Pseudo-random generator

A pseudo-random generator is a deterministic function f which takes a uniform random bit string as input and outputs a bit string which cannot be distinguished from a uniform random string.

In more detail, this means that for starting value (seed) u_0 and any n, the sequence

$$\{u_0, f(u_0), f(f(u_0)), f(f(f(u_0))), \ldots, f^n(u_0)\}$$

behaves statistically like an $\mathcal{U}(0,1)$ sequence (when appropriately scaled).

Random number generator

From now on, we assume to have a random generator on the unit interval available.

Simulation from discrete distributions

- Let x be a stochastic variable, where $x \in \{x_1, \ldots, x_k\}$ and

$$P(x = x_i) = p_i \quad \text{where} \quad \sum_{i=1}^{k} p_i = 1$$

- Define $F_i = \sum_{j=1}^{i} p_j$ for $i = 0, 1, \ldots, k$
- General simulation algorithm:

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

- Note:
  - can be used for any discrete distribution, may be inefficient
  - efficent searching algorithm can make the algorithm faster
  - used known relations: binomial, geometric, poisson

Simulation from continuous distributions

- Probability integral transform (inversion method)
- Let x have density $f(x), x \in \mathbf{R}$, $F(x) = \int_{-\infty}^{x} f(z)\mathrm{d}z$
- General simulation algorithm:

  $u \sim U[0,1]$
  $x = F^{-1}(u)$
  return x
- We have:
  - proved that the algorithm is correct
  - looked at an examples: exponential distribution
  - understood when the algorithm can be used

Today: More on continuous distributions

- Use known relations: gamma, location/scale parameters
- Bivariate techniques:

$$(x_1, x_2) \sim f_x(x_1, x_2) \quad \Rightarrow \quad (y_1, y_2) = g(x_1, x_2) \sim f_y(y_1, y_2)$$

  - note: Probability integral transform is a univariate technique:

$$x \sim U[0, 1] \quad \Rightarrow \quad y = F^{-1}(x) \sim f(y)$$

  - standard normal distribution (tomorrow)
  - ratio-of-uniforms method
- If we have time:
  - Methods based on mixtures: $f_x(x) = \int f_{x|y}(x|y) f_y(y) \mathrm{d}y$
  - Multivariate normal