# Lecture 3: Review

- Inversion Method:
  - Discrete RV
  - Continuous RV (where it is possible to compute $F^{-1}(x)$)
- Use known relationship between RV
  - Examples: Gamma, $\chi^2$ 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

# Review: Bivariate techniques

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

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

# Review: 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.

# Review: 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 $f(y) = \frac{f^\star(y)}{\int_{-\infty}^{\infty} f^\star(u)du}$

**In general, it can be difficult to sample uniformly from $C_f$....**

# How to sample 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

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

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

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 $\mu$ and scale $\sigma^2$ 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

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

It can be shown that then

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

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

Another application is, 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

- $\boldsymbol{x} \in \mathbb{R}^d$
- $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^\top$
- $\Sigma \in \mathbb{R}^{d \times d}$, $\Sigma$ 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 $\mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$?

# Simulation from the multivariate normal

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

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

# 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 $\Sigma$.

# 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

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 $\alpha$ 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 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 $x \sim g(x)$

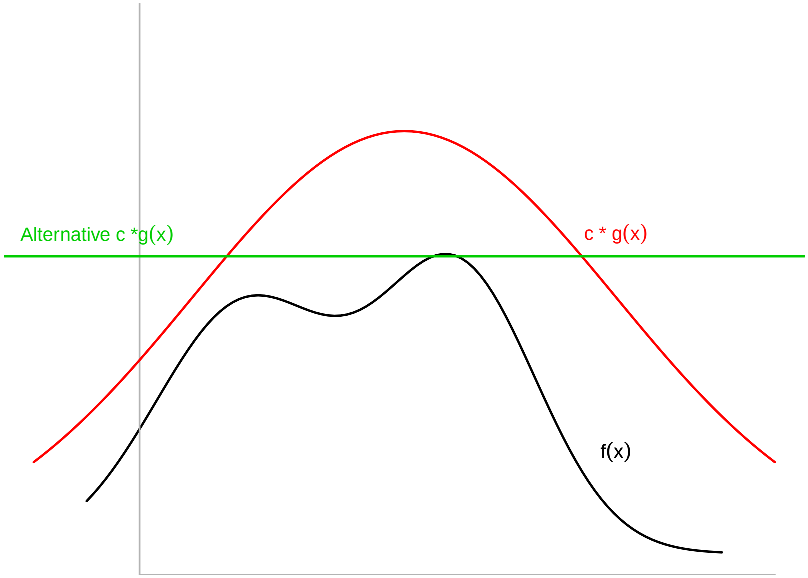
  compute $\alpha = \frac{1}{c} \cdot \frac{f(x)}{g(x)}$

  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??

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

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.

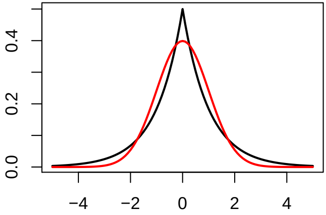
# Example: Sample from $N(0,1)$ with rejection sampling

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 $x \sim \exp(\lambda)$

Simulate

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

**if** $y = 0$ **then**

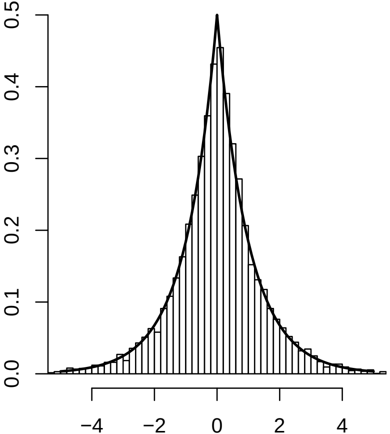
$x = z$

**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 $\lambda$ should we choose?

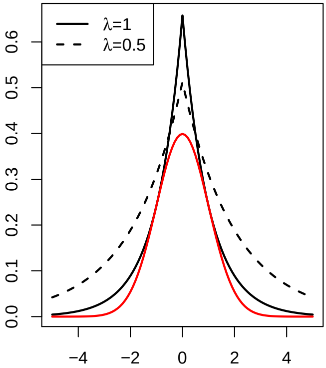
# Example: Find an efficient bound c

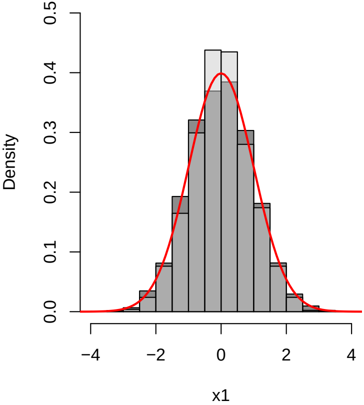
$$\frac{f(x)}{g(x)} \leq \sqrt{\frac{2}{\pi}}\lambda^{-1}\exp\left(\frac{1}{2}\lambda^2\right) \leq c$$

Which value of $\lambda$ should we choose?

We need to choose the smallest possible value for $c$

# Example: Illustration




- Left: Comparison of $f(x)$ versus $c \cdot g(x)$ when $\lambda = 1$ 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 $\lambda = 0.5$.