Lecture 4: How to sample from C_f?

We have $C_f = \{(x_1, x_2) \mid 0 \le x_1 \le \sqrt{f^\star\left(\frac{x_2}{x_1}\right)}\}$. If $f^\star(x)$ and $x^2 f^\star(x)$ are bounded we have

$$C_f \subset [0, a] \times [b_-, b_+], \qquad \text{with}$$

- $a = \sqrt{\sup_x f^\star(x)} > 0$
- $b_+ = \sqrt{\sup_{x \ge 0}(x^2 f^\star(x))}$
- $b_- = -\sqrt{\sup_{x \le 0}(x^2 f^\star(x))}$

Proof: see blackboard

Use Rejection sampling to sample from C_f.

Methods based on mixtures

Remember: $f(x_1, x_2) = f(x_1|x_2) f(x_2)$

Thus: To generate $(x_1, x_2) \sim f(x_1, x_2)$ we can

- generate $x_2 \sim f(x_2)$
- generate $x_1 \sim f(x_1|x_2)$, where x_2 is the value just generated.

Note: This mechanism automatically provides a value x_1 from its marginal distribution, i.e. $x_1 \sim f(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2$.

$\Rightarrow$ We are able to generate a value for x_1 even when its marginal density is awkward to sample from directly.

Example: Simulation from Student-t (I)

The density of a Student t distribution with $n > 0$ degrees of freedom, mean μ and scale σ^2 is

$$f_t(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi\sigma^2}} \left[1 + \frac{1}{n}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{n+1}{2}}, \quad -\infty < x < \infty.$$

Let

$$x_2 \sim \text{Ga}\left(\frac{n}{2}, \frac{nS}{2}\right)$$
$$x_1 | x_2 \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{x_2}\right)$$

It can be shown that then

$$x_1 \sim t_n(\mu, S\sigma^2) \quad \text{(show yourself)}$$

Example: Simulation from Student-t (II)

Thus, we can simulate $x_1 \sim t_n(\mu, \sigma^2)$ by

$$x_2 \sim \text{Ga}\left(\frac{n}{2}, \frac{n}{2}\right)$$
$$x_1 \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{x_2}\right)$$

return x_1.

Another application is sampling from a mixture distribution, i.e. mixture of two normals.

Multivariate normal distribution

$\boldsymbol{x} = (x_1, \ldots, x_d)^\top \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ if the density is

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \cdot \frac{1}{\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

with

- $\boldsymbol{x} \in \mathbb{R}^d$
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top$
- $\Sigma \in \mathbb{R}^{d \times d}$, Σ must be positive definite.

Important properties (I)

Important properties of $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ (known from "Linear statistical models")

i) Linear transformations:
$\boldsymbol{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma) \Rightarrow \boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{b} \sim \mathcal{N}_r(\mathbf{A}\boldsymbol{\mu} + \boldsymbol{b}, \mathbf{A}\Sigma\mathbf{A}^\top)$, with $\mathbf{A} \in \mathbb{R}^{r \times d}$, $\boldsymbol{b} \in \mathbb{R}^r$.

ii) Marginal distributions:
Let $\boldsymbol{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ with

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

Then

$$\boldsymbol{x}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_{11})$$
$$\boldsymbol{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_{22})$$

Important properties (II)

iii) Conditional distributions:
With the same notation as in ii) we also have

$$\boldsymbol{x}_1 | \boldsymbol{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

iv) Quadratic forms:
$\boldsymbol{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma) \Rightarrow (\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \sim \chi^2_d$

Simulation from the multivariate normal

How can we simulate from $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$?

Let $\boldsymbol{x} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I})$

$$\boldsymbol{y} = \boldsymbol{\mu} + \mathbf{A}\boldsymbol{x} \quad \overset{\text{i)}}{\Rightarrow} \quad \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{A}\mathbf{A}^\top)$$

Thus, if we choose $\mathbf{A}$ so that $\mathbf{A}\mathbf{A}^\top = \Sigma$ we are done.

Note: There are several choices of $\mathbf{A}$. A popular choice is to let $\mathbf{A}$ be the Cholesky decomposition of Σ.

Rejection sampling

We discuss a general approach to generate samples from some target distribution with density $f(x)$, called rejection sampling, without actually sampling from $f(x)$.

Rejection sampling

The goal is to effectively simulate a random number $X \sim f(x)$ using two independent random numbers

- $U \sim \mathrm{U}(0,1)$ and
- $X \sim g(x)$,

where $g(x)$ is called proposal density and can be chosen arbitrarily under the assumption that there exists an $c \geq 1$ with

$$f(x) \leq c \cdot g(x) \quad \text{for all } x \in \mathbb{R}\,.$$

Rejection sampling - Algorithm

Let $f(x)$ denote the target density.

1. Generate $x \sim g(x)$
2. Compute $\alpha = \frac{1}{c} \cdot \frac{f(x)}{g(x)}$.
3. Generate $u \sim \mathcal{U}(0,1)$.
4. If $u \leq \alpha$ return x (acceptance step).
5. Otherwise go back to (1) (rejection step).

Note $\alpha \in [0,1]$ and α is called acceptance probability.

Claim: The returned x is distributed according to $f(x)$.

Proof

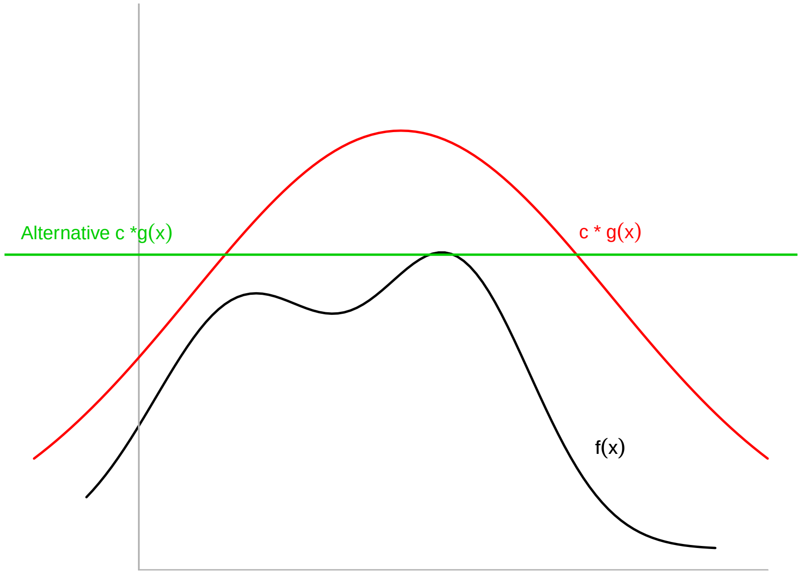
Rejection sampling

- We want $x \sim f(x)$ (density).
- We know how to generate realisations from a density $g(x)$
- We know a a value $c > 1$, so that $\frac{f(x)}{g(x)} \leq c$ for all x where $f(x) > 0$.

Algorithm:

finished $= 0$
while (finished $= 0$)
    generate $x \sim g(x)$
    compute $\alpha = \frac{1}{c} \cdot \frac{f(x)}{g(x)}$
    generate $u \sim U[0,1]$
    if $u \leq \alpha$ set finished $= 1$
return x

Rejection sampling



Example: Setting

Suppose we want to sample standard normal random numbers.

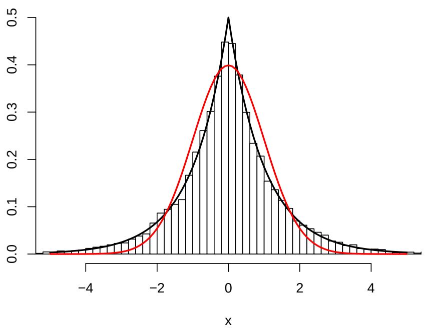
Then

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

As proposal distribution we use a double exponential distribution:

$$g(x) = \frac{\lambda}{2} \exp(-\lambda|x|), \lambda > 0$$

```
> g <- function(x, lambda=1){
+   return(lambda/2 *
+     exp(-lambda * abs(x)))
+ }
> rg <- function(n, lambda){
+   z = rexp(n, lambda)
+   y = sample(c(0,1), n,
+     prob=c(0.5,0.5), replace=TRUE)
+   x = c(z[y==0], -z[y==1])
+   return(x)
+ }
```



Rejection sampling

The overall acceptance probability is

$$P(c \cdot U \cdot g(x) \leq f(x)) = \int_{-\infty}^{\infty} \frac{f(x)}{c \cdot g(x)} g(x)\, dx = \int_{-\infty}^{\infty} \frac{f(x)}{c}\, dx = c^{-1}.$$

The single trials are independent, so the number of trials up to the first success is geometrically distributed with parameter $1/c$. The expected number of trials up to the first success is therefore c.

Problem:

In high-dimensional spaces c is generally large so that many samples will get rejected.

Example: Find an efficient bound c

$$\begin{aligned}
\frac{f(x)}{g(x)} &= \frac{\frac{1}{\sqrt{2\pi}} \exp(-1/2x^2)}{\frac{\lambda}{2}\exp(-\lambda|x|)} \\
&= \sqrt{\frac{2}{\pi}} \lambda^{-1} \exp\left(-\frac{1}{2}x^2 + \lambda|x|\right) \\
&\leq \sqrt{\frac{2}{\pi}} \lambda^{-1} \exp\left(\max_{x\in\mathbb{R}}\{-\frac{1}{2}x^2 + \lambda|x|\}\right) \\
&\stackrel{|x|=\lambda}{=} \sqrt{\frac{2}{\pi}} \lambda^{-1} \exp\left(\frac{1}{2}\lambda^2\right) \\
&\equiv c
\end{aligned}$$

Example: Acceptance probability

Thus the acceptance probability becomes

$$\alpha = \frac{1}{c}\frac{f(x)}{g(x)} = \frac{\sqrt{\frac{2}{\pi}}\lambda^{-1}\exp\left(-\frac{1}{2}x^2+\lambda|x|\right)}{\sqrt{\frac{2}{\pi}}\lambda^{-1}\exp\left(\frac{1}{2}\lambda^2\right)}$$

$$= \exp\left\{-\frac{1}{2}x^2 - \frac{1}{2}\lambda^2 + \lambda|x|\right\}$$

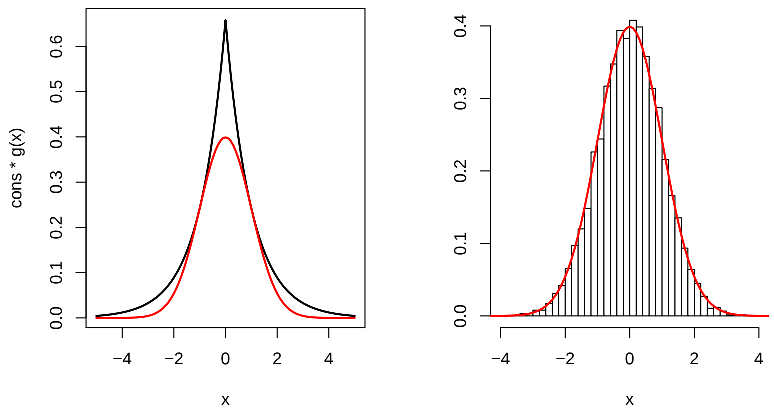
Note, the algorithm is correct for all values of $\lambda > 0$. However, we should choose $\lambda > 0$ so that c becomes as small as possible and consequently α.

$\Rightarrow$ Choose the λ that minimises c which is $\lambda = 1$

$$\frac{f(x)}{g(x)} \leq \sqrt{\frac{2}{\pi}}\lambda^{-1}\exp\left(\frac{1}{2}\lambda^2\right) \overset{\lambda=1}{=} \sqrt{\frac{2}{\pi}}\exp\left(\frac{1}{2}\right) \approx 1.32$$

$(1/1.32 \approx 0.7602)$.

Example: Illustration



- Left: Comparison of $f(x)$ versus $c \cdot g(x)$ when $\lambda = 1$.
- Right: Distribution of accepted samples compared to $f(x)$. 10000 samples were generated and 7582 accepted.

Continuation: Standard Cauchy

How can we sample from the semi-unit circle?

Rejection sampling also works when x is a vector:

$$C_f = \{(x_1, x_2) \mid x_1^2 + x_2^2 \leq 1, x_1 > 0\}$$

with

$$f(x_1, x_2) = \frac{1}{\text{area}(C_f)}, \quad (x_1, x_2) \in C_f$$

Let the proposal density be

$$g(x_1, x_2) = \begin{cases} \frac{1}{2} & x_1 \in [0,1], x_2 \in [-1,1] \\ 0 & \text{otherwise} \end{cases}$$

Thus the density g is that $x_1 \in \mathcal{U}(0,1)$, $x_2 \in \mathcal{U}(-1,1)$ independently.

Standard Cauchy: Rejection sampling algorithm

finished $= 0$

while finished $= 0$ **do**

  generate $(x_1, x_2) \sim g(x_1, x_2)$

  compute

$$\alpha = \frac{1}{c}\frac{f(x_1,x_2)}{g(x_1,x_2)} = \begin{cases} \frac{1}{c}\cdot\frac{2}{\text{area}(C_f)} \overset{c=\frac{2}{\text{area}(C_f)}}{=} 1, & (x_1,x_2) \in C_f \\ 0, & \text{otherwise} \end{cases}$$

  generate $u \sim \mathcal{U}(0,1)$

  if $u \leq \alpha$ **then** finished $= 1$

  end if $\triangleright$ i.e. If $(x_1, x_2) \in C_f$ finished $= 1$

end while

return x_1, x_2

Standard Cauchy: Summary

Note: To do this algorithm we do not need to know the value of the normalising constant area(C_f).

This is always true in rejection sampling.

Rejection sampling - Acceptance probability

Note: For c to be small, $g(x)$ must be similar to $f(x)$.
The art of rejection sampling is to find a g(x) that is similar to f(x) and which we know how to sample from.

Issues: c is generally large in high-dimensional spaces, and since the overall acceptance rate is $1/c$, many samples will get rejected.