Short reminder

What are pseudo-random numbers?

A deterministic sequence of numbers in $[0,1]$ with the same statistical properties as a sequence of independent $\mathcal{U}(0,1)$ numbers.

Random variables

We shall not always be interested in an experiment itself, but rather in some consequence of its random outcome.

Example: Rolling two dice

Let

$$X := \text{“Sum of the two dice”}$$

Such consequences, when real-valued, may be thought of as functions which maps the state space S into the real line $\mathbb{R}$, and are called random variables.

Discrete random variables

A random variable X is discrete, if they can either take a finite or countable number of values.

We define the probability mass function $\mathrm{p}(x)$ of X by

$$\mathrm{p}(x) = \mathrm{P}(X = x)$$

The following properties have to be fulfilled:

$$\sum_{i=1}^{\infty} \mathrm{p}(x_i) = 1, \quad \mathrm{p}(x_i) \geq 0.$$

The cumulative distribution functions F can be expressed in terms of $\mathrm{p}(a)$ by

$$F(a) = \mathrm{P}(X \leq a) = \sum_{i:x_i \leq a} \mathrm{p}(x_i)$$

Properties of the cumulative distribution function (CDF)

- $F(x)$ is monotone increasing (“step function”) .
- $F(x)$ is piece-wise constant with jumps at elements x_i, where $\mathrm{p}(x_i) > 0$.
- $\lim_{x\to\infty} F(x) = 1$.
- $\lim_{x\to-\infty} F(x) = 0$.

Examples of discrete distributions

- Bernoulli distribution, $\text{Bin}(1, p)$
- Binomial distribution, $\text{Bin}(n, p)$
- Geometric/Negative binomial distribution, $\text{NB}(r, p)$
- Poisson distribution, $\text{Po}(\lambda)$
- ...

Definition of continuous random variables

- Idea: A random variable X is called continuous, if for two arbitrary values $a < b$ from the support of X, every intermediate value in the interval $[a, b]$ is possible.

Continuous distributions

A random variable X whose set of possible values is uncountable, is called a continuous random variable.
A random variable is called continuous, if there exists a function $f(x) \geq 0$, so that the cumulative distribution function $F(x)$ can be written as

$$F(a) = \text{P}(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Some consequences:

- $\text{P}(X = a) = \int_a^a f(x)dx = 0$
- $\text{P}(X \in B) = \int_B f(x)dx$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

The CDF $F(x)$ of continuous random variables

Properties:

- $F(a)$ is a nondecreasing function of a,
  i.e. if $x < y$ then $F(x) < F(y)$
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $\frac{d}{da} F(a) = f(a)$
- $\text{P}(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$
- $\text{P}(X \geq a) = 1 - F(a)$

Normalising constant

A normalising constant c is a multiplicative term in $f(x)$, which does not depend on x. The remaining term is called core:

$$f(x) = c \underbrace{g(x)}_{\text{core}}$$

We often write $f(x) \propto g(x)$.

Examples of continuous distributions

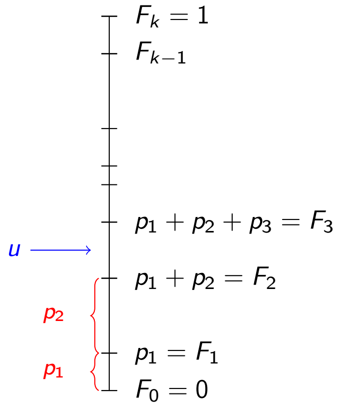
- Uniform distribution $\mathcal{U}[0, 1]$
- Exponential distribution $\text{Exp}(\lambda)$
- Gamma distribution $\text{Ga}(\text{shape} = \alpha, \text{rate} = \beta)$
- Normal distribution $\mathcal{N}(\mu, \sigma^2)$.
- ...

Discrete distributions

Let X be a stochastic variable with possible values $\{x_1, \ldots, x_k\}$ and $\text{P}(X = x_i) = \text{p}_i$. Of course $\sum_{i=1}^{k} p_i = 1$.

An algorithm for simulating a value for x is then:

$u \sim U[0, 1]$
for $i = 1, 2, \ldots, k$ **do**
  if $u \in (F_{i-1}, F_i]$ **then**
    $x \leftarrow x_i$
  end if
end for



Each interval $I_i = (F_{i-1}, F_i]$ corresponds to single value of x.

Proof & Note

Proof.

$$\begin{aligned} \text{P}(X = x_i) &= \text{P}(u \in (F_{i-1}, F_i]) \\ &= \text{P}(u \leq F_i) - \text{P}(u \leq F_{i-1}) \\ &= F_i - F_{i-1} = (p_1 + \ldots + p_i) - (p_1 + \ldots + p_{i-1}) = p_i \end{aligned}$$

□

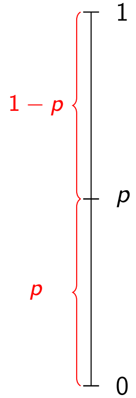
Note: We may have $k = \infty$

- The algorithm is not necessarily very efficient. If k is large, many comparisons are needed.
- This generic algorithm works for any discrete distribution. For specific distributions there exist alternative algorithms.

Bernoulli distribution

Let $S = \{0, 1\}$, $\mathrm{P}(X = 0) = 1 - p$, $\mathrm{P}(X = 1) = p$.

Thus $X \sim \mathrm{Bin}(1, p)$.



The algorithm becomes now:

$u \sim U[0, 1]$
$x = I(u \leq p)$

Binomial distribution

Let $X \sim \mathrm{Bin}(n, p)$.

The generic algorithm from before can be used, but involves tedious calculations which may involve overflow difficulties if n is large.

An alternative is:

$x = 0$
for $i = 1, 2, \ldots, n$ **do**
  generate $u \sim U[0, 1]$
  if $u \leq p$ **then**
    $x \leftarrow x + 1$
  end if
end for
return x

Geometric and negative binomial distribution

The negative binomial distribution gives the probability of needing x trials to get r successes, where the probability for a success is given by p. We write $X \sim \mathrm{NB}(r, p)$.

The generic algorithm can still be used, but an alternative is:

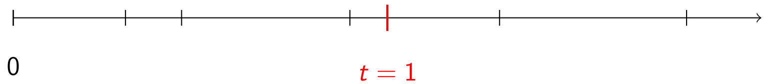
$s = 0$ ▷ (# of successes)
$x = 0$ ▷ (# of tries)
while $s < r$ **do**
  $u \sim U[0, 1]$
  $x \leftarrow x + 1$
  if $u \leq p$ **then**
    $s \leftarrow s + 1$
  end if
end while
return x

Poisson distribution

Let $X \sim \mathrm{Po}(\lambda)$, i.e. $f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, 2, \ldots$.

An alternative to the generic algorithm is:

$x = 0$ ▷ (# of events)
$t = 0$ ▷ (time)
while $t < 1$ **do**
  $\Delta t \sim \mathrm{Exp}(\lambda)$
  $t \leftarrow t + \Delta t$
  $x \leftarrow x + 1$
end while
$x \leftarrow x - 1$
return x



It remains to decide how to generate $\Delta t \sim \mathrm{Exp}(\lambda)$.

Change of variables formula

Let X be a continuous random variable with density $f_X(x)$.
Consider now the random variable $Y = g(X)$, where for example $Y = \exp(X)$, $Y = X^2$, ...

Question: What is the density $f_Y(y)$ of Y?

For a strictly monotone and differentiable function g we can apply the change of variables formula:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \underbrace{\left| \frac{dg^{-1}(y)}{dy} \right|}_{g^{-1'}(y)}$$

Proof over cumulative distribution function (CDF) $F_Y(y)$ of Y (blackboard).

Example

Consider $X \sim \mathcal{U}[0,1]$ and $Y = -\log(X)$, i.e. $y = g(x) = -\log(x)$.
The inverse function and its first derivative are:

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

Application of the change of variables formula leads to:

$$f_Y(y) = 1 \cdot |-\exp(-y)|$$

It follows: $Y \sim \text{Exp}(1)$! Thus, this is a simple way to generate exponentially distributed random variables!
More generally, leads $Y = -\frac{1}{\lambda}\log(x)$ to random variables from an exponential distribution with parameter λ: $Y \sim \text{Exp}(\lambda)$.

Inverse cumulative distribution function

More generally, inversion method or the probability integral transform approach can be used to generate samples from an arbitrary continuous distribution with density $f(x)$ and CDF $F(x)$:

1. Generate random variable U from the standard uniform distribution in the interval $[0,1]$.
2. Then is

$$X = F^{-1}(U)$$

a random variable from the target distribution.

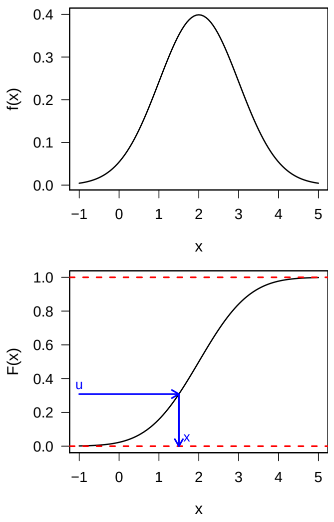
Proof.

$$f_X(x) = \underbrace{f_U(F(X))}_{1} \cdot \underbrace{F'(x)}_{f(x)} = f(x)$$

□

Inverse cumulative distribution function (II)

Let X have density $f(x)$, $x \in \mathbb{R}$ and CDF $F(x) = \int_{-\infty}^{x} f(z)dz$:



Simulation algorithm:

$u \sim U[0,1]$

$x = F^{-1}(u)$

return x

Standard Cauchy distribution

Density and CDF of the standard Cauchy distribution are:

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \text{and} \quad F(X) = \frac{1}{2} + \frac{\arctan(x)}{\pi}$$

The inverse CDF is thus:

$$F^{-1}(y) = \tan\left[\pi\left(y - \frac{1}{2}\right)\right]$$

Random numbers from the standard Cauchy distribution can easily be generated, by sampling $U_1, \ldots, U_n$ from $\mathcal{U}[0,1]$, and then computing $\tan[\pi(U_i - \frac{1}{2})]$.

Note

The inversion method cannot always be used! We must have a formula for $F(x)$ and be able to find $F^{-1}(u)$. This is for example not possible for the normal, χ^2, gamma and t-distributions.

Gamma distribution

Let $X \sim \text{Ga}(\text{shape}{=}\alpha, \text{rate}{=}\beta)$, i.e.

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta \cdot x}, x > 0.$$

From stochastic processes we know that if $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, then $X_1 + \ldots + X_n \sim \text{Ga}(n, \lambda)$.

This gives how to simulate when α is an integer.

Gamma distribution

Further remember: $\chi^2_\nu = \text{Ga}(\frac{\nu}{2}, \frac{1}{2})$,

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \Rightarrow \sum_{i=1}^n X_i^2 \sim \chi^2_n$.

Thus, we can simulate $X \sim \text{Ga}(\frac{n}{2}, \frac{1}{2})$ by

$x = 0$
for $i = 1, 2, \ldots, n$ **do**
  generate $y \sim \mathcal{N}(0,1)$ ▷ Still have to learn how
  $x \leftarrow x + y^2$
end for
return x

Gamma distribution (II)

β is a rate (inverse scale) parameter, i.e.

$$X \sim \text{Ga}(\alpha, 1) \quad \Leftrightarrow \quad X/\beta \sim \text{Ga}(\alpha, \beta)$$

Thus, we can simulate $X \sim \text{Ga}(\frac{n}{2}, \beta)$ by the algorithm

$x = 0$
for $i = 1, 2, \ldots, n$ **do**
  generate $y \sim \mathcal{N}(0,1)$ ▷ Still have to learn how
  $x \leftarrow x + y^2$
end for
$x \leftarrow \frac{1}{2}x$ ▷ $\text{Ga}(\frac{n}{2}, 1)$
$x \leftarrow \frac{1}{\beta}x$ ▷ $\text{Ga}(\frac{n}{2}, \beta)$
return x

Thus, we know how to simulate from a $\text{Ga}(\alpha, \beta)$ whenever $\alpha = \frac{n}{2}$ where n is an integer.

Linear transformations

Many distributions have scale parameters, for example

$$X \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad \sigma X \sim \mathcal{N}(0, \sigma^2)$$

$$X \sim \text{Exp}(1) \quad \Leftrightarrow \quad \frac{1}{\lambda} X \sim \text{Exp}(\lambda)$$

$$X \sim \mathcal{U}[0,1] \quad \Leftrightarrow \quad \beta X \sim \mathcal{U}[0, \beta]$$

Adding a constant can also helping us in some situations

$$X \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad X + \mu \sim \mathcal{N}(\mu, 1)$$

and thereby

$$X \sim \mathcal{N}(0,1) \quad \Leftrightarrow \quad \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$$

Review: inverse transform technique

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

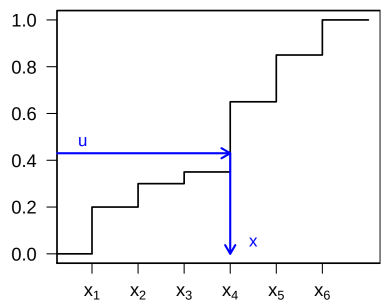
$$X = x_i \quad \text{if and only if} \quad F_{i-1} < u \leq F_i$$

has distribution function F.

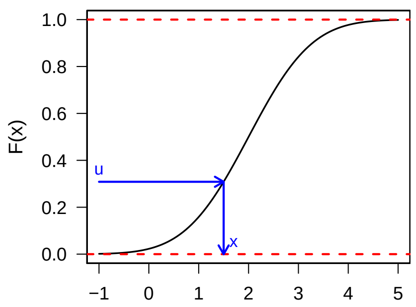
b) If F is a continuous function, the random variable $X = F^{-1}(u)$ has distribution function F.

Review: inverse transform technique (II)

a) Discrete case:



b) Continuous case:



The inverse transform technique is conceptually easy, but

- in the discrete case, a large number of comparisons may be necessary.
- in the continuous case, F^{-1} must be available.

Bivariate techniques

Remember: If $(x_1, x_2) \sim f_X(x_1, x_2)$

and $(y_1, y_2) = g(x_1, x_2)$

$$\Updownarrow$$

$$(x_1, x_2) = g^{-1}(y_1, y_2)$$

where g is a one-to-one differentiable transformation. Then

$$f_Y(y_1, y_2) = f_X(g^{-1}(y_1, y_2))|\mathbf{J}|$$

with the determinant of the Jacobian matrix $\mathbf{J}$

$$|\mathbf{J}| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

$\Rightarrow$ Multivariate version of the change-of-variables transformation

Bivariate techniques (II)

If we know how to simulate from $f_X(x_1, x_2)$ we can also simulate from $f_Y(y_1, y_2)$ by

$$(x_1, x_2) \sim f_X(x_1, x_2)$$
$$(y_1, y_2) = g(x_1, x_2)$$

Return (y_1, y_2).

Example: Normal distribution (Box-Muller)

see blackboard