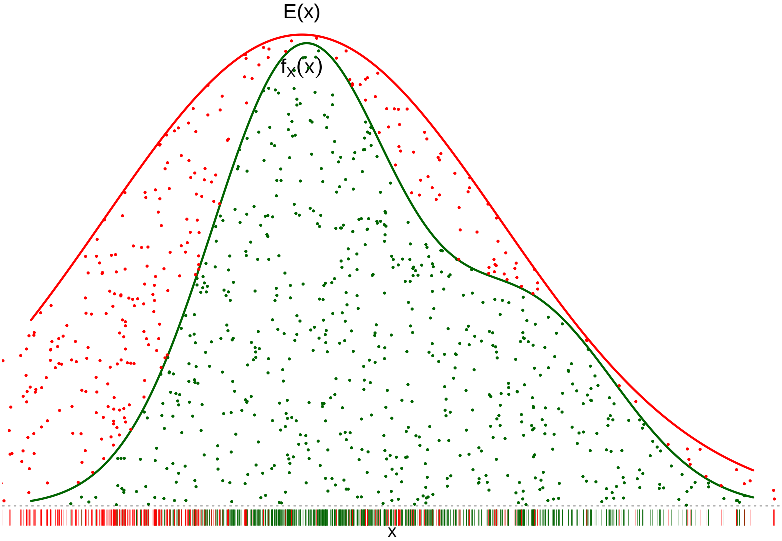
$f_{X,Y}(x, y) = f_X(x) f_{Y|X}(y|x) = \begin{cases} \frac{1}{\int g(t)dt} & 0 \leq y \leq g(x) \\ 0 & \text{otherwise} \end{cases}$

$f_{X,Y}(x, y)$ is non-zero and constant precisely over the region of interest.

But then ...

If we knew how to sample from f_X,
why would we try to design a method for the same purpose?

The missing piece: Envelope function $E(x)$



Rejection method

The complete prescription

- Suppose $f_X(x)$: Target or desired PDF we wish to sample from.
- Choose $E(x) \geq f_X(x)$ such that $E(x) = ke(x)$, where $e(x)$ is some PDF which we know how to sample from. Let us call $E(x)$ the *envelope function*, and $e(x)$ the *envelope* PDF.
- Sample $X \sim e$ $\qquad \Longrightarrow x$
- Sample $Y|(X = x) \sim \text{Uniform}(0, E(x))$ $\qquad \Longrightarrow y$
- Reject those points (x, y) for which $y > f_X(x)$.
- x-coordinates of points not rejected form a sample from f_X.

Pseudocode: One (random) number at a time

Algorithm 1 Rejection sampling, one (random) number at a time

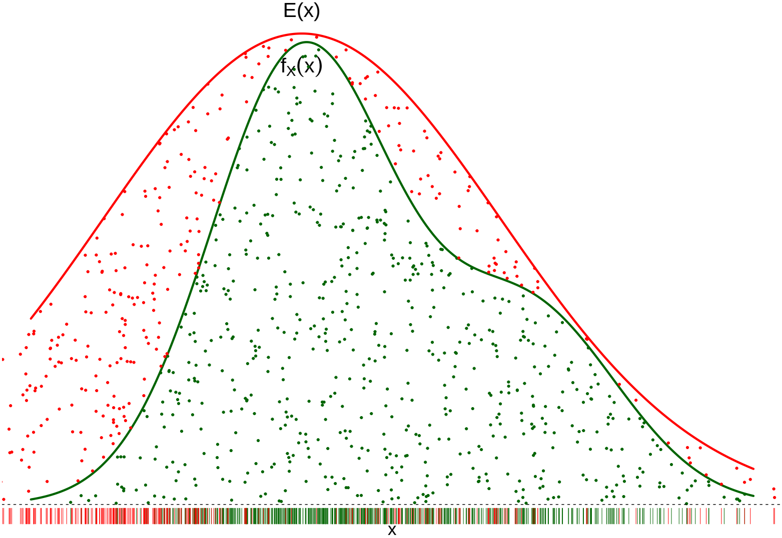
```
Require: Target PDF $f_X$; envelope PDF $e$; constant $k > 0$ s.t. $E(x) = ke(x) \geq f_X(x)\ \forall x$; number $N$ of random numbers required.
$i \leftarrow 0$
while $i \neq N$ do
    Sample $x \sim e$
    Sample $u \sim \text{Uniform}(0, E(x))$
    if $u \leq f_X(x)$ then
        Store $x$
        $i \leftarrow i + 1$
    else
        Reject $x$
    end if
end while
```

A schematic vectorized implementation in R

```
# The code below is for illustrative purposes.

# Require
#  -- density():   R function, the target univariate PDF
#  -- denvelope(): R function, the chosen envelope PDF
#  -- k:           k > 0, such that k * denvelope(x) >= density(x) for all x
#  -- renvelope(): R function, a random number generator for the envelope PDF
#  -- n:           n > 0, the # of random numbers generated from the envelope

# densities from the previous figure
#
#  density   <- function( x ) { 0.7*dnorm( x ) + 0.3*dnorm( x, mean = 2.5 ) }
#  renvelope <- function( n ) { rnorm( n, sd = 2.2 ) }
#  denvelope <- function( n ) { dnorm( n, sd = 2.2 ) }
#  k         <- 1.6

# (x,y) sampled below fill the area below k*denvelope(x) with uniform density
#
 x <- renvelope( n )
 y <- runif( n, 0, k * denvelope( x ) )

# a random sample from the target density
#
 x[y <= density( x )] # length <= n
```

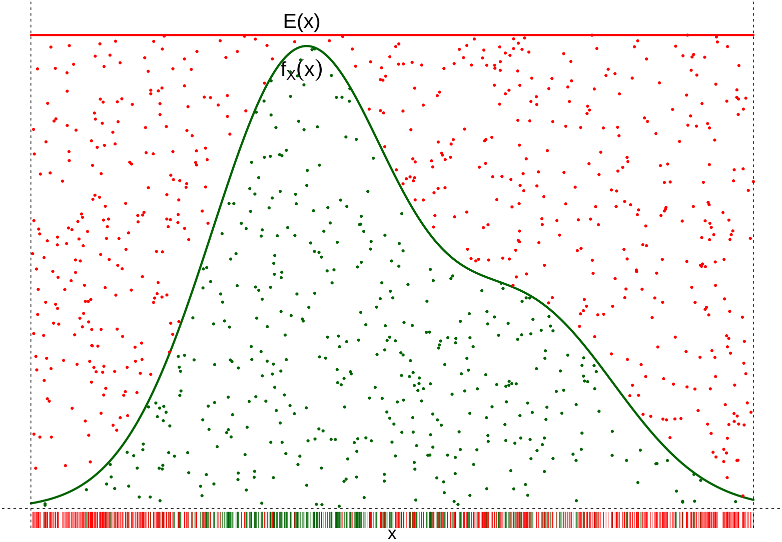
How efficient is a rejection method?

Compare this envelope ...



How efficient is a rejection method?

... with this one

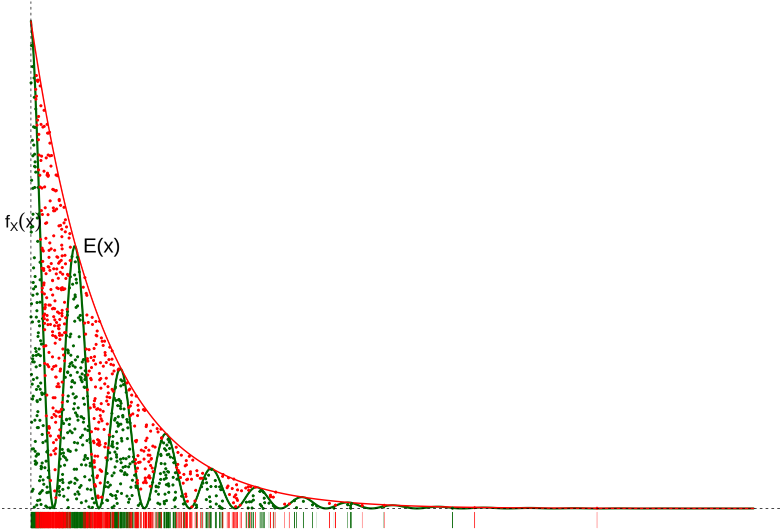


How efficient is a rejection method?

Which of the two envelopes is “better” for the target PDF, and why?

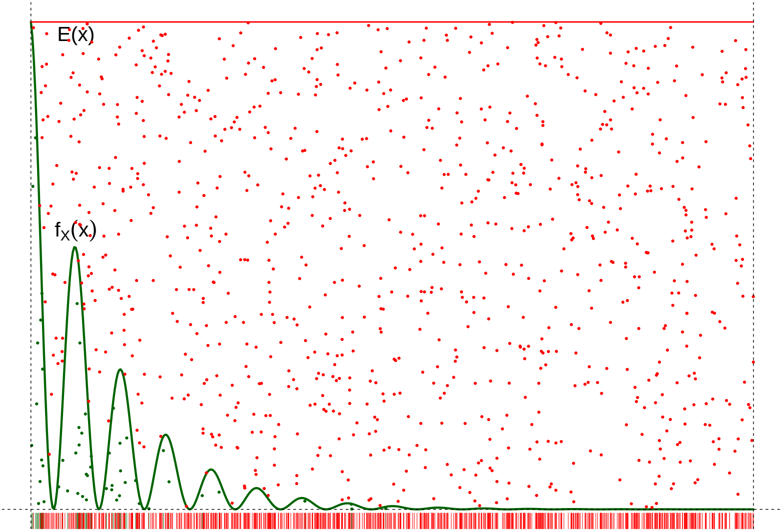
How efficient is a rejection method?

Compare this envelope ...



How efficient is a rejection method?

... with this one



How efficient is a rejection method?

Which of the two envelopes is “better” for the target PDF, and why?

How efficient is a rejection method?

Efficiency of a rejection sampler; i.e., the probability of acceptance is

$$\frac{\int f_X(x)dx}{\int E(x)dx} = \frac{1}{k\int e(x)dx} = \frac{1}{k}.$$

$k \to 1$ when $e(x) \to f_X(x)\ \forall x$.

In practice, choose e that mimics the shape of f_X as closely as possible, and choose the smallest k such that $E(x) = ke(x) \geq f_X(x)\ \forall x$.

The rejection method in its general form is attributed to von Neumann, but similar ideas existed; e.g., Buffon's needle, estimating π by throwing darts, etc.

A refinement of the vanilla rejection sampler

Adaptive rejection sampling; see details in, e.g.,

- `http://www1.maths.leeds.ac.uk/~wally.gilks/adaptive.rejection/web_page/Welcome.html`
- `http://www2.stat.duke.edu/~cnk/Links/slides.pdf`

etc.