Proof of theorem

see blackboard

Review scaling: Change of variables

$X \sim \text{Exp}(1)$. We are interested in $Y = \frac{1}{\lambda}X$, i.e. $y = g(x) = \frac{1}{\lambda}x$.

$$g^{-1}(y) = \lambda y \qquad \frac{dg^{-1}(y)}{dy} = \lambda$$

Application of the change of variables formula leads to:

$$f_Y(y) = \exp(-\lambda y)\lambda$$

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

Exercise: Check other transformations, we mentioned.

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

How 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 $x_1 \sim f(x_1|x_2)$

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

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

Let

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

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

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

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

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

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.

Read chapter 1.4.2 and 1.4.3 in Gamerman & Lopes yourself

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.

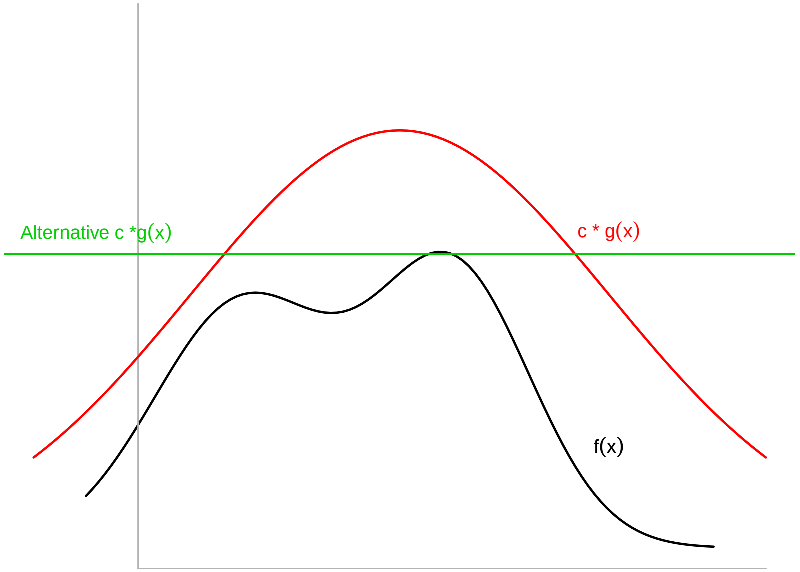
1. Generate independent random variables $X \sim g(x)$ and $U \sim \mathcal{U}(0, 1)$.
2. If $U \leq \frac{f(x)}{c \cdot g(x)}$, you are finished (acceptance step).
3. Otherwise go back to (1) (rejection step).

The quantity $\alpha(x) = f(x)/(c \cdot g(x))$ is called acceptance probability.

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

Proof

Rejection sampling - Illustration



Rejection sampling - Acceptance probability

The acceptance probability can be written as

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

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.

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.