Sampling from unknown distributions

Feng Li
feng.li@cufe.edu.cn

School of Statistics and Mathematics
Central University of Finance and Economics

Revised on April 17, 2017

Today we are going to learn...

Only need Uniform

- Assume that we have a way to simulate from a uniform distribution between 0 and 1, $u \sim U(0, 1)$
- If this is available, it is possible to simulate many other probability distributions.
- The most simple method is the **Direct Method**

↬ Discrete Case: Example 1

- Assume that we want to simulate a binary variable X with $\Pr(X = 0) = 0.3$ and $\Pr(X = 1) = 0.7$
- Let $u \sim U(0, 1)$. Then the following rule can be used

$$x = \begin{cases} 0 & \text{if } u < 0.3 \\ 1 & \text{if } u > 0.3 \end{cases} \quad (1)$$

- It is expected that if this is repeated many times 30% $X = 0$ and 70% $X = 1$

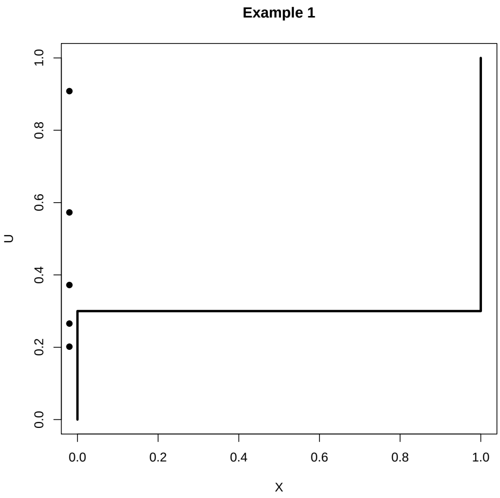
↪ Discrete Case: Example 2

- Assume that we want to **simulate a discrete variable** X **with probability** $\Pr(X = 0) = 0.3$ and $\Pr(X = 1) = 0.25$ and $\Pr(X = 2) = 0.45$
- Let $u \sim U(0, 1)$. Then the following rule can be used

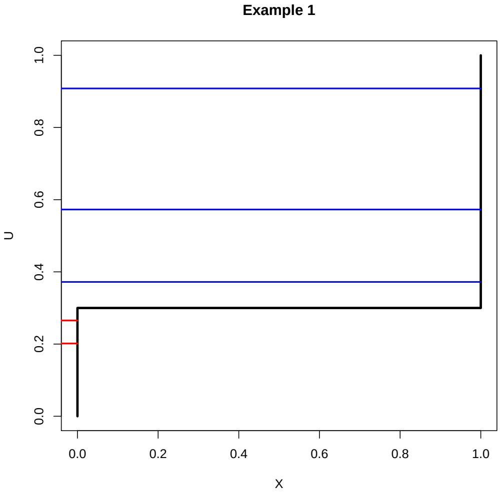
$$x = \begin{cases} 0 & \text{if } u < 0.3 \\ 1 & \text{if } 0.3 < u < 0.55 \\ 2 & \text{if } u > 0.55 \end{cases} \tag{2}$$

- It is expected that if this is repeated many times, roughly 30% $X = 0$, 25% $X = 1$ and 45% $X = 2$

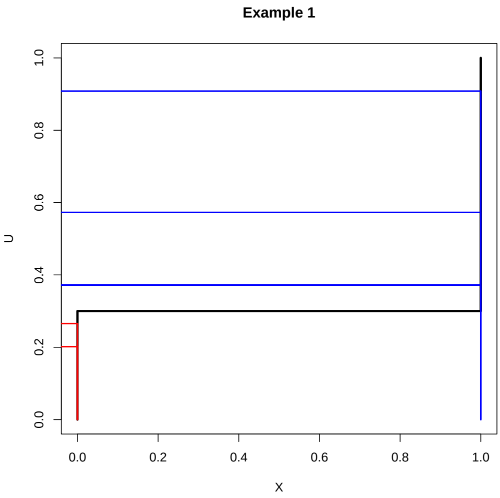
Direct Methods
↪ Visualization: Example 1



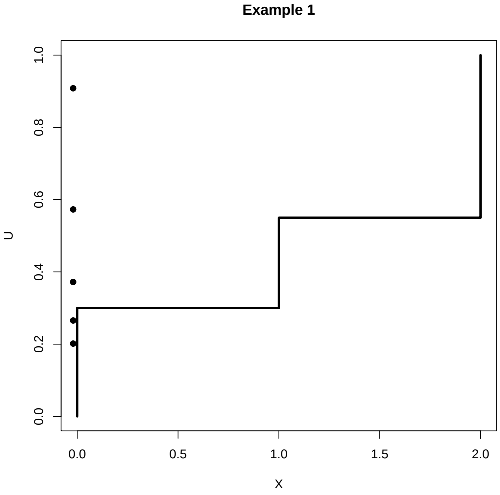
Direct Methods
↪ Visualization: Example 1



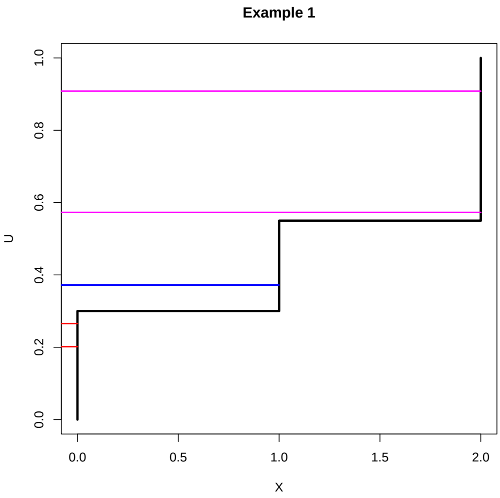
Direct Methods
↪ Visualization: Example 1



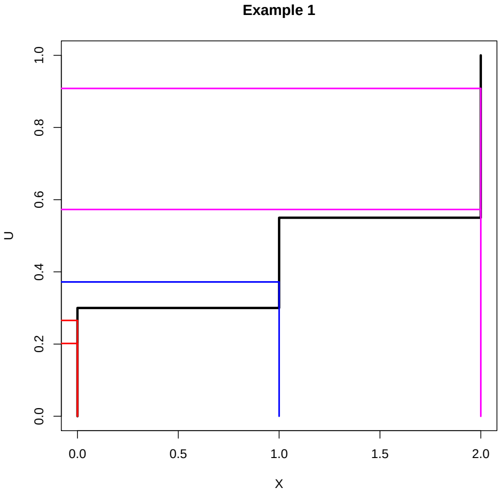
Direct Methods
↪ Visualization: Example 2



Direct Methods
↪ Visualization: Example 2



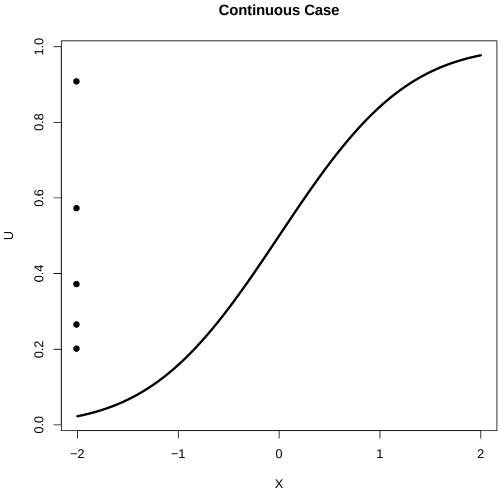
Direct Methods
↪ Visualization: Example 2



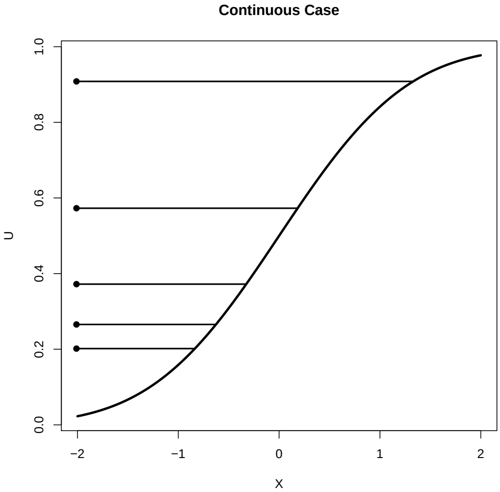
Direct Methods
↪ Continuous Case

- How do we extend this idea to the continuous case?
- What was the step function in our discrete example?
- It is the **cumulative distribution function (cdf)**
- Can we replace the discrete cdf with a continuous cdf?
- Yes!

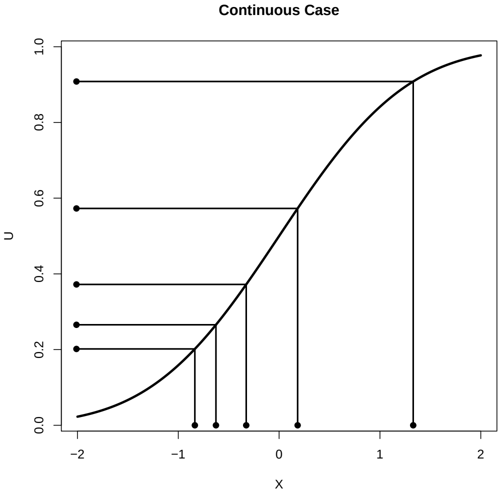
Direct Methods
↪ Visualization: Continuous



Direct Methods
↪ Visualization: Continuous



Direct Methods
↪ Visualization: Continuous



Direct Methods
↪ Continuous Case

- The cdf, $F(X)$ takes values of X and gives a value between 0 and 1
- Here we take values between 0 and 1 and get a value of X
- What function do we use?
- We use the **Inverse cdf**

Direct Methods
↛ Probability Integral Transform

- If Y is a continuous random variable with cdf $F(y)$, then the random variable $F_y^{-1}(U)$, where $U \sim uniform(O, 1)$, has distribution $F(y)$.
- **Example:** If $Y \sim exponential(A)$, then the probability density function (PDF) is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0, \\ 0 & x < 0. \end{cases}$$

and the cumulative distribution function (CDF) is given by

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geqslant 0, \\ 0 & x < 0. \end{cases}$$

Then

$$F_Y^{-1}(U) = -\log(1 - U)/\lambda$$

is an exponential random variable.

- Thus, if we generate $U_1, ..., U_n$ as iid uniform random variables, $-\lambda(1 - U_i)$, are iid exponential random variables with parameter λ.

Indirect simulation

- What if the cumulative distribution function is difficult to invert, or not even available?
- How to invert the cdf of a standard normal distribution?

$$F(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} e^{-x^2/2} dx \tag{3}$$

- It is still possible to simulate from this distribution?
- If the **density** is available, then the answer is YES!

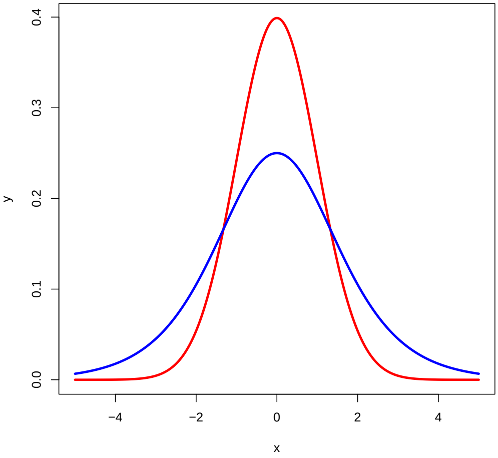
The student's *t* distribution

- The student's *t* distribution has density function

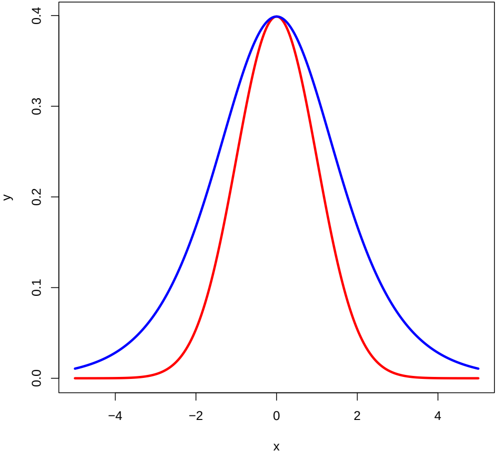
$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

- It is similar to the normal but has fatter tails.
- It is not so easy to simulate from this distribution using the direct method.

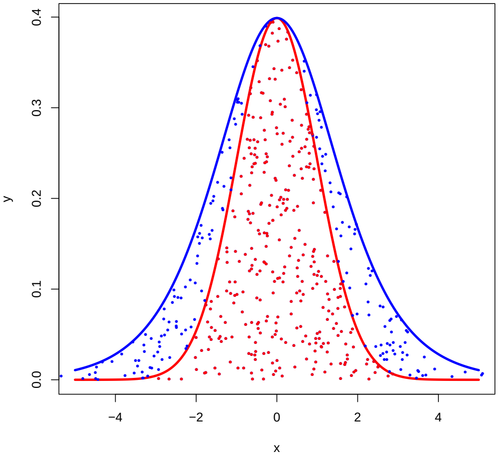
Normal (red) vs Student's t (blue)



Normal vs Student's t



Normal vs Student's t



The idea

- Let $f_y(x)$ be the **target distribution** and $f_v(v)$ be the **proposal distribution**
- Simulate an x-coordinate from the proposal $f_v(x)$
- Simulate a y-coordinate from $U(0, M * f_v(x))$
- Reject any points that are not 'inside' $f_y(x)$

Indirect Methods
↪ Reject and accept method

- **Theorem:** Let $Y \sim f_Y(y)$ and $V \sim f_\nu(\nu)$, where f_Y and f_V have common support with

$$M = \sup_y f_Y(y)/f_V(y) < \infty.$$

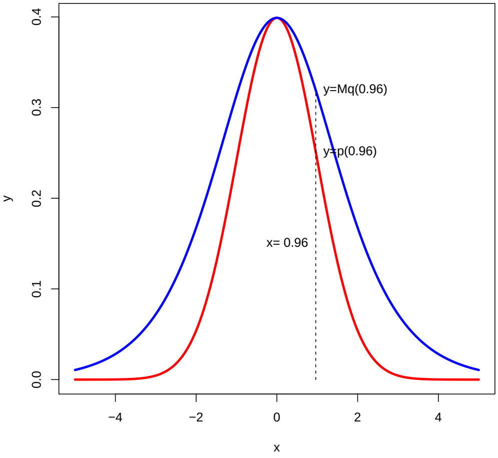
To generate a random variable $Y \sim f_Y$, we do the following steps

1. Generate $U \sim \mathrm{uniform}(0,1)$ and $V \sim f_V$ independently.
2. If

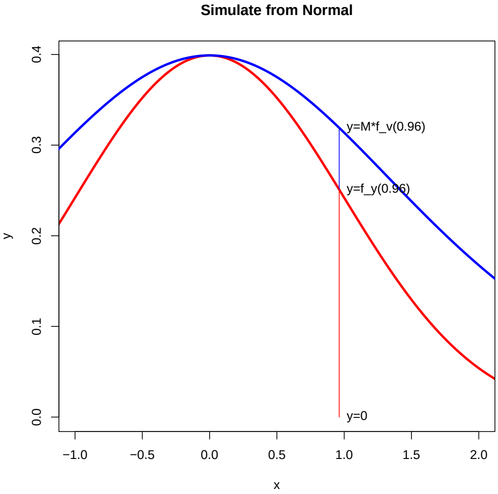
$$U < \frac{1}{M}\frac{f_Y(V)}{f_V(V)}$$

set $Y = V$; otherwise, return to step 1.

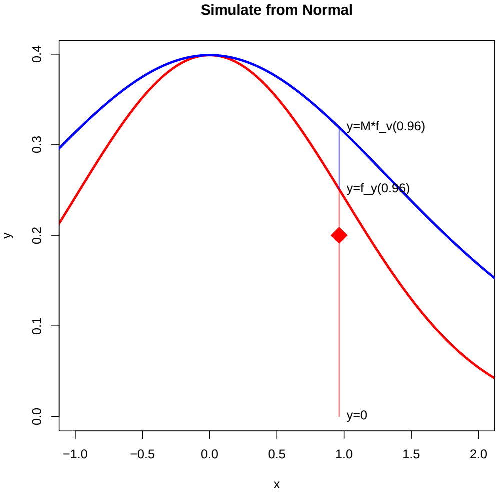
Indirect Methods
↪ Reject and accept method



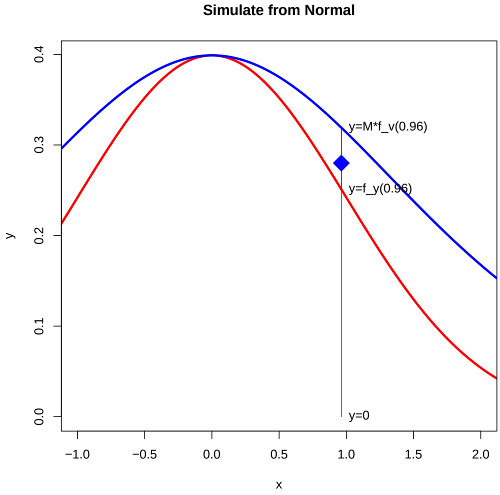
Indirect Methods
↪ Reject and accept method



Indirect Methods
↪ Reject and accept method



Indirect Methods
↪ Reject and accept method



Indirect Methods

↬ Reject and accept method: the proof

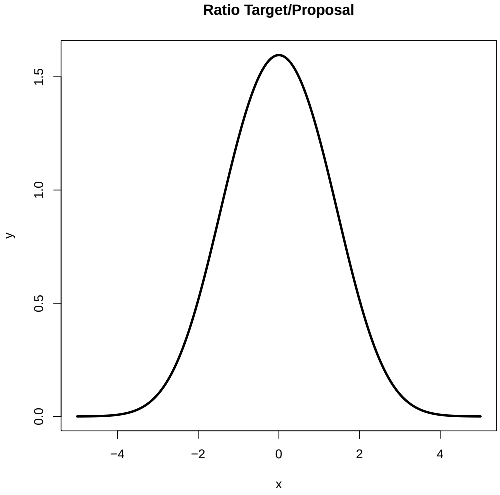
$$
\begin{aligned}
P(V \leqslant y|\text{stopping rule}) &= P(V \leqslant y|U < \frac{1}{M}\frac{f_Y(V)}{f_V(V)}) \\
&= \frac{P(V \leqslant y, U < \frac{1}{M}\frac{f_Y(V)}{f_V(V)})}{P(U < \frac{1}{M}\frac{f_Y(V)}{f_V(V)})} \\
&= \frac{\int_{-\infty}^{y}\int_{0}^{\frac{1}{M}\frac{f_Y(V)}{f_V(V)}} du\ f_V(v)dv}{\int_{-\infty}^{\infty}\int_{0}^{\frac{1}{M}\frac{f_Y(V)}{f_V(V)}} du\ f_V(v)dv} \\
&= \frac{\int_{-\infty}^{y}\frac{1}{M}\frac{f_Y(V)}{f_V(V)} f_V(V)dv}{\int_{-\infty}^{\infty}\frac{1}{M}\frac{f_Y(V)}{f_V(V)} f_V(V)dv} \\
&= \int_{-\infty}^{y} f_Y(v)dv
\end{aligned}
$$

- Can have any M? In fact M shows the efficiency of the sampling algorithm, i.e. $M = 1/P(\text{stopping rule})$.

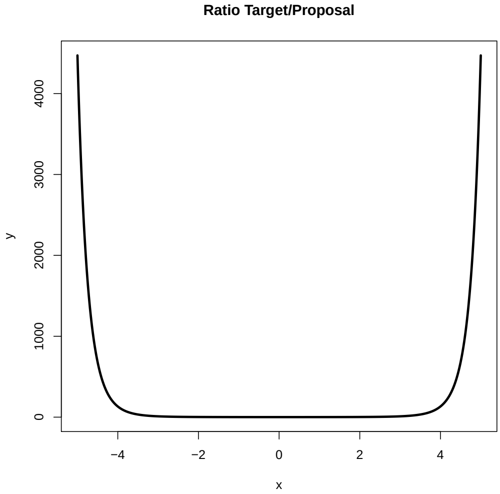
↛ Choosing $f_\nu(x)$ and M

- Two things are necessary for the accept/reject algorithm to work:
  1. The domain of $p(x)$ and the domain of $q(x)$ MUST be the same.
  2. The value of M must satisfy $M \geqslant \sup_x p(x)/q(x)$
- The algorithm will be more **efficient** if:
  1. The proposal $q(x)$ is a good approximation to $p(x)$
  2. The value of M is $M = \sup_x p(x)/q(x)$

Indirect Methods
↛ A look at $f_y(x)/f_v(x)$



Indirect Methods
⇸ A look at $f_y(x)/f_v(x)$



Normalizing Constant and Kernel

- Today we saw many density functions $f(x)$
- Many density functions can written as $f(x) = k\tilde{f}(x)$
- The part k is called the normalizing constant and the part $\tilde{f}(x)$ is called the kernel.
- For example the standard normal distribution is

$$p(x) = (2\pi)^{-1/2} e^{-x^2/2} \tag{4}$$

Normalizing Constant and Kernel

- Today we saw many density functions $f(x)$
- Many density functions can written as $f(x) = k\tilde{f}(x)$
- The part k is called the normalizing constant and the part $\tilde{f}(x)$ is called the kernel.
- For example the standard normal distribution is

$$p(x) = (2\pi)^{-1/2} e^{-x^2/2} \tag{4}$$

Normalizing Constant and Kernel

What are the normalizing constant and kernel of the Beta function?

$$\text{Beta}(x; a, b) = \frac{\Gamma(a+b)}{(\Gamma(a)\Gamma(b))} x^{a-1}(1-x)^{b-1} \tag{5}$$

Accept/Reject and the normalising constant

- An advantage of the Accept/Reject algorithm is that is works, even if the normalizing constant is unavailable.
- Only the kernel is needed.
- There are many examples where the normalising constant is either unavailable or difficult to compute.
- This often happens in Bayesian analysis

Suggested Reading

- Jones (2009), **Chapter 18**