Lecture 3: What have we learned..

Sampling from random distribution

- Inversion Method:
  - Discrete RV
  - Continuous RV (where it is possible to compute $F^{-1}(x)$)
- Use known relationship between RV
  - Examples: Gamma, χ^2 distributions
- Change of variables
  - Univariate: scale and location parameters
  - Bivariate: Box-Muller algorithm
- Ratio of uniforms method
  - Don't need to know the normalising constant
  - Example: Cauchy distribution

Inversion method

Let F be a distribution, and let $U \sim \mathcal{U}[0,1]$.

a) Let F be the distribution function of a random variable taking non-negative integer values. The random variable X given by

$$X = x_i \quad \text{if and only if} \quad F_{i-1} < u \leq F_i$$

has distribution function F.

b) If F is a continuous function, the random variable $X = F^{-1}(u)$ has distribution function F.

Change of variables

- Can sample from $X \sim \text{Exp}(1)$.
- Interested in $Y = \frac{1}{\lambda}X$
- Method: sample $X \sim \exp(1)$, return $Y = \frac{1}{\lambda}X$

Why does it work?? We have that: $y = g(x) = \frac{1}{\lambda}x$. Application of the change of variables formula leads to:

$$f_Y(y) = f_X(g^{-1}(x))|\frac{d\ g^{-1}(x)}{d\ x}| = \exp(-\lambda y)\lambda$$

It follows: $Y \sim \text{Exp}(\lambda)$.

Bivariate techniques

Generalization of the previous technique

- $(x_1, x_2) \sim f_X(x_1, x_2)$
- $(y_1, y_2) = g(x_1, x_2) \Leftrightarrow (x_1, x_2) = g^{-1}(y_1, y_2)$
- $f_Y(y_1, y_2) = f_X(g^{-1}(y_1, y_2)) \cdot |\mathrm{J}|$

Example: Box-Muller to simulate from $\mathcal{N}(0, 1)$

Box-Muller algorithm

Let

$$X_1 \sim \mathcal{U}[0, 2\pi] \text{ and } X_2 \sim \text{Exp}\left(\frac{1}{2}\right)$$

independently (We already know how to do this).

Let

$$\left.\begin{aligned} y_1 &= \sqrt{x_2}\cos x_1 \\ y_2 &= \sqrt{x_2}\sin x_2 \end{aligned}\right\} \Leftrightarrow \begin{cases} x_1 = \tan^{-1}\left(\frac{y_2}{y_1}\right) \\ x_2 = y_1^2 + y_2^2 \end{cases}$$

This defines a one-to-one function g.

Then hat $y_1 \sim \mathcal{N}(0, 1)$ and $y_2 \sim \mathcal{N}(0, 1)$ independently.

Graphical interpretation:

Relationship between polar and Cartesian coordinates.

Box-Muller algorithm

Generate

$x_1 \sim U(0, \pi)$

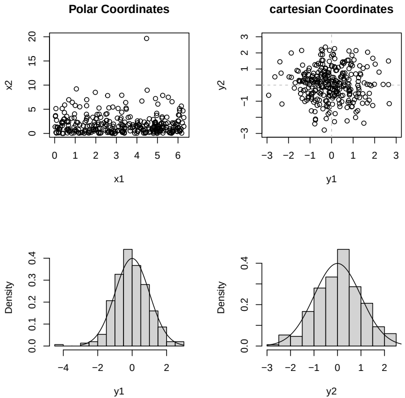
Generate

$x_2 \sim \exp(0.5)$

$y_1 \leftarrow \sqrt{(}x_2) \cos(x_1)$

$y_2 \leftarrow \sqrt{(}x_2) \sin(x_1)$

return (y_1, y_2)



Ratio-of-uniforms method

- $f^\star(x)$ non-negative function with $\int_{-\infty}^{\infty} f^\star(x)dx < \infty$
- $C_f = \{(x_1, x_2)|0 \leq x_1 \leq \sqrt{f^\star(x_2/x_1)}\}$

Thus

a) then C_f has finite area.

Let (x_1, x_2) be uniformly distributed on C_f.

b) Let $y = \frac{x_2}{x_1}$, then $f(y) = \frac{f^\star(y)}{\int_{-\infty}^{\infty} f^\star(u)du}$

Plan for today

- More on ratio of uniform method
- Methods based on mixtures
- The multivariate normal distribution
- Rejection sampling

How to sample from C_f?

- Sometimes we can find an easy way to sample directly from C_f (example: half circle.)
- In general, it can be difficult to sample uniformly from C_f but it is easy in some special cases....

We have $C_f = \{(x_1, x_2) \mid 0 \leq x_1 \leq \sqrt{f^\star\left(\frac{x_2}{x_1}\right)}\}$.

If $f^\star(x)$ and $x^2 f^\star(x)$ are bounded we have

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

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

Ratio of uniform method

If $f^{\star}(x)$ and $x^2 f^{\star}(x)$ are bounded we have

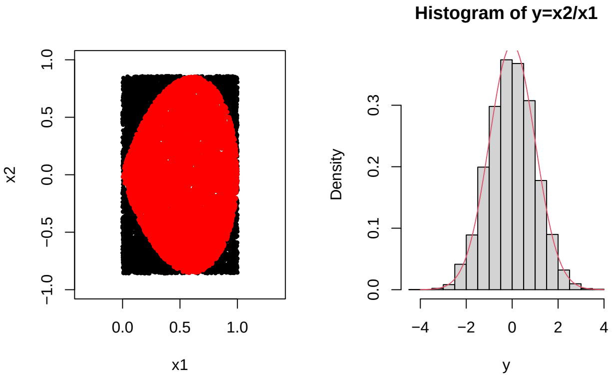
$$C_f \subset [0, a] \times [b_-, b_+], \qquad \text{with}$$

- Proof: see notes from last class
- Use Rejection sampling to sample from C_f.

Example: Normal distribution

see notes from last class

Example: Normal distribution



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.

Here x_2 is called the mixing element.

Methods based on mixtures

Note: This mechanism automatically provides a value x_1 from its marginal distribution,

$$x_1 \sim f(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2.$$

$\Rightarrow$ 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

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

It can be shown that then

$$x_2 \sim t_n(\mu, \sigma^2) \quad \text{(show yourself)}$$

Example: Simulation from Student-t (II)

Thus, we can simulate

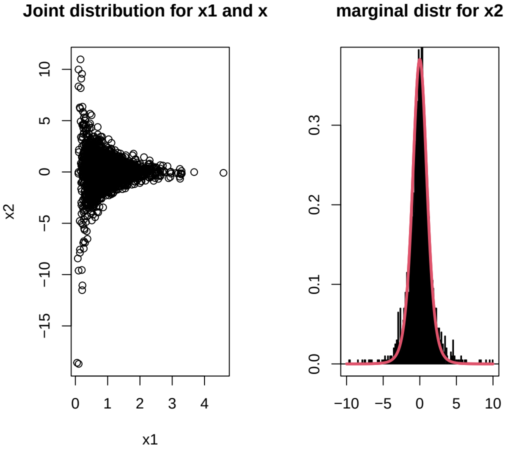
$x_1 \sim t_n(\mu, \sigma^2)$ by

Generate $x_1 \sim \text{Ga}\left(\frac{n}{2}, \frac{n}{2}\right)$

Generate

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

return x_2



Another example: Mixture of normal distributions

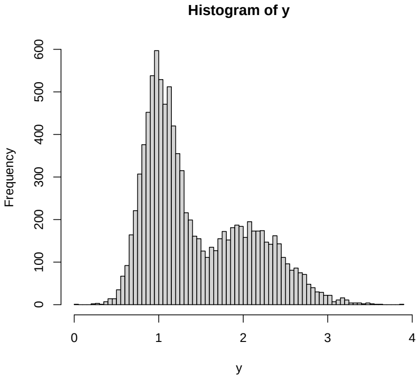
If the mixing element is discrete $f(a_i) = P(X_2 = a_i) = p_i$, $i = 1, \dots, k$, then:

$$f(x_2) = \sum_i f(x_2 | X_1 = a_i) * P(X_1 = a_i)$$

Say we have that $X_1 = \{1, 2\}$ with $p_1 = 0.2$ and $p_2 = 0.8$ and $f(x_2 | X_1 = i) = \mathcal{N}(\mu_i, \sigma_i)$ for $i = 1, 2$

Mixture of normal distribution

Generate $x_1 \sim \text{Binom}(1, p_1)$
if $x_1 = 0$ **then**
  Generate $x_2 \sim \mathcal{N}(\mu_1, \sigma_1)$
else
  Generate $x_2 \sim \mathcal{N}(\mu_2, \sigma_2)$
end if
return x_2



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:

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

ii) Marginal distributions:

Let $\boldsymbol{x} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ with $\Sigma^{-1} = \mathrm{Q}$, and:

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

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,

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

and,

$$\boldsymbol{x}_1|\boldsymbol{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_1 - \boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{Q}_{11}^{-1}).$$

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

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

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

Note: There are several choices of A. A popular choice is to let 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. Generate $u \sim \mathcal{U}(0,1)$.
3. Compute $\alpha = \frac{1}{c} \cdot \frac{f(x)}{g(x)}$.
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

See blackboard

Rejection sampling

- We want $x \sim f(x)$ (density).
- We know how to generate realisations from a density $g(x)$
- We know 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 $x \sim g(x)$

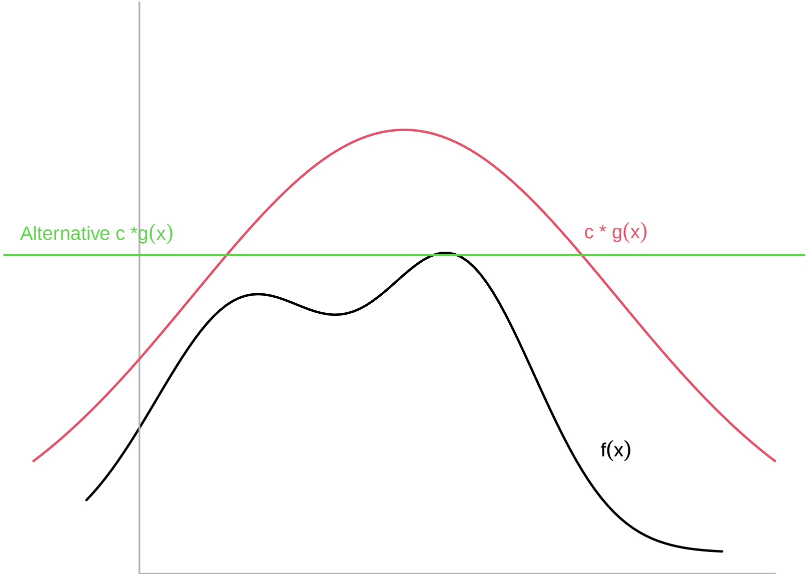
    compute $\alpha = \frac{1}{c} \cdot \frac{f(x)}{g(x)}$

    generate $u \sim U[0, 1]$

    if $u \leq \alpha$ set finished $= 1$

return x

Rejection sampling



Rejection sampling

What is the overall acceptance probability??

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

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 many samples will get rejected.

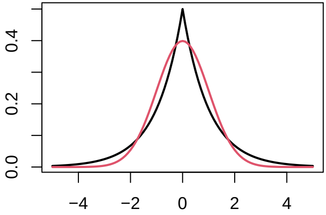
Example: Sample from $N(0,1)$ with rejection sampling

Target distribution:

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

Proposal distribution:

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



Sampling from a double exponential

Proposal distribution:

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

How to sample from $g(x)$:

Sampling from a double exponential

Proposal distribution:

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

How to sample from $g(x)$:

Simulate $z \sim \exp(\lambda)$

Simulate

$y \sim \text{Bern}(p = 0.5)$

if $y = 0$ **then**

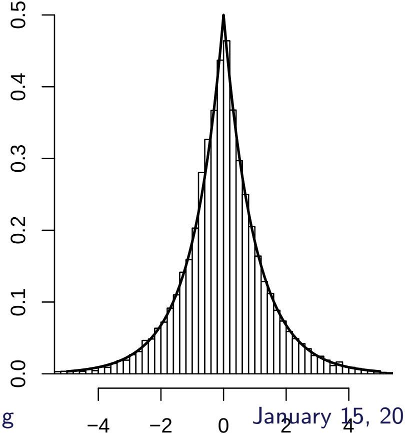
  $x = z$

else

  $x = -z$

end if

return x



Example: Sample from $N(0,1)$ with rejection sampling

Target distribution:

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

Proposal distribution:

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

- Need to find c such that $\frac{f(x)}{g(x)} < c$, $\forall x$ where $f(x) > 0$

Example: Find an efficient bound c

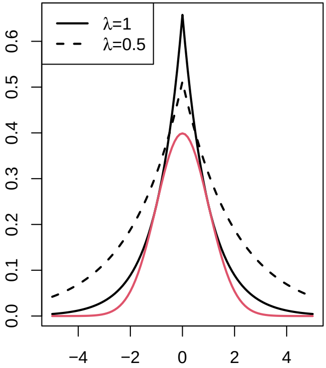
$$\frac{f(x)}{g(x)} \leq \sqrt{\frac{2}{\pi}}\lambda^{-1} \exp\left(\frac{1}{2}\lambda^2\right) \leq c$$

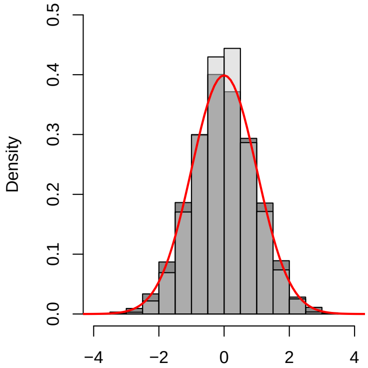
Which value of λ should we choose for proposal density?

We need to choose the smallest possible value for c

Do this yourself

Example: Illustration





- Left: Comparison of $f(x)$ versus $c \cdot g(x)$ when $\lambda \stackrel{\text{x1}}{=} 1$ and $\lambda = 1$.
- Right: Distribution of accepted samples compared to $f(x)$.

  10000 samples were generated and 7582 accepted for $\lambda = 1$.

  10000 samples were generated and 4774 accepted for $\lambda = 0.5$.

Temporary page!

LaTeX was unable to guess the total number of pages correctly. A
there was some unprocessed data that should have been added t
the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page
will go away, because LaTeX now knows how many pages to expe
for this document.