Monte Carlo Methods: Lecture 2 : Transformation and Rejection

Nick Whiteley

Overview of this lecture

What we have seen ...

How to generate uniform $\mathrm{U}[0,1]$ pseudo-random numbers.

This lecture will cover ...

Generating random numbers from any distribution using

- transformations (CDF inverse, Box-Muller method).
- rejection sampling.

2.1 Transformation Methods

Transformation methods: Idea

- We can generate
$$U \sim \mathsf{U}[0,1].$$
- Can we find a transformation T such that
$$T(U) \sim F$$
for a distribution of interest with CDF F?
- One answer to this question: inversion method.

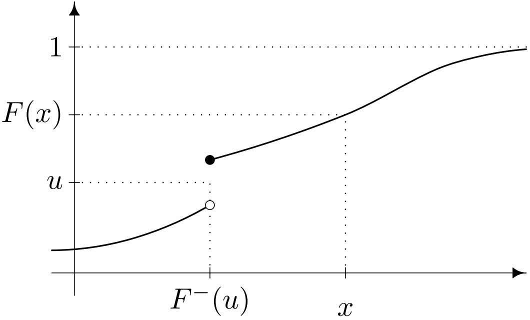
The CDF and its generalised inverse (1)

Cumulative distribution function (CDF)

$$F(x) = \mathbb{P}(X \leq x)$$

Generalised inverse of the CDF

$$F^{-}(u) := \inf\{x : \; F(x) \geq u\}$$



The CDF and its generalised inverse (2)

Properties of F^- (taken without proof)

1. $F^-(F(x)) \leq x, \quad \forall x \in F^-([0,1])$
2. $F(F^-(u)) \geq u, \quad \forall u \in [0,1]$

CDF inversion method (1)

Theorem 2.1: Inversion method

Let $U \sim \mathsf{U}[0,1]$ and F be a CDF. Then $F^{-}(U)$ has the CDF F.

Proof: From the definition of the CDF, $F(x) = \mathbb{P}(U \leq F(x))$, so we need to prove that

$$\mathbb{P}(F^{-}(U) \leq x) = \mathbb{P}(U \leq F(x)), \quad \forall x.$$

It is sufficient to prove the equivalence:

$$F^{-}(U) \leq x \Leftrightarrow U \leq F(x).$$

CDF inversion method (2)

We start by proving that

$$U \leq F(x) \Rightarrow F^{-}(U) \leq x.$$

For all $(v, w) \in [0,1] \times [0,1]$ such that $v \leq w$,

$$\begin{aligned}
&\{x : F(x) \geq w\} \subset \{x : F(x) \geq v\} \\
&\Rightarrow \inf\{x : F(x) \geq w\} \geq \inf\{x : F(x) \geq v\} \\
&\Leftrightarrow F^{-}(w) \geq F^{-}(v),
\end{aligned}$$

so in words, F^{-} is non-decreasing. We then have that

$$U \leq F(x) \Rightarrow F^{-}(U) \leq F^{-}(F(x)).$$

CDF inversion method (3)

Next, by the given property $F^{-}(F(x)) \leq x$,

$$U \leq F(x) \Rightarrow F^{-}(U) \leq x,$$

as required. It remains to prove the implication

$$F^{-}(U) \leq x \Rightarrow U \leq F(x).$$

As F is non-decreasing by definition,

$$F^{-}(U) \leq x \Rightarrow F(F^{-}(U)) \leq F(x).$$

To make the final step we use the property that $F(F^{-}(U)) \geq U$, yielding

$$F^{-}(U) \leq x \Rightarrow U \leq F(x). \quad \square$$

CDF inversion method (4)

So we have a simple algorithm for drawing $X \sim F$:

1. Draw $U \sim \mathsf{U}[0, 1]$.
2. Set $X = F^{-}(U)$.

(requires that $F^{-}(\cdot)$ can be evaluated efficiently)

Example 2.1: Exponential distribution

The exponential distribution with rate $\lambda > 0$ has the CDF $(x \geq 0)$

$$\begin{aligned} F_\lambda(x) &= 1 - \exp(-\lambda x) \\ F_\lambda^-(u) &= F_\lambda^{-1}(u) = -\log(1-u)/\lambda. \end{aligned}$$

So we have a simple algorithm for drawing Expo(λ):

1. Draw $U \sim \mathsf{U}[0,1]$.
2. Set $X = -\dfrac{\log(1-U)}{\lambda}$, or equivalently $X = -\dfrac{\log(U)}{\lambda}$.

Example 2.2: Box-Muller method for generating Gaussians

- Consider a bivariate real-valued random variable (X_1, X_2) and its polar coordinates (R, θ), i.e.

$$X_1 = R \cdot \cos(\theta), \qquad X_2 = R \cdot \sin(\theta) \tag{1}$$

- Then the following equivalence holds: $X_1, X_2 \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,1) \iff \theta \sim \mathsf{U}[0, 2\pi]$ and $R^2 \sim \mathsf{Expo}(1/2)$ indep.
- Suggests following algorithm for generating two Gaussians $X_1, X_2 \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,1)$:
  1. Draw angle $\theta \sim \mathsf{U}[0, 2\pi]$ and squared radius $R^2 \sim \mathsf{Expo}(1/2)$.
  2. Convert to Cartesian coordinates as in (1)
- From $U_1, U_2 \overset{\text{i.i.d.}}{\sim} \mathsf{U}[0,1]$ we can generate R and θ by

$$R = \sqrt{-2\log(U_1)}, \qquad \theta = 2\pi U_2,$$

giving

$$X_1 = \sqrt{-2\log(U_1)} \cdot \cos(2\pi U_2), \qquad X_2 = \sqrt{-2\log(U_1)} \cdot \sin(2\pi U_2)$$

Example 2.2: Box-Muller method for generating Gaussians

Box-Muller method

1. Draw
$$U_1, U_2 \overset{\text{i.i.d.}}{\sim} \mathsf{U}[0,1].$$
2. Set
$$\begin{aligned} X_1 &= \sqrt{-2\log(U_1)} \cdot \cos(2\pi U_2), \\ X_2 &= \sqrt{-2\log(U_1)} \cdot \sin(2\pi U_2). \end{aligned}$$

Then $X_1, X_2 \overset{\text{i.i.d.}}{\sim} \mathsf{N}(0,1)$.

2.2 Rejection sampling

Basic idea of rejection sampling

- Assume we cannot directly draw from density f.
- Tentative idea:
  1. Draw X from another density g (similar to f, easy to sample from).
  2. Only keep some of the X depending on how likely they are under f.

Basic idea of rejection sampling

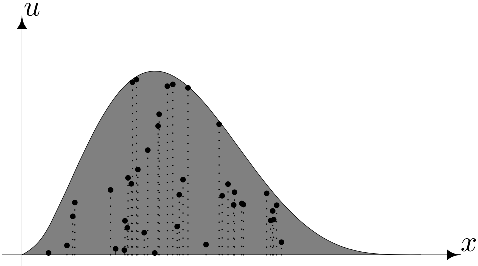
- Consider the identity

$$f(x) = \int_0^{f(x)} 1 \, du = \int \underbrace{1_{0<u<f(x)}}_{=f(x,u)} du.$$

- $f(x)$ can be interpreted as the marginal density of a uniform distribution on the area under the density $f(x)$:

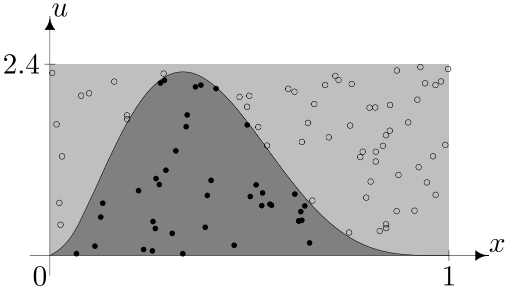
$$\{(x,u): \; 0 \le u \le f(x)\}.$$

- Sample from f by sampling from the area under the density.



Example 2.3: Sampling from a Beta$(3, 5)$ distribution (1)

- How can we draw points from the area under the density?
  1. Draw (X, U) from the grey rectangle, i.e. $X \sim \mathsf{U}(0, 1)$ and $U \sim \mathsf{U}(0, 2.4)$.
  2. Accept X as a sample from f if (X, U) lies under the density (dark grey area).



- Step 2 equivalent to: Accept X if $U < f(X)$, i.e. accept X with probability $\mathbb{P}(U < f(X) | X = x) = f(X)/2.4$.

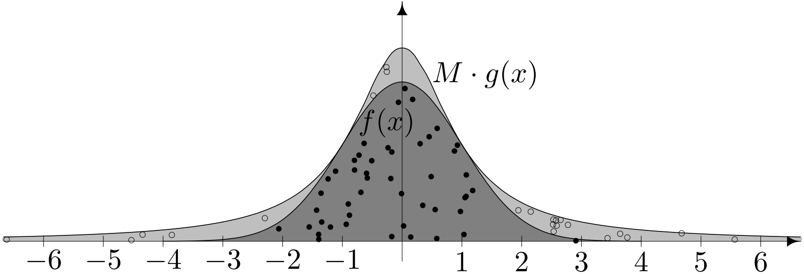
Example 2.3: Sampling from a Beta$(3, 5)$ distribution (2)

- Resulting algorithm:
  1. Draw $X \sim \mathrm{U}(0, 1)$.
  2. Accept X as a sample from Beta$(3, 5)$ with probability

$$\frac{f(X)}{2.4}$$

.

- Not every density can be bounded by a box. How can we generalise the idea?
  $\rightsquigarrow$ Bounding f by M times another density g.



The rejection sampling algorithm (1)

Algorithm 2.1: Rejection sampling

Given two densities f, g with $f(x) < M \cdot g(x)$ for all x, we can generate a sample from f by

1. Draw $X \sim g$.
2. Accept X as a sample from f with probability

$$\frac{f(X)}{M \cdot g(X)},$$

   otherwise go back to step 1.

Note: $f(x) < M \cdot g(x)$ implies that f cannot have heavier tails than g.

The rejection sampling algorithm (2)

Remark 2.1

If we know f only up to a multiplicative constant, i.e. if we only know $\pi(x)$, where $f(x) = C \cdot \pi(x)$, we can carry out rejection sampling using

$$\frac{\pi(X)}{M \cdot g(X)}$$

as probability of rejecting X, provided $\pi(x) < M \cdot g(x)$ for all x.

Can be useful in Bayesian statistics:

$$f^{\text{post}}(\theta) = \frac{f^{\text{prior}}(\theta) l(\mathbf{y}_1, \ldots, \mathbf{y}_n|\theta)}{\int_{\Theta} f^{\text{prior}}(\vartheta) l(\mathbf{y}_1, \ldots, \mathbf{y}_n|\vartheta)\; d\vartheta} = C \cdot f^{\text{prior}}(\theta) l(\mathbf{y}_1, \ldots, \mathbf{y}_n|\theta)$$

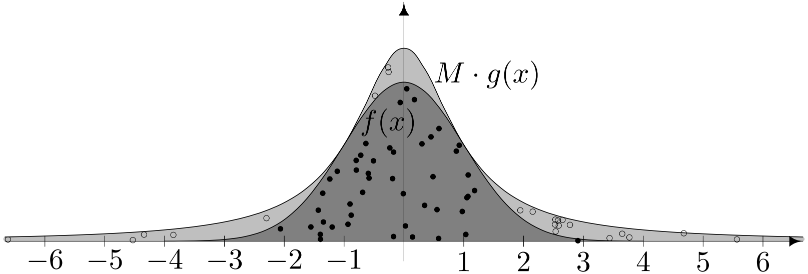
Example 2.4: Rejection sampling from the $\mathsf{N}(0,1)$ distribution using a Cauchy proposal (1)

- Recall the following densities:

$$\mathsf{N}(0,1) \qquad f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\text{Cauchy} \qquad g(x) = \frac{1}{\pi(1+x^2)}$$

- For $M = \sqrt{2\pi} \cdot \exp(-1/2)$ we have that $f(x) \leq Mg(x)$.
  $\leadsto$ We can use rejection sampling to sample from f using g as proposal.



Example 2.4: Rejection sampling from the N$(0, 1)$ distribution using a Cauchy proposal (2)

- We cannot sample from a Cauchy distribution (g) using a Gaussian (f) as instrumental distribution.
- Whe Cauchy distribution has heavier tails than the Gaussian distribution: there is no $M \in \mathbb{R}$ such that

$$\frac{1}{\pi(1+x^2)} < M \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2}\right).$$