Rejection Sampling
(Chapter 4)

Michael Tsiang

Stats 102C: Introduction to Monte Carlo Methods

UCLA

Acknowledgements: Qing Zhou

Outline

Introduction

- The inverse CDF method applies very generally to many distributions.
- The method relies on a closed form expression for $F^{-1}(u)$: Given $F(x) = u$, we assume we can derive $x = F^{-1}(u)$.
- However, there are random variables for which this is not possible.

Introduction

- Consider the standard normal distribution $\mathcal{N}(0,1)$, with density
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$
- The CDF of $\mathcal{N}(0,1)$ is given by
$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,\mathrm{d}t = u.$$
- There is no closed form expression for $F(x)$ or $F^{-1}(u)$, so we cannot use the inverse CDF method.
- We need a different method to sample from distributions that does not rely on the CDF or inverse CDF.

Outline

1 Introduction

2 Rejection Sampling (Uniform Case)

- Example 1: $f(x) = \sin x$
- Example 2: $\mathrm{Beta}(2, 2)$ $\alpha = \beta = 2$

3 Rejection Sampling (General Case)

- Example 2: Folded Normal Distribution
- Example 2a: Standard Normal Distribution

Rejection Sampling (Uniform Case)

Consider a PDF $f(x)$, defined for $x \in [a, b]$, such that there is a constant M such that

$$M \geq f(x), \text{ for all } x \in [a, b],$$ the most pdf is M

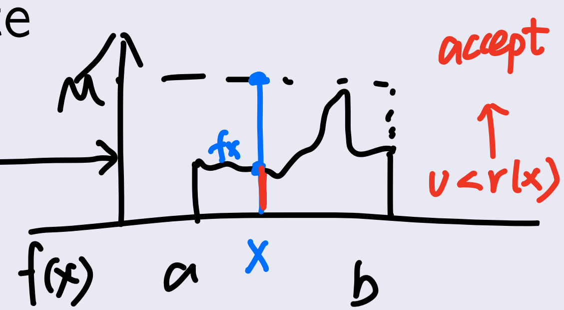
i.e., $f(x)$ is bounded above by M.

1. Generate $X \sim \mathrm{Unif}(a, b)$, and compute

   $\left(\frac{f(x)}{M}\right)$

   $0 < \boxed{M > f(x)} < 1$

   $$r(X) = \frac{f(X)}{M}.$$

   Note that $r(X) \leq 1$ by choice of M.

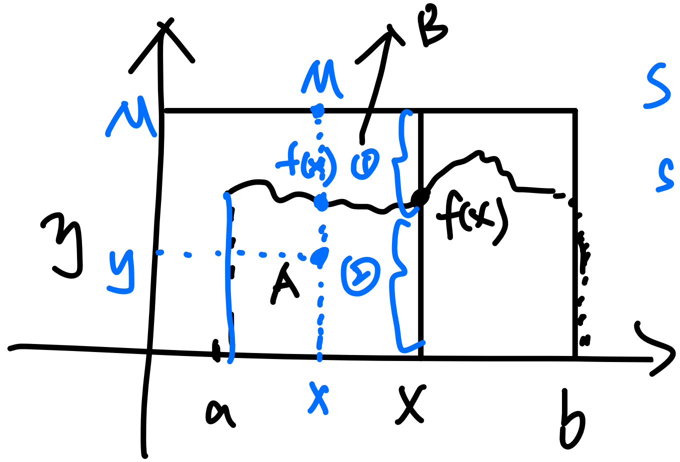
   $0 < r(x) \leq 1 \longrightarrow U$

2. Generate $U \sim \mathrm{Unif}(0, 1)$.
3. If $U \leq r(X)$, then accept X as a sample from $f(x)$. Otherwise, repeat (1) and (2).

   $\begin{cases} U \leq r(x) & ✓ \\ U > r(x) & ✗ \longrightarrow \text{repeat} \end{cases}$

Rejection Sampling (Uniform Case)

- This is similar to Example 1 in Chapter 1, where we considered the area of a region D in $\mathbb{R}^2$.
- D is now the area under the curve $f(x)$ over the interval $[a, b]$.
- We generate $X \sim \mathrm{Unif}(a, b)$ and $Y \sim \mathrm{Unif}(0, M)$, so (X, Y) is drawn uniformly from the bounding rectangle $[a, b] \times [0, M]$.



$S_A = 1$

$S_B = b \times m$

$$\frac{②}{① + ②}$$

$$\frac{f(x)}{M}$$

Rejection Sampling (Uniform Case)

$X \sim \text{unifor}(a, b)$ $\quad x = 2 \quad y = 0.1$

$f(x) = \quad < M \rightarrow \text{accept}$ $\quad$ if $y < f(x) \rightarrow$ accept $(2, 0.1)$

- If $Y \leq f(X)$ (i.e., (X, Y) is below the curve), accept X as a sample from $f(x)$.

$\rightarrow r(x)$ $\quad 0 \leq \frac{Y}{M} \leq 1$

- The criterion $Y \leq f(X)$ is equivalent to $\dfrac{Y}{M} \leq \dfrac{f(X)}{M} = r(X)$.

  Since $Y \sim \text{Unif}(0, M)$, then $U = \dfrac{Y}{M} \sim \text{Unif}(0, 1)$.

  $X \sim \text{unifor}(a, b)$

Rejection Sampling (Uniform Case)

To make sure this rejection sampling method is generating samples from $f(x)$, we need to show that $P(X = x | X \text{ is accepted}) = f(x)$.

Proof (Rejection Sampling, Uniform Case, Part 1).

By Bayes Theorem,

$$\begin{aligned} P(X = x | X \text{ is accepted}) &= \frac{P(X \text{ is accepted} | X = x) P(X = x)}{P(X \text{ is accepted})} \\ &= \frac{P[U \le r(x)] \cdot \frac{1}{b-a}}{P(X \text{ is accepted})} \\ &= \frac{\frac{f(x)}{M} \cdot \frac{1}{b-a}}{P(X \text{ is accepted})}. \end{aligned}$$

$U \le r(x) = \frac{r}{M}$

$U \sim \text{unifor}(0,1)$

$x \sim \text{unif}[a,b]$

$= f(x)$

?

We need to compute $P(X \text{ is accepted})$.

Rejection Sampling (Uniform Case)

Proof (Rejection Sampling, Uniform Case, Part 2).

By the Law of Total Probability,

$$
\begin{aligned}
P(X \text{ is accepted}) &= \int_a^b P(X \text{ is accepted} | X = x) P(X = x)\,\mathrm{d}x \\
&= \int_a^b P[U \leq r(x)] \cdot \frac{1}{b-a}\,\mathrm{d}x \\
&= \int_a^b \frac{f(x)}{M} \cdot \frac{1}{b-a}\,\mathrm{d}x \\
&= \frac{1}{(b-a)M} \int_a^b f(x)\,\mathrm{d}x \\
&= \frac{1}{(b-a)M}. \quad \times|
\end{aligned}
$$

Rejection Sampling (Uniform Case)

Proof (Rejection Sampling, Uniform Case, Part 3).

In Part 2, we have shown that $P(X \text{ is accepted}) = \dfrac{1}{(b-a)M}$.

(a constant)

Combining this with Part 1, we have

$$
\begin{aligned}
P(X = x | X \text{ is accepted}) &= \frac{\dfrac{f(x)}{M} \cdot \dfrac{1}{b-a}}{P(X \text{ is accepted})} \\
&= \frac{\dfrac{f(x)}{M} \cdot \dfrac{1}{b-a}}{\dfrac{1}{(b-a)M}} \\
&= f(x),
\end{aligned}
$$

which is what we wanted to show. □

Rejection Sampling (Uniform Case)

- The probability of acceptance can also be interpreted as the **efficiency** of the rejection sampling method:

$$
\begin{aligned}
P(X \text{ is accepted}) &= \boxed{\frac{1}{(b-a)M}} \\
&= \frac{\text{Area under } f(x)}{\text{Total Area of Bounding Rectangle}}
\end{aligned}
$$

  On average, one sample value of X takes $(b-a)M$ iterations.

- Since we require $f(x) \leq M$ for $x \in [a, b]$, the highest acceptance rate (most efficient) is when $M = \max_{x \in [a,b]} f(x)$.

define M

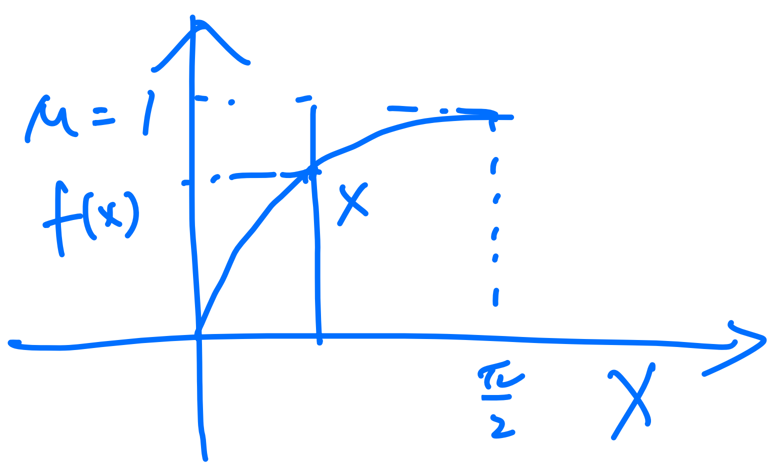
a, b

Example 1: $f(x) = \sin x$

$\sin(\frac{\pi}{2}) = 1 \quad \sin(\cdot) = 0 \quad \rightarrow \quad M \rightarrow 1$

- Let $f(x) = \sin x$, for $x \in [0, \frac{\pi}{2}]$.
- We verify that $\int_0^{\pi/2} \sin x \, \mathrm{d}x = 1$, so $f(x)$ is a PDF on $[0, \frac{\pi}{2}]$.
- The most efficient choice of $M = \max_{x \in [0, \frac{\pi}{2}]} \sin x = 1$.

$x = \frac{\pi}{2} \quad \sin(x) = 1$ (most)



$$p(x = \text{accept}) = \frac{1}{M(b-a)} = \frac{1}{1 \cdot \frac{\pi}{2}} = \frac{2}{\pi} = 0.64$$

Example 1: $f(x) = \sin x$

Rejection sampling algorithm to sample from $f(x) = \sin x$:

{ X is accepte ★
{ X is not accept

1. Generate $X \sim \text{Unif}(0, \frac{\pi}{2})$, and compute

accept or not $$r(X) = \frac{f(X)}{M} = \sin X. = \frac{\sin(x)}{1}$$ $M = 1$

2. Generate $U \sim \text{Unif}(0, 1)$
3. If $U \leq \sin X$, accept X as a sample from $f(x)$. Otherwise, repeat (1) and (2).

The acceptance rate is $\frac{1}{M(b-a)}$

$M = 1$

accept rate

$$P(X \text{ is accepted}) = \frac{1}{\frac{\pi}{2} \cdot 1} = \frac{2}{\pi} \approx 0.64.$$

$a = 0$

$b = \frac{\pi}{2}$

How often

$$P(X \text{ is accepted}) = \frac{1}{(b-a)M}$$

accept X $\rightarrow$ X distribution $= \dfrac{\text{Area under } f(x) = 1}{\text{Total Area of Bounding Rectangle}}$

Example 1: $f(x) = \sin x$

R Code to sample from $\sin x$ on $[0, \frac{\pi}{2}]$:

find M

$M = \max_{x \in (a,b)} f(x)$

```
> # Set the seed for reproduceability
> set.seed(9999)

> n <- 2000 # Specify the number of points to generate

> # Generate n points from Unif(0, pi / 2)
> X <- runif(n, 0, pi / 2)
```

a b

accepted

not accepted

```
> # Compute r(X)
> r <- sin(X)
```

$r(x) < \frac{\sin x}{1} = \frac{f(x)}{M}$, $M = \max f(x) = \sin(\frac{\pi}{2}) = 1$

```
> # Generate n points from Unif(0, 1)
> U <- runif(n, 0, 1)
```

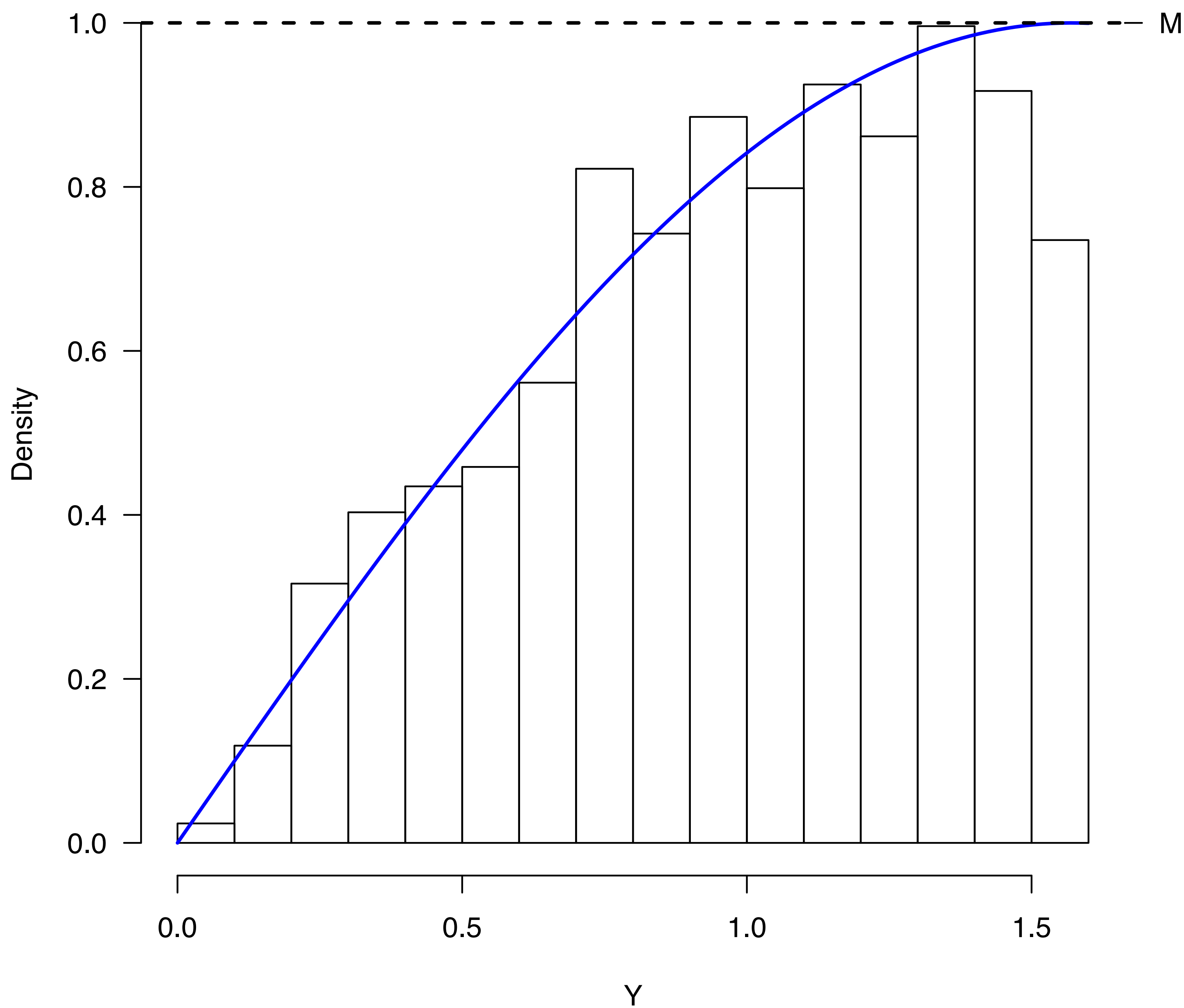
$r = \frac{f(x)}{M}$

```
> # Accept points if U <= r(X)
> Y <- X[U <= r]
```

$X = [\quad] \to 2000$

$y = [\quad] \to$ accept point

Example 1: $f(x) = \sin x$



Example 1: $f(x) = \sin x$

R Code for the plot: $Y =$ accepted x

```
> hist(Y, prob = TRUE, las = 1, main = "")
> curve(sin, add = TRUE)
> abline(h = 1, lty = 2)
> axis(4, at = 1, labels = "M", las = 1)
```

→ esitmate

→ true

→ M

R Code for acceptance rates: $p(X = \text{accept}) = \frac{1}{M(b-a)}$

```
> length(Y) / n # empirical acceptance rate
[1] 0.6325
> 2 / pi # theoretical acceptance rate
[1] 0.6366198
```

$\frac{2}{\pi}$

$$\int_0^{\frac{\pi}{2}} \sin x \, dx = \frac{2}{\pi}$$

Example 2: $\mathrm{Beta}(2,2)$

$$f(x) = 6\, x^{1}(1-x)^{1} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \longrightarrow \begin{cases} a-1=1 \\ b-1=1 \end{cases} \quad a=b=2 \sim \mathrm{Beta}(2,2)$$

$$\frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} = \frac{3\times2\times1}{1\times1\times1\times1} = 6$$

- The $\mathrm{Beta}(2,2)$ distribution has PDF

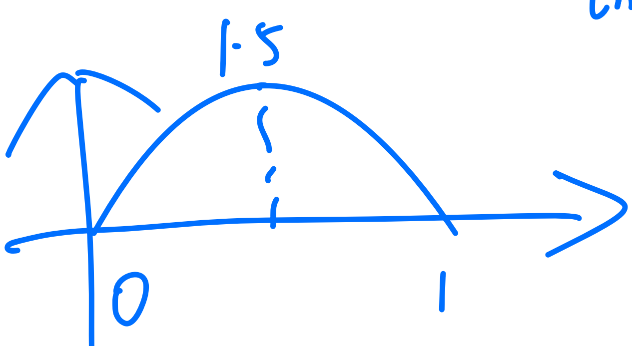
Max $x = 1.5 = M$

$$f(x) = 6x(1-x), \text{ for } x \in [0,1].$$

- We want to use rejection sampling to sample from $f(x)$.
- The most efficient choice of $M = \max\limits_{x\in[0,1]} 6x(1-x) = 1.5$.

therical: $f(x = \text{accept}) = \dfrac{1}{M(b-a)}$ $\quad M=1.5$

$= 0.667$



Example 2: $\mathrm{Beta}(2,2)$

(0.1)

Rejection sampling algorithm to sample from $f(x) = 6x(1-x)$:

1. Generate $X \sim \mathrm{Unif}(0,1)$, and compute
$$\boxed{r(X) = \frac{f(X)}{M} = \frac{6X(1-X)}{1.5} = 4X(1-X).}$$
acept or not
2. Generate $U \sim \mathrm{Unif}(0,1)$.
3. If $U \leq 4X(1-X)$, accept X as a sample from $f(x)$. Otherwise, repeat (1) and (2). $\longrightarrow$ sample point

The acceptance rate is

$$P(X \text{ is accepted}) = \frac{1}{1 \cdot 1.5} = \frac{2}{3} \approx 0.667.$$

point rate

$$= \int_0^1 6x(1-x)\,dx = \frac{2}{3}$$

Example 2: $\text{Beta}(2,2)$

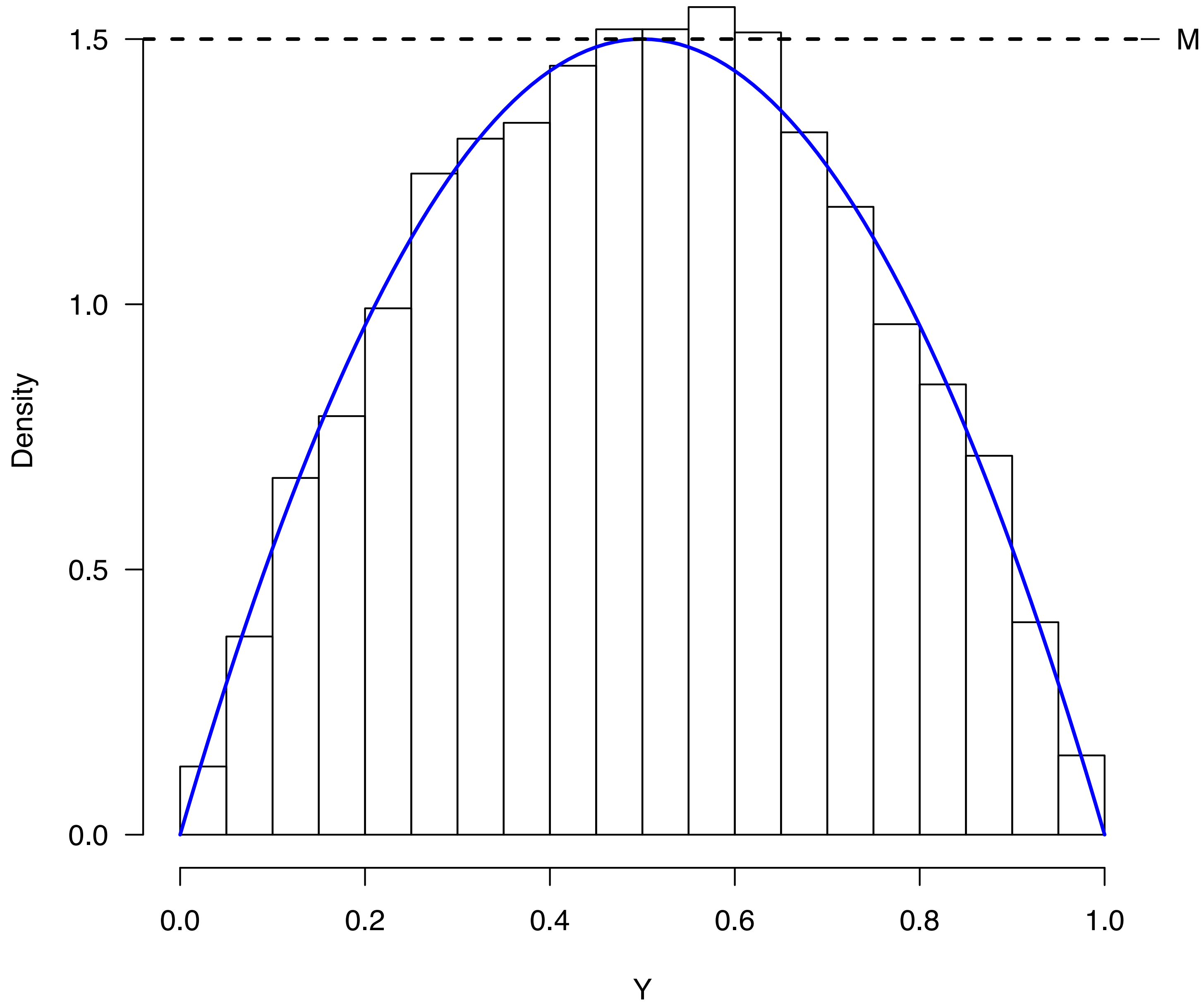
R Code to sample from $f(x) = 6x(1-x)$ on $[0,1]$:

```
> # Set the seed for reproduceability
> set.seed(9999)

> n <- 10000 # Specify the number of points to generate

> # Generate n points from Unif(0, 1)
> X <- runif(n, 0, 1)

> # Compute r(X)
> r <- 4 * X * (1 - X)

> # Generate n points from Unif(0, 1)
> U <- runif(n, 0, 1)

> # Accept points if U <= r(X)
> Y <- X[U <= r]
```

$r(X) = \frac{f(x)}{M}$

$\frac{\text{length}(Y)}{n} = P(\text{accept}) = \frac{1}{M(b-a)}$

Example 2: $\text{Beta}(2, 2)$



Example 2: $\text{Beta}(2, 2)$

R Code for the plot:

```
> hist(Y, prob = TRUE, las = 1, main = "", breaks = 30)
> curve(6 * x * (1 - x), add = TRUE, col = "blue", lwd = 2)
> abline(h = 1.5, lty = 2, lwd = 2)
> axis(4, at = 1.5, labels = "M", las = 1)
```

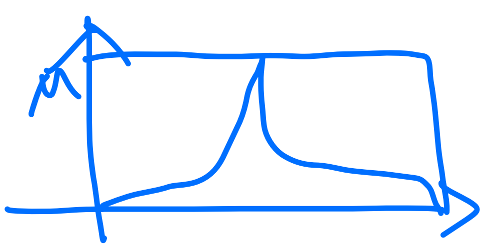
R Code for acceptance rates:

```
> length(Y) / n # empirical acceptance rate
[1] 0.6691
> 2 / 3 # theoretical acceptance rate
[1] 0.6666667
```

$\frac{1}{1.5(1-0)} = \frac{2}{3}$

$\frac{1}{M(b-a)}$

Outline



p(accept) = small

M is not meaningful

Rejection Sampling (General Case)

The rejection sampling method described above leverages our ability to sample from the uniform distribution to sample from distributions with PDFs that are defined on intervals $[a, b]$.

However, there are some limitations:

low accept rate

- If $f(x)$ is far from uniform (i.e., the area under $f(x)$ is much smaller than the area of the bounding rectangle), the acceptance rate when sampling from $\mathrm{Unif}(a, b)$ can be low.

unbounded domain

- Many PDFs are defined on an infinite (unbounded) domain, such as $(0, \infty)$ or $(-\infty, \infty)$, but we cannot sample uniformly from an unbounded range (Why not?).

How can we modify the rejection sampling method to sample from more general PDFs and increase efficiency (the acceptance rate)?

tagert and trial distribution

$f(x)$ $g(x)$ in region D

Rejection Sampling (General Case)

Consider a PDF $f(x)$, defined for x on a region D.

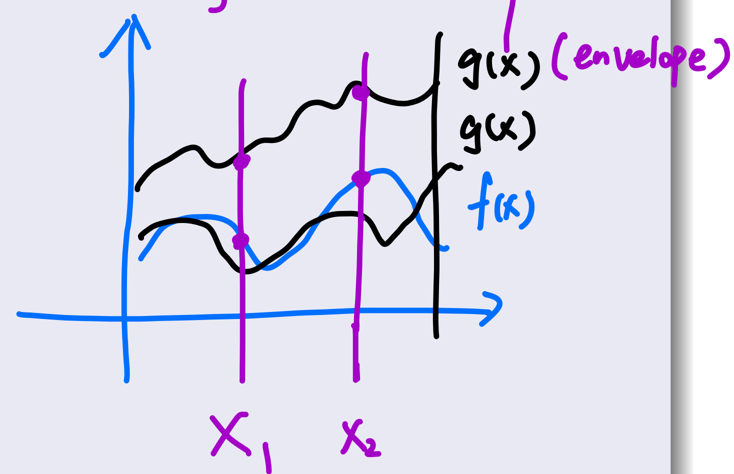
1. Find a **trial** or **candidate distribution** $g(x)$ such that:
   1. There is a constant M such that $Mg(x) \geq f(x)$ for $x \in D$.
   2. We can sample from $g(x)$.

area under envelope

$= \int Mg(x)\,dx = M$

2. Generate $X \sim g(x)$, and compute

$$r(X) = \frac{f(X)}{Mg(X)}.$$



   Note that $r(X) \leq 1$ by choice of M.

3. Generate $U \sim \mathrm{Unif}(0, 1)$.

find $g(x)$ same as $f(x)$ (shape)

4. If $U \leq r(X)$, then accept X as a sample from $f(x)$. to max proportion
   Otherwise, repeat (2) and (3).

Rejection Sampling (General Case)

To make sure this rejection sampling method is generating samples from $f(x)$, we need to show that $P(X = x | X \text{ is accepted}) = f(x)$.

Proof (Rejection Sampling, General Case).

By Bayes Theorem,

$$
\begin{aligned}
P(X = x | X \text{ is accepted}) &= \frac{P(X \text{ is accepted} | X = x) P(X = x)}{P(X \text{ is accepted})} \\
&= \frac{P[U \leq r(x)] \cdot g(x)}{\int_D P[U \leq r(x)] \cdot g(x) \, \mathrm{d}x} \\
&= \frac{\frac{f(x)}{Mg(x)} \cdot g(x)}{\int_D \frac{f(x)}{Mg(x)} \cdot g(x) \, \mathrm{d}x} \\
&= \frac{\frac{f(x)}{M}}{\frac{1}{M}} \\
&= f(x),
\end{aligned}
$$

which is what we wanted to show. □

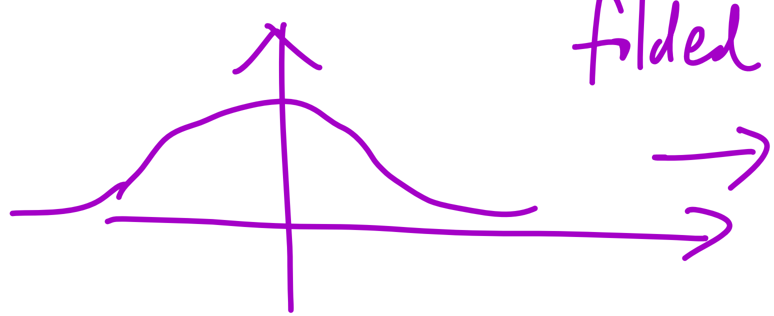
- The function $Mg(x)$ is sometimes called the **envelope**.
- The **efficiency** of the rejection sampling method is

$$\frac{1}{M(b-a)}$$

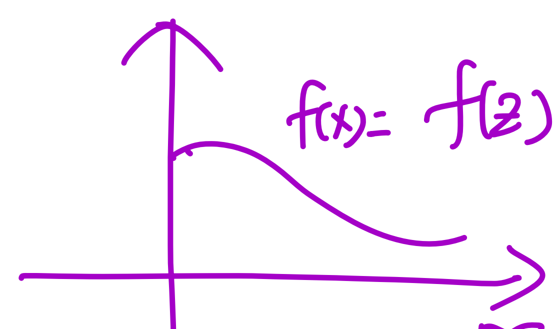
$$\begin{aligned} P(X \text{ is accepted}) &= \frac{1}{M} \\ &= \frac{\text{Area under } f(x)}{\text{Area under envelope}} \end{aligned}$$

$M \downarrow$ rate $\uparrow$

- A smaller M will result in a higher acceptance rate.



folded



$f(x) = f(z) \cdot 2$

Example 2: Folded Normal Distribution

$Z \sim N(0,1)$

$X = |Z|$: folded normal

- Let $Z \sim \mathcal{N}(0,1)$ with PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \text{for } z \in (-\infty, \infty).$$

- Suppose we want to sample from $X = |Z|$, which has PDF

$$f_X(x) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \sqrt{\frac{2}{\pi}} e^{-x^2/2}, \quad \text{for } x \geq 0.$$

  X has a **folded normal distribution**.

- How can we use rejection sampling to sample from X? $|X| = Z$

$g(x)$?

Example 2: Folded Normal Distribution

Goal: Sample from $f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}$, for $x \geq 0$.

1. Find a trial or candidate distribution $g(x)$ such that:
   1. There is a constant M such that $Mg(x) \geq f(x)$ for $x \in D$.
   2. We can sample from $g(x)$.

- Consider the PDF of $\mathrm{Exp}(\lambda = 1)$, given by random $g(x)$

$$g(x) = e^{-x}, \quad \text{for } x \geq 0.$$ same region

  Note that $g(x)$ is defined (i.e., $g(x) > 0$) on the same region as $f(x)$ (namely $D = [0, \infty)$).

- Does $g(x)$ satisfy the conditions to be a suitable trial distribution?

Example 2: Folded Normal Distribution

(i) There is a constant M such that $Mg(x) \ge f(x)$ for $x \in D$.

- We want to find an M such that $M \ge \max_{x \ge 0} \frac{f(x)}{g(x)}$.
- Find M such that

min $x=1$

$$M = \max_{x \ge 0} \frac{f(x)}{g(x)} = \max_{x \ge 0} \frac{\sqrt{\frac{2}{\pi}} e^{-x^2/2}}{e^{-x}} = \max_{x \ge 0} \sqrt{\frac{2}{\pi}} e^{-\left(\frac{x^2}{2} - x\right)}.$$

- Finding the maximum of $e^{-\left(\frac{x^2}{2} - x\right)}$ is equivalent to finding the minimum of $\left(\frac{x^2}{2} - x\right)$ for $x \ge 0$, which occurs at $x^* = 1$.
- So $M = \sqrt{\frac{2}{\pi}} e^{-\left(\frac{1}{2} - 1\right)} = \sqrt{\frac{2e}{\pi}} \approx 1.32$.

$\frac{1}{1.32}$

$m = \max_{x \in D} \frac{f(x)}{g(x)}$

Example 2: Folded Normal Distribution

$g(x) = e^{-x}$

$G(x) = 1 - e^{-x} = V$

$x = -\log(1-V)$

$= -\log(V)$

(ii) We can sample from $g(x)$. (by inverse cdf)

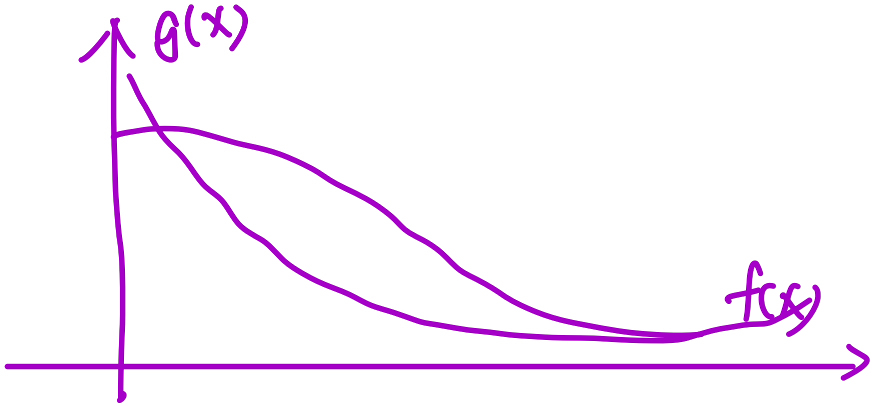
- Use the inverse CDF method for $\mathrm{Exp}(\lambda = 1)$:

  1. Generate $V \sim \mathrm{Unif}(0, 1)$.

     $F^{-1}(V) = -\log(V)$

  2. Then $X = -\log V \sim \mathrm{Exp}(\lambda = 1)$. $\sim g(x)$

- Since (i) and (ii) hold for $g(x) \sim \mathrm{Exp}(\lambda = 1)$, then $g(x)$ is a suitable trial distribution to use for rejection sampling.



Example 2: Folded Normal Distribution

Rejection sampling to sample from $f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}$, for $x \geq 0$:

1. Find a trial or candidate distribution $g(x)$ such that:
   1. There is a constant M such that $Mg(x) \geq f(x)$ for $x \in D$.
   2. We can sample from $g(x)$.

$M \geq \frac{f(x)}{g(x)}$

$x \geq 0$

$$g(x) = e^{-x}, \quad \text{with } M = \sqrt{\frac{2e}{\pi}}.$$

find $M = \max_{x \in D} \frac{f(x)}{g(x)}$

2. Generate $X \sim g(x)$, and compute U

$$r(X) = \frac{f(X)}{Mg(X)} = \sqrt{\frac{\pi}{2e}} \cdot \sqrt{\frac{2}{\pi}} e^{-\left(\frac{X^2}{2} - X\right)} = \exp\left(X - \frac{X^2}{2} - \frac{1}{2}\right).$$

3. Generate $U \sim \text{Unif}(0, 1)$.

4. If $U \leq r(X)$, then accept X as a sample from $f(x)$. Otherwise, repeat (2) and (3).

M. $g(x) \to G(x) \to F^{-1}(U)$

Example 2: Folded Normal Distribution

R Code to sample from $f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}$, for $x \geq 0$:

$M = \max_x \frac{f(x)}{g(x)} \to$ find x plug into $\frac{f(x)}{g(x)}$

```
> # Set the seed for reproduceability
> set.seed(9999)

> n <- 2000 # Specify the number of points to generate

> # Generate n points from Exp(lambda = 1)
> V <- runif(n, 0, 1)
> X <- -log(V)
```

$\sim g(x)$

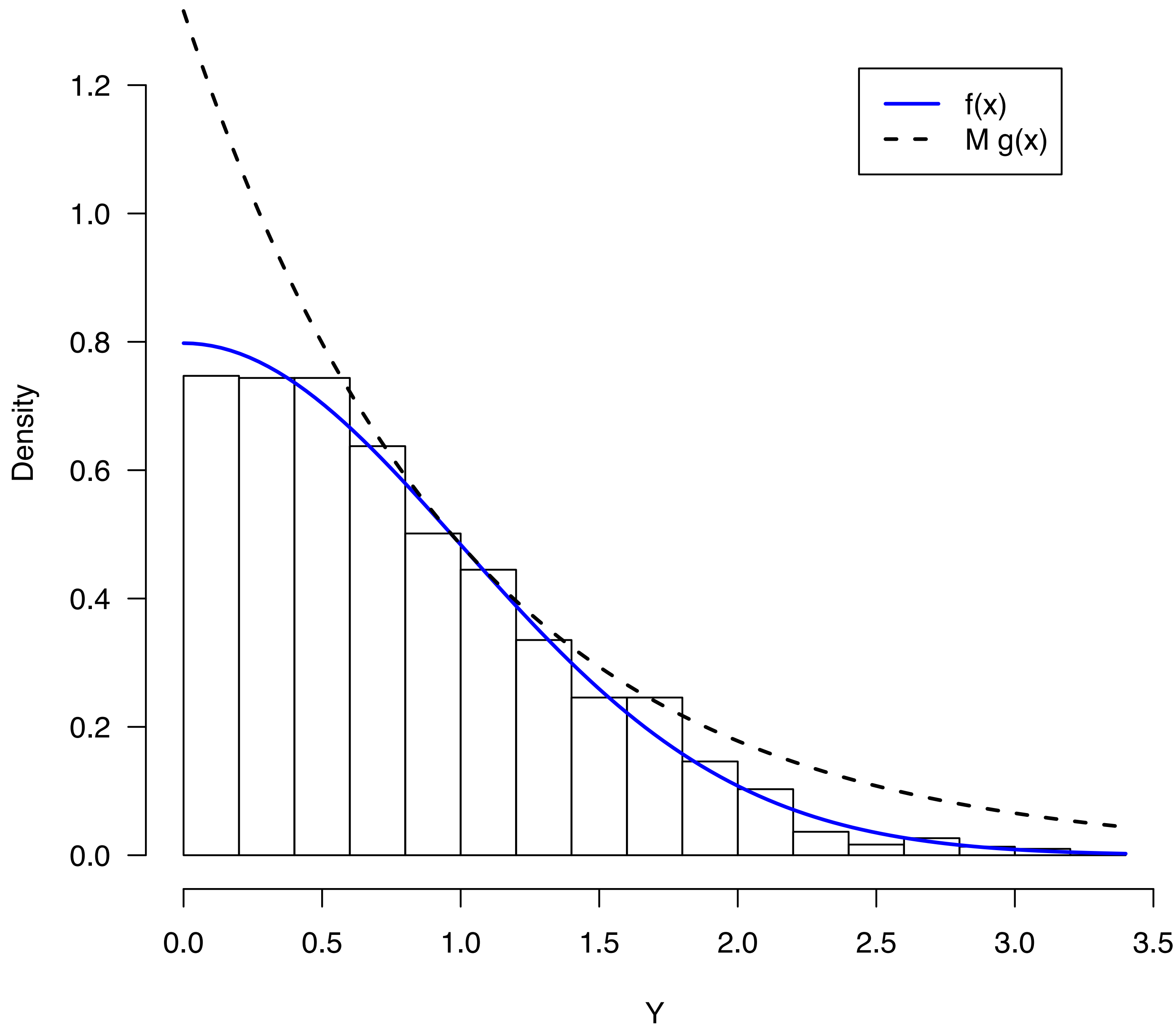
V

```
> # Compute r(X)
> r <- exp(X - X^2 / 2 - 0.5)
```

$r(x) = \frac{f(x)}{g(x) \cdot M}$

```
> # Generate n points from Unif(0, 1)
> U <- runif(n, 0, 1)

> # Accept points if U <= r(X)
> Y <- X[U <= r]
```

Example 2: Folded Normal Distribution



Example 2: Folded Normal Distribution

R Code for the plot:

```
> M <- sqrt(2 * exp(1) / pi) # envelope multiplier

> hist(Y, prob = TRUE, las = 1, main = "", ylim = c(0, M))
> curve(sqrt(2 / pi) * exp(-x^2 / 2),
+       col = "blue", lwd = 2, add = TRUE
+       )
> curve(M * exp(-x), lty = 2, lwd = 2, add = TRUE)
> legend("topright", c("f(x)", "M g(x)"), inset = 0.1
+        lty = 1:2, lwd = 2, col=c("blue", "black")
+        )
```

R Code for acceptance rates:

```
> length(Y) / n # empirical acceptance rate
[1] 0.753
> 1 / M # theoretical acceptance rate
[1] 0.7601735
```

Example 2a: Standard Normal Distribution

- Let $Z \sim \mathcal{N}(0,1)$ with PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \text{for } z \in (-\infty, \infty).$$

Suppose we want to sample from Z.

- Using rejection sampling, we can now sample from $X = |Z|$, which has PDF

$$f_X(x) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \sqrt{\frac{2}{\pi}} e^{-x^2/2}, \quad \text{for } x \geq 0.$$

- How can we use $X = |Z|$ to generate samples from Z?

first $U \to g(x) \to f(x) \to Z$

$Z = S \cdot X \begin{cases} S = 1 \\ S = -1 \end{cases}$

Example 2a: Standard Normal Distribution

Define:

$$S = \begin{cases} 1 & \text{with probability } 0.5 \\ -1 & \text{with probability } 0.5 \end{cases}$$

1. Generate $X \sim |Z|$.
2. Then $Z = S \cdot X \sim \mathcal{N}(0, 1)$.