Lab 9

November 10, 2016

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Rejection sampling

Idea: Simulate random variables (pair of coordinates) distributed uniformly over the region under the density curve (F). However doing so directly may be difficult.

Simulate a random variables uniformly in a region that encapsulates the target density.

i.e. simulate from $Y \sim g$, $U \sim \mathrm{U}(0, Mg(Y))$, where $Mg(t) \geq f(t)$ everywhere.

Algorithm (Rejection Sampling)

1. Generate $Y \sim g$ and generate $U \sim \mathrm{U}(0, Mg(Y))$.
2. If $U \leq f(Y)$, Stop and return Y.
3. Otherwise, reject Y and go back to Step 1.

Here is an equivalent version:

Algorithm (Rejection Sampling)

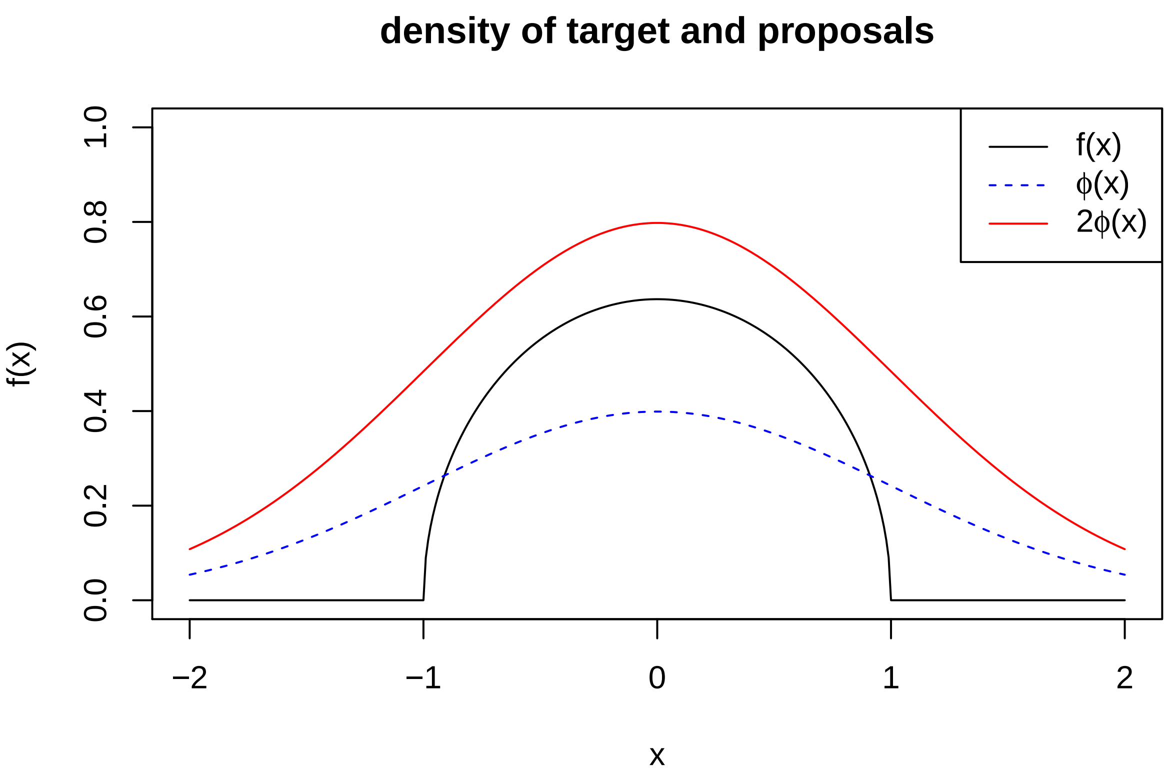
1. Generate Y g and generate $U \sim \mathrm{U}(0, 1)$.
2. If $UMg(Y) \leq f(Y)$, Stop and return Y.
3. Otherwise, reject Y and go back to Step 1.

Overall efficiency of the algorithm is $1/M$. Therefore we would like the constant M to be as small as possible (M cannot be less than 1; why?).

Example

simulate circular distribution on $[-1, 1]$ using standard normal as the proposal distribution.

```
xseq <- seq(-2,2,0.01)
f <- function(x){
  2/pi*sqrt(pmax(1-x^2,rep(0,length(x))))
}
plot(xseq,f(xseq),'l',ylim=c(0,1),ylab = "f(x)",xlab="x",main = "density of target and proposals")
lines(xseq,dnorm(xseq),'l',col=4,lty=2)
lines(xseq,2*dnorm(xseq),'l',col=2)
legend("topright",
       legend=c('f(x)',
                expression(paste(phi,'(x)')),
                expression(paste('2',phi,'(x)'))),
       col=c(1,4,2),lty = c(1,2,1))
```

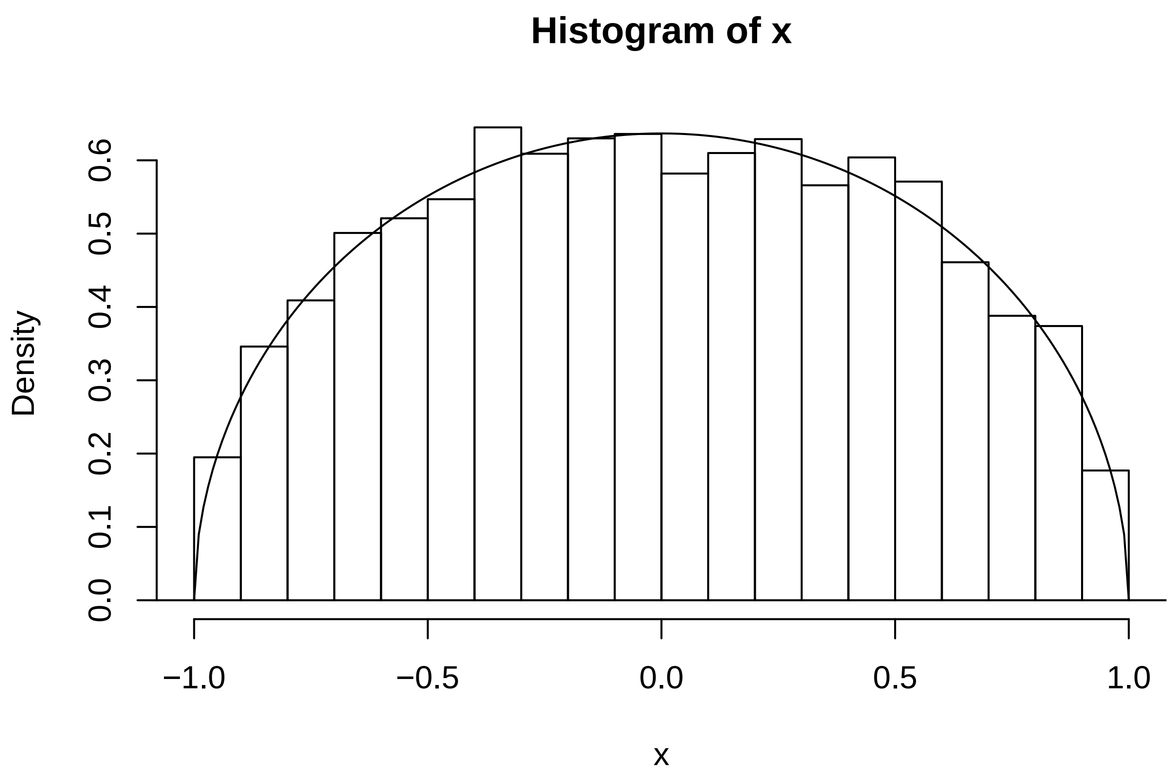



```
rejection_sampling <- function(n)
{
  ## constant M
  M <- 2

  ## Initialize the output
  x <- rep(0, n)

  k <- 0 ## Number of accepted samples
  ## While loop until there are n accepted samples
  while (k <= n)
  {
    y <- rnorm(1)
    u <- runif(1)
    if (u*M*dnorm(y) <= f(y))
    {
      k <- k + 1
      x[k] <- y
    }
  }
  return(x)
}

x <- rejection_sampling(10000)
hist(x, breaks=20,prob = TRUE)
lines(xseq,f(xseq),'l')
```



Remarks

- We only need to know f up to a constant factor (why?). This is helpful when the normalizing constant for f(x) is expensive to compute.
- Best rejection sampling can do: choose a proper M such that $f(x) \leq Mg(x)$ but f(x) and M g(x) as close as possible. i.e. M as small as possible; best choice of M is

$$M^* = \sup_x \frac{f(x)}{g(x)}$$

Monte Carlo integration

- Goal: Approximate $I := \int_\Omega f(x)\mathrm{d}x$, by Monte Carlo.
- Why Monte Carlo integration?
  - algebraic solution not available
  - high dimensional integration

How does it work?

Simple case: integration over region of finite volume.

If $V(\Omega) < \infty$, we can simulate a random variable uniformly over the region Ω, with density $\frac{1}{V(\Omega)}$.

$$\begin{aligned} I = \int_\Omega f(x)\mathrm{d}x &= \int_\Omega \frac{1}{V(\Omega)} f(x) V(\Omega)\mathrm{d}x \\ &= V(\Omega)\int_\Omega \frac{1}{V(\Omega)} f(x)\mathrm{d}x \\ &= V(\Omega)\mathrm{E}[f(X)] \quad \text{where} \quad X \sim \mathrm{Uniform}(\Omega) \end{aligned}$$

Where we know (by Law of Large Numbers) that $\mathrm{E}[f(X)]$ can be approximated by

$$\frac{1}{N}\sum_{i=1}^{N} f(X_i) \quad \text{where} \quad X_i \overset{iid}{\sim} \text{Uniform}(\Omega)$$

So our estimator of the integral is

$$\hat{I} = \frac{V(\Omega)}{N}\sum_{i=1}^{N} f(X_i) \quad \text{where} \quad X_i \overset{iid}{\sim} \pi$$

More generally: when the region has infinite volume

When $V(\Omega)$ is $+\infty$, we can no longer simulate a uniform random variable over Ω.

We can, however, simulate a random variable over the region Ω with density π, such that $\int_\Omega \pi(x)\mathrm{d}x = 1$.

$$\begin{aligned} I = \int_\Omega f(x)\mathrm{d}x &= \int_\Omega \frac{f(x)}{\pi(x)}\pi(x)\mathrm{d}x \\ &= \mathrm{E}\frac{f(X)}{\pi(X)} \quad \text{where} \quad X \sim \pi \end{aligned}$$

By the same token as in the simple case, $\mathrm{E}\frac{f(X)}{\pi(X)}$ can be approximated by

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} \frac{f(X_i)}{\pi(X_i)} \quad \text{where} \quad X_i \overset{iid}{\sim} \pi$$

How good are the estimates?

Let's look at the variance of the estimator

$$\begin{aligned} \mathrm{Var}(\hat{I}) &= \mathrm{Var}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{\pi(X_i)}\Big) \\ &= \frac{1}{N^2}\mathrm{Var}\Big(\sum_{i=1}^{N}\frac{f(X_i)}{\pi(X_i)}\Big) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}\Big(\frac{f(X_i)}{\pi(X_i)}\Big) \\ &= \frac{1}{N}\mathrm{Var}\Big(\frac{f(X)}{\pi(X)}\Big) \end{aligned}$$

Now,

$$\begin{aligned} \mathrm{Var}\Big(\frac{f(X)}{\pi(X)}\Big) &= \mathrm{E}\Big[\Big(\frac{f(X)}{\pi(X)}\Big)^2\Big] - \Big(\mathrm{E}\Big[\frac{f(X)}{\pi(X)}\Big]\Big)^2 \\ &= \int_\Omega \Big(\frac{f(x)}{\pi(x)}\Big)^2 \pi(x)\mathrm{d}x - I^2 \\ &= \int_\Omega \frac{f(x)^2}{\pi(x)}\mathrm{d}x - I^2 \end{aligned}$$

If the first integral is finite, variance of $\frac{f(X)}{\pi(X)}$ is bounded, and the variance of the average goes to 0 as N increases.

Algorithm

To approximate $I = \int_\Omega f(x)\mathrm{d}x$.

1. Simulate $X_i \sim \pi$
2. Calculate $\hat{I} = \frac{1}{N}\sum_{i=1}^N \frac{f(X_i)}{\pi(X_i)}$

Example

Let

$$f(x) = \sin\left(\frac{\cos(x)}{x^3}\right)$$

We wish to approximate the following integral

$$I = \int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-\infty}^{\infty} \sin\left(\frac{\cos(x)}{x^3}\right)\mathrm{d}x$$

- $\sin(\cos(x)/x^3)$ is integrable.

$$\begin{aligned}
\int_{-\infty}^{\infty} \sin\left(\frac{\cos(x)}{x^3}\right)\mathrm{d}x &\le \int_{-\infty}^{\infty} \left|\sin\left(\frac{\cos(x)}{x^3}\right)\right|\mathrm{d}x \\
&\le \int_{-2}^{2} 1\mathrm{d}x + 2\int_{2}^{\infty} \left|\sin\left(\frac{\cos(x)}{x^3}\right)\right|\mathrm{d}x \\
&\le \int_{-2}^{2} 1\mathrm{d}x + 2\int_{2}^{\infty} \sin(1/x^3)\mathrm{d}x \quad \left(\text{since} \quad \frac{\cos(x)}{x^3} < \frac{1}{x^3} < \frac{\pi}{2} \quad \forall x > 2\right) \\
&\le 4 + 2\int_{2}^{\infty} \frac{1}{x^3} - \frac{1}{6x^9}\mathrm{d}x \quad \left(\text{since} \quad \sin(1/x^3) < \frac{1}{x^3} - \frac{1}{6x^9}\right) \\
&< \infty
\end{aligned}$$

In practice we usually are not able to figure integrability of the target function.

- f is an odd function (why?), so the integral is 0.

But imagine we do not know the value of the integral, and we want to approxiamte via Monte Carlo methods. (That's the purpose of Monte Carlo integration. We mention the true value of the integral here for the purpose of illustrating the choice of π; more on this later.)

- choice of π.

The region to integrate over is infinite volume, so we resort to Monte Carlo integration in the general setting.

Here we have the freedom to choose π. Consider two choices of $\pi(x)$:

1. Cauchy: $\pi(x) = \frac{1}{\pi(1+x^2)}$

Here $\lim_{x\to\infty} f(x)/\pi(x) = 0$. The estimator $\hat{I}$ is "stable" in the sense that $\mathrm{Var}(\hat{I}) = \frac{1}{N}\mathrm{Var}\left(\frac{f(X)}{\pi(X)}\right) \to 0$ as $N \to \infty$, because.

$$
\begin{aligned}
\mathrm{Var}\Big(\frac{f(X)}{\pi(X)}\Big) &= \int_{-\infty}^{\infty} \frac{f(x)^2}{\pi(x)} \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \sin^2\Big(\frac{\cos(x)}{x^3}\Big)\pi(1+x^2)\mathrm{d}x \\
&\le 2\pi \int_2^{\infty} \Big|\sin^2\Big(\frac{\cos(x)}{x^3}\Big)\Big|(1+x^2)\mathrm{d}x + \pi\int_{-2}^{2}(1+x^2)\mathrm{d}x \\
&\le 2\pi \int_2^{\infty} O\Big(\frac{1}{x^6}\Big)(1+x^2)\mathrm{d}x + 20\pi \\
&< +\infty
\end{aligned}
$$

2. Gaussian: $\pi(x) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{x^2}{2}\}$

Things are not so pretty if you choose π to be a Gaussian density.

$$
\begin{aligned}
\mathrm{Var}\Big(\frac{f(X)}{\pi(X)}\Big) &= \int_{-\infty}^{\infty} \frac{f(x)^2}{\pi(x)} \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \sin^2\Big(\frac{\cos(x)}{x^3}\Big)\sqrt{2\pi}\exp\{x^2/2\}\mathrm{d}x \\
&\sim \sqrt{2\pi}\int_{-\infty}^{\infty} \frac{\cos^2(x)}{x^6} e^{x^2/2}\mathrm{d}x = \infty
\end{aligned}
$$

The estimator $\hat{I}$ is not stable.

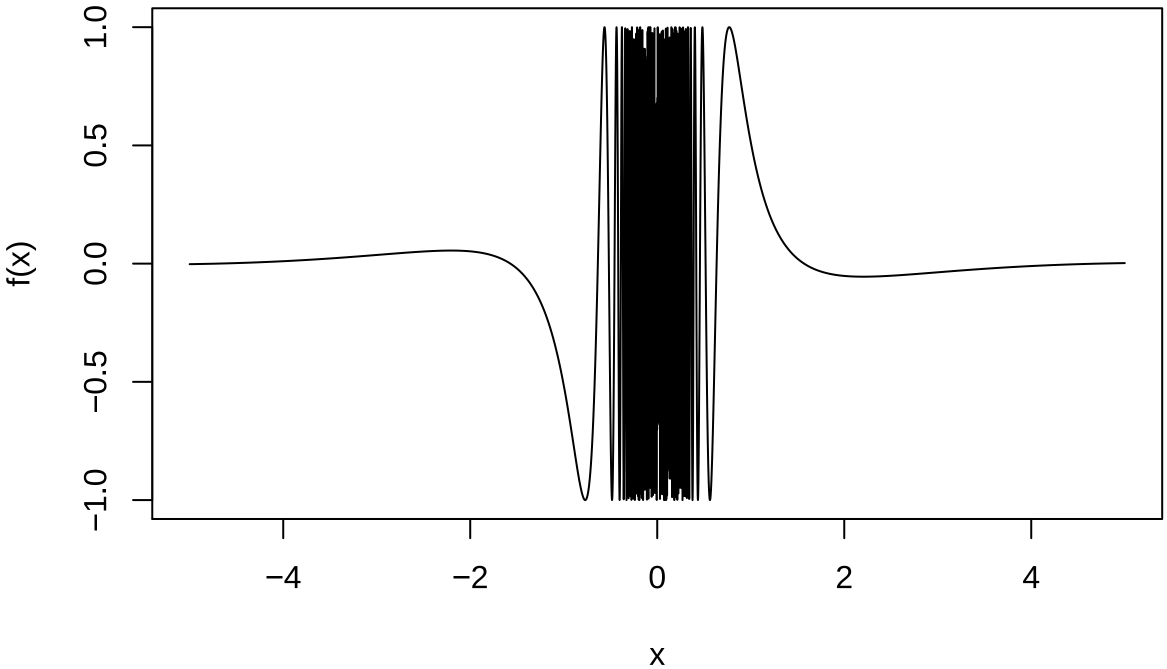
Numerical results

```r
### Monte Carlo integration

rm(list=ls())

# define the integrand
f <- function(x){
  return(sin(cos(x)/x^3))
}

x <- seq(-5.0005,5.0005,0.001)
plot(x,f(x),'l')
```




```r
# start Monte Carlo integration
set.seed(2016)
n <- 1e7 # INCREASE n by adding 0's

## Approach 1(good): using pi(x) = PDF(Cauchy; x)

sample.cauchy <- rcauchy(n)
sample.integrand.cauchy <- f(sample.cauchy) / dcauchy(sample.cauchy)
I.cauchy <- mean(sample.integrand.cauchy)
var.I.cauchy <- var(sample.integrand.cauchy)#/sqrt(n)

cat(paste('mean=', round(I.cauchy, 3), ',  var=', round(var.I.cauchy, 3), '\n', sep=''))
```

```
## mean=0.001,  var=5.177
```

```r
## Approach 2(bad): using pi(x) = PDF(Normal x)
sample.norm <- rnorm(n)
sample.integrand.norm <- f(sample.norm) / dnorm(sample.norm)
I.norm <- mean(sample.integrand.norm)
var.I.norm <- var(sample.integrand.norm)#/sqrt(n)

cat(paste('mean=', round(I.norm, 3), ',  var=', round(var.I.norm, 3), '\n', sep=''))
```

```
## mean=-0.004,  var=99.268
```

```r
## NOTE: normally we report standard error of Monte Carlo estimate instead of sd of the samples,
## here we report the latter to emphasize the point that using a normal sample has infinite variance.

# results:
# n = 1e7:
# (cauchy): mean=0.001,  var=5.177
# (normal): mean=-0.004,  var=99.268
```