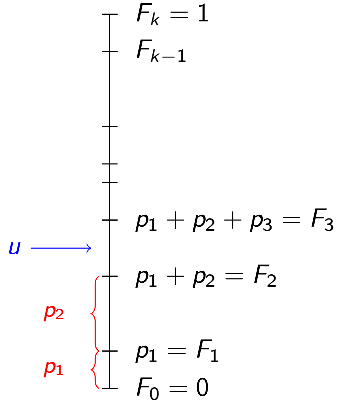
Discrete distributions

Let X be a stochastic variable with possible values $\{x_1, \ldots, x_k\}$ and $\mathrm{P}(X = x_i) = \mathrm{p}_i$. 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.

$$\mathrm{P}(X = x_i) = \mathrm{P}(u \in (F_{i-1}, F_i])$$
$$= \mathrm{P}(u \le F_i) - \mathrm{P}(u \le F_{i-1})$$
$$= F_i - F_{i-1} = (p_1 + \ldots + p_i) - (p_1 + \ldots + p_{i-1}) = p_i$$

□

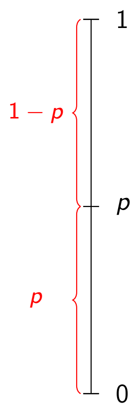
Note: We may have $k = \infty$

- 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:

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

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 $u \le p$ **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 $X \sim \text{NB}(r, p)$.

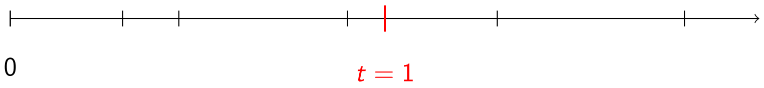
The generic algorithm can still be used, but an alternative is:

$s = 0$ ▷ (# 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 $X \sim \text{Po}(\lambda)$, 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**
  $\Delta t \sim \text{Exp}(\lambda)$
  $t \leftarrow t + \Delta t$
  $x \leftarrow x + 1$
end while
$x \leftarrow x - 1$
return x



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

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 $Y = \exp(X)$, $Y = 1/X$, ...

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.

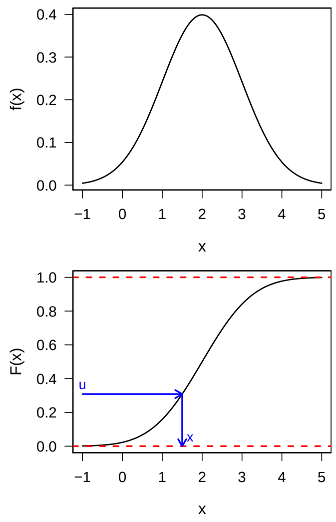
Proof.

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

□

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.