LECTURE 3: Generation of Random Variables

Assume that an infinite series of iid uniform random variables on $(\Omega, \mathcal{F})$, $\{U_i\}$ iid $U_i \sim U(0,1)$ is "available". Usually programming languages come with a "random" function that is supposed to generate such a sequence.

Def: A random number generator, or more correctly a pseudo-random number generator is an algorithm of the form $U_{n+1} = \phi_n(U_n, U_{n-1}, \ldots ; w_0)$ such that for every w_0, the sequence $\{U_n\}_{n\geq 1}$ is statistically an iid sequence of rv's on $(0,1)$.

As well, if w_0 has lebesgue measure on $\Omega \subset \mathbb{R}$, (uniformly distributed) then $U_1 = \phi_0(w_0) \sim U(0,1)$.

"Random variates".

Remarks:

(1) Algorithms to make output look "erratic": complexity

(2) Fast operation: simplicity

(3) Statistical hypothesis testing

Fields of mathematics: number theory, algebra, cryptography, statistics.

Def: A generator of a random variable with distribution F, or simply a generator of distribution F, is an algorithm of the form:

$$X = \phi(U_1, \ldots, U_\tau)$$

where $\{U_i\}_{i\geq 1}$ on $(\Omega, \mathcal{F})$ are iid uniform $U(0,1)$ and τ is a random stopping time adapted to $\mathcal{F}_n = \sigma(U_1, \ldots, U_n)$. The generator must satisfy

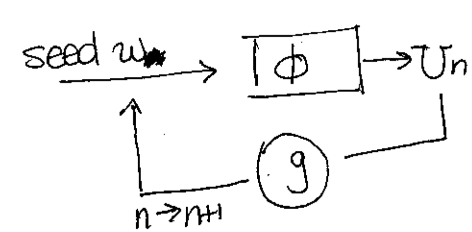
(i) $\mathbb{P}[\phi(U_1, \ldots, U_\tau) \leq x] = F(x) \quad \forall x \in \mathbb{R}$

(ii) $\tau < \infty$ w.p.1 (desirable)

(iii) $\mathbb{E}(\tau)$ "small" (desirable)

Methods for Generation of RV's

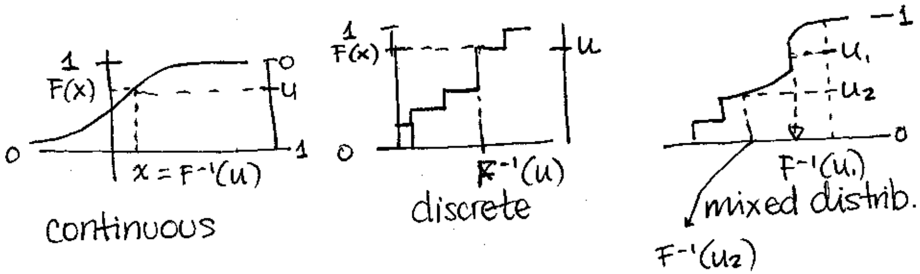
- Inverse Function Method
- Acceptance/Rejection Method
- Composition, Convolution, Transformations



(a). The Inverse Function Method

<u>Def</u>: Let $F: \mathbb{R} \to [0,1]$ be a distribution function (thus a non-decreasing function). The <u>inverse function</u> F^{-1} is:

$$F^{-1}(u) = \inf(x : F(x) \geq u), \quad u \in [0,1].$$



<u>Theorem</u>: Let $U \sim U(0,1)$ be a rv on $(\Omega, \mathcal{F})$ and define $X(U) = F^{-1}(U)$, where F is a distribution function. Then the rv X, defined on $(\Omega, \mathcal{F})$, has distribution F.

<u>Proof</u>: if F is a continuous distribution then:

$$\mathbb{P}(X(U) \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x))$$
$$= F(x), \quad \forall x \in \mathbb{R},$$

if F is a discrete distribution then by construction of F^{-1}:

$$\{X(U) = k\} \iff \{F(k) \leq U < F(k+1)\}$$

so that

$$\mathbb{P}(X(U) = k) = F(k+1) - F(k), \quad (\text{because } U \sim U(0,1))$$

which yields the desired result. The mixed distribution case is shown using the above results. QED.

<u>Exercise</u>: Show the Skorohod representation theorem, using the inverse function method.

See examples in Ross (Simulation): UR.

<u>Remark</u>: if a rv is defined via the inverse function method, so that $X(U) = F^{-1}(U)$, for $U \sim U(0,1)$, then X is a non-decreasing function of U.

<u>Example</u> (SimSpiders) Let $\{Y_1, \ldots Y_n\}$ be iid random vars with known distribution F and suppose that we wish to simulate the extreme values:

$$R = \max(Y_1, \ldots Y_n), \qquad S = \min(Y_1, \ldots Y_n).$$

If we generate each $Y_n = F^{-1}(U_n)$ and then perform the search for the extreme values, then the computational effort is $O(n \log n)$. A clever way to generate R and S more efficiently use monotonicity of $F^{-1}(U_n)$. Indeed

$$\max(F^{-1}(U_1), \ldots, F^{-1}(U_n)) = F^{-1}(\max(U_1, \ldots, U_n)) = R$$

and similarly for S. Because $\{U_1, \ldots U_n\}$ are iid $U(0,1)$, we know that the k-th order statistic $U_{(k)} \sim \text{Beta}(k, n+1-k)$. In particular:

$$U_{(n)} \sim W^{1/n}, \quad W \sim U(0,1)$$
$$U_{(1)} \sim 1 - V^{1/n}, \quad V \sim U(0,1)$$

So we only need to generate <u>one</u> random variable and set $R = F^{-1}(W^{1/n})$, $S = F^{-1}(1 - W^{1/n})$. When n is large this can save considerable execution time.

[See other examples in references].

For the general discrete distribution, the inverse function method considers:

$$X(U) = \min \left(k : F(k) \le U \le F(k) + p_k \right)$$

where $p_k = \mathbb{P}(X = k+1)$. If a sequential search is used:

```
F = p(0) ; N=0
U ~ U(0,1)
while ( F < U ) do
    N++;
    F = F + p(N);
end (while);
Return (N).
```

then the number of iterations ν of the while loop is a random number (depending on U), and it satisfies $\mathbb{E}\nu = \mathbb{E}X + 1$ when $X \in \{0, 1, 2, \dots\}$.

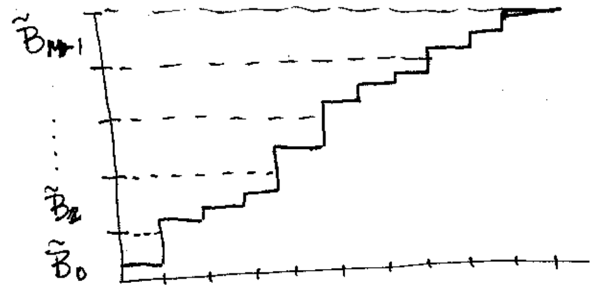
Proof: Let ν be the number of iterations, and notice that if $\{X(U) = n\}$ then $\nu = n+1$. comparisons between F and U. Thus $\nu = X + 1$ a.s., which also implies the result.

Accelerated search: to increase the efficiency, one can use better search methods, depending on the distribution.

Ex: Method of Buckets: consider $M > 0$ an integer, and define the buckets:

$$B_m = \left\{ k : F(k) \in \left[\frac{m}{M}, \frac{m+1}{M}\right) \right\} \quad k = 0, \dots, M-1.$$

Assume first that $\nexists m : B_m = \phi$.



$B_1 = \{0\}$ $B_2 = \{1, 2, 3\}$, $B_3 = \{4\}$

$B_4 = \{5, 6, 7\}$, $B_5 = \{8, 9, 10\}$.

Let $x_m = \min k \in B_m$ be the first element.

The method of buckets generates first the bucket uniformly:

$U \sim U(0,1)$ $\quad m = \lfloor MU \rfloor$ $\quad$ (fast operation)

Then it finds:

$$\min \left(k \geq x_m : F(k) > U \right).$$

→ Is this method more efficient?

By construction, $X = x_m + t_m - 1$, where t_m = # iterations when searching in bucket B_m, thus:

$$\mathbb{E}(T) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}(t_m)$$

and

$$\mathbb{E}(X) + 1 = \frac{1}{M} \sum_{m=1}^{M} \left(x_m + \mathbb{E}(t_m) \right) \cancel{\mathbb{P}(X \in B_m)}$$

$$= \mathbb{E}(T) + \frac{1}{M} \sum_{m=1}^{M} x_m$$

which shows that $\mathbb{E}(X) + 1 \geq \mathbb{E}(T)$ when no B_m is empty.

Exercise: consider the general case where some $B_m = \phi$, and argue that the method may be worse than direct search.

<u>Examples of Inverse Function Method:</u>

Exponential: $f(x) = \lambda e^{-\lambda x}$

Set $u = F(x) = 1 - e^{-\lambda x}$, so that

$$e^{-\lambda x} = 1 - u$$

$$-\lambda x = \log(1-u)$$

$$\Rightarrow F^{-1}(u) = -\frac{1}{\lambda} \overset{\ln}{\cancel{\log}}(1-u)$$

Weibull (α, β): $x > 0$

$$\cancel{f(x) = \frac{\alpha}{(x+1)^{\alpha+1}}}$$

$$f(x) = \alpha\beta e^{-x^{\alpha}/\beta}, \quad F(x) = 1 - x^{\alpha/\beta}$$

$$F^{-1}(u) = [-\beta \ln(1-u)]^{1/\alpha}.$$

Geometric distribution: $\mathbb{P}(X=k) = p(1-p)^k$, so that $\mathbb{P}(X \le k) = 1-(1-p)^{k+1}$, and the inverse function can be evaluated analytically:

$$F^{-1}(u) = \left\lfloor \frac{\ln(1-u)}{\ln(1-p)} \right\rfloor$$

~~(exercise left for students)~~. See Ross.

General discrete distribution via inverse function method: use a linear search method to find the inverse function, $\mathbb{P}(X=k) = p(k)$.

$$F^{-1}(U) = \min(k : F(k) \geq U)$$

```
F = p(0) ; N = 0;
U = RANDOM(seed);
While (F < U)
      N = N+1;
      F = F+p(N);
end (while)
RETURN (N);
```

<u>Exercise</u>: write your program to generate a Poisson r.v. with $\mathbb{P}(X=n) = e^{-\lambda}\frac{\lambda^n}{n!}$, $n \geq 0$.

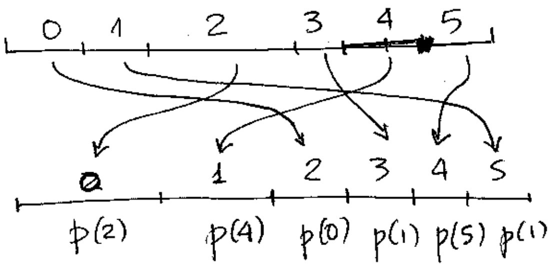
<u>Excercise</u>: how would you change your code if the random variable is defined with values $\{1,3,4\}$ only?

The number of iterations for the Poisson r.v. is ~~λ+~~ $\mathbb{E}(X)+1$.

Ross p. 55: use binary search instead of linear search to accelerate the procedure, because now the expected number of iterations is reduced to $1 + 0.798\sqrt{\lambda}$, which is much better for large λ.

<u>~~Aliasing~~ Reordering method</u>: Instead of performing a linear search, the idea is to re-order the probabilities $\{\mathbb{P}(X=k) = p(k)\}$ in decreasing order to perform the search. Accordingly, one must keep the label of the original value.

Example:



X rv that we want to generate

Y: rv that contains the labels

~~$Y(4) = 0,\ Y(2) = 0$, etc.~~

As well, ~~$Y = 0 \Leftrightarrow X = 2$~~

~~Aliasing~~

$X(0) = 2,\ X(1) = 4,\ X(2) = 0$, etc.

Generate Y and then set $X(Y)$ as the required random variable. Because the probabilities are re-ordered, it follows that $\mathbb{E}(Y) \leq \mathbb{E}(X)$.

Proof: If $X \in \{0, 1, \ldots\}$ then $\mathbb{E}X = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$

$$\mathbb{E}X = \sum_{k=1}^{\infty} \left(\sum_{m=k+1}^{\infty} p(m) \right) \cancel{= \sum_{k=1}^{\infty} \left(1 - \sum_{m=0}^{k} p(m)\right)}$$

$$\mathbb{E}Y = \cancel{\sum_{k=1}^{\infty} \left(1 - \sum_{m=0}^{k} p'(m)\right)} = \sum_{k=1}^{\infty} \sum_{m=k+1}^{\infty} p'(m)$$

We know that $p'(j) \geq p(j)\ \forall j$

(...) to finish

American Statistician, 1979 Alias Method proof (get paper)

Aliasing method

A general method for generating a rv is to use a rv that is "easier" or "faster" to generate, and define a function to recover the original values.

Set-up: X discrete rv, $\mathbb{P}(X=k) = p_k$, $k \in \{0, \ldots, n\}$

Let $R(k)$ be a vector, $R(k) \in [0,1]$ for each k,

Let $A : \{0, \ldots, n\} \to \{0, \ldots n\}$ be a permutation function $A(j) = k$ is the "alias".

- Generate $Y \sim U\{0, \ldots, n\}$ and $U \sim U(0,1)$
- If $U \leq R(Y) \Rightarrow X = Y$
  else $X = A(Y)$

what is the outcome of this algorithm?

$$\mathbb{P}(X=k) = \frac{1}{n+1} R(k) + \frac{1}{n+1} \sum_{\{j : A(j) = k\}} (1 - R(k))$$

Thus, it suffices to define appropriate (not unique) vectors R and A such that $\mathbb{P}(X=k) = p_k$.

There are ways to create "optimal" tables for speed up of the algorithm (1979, Kronmal & Peterson Jr)

Example: Poisson rv:

(a) Direct inverse function

(b) Let $\{X_1, \ldots X_n, \ldots\}$ iid $\exp(\lambda)$, then $N = \min(n : \sum_{i=1}^{n} X_i \leq 1) \sim$ Poisson
p.20-21 simulación

(c) Binary search (Ross UR p.56)

(b) Acceptance / Rejection Method

Motivation $X \sim N(\mu, \sigma^2)$

$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-(y-\mu)^2/\sigma^2} dy$$ no analytical inverse...

$X \sim \beta(\alpha, \beta)$ (Beta distribution)

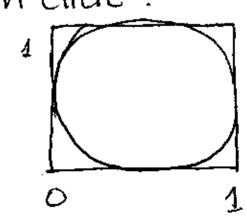
$$F(x) = \int_0^x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} dy$$

and many others may fail to have an analytical inverse.

A/R method is an extension of the Alias method that is based on the assumption that a random distribution G is "easy" to generate. Let $Y \sim G$, and now try to find a way to generate the required distribution F using a particular function of Y.

Example: Uniform in circle?

Use $X \sim U(0,1)$
$Y \sim U(0,1)$



as coordinates and only take those points that lie inside the circle.

Does this work? Why?

Thm: Let $\{(Y_n, U_n)\}_{n \geq 1}$ be independent, $Y_n \perp\!\!\!\perp U_n$, where $U_n \sim (0,1)$ and $Y_n \sim G$. (density g is assumed continuous and bounded). Assume that:

$$c = \sup_{x \in \mathbb{R}} \frac{f(x)}{g(x)} < \infty$$

for the density f, and let

$$R(y) = \frac{f(y)}{c\,g(y)} \leq 1.$$

Define a random stopping time w.r.t. $\mathcal{F}_n = \sigma((U_i, Y_i)\, i \leq n)$ by:

$$\tau = \min(n : U_n \leq R(Y_n)).$$

Then $X = Y_\tau$ has a probability density f.

REMARK: The pseudo-code to express the definition of X in the theorem is:

1 Generate $Y \sim g$, Generate $U \sim U(0,1)$
2 If $U \leq R(Y)$ then $X = Y$
Else GOTO 1.

Proof: By construction,

$$\mathbb{P}(X \leq x) = \mathbb{P}(Y_\tau \leq x) = \mathbb{P}(Y_n \leq x \mid U_n \leq R(Y_n))$$

$$= \frac{\mathbb{P}(Y \leq x, U \leq R(Y))}{\mathbb{P}(U \leq R(Y))} \quad (**)$$

The denominator can be calculated by conditioning:

$$\mathbb{P}(U \le R(Y)) = \int_{\mathbb{R}} \mathbb{P}(U \le R(y))\, g(y)\, dy = \frac{1}{c}\int_{\mathbb{R}} \frac{f(y)}{g(y)} g(y)\, dy$$

$$= \frac{1}{c} \quad (\text{because } U \sim U(0,1))$$

We do similarly for the numerator; condition on $\{Y=y\}$

$$\mathbb{P}(Y \le x, U \le R(Y)) = \int_{\mathbb{R}} \mathbb{P}(y \le x, U \le R(y) \mid Y=y)\, g(y)\, dy$$

$$= \int_{-\infty}^{x} R(y)\, g(y)\, dy = \frac{1}{c}\int_{-\infty}^{x} f(y)\, dy$$

$$= \frac{1}{c}\mathbb{P}(X \le x),$$

which, using $(**)$, yields the desired result. QED.

See examples from Ross, and "~~meta paper~~ Gaussian" EJOR 2008

<u>REMARK</u>: The ratio f/g of two densities is called in statistics the "likelihood ratio". When interpreted as a random variable ($\sigma(Y)$-mbl), the ratio is a Radon-Nykodim derivative of f w.r.t. g. RN derivatives are used to change the probability measures in analogous way that one uses change of variables in ordinary calculus. In simulation, change of measure techniques are very important and we will be seeing some of this lateron.

<u>Thm</u>: The expected number of iterations $\mathbb{E}(\tau)$ in the A/R method, is $\mathbb{E}(\tau) = c$. Furthermore, τ has a geometric distribution.

<u>Proof</u>: Let $I_n = \mathbb{1}(U_n \le R(Y_n))$. Because $\{(U_n, Y_n)\}$ are independent, then $\{I_n\}$ form a sequence of indep. Bernoulli trials with parameter $\mathbb{P}(I_n = 1) = \mathbb{P}(U_n \le R(Y_n)) = 1/c$ (as calculated in proof of Thm). Because τ represents the first "success" in a Bernoulli trial, it follows that $\tau \sim \text{Geom}(1/c)$. QED.

The method is also applicable to discrete rv's.

<u>Thm</u>: Let X be a discrete rv with distribution

$$\mathbb{P}(X=k) = p_k.$$

Let $\{q_j\}$ be another discrete dist $\sum_{k\in\mathbb{N}} q_k = 1$, and suppose that

$$c = \max_{j\in\mathbb{N}} \frac{p_j}{q_j} < \infty, \quad R(y) = \frac{p_y}{c\, q_y}$$

Let $\{U_n\} \sim$ iid $U(0,1)$, $\perp\!\!\!\perp$ $\{Y_n\}$ iid $\sim Q = \{q_j\}$ and define $\tau = \min(n : U_n \le R(Y_n))$. Then $X \overset{d}{=} Y_\tau \sim \{p_j\}$ and $\tau \sim \text{Geom}(1/c)$.

~~Buckets for speed up.~~

Ex: Consider the hyperexponential distribution:

$$f(x) = \sum_{i=1}^{n} p_i \lambda_i e^{-\lambda_i x}, \text{ where } p_i \in [0,1], \sum_i p_i = 1$$

$$F(x) = \int_0^x f(y)\,dy = \sum p_i (1 - e^{-\lambda_i x}).$$

This distribution can be interpreted as a composition of n random variables, each with an exponential distribution:

Let

$\Rightarrow$

Let $J \in \{1, \dots, n\}$, $\mathbb{P}(J=i) = p_i$, and $X = X_J$

where $\{X_i\} \sim \exp(\lambda_i)$.

Then $$\mathbb{P}(X \le x) = \sum_{k=1}^{n} \mathbb{P}(X_i \le x \mid J=i)\, p_i$$

$$= \sum_{k=1}^{n} p_i (1 - e^{-\lambda_i x}).$$

$$\left.\begin{array}{l} J = 1 \\ U \sim U(0,1) \\ \text{while } (U > P(J)) \\ \quad J = J+1; \ \text{\% Find } J \end{array}\right\} \text{Generates } J \sim \{P_k\}.$$

$U \sim U(0,1)$

$$X = -\frac{1}{\lambda_J} \ln(1-U)$$

Convolution:

Ex: $\{Y_1, Y_2, \dots\}$ iid $\exp(\lambda) \Rightarrow X = \sum_{i=1}^{n} Y_i \sim \Gamma(\lambda, n)$

combine with inverse:

$$X \stackrel{d}{=} -\frac{1}{\lambda} \sum_{i=1}^{n} \ln(1-U_i) = -\frac{1}{\lambda} \ln\left(\prod_{i=1}^{n} U_i\right)$$

so just one time applying logarithms (can speed up).

Transformations

Examples:
- Box-Muller (see Ross UR)
- $$\frac{\Gamma(\alpha,\lambda)}{\Gamma(\alpha,\lambda)+\Gamma(\beta,\lambda)} \sim \text{Beta}(\alpha,\beta)$$

use $\lambda = 1$ for example.

If time permits, show Von Neumann's method for generating exponentials.

Von Neumann's method to generate an exponential dist.

Notice that A/R method is based on the following:

$Y \sim G$ is a rv that can be generated

$X \sim F$ is the desired rv to be generated

Let A be an event such that:

$$\mathbb{P}(X \le x) = \mathbb{P}(Y \le x \mid A),$$

where $x \in \mathbb{R}$. Then A/R method is based on the joint simulation of Y and A (the acceptance event).

The method for generating $X \sim \exp(1)$ without the use of logarithms or exponentials, is the following:

STEP 1: Generate iid $U(0,1)$ sequence $U_1, U_2, \dots$ and stop at:

$$N = \min(n : U_n < U_{n-1}).$$

STEP 2: If N is even then go to STEP 3. Otherwise reject the run and go to STEP 1.

STEP 3: X = # failed runs plus U_1 (the first uniform in the succesful run).

Theorem: The resulting random variable X has distribution $F(x) = 1 - e^{-x}$.

Let $\{N_i\}$ be the consecutive numbers and $\tau = \min(n : N_n \text{ is even})$ we saw that $\tau \sim Geom(1-e^{-1})$. Ross p.688 shows that $\mathbb{E}\sum_{i=1}^{\tau} N_i = \frac{e}{1-e^{-1}} \approx 4.3$

Proof: (Ross pp. 686-688 gives a different one from mine).

Let $X \sim \exp(1)$.

We first use a composition method with subintervals $A_k = [k, k+1)$, $k = 0, 1, 2, \dots$

Then by conditioning:

$$F(x) = \sum_{k \ge 0} \alpha_k F_k(x) = \sum_{k \ge 0} \alpha_k F(x \mid A_k),$$

where

$$\alpha_k = \mathbb{P}(X \in A_k) = \int_k^{k+1} e^{-x} dx = e^{-k}(1 - e^{-1}) \quad (1)$$

so that α_k are the geometric weights for $p = (1 - e^{-1})$.

$$\mathbb{P}(X \le k + y \mid X \in A_k) = \int_0^y \frac{e^{-k} e^{-s} ds}{\alpha_k} = \frac{1}{1-e^{-1}} \int_0^y e^{-s} ds$$

$$= \frac{1 - e^{-y}}{1 - e^{-1}}, \quad (2)$$

which is independent of k. therefore,

$$X \stackrel{d}{=} M + Y \quad (3)$$

$M \sim$ Geometric $(1 - e^{-1})$

Y has distribution $\mathbb{P}(Y \le y) = \frac{1 - e^{-y}}{1 - e^{-1}}$.

See Ross p. 686-687 for the following result:

$$\mathbb{P}(N \text{ is even}, U_1 \le y) = 1 - e^{-y}, \text{ where} \quad (4)$$

$$N = \min(n : U_n < U_{n-1}).$$

From (4) it follows that $\mathbb{P}(N \text{ is even}) = 1 - e^{-1}$, so M is the number of failed runs, and $\mathbb{P}(U_1 \le y \mid N \text{ is even}) = \frac{1 - e^{-y}}{1 - e^{-1}}$, which sets $Y = U_1$ in (3).