Random Number Generation

Harry Perros

Computer Science Department

NC State University

hp@csc.ncsu.edu

http://www.csc.ncsu.edu/faculty/perros//index.html

Generating random numbers

- In the hand simulation example we assumed that the operational times and repair times are all constant.
- However, in real life, they may follow an empirical distribution
- We need to have a technique for generating numbers from empirical and theoretical distributions in a random fashion.

Random numbers

- *Pseudo-random numbers*:
  - These are uniformly distributed random numbers in [0,1]. They are usually referred to as *random numbers*
- *Random variates* or *stochastic variates:*
  - These are random numbers that follow any distribution other than the above uniform distribution in [0,1].

Part 1: Pseudo-random numbers

- In a sense, there is no such a thing as a single random number.
- Rather, we speak of a sequence of random numbers which pass a battery of statistical tests.
- Pseudo random numbers are generated using different methods, such as:
  - *Linear congruential generator*
  - *Tausworthe generators*
  - *The Mercenne Twister*

Linear congruential methods

- A very popular technique. It uses the relationship:

$$x_{i+1} = ax_i + c \pmod{m}$$

The starting value x_0 is known as the *seed*.

- *Example:* if $x_0 = 0$, a = c = 7, and m = 10 then we can obtain the following sequence of numbers: 7,6,9,0,7,6,9,0,...

The period of the generator

- The numbers generated are in [0, m-1]. Numbers in [0,1] are then obtained by dividing by m.
- The number of successively generated numbers after which the sequence starts repeating itself is called the *period*.
- A generator has a full period if its period is equal to m.
- Generators should have very large periods, so that the numbers are never repeated in a simulation.

Selection of values

- Values for a, c and m should not be chosen arbitrarily.
- Specific conditions from num have to apply.
- Systematic testing have led to generators which have a full period and which are statistically satisfactory. A set of such values is: a = 314, 159, 269, c = 453, 806, 245, and $m = 2^{31}$ (for a 32 bit machine)

Which random number generator?

- Use the one from the high-level language you use to write the simulation, or use one from MATLAB, Mathematica, and other packages.
- Make sure you read the instructions as how to use it!! Particularly, as to what kind of seed it expects.
- Test its statistical properties first!!
- Simulation languages have their own built-in generators which you can trust !

Statistical tests

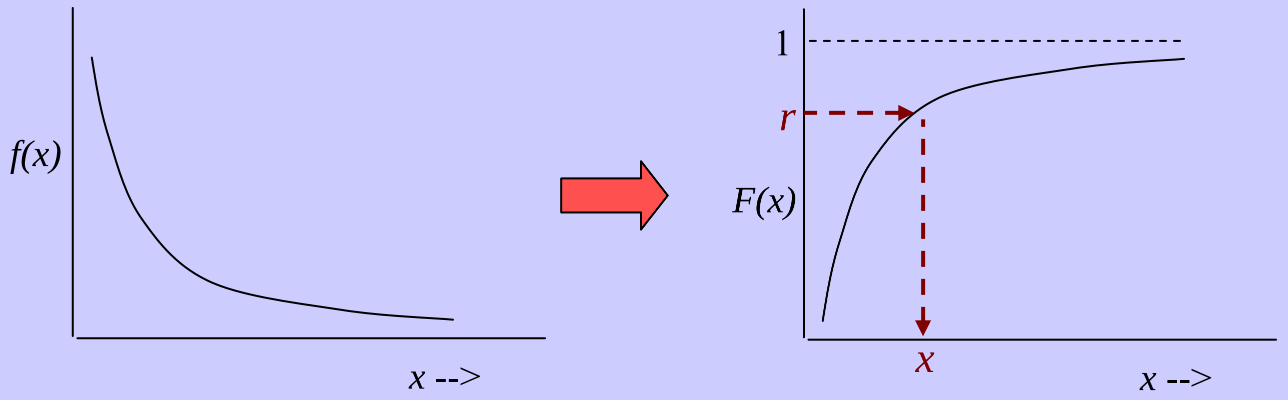
- A number of statistical tests are available to test how random is the deterministically created sequence of numbers !
- Runs test: It tests whether the successive numbers are independent from each other
- Chi-square test for goodness of fit : It tests whether the sequence of numbers is uniformaly distributed in [0,1].

Part 2: Random variates

- Depending upon the situation, we may need random variates from a given histogram (calculated using real-life observations) or from a theoretical distribution.
- Random variates are generated using the pseudo-random numbers in a variety of ways.

Sampling from a continuous-time probability distribution

- The *inverse transformation method* is is a popular technique, and it requires the ability to invert a cumulative function.



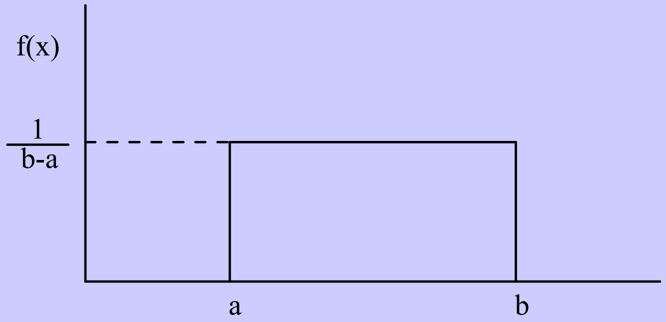
The inverse transformation method

- Let *f(.)* be the pdf of a random variable *X* and *F(x)* its cumulative distribution function (CDF):

$$F(x) = \int_{o}^{x} f(x)dx$$

- Generate a pseudo-random number r.
- Set $r = F(x)$, and solve for x, i.e. $x = F^{-1}(r)$

Sampling from the Uniform distribution



- The CDF is: $F(x) = (x-a)/(b-a)$.
  - Generate a pseudo-random number r.
  - $r = (x-a)/(b-a)$
  - Solving for x we have: $x = a + (b-a)*r$.
- (We can get the same result using a simple linear transformation)

Sampling from the exponential distribution

- Prob. density function: $f(x) = ae^{-ax}$, $a > 0$, $x \geq 0$.
- Cumulative density function: $F(x) = 1 - e^{-ax}$.
- Generate a pseudo-random number r. Then

$$r = F(x) = 1 - e^{-ax}$$

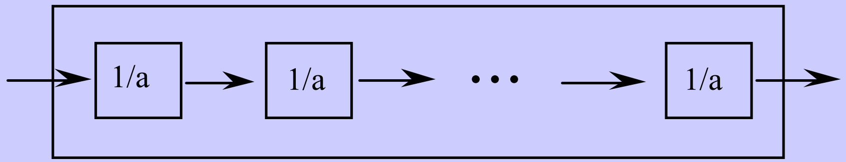
or

$$1 - r = e^{-ax}$$

or

$$x = - \log_e(1-r) = - (1/a)\log_e(l-r).$$

Sampling from the Erlang distribution



- The Erlang distribution is a convolution of *k* exponential distributions, all with the same mean.
- Generate *k* variates from the same distribution, i.e. x_1, x_2, ..., x_k
- Then, Erlang variate is the sum $x_1 + x_2 + \ldots + x_k$

Sampling from a discrete-time distribution

- The inverse transformation method can also be used to generate variates from a discrete-time distribution.
- Other techniques which explore the nature of the underlying process that gives rise to a discrete-time distribution have also been used.

Sampling from a geometric distribution

- The pdf of the geometric distribution is:

$$p(n) = pq^n, \quad n = 0,1,2, ...$$

  where p = prob. of success, and q = 1-p.
- The cumulative probability is:

$$F(n) = \sum_{s=0}^{n} pq^s, n = 0,1,2,...$$

- After some calculations, we have:

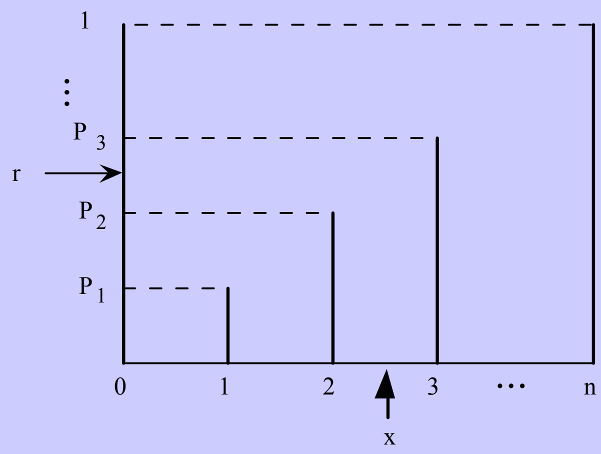
$$F(n) = p\frac{1-q^{n+1}}{1-q} = 1-q^{n+1}, n = 0,1,2,\ldots$$

- Procedure:
  - Generate a pseudo-random number *r*.
  - Set *r=F(n)*, and solve for *n*. We have:

$$n = \frac{\log r}{\log q} - 1$$

Sampling from an empirical discrete-time probability distribution

- X is a discrete random variable with p(X = i) = p_i. Let p($X \leq i$) = P_i be the cumulative probability.



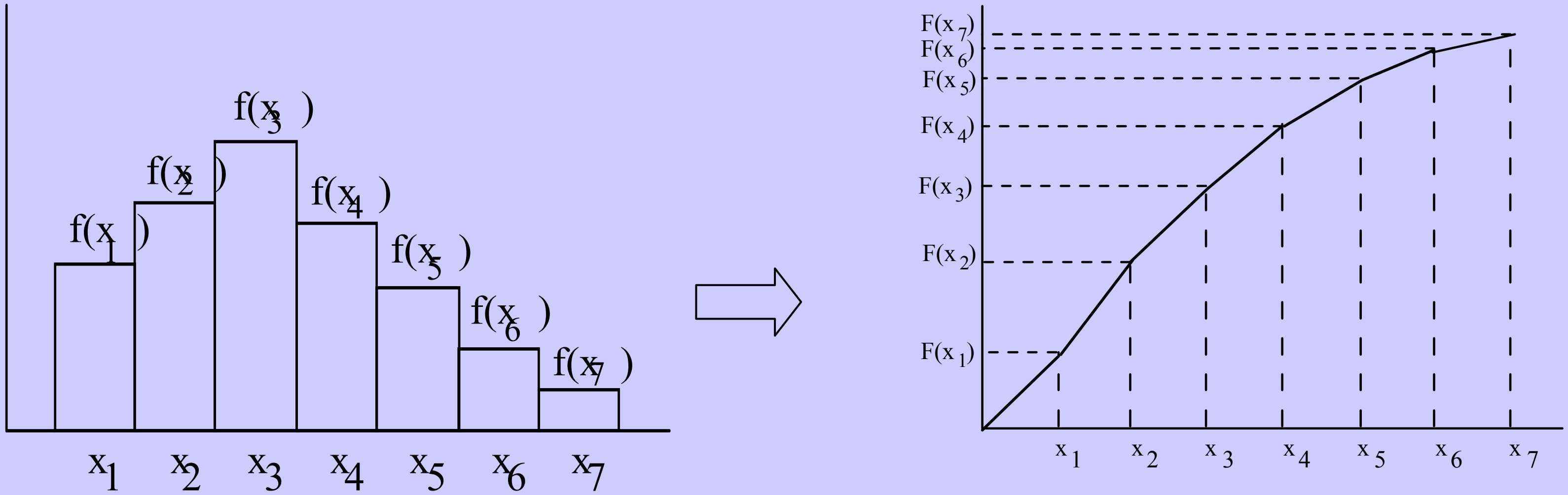
Example: The newsboy problem

X	1	2	3	4	5
f(x)	0.20	0.20	0.30	0.15	0.15

X	1	2	3	4	5
F(x)	0.20	0.40	0.70	0.85	1.00

- Generate a random number r. Then:
  - If $0.85 < r \leq 1.00$ then $x = 5$
  - If $0.70 < r \leq 0.85$ then $x = 4$
  - If $0.40 < r \leq 0.70$ then $x = 3$
  - If $0.20 < r \leq 0.40$ then $x = 2$
  - Otherwise $x = 1$

Sampling from an empirical continuous-time probability distribution



- Same as before, after the cumulative probability has been constructed