Construction Operation Simulation

Lecture #4

Stochastic simulation and random number generation

Amin Alvanchi, PhD

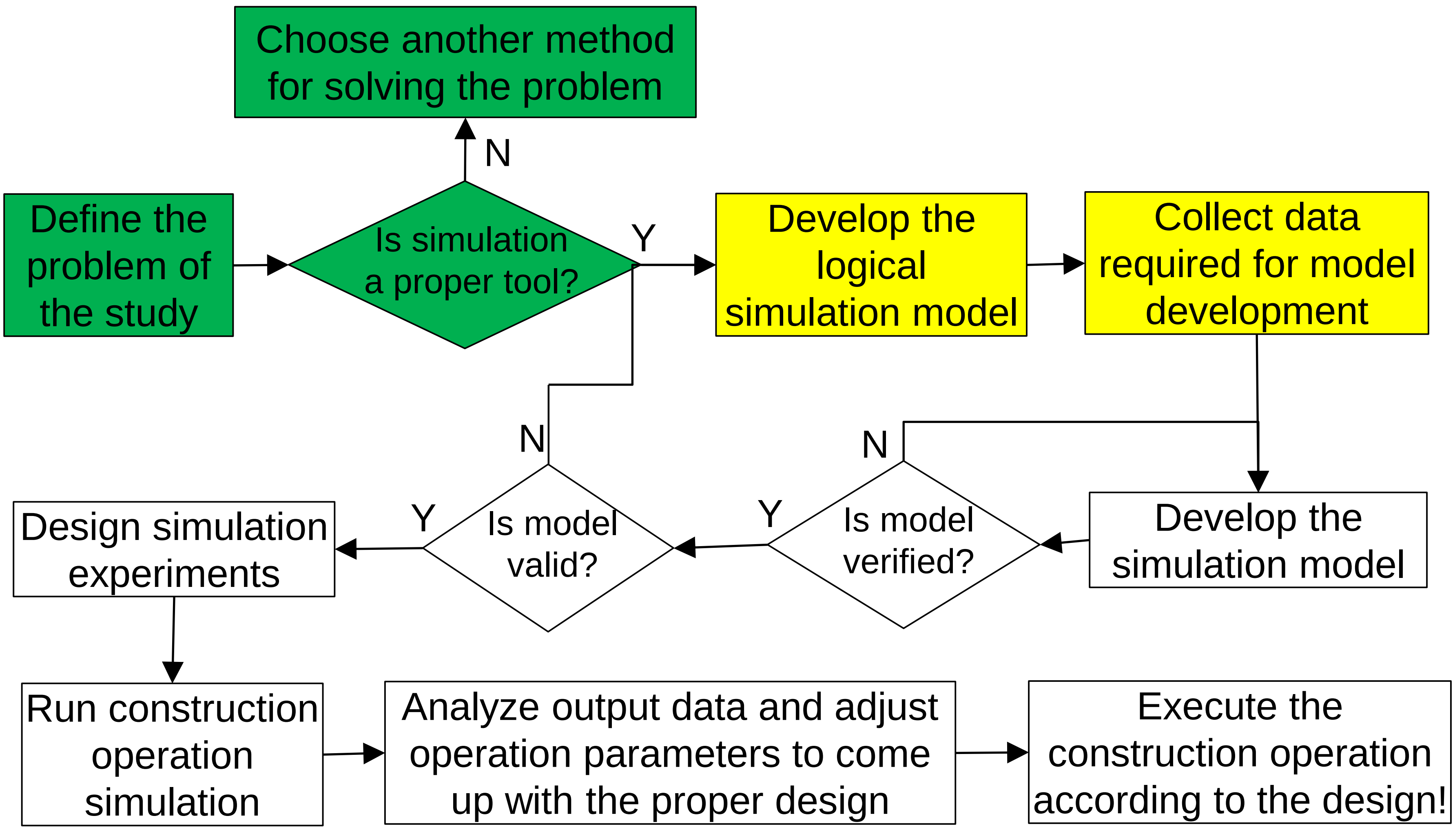
Construction Engineering and Management

LinkedIn Instagram WebPage

Outline

- Introduction
- Statistical models in simulation
- Random variable generation

Main steps in simulation studies



Statistical models in simulation

- In many construction operation simulation studies simulation model element specifications (e.g., activity durations, capacity of resources and entity attributes) are described in stochastic fashions; so called *stochastic variables* or *random variables*.
- We usually capture behaviour of stochastic variables in statistical distributions!
- We are going to have a quick review on different types of stochastic variables and statistical distributions.

Statistical models in simulation

1. Discrete versus Continuous random variables:

- *Discrete random variable*: If the number of possible values of a random variable (X) is finite, or countably infinite, X is called a discrete random variable. (Banks et al. 2004, p132)

  Example: We are receiving asphalt for our project from an asphalt plant whihc has 10 asphalt trucks with capacity of 8 tonnes, 30 trucks with capacity of 10 tonnes and 10 with the capacity of 12 tonnes. Plant concurrently serves several projects and loads and sends asphalt trucks to the projects based on their arrival time to the plant. Capacity of asphalt trucks arrive to our project site is a discrete variable with the following capacity:

  P(Truck capacity: 8 tonnes)= 10 / 50 = 0.2

  P(Truck capacity: 10 tonnes)= 30 / 50 = 0.6

  P(Truck capacity: 12 tonnes)= 10 / 50 = 0.2

Statistical models in simulation

1. Discrete versus Continuous random variables:

- *Continuous random variable*: If the range (or definition) space (Rx) of a random variable (X) is an interval or a collection of intervals, X is called a continuous random variable. (Banks et al. 2004, p132)

Unlike discrete variables which have probability values of P(x) assigned to them at every variable point defined for the variable, probability value of every single point of a random continuous variable within the range space of the variable is 0. Though, probability density of f(x) is defined for random continuous variables.

- **Example**: In our earthmoving example the fastest time a truck can finish its dumping trip (from leaving the loading site until in returns for another loading) is 10 minutes and maximum time that it might take is 15 minutes.

f(Truck trip duration=x)~ [10 minute, 15 minute]= 1/(15-10)= 1/5=0.2; for x between 10 &15

f(x)=0; otherwise

Statistical models in simulation

2. Cumulative distribution:

- The cumulative distribution function (CDF), denoted by F(x), measures the probability that the random variable X assumes a value less than or equal to x, that is, $F(x) = P(X \leq x)$. (Banks et al. 2004, p132)
- CDF value is between 0 and 1.

Example 1: The cumulative distribution function for the asphalt truck capacity in our last paving project example is:

$$F(x)=\begin{cases} 0; \text{ if } x<8 \\ 0.2; \text{ if } x>=8 \text{ and } <10 \\ 0.8; \text{ if } x>=10 \text{ and } <12 \\ 1; \text{ if } x\geq 12 \end{cases}$$

Example 2: In our last earthmoving example, cumulative distribution function for trip duration is:

$$F(x)=\begin{cases} 0; \text{ if } x<10 \\ (x-10)/(15-10); \ x\in[10,15] \\ 1; \text{ if } x>15 \end{cases}$$

Statistical models in simulation

3. Several useful discrete distributions:

- *Bernoulli trial and the Bernoulli distribution* (Banks et al. 2004, p141)
  - Consider an activity (or trial), we call this activity a Bernoulli trial if it has a chance of success (x=1) of p and a chance of failure (x=0) of 1-p = q on each time of trial.
  - Bernoulli trial has a Bernoulli distribution as follows:

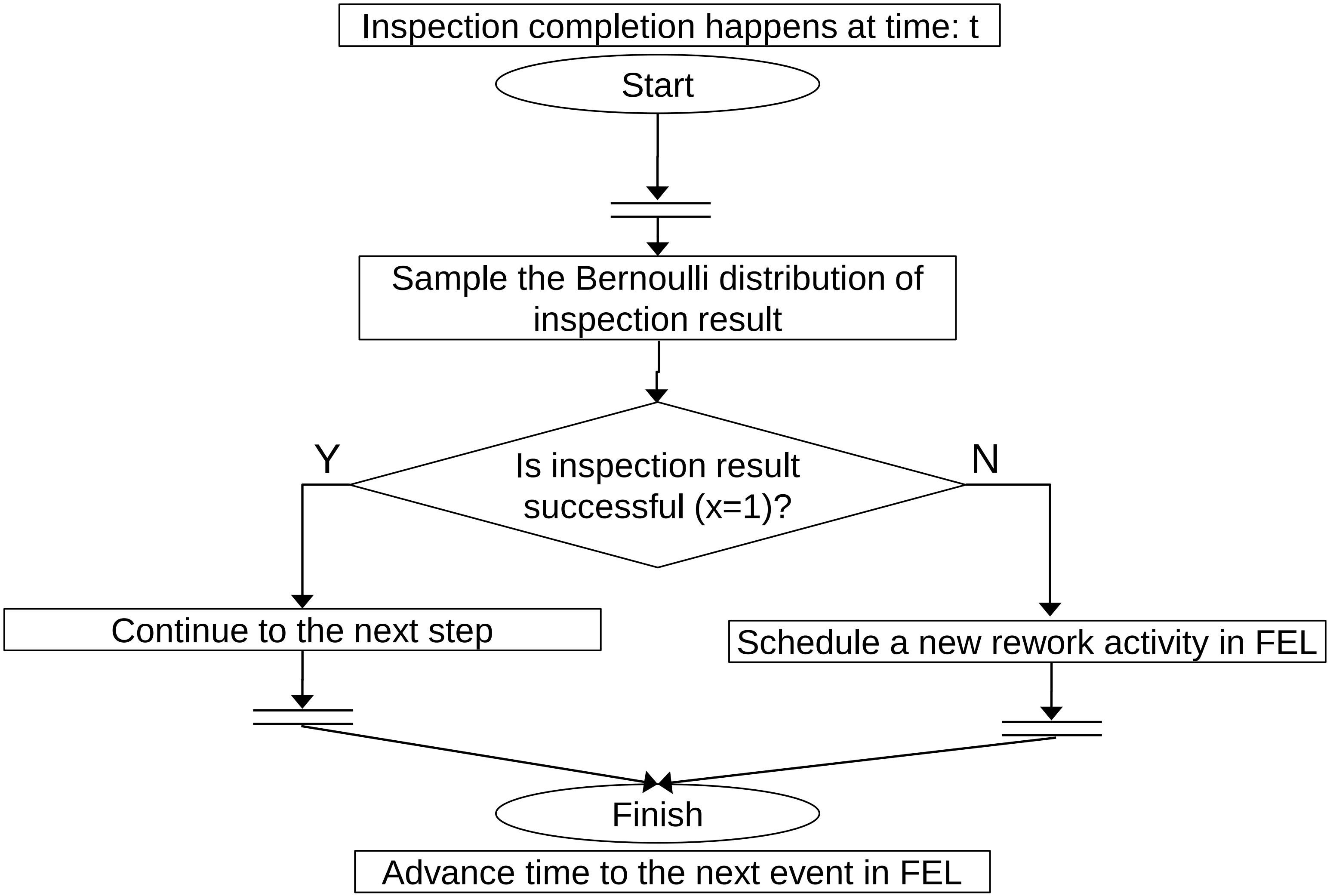
$$P(x)=\begin{cases} p; \text{ if } x=1 \\ 1-p=q; \text{ if } x=0 \\ 0; \text{ otherwise} \end{cases}$$

$$V(X) = p(1-p)$$

$$E(X) = p$$

  - **Example:** Inspection results in construction operation usually follow Bernoulli trial:

Statistical models in simulation



Statistical models in simulation

3. Several useful discrete distributions :

- *Binomial distribution* (Banks et al. 2004, p142)
  - The random variable X that denotes the **number of successes** in n Bernoulli trials has a Binomial distribution given by p(x), where:

$$p(x)=\begin{cases}\binom{n}{x}p^x q^{n-x}, & x=0,1,2,\ldots,n\\ 0, & \text{otherwise}\end{cases} \qquad \binom{n}{x}=\frac{n!}{x!(n-x)!} \qquad \begin{aligned}E(X)&=np\\ V(X)&=npq\end{aligned}$$

  - **Example:** Client accepts quality of fabricated structural steels, if third party inspection of 20 samples of welded points does not result in more than 2 reworks. In case of more than 2 reworks caught third party inspector will run a 100% inspection (e.g., on 200 welding pints) and the extra cost is upon the contractor. If there is a chance of 1% error (failure) on each welded point, what is the chance that contractor faces 100% inspection in each third party inspection activity? (x= number of success)

    Chance of 100% inspection = P(x<18)= 1-[P(x=20)+P(x=19)+P(x=18)]

    P(x=20)=20!/20! * 0.99^20 * 0.01^0 = 0.818; P(x=19)= 0.165; P(X=18)=0.016

    Chance of 100% inspection = 1- 0.818 - 0.165 – 0.016 = 0.001 = 0.1%

In class practice 1

- A tile installation contractor has a contract for procurement and installation of premium grade tiles. In the contract, it is determined that after installation of every 100 tiles client's inspector inspects installed tiles quality. If there are less than 2 lower quality tiles found, client will approve the installation and accept the installation. Otherwise, all lower quality tiles are removed and replaced with premium grade tiles. If there is chance of 0.5% that tiles arrive from tile factory have a different quality than premium, what is the chance of tile installation rework?

$$p(x)=\begin{cases}\binom{n}{x}p^{x}q^{n-x}, & x=0,1,2,\ldots,n\\ 0, & \text{otherwise}\end{cases}$$

$$\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

Statistical models in simulation

3. Several useful discrete distributions:

- *Poisson distribution* (Banks et al. 2004, p144)
  - Poisson probability mass function (pmf) represents probability of x number of random events happening within determined period of time. :

$$p(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & x = 0, 1, \ldots \\ 0, & \text{otherwise} \end{cases}$$

$$\lambda = E(x) = Var(x)$$

  - An interesting property of Poisson distribution is that λ (the only parameter of the distribution) represents the average and variance of the distribution.

  - **Example:** Suppose that you are going to simulate maintenance operation of a construction company in which the maintenance department sends its crew to different construction sites on weekly basis to collect broken tools to be fixed on the central maintenance shop (i.e., new jobs for the maintenance department). Number of broken tools collected from each construction site varies in each week. It is possible that there is no broken tool on a site. However, number of broken tools can go up to more than 10.

This can be a good example for modeling number of weekly broken tools with Poisson distribution function.

Statistical models in simulation

3. Several useful discrete distributions :

- *Poisson distribution* (Banks et al. 2004, p144)
  - **Example (cont'd):**

According to the past data from last 30 weeks, average number of weekly broken tools collected from a construction site is 2.3. Chance of collecting 3 broken tools for a week can be calculated as in below:

Distribution function for number of broken tools collected on weekly basis for the site will be:

$$P(x)=\begin{cases} e^{-2.3}\, 2.3^{x} / x!; & x=0,1,2,\ldots \\ 0; & \text{otherwise} \end{cases}$$

Chance of collecting 3 tools= $P(X=3) = e^{-2.3}\, 2.3^{3} / 3! = 0.203$

- **Poisson process:** Number of events within specific period of time follows a Poisson distribution if it pursues Poisson process properties:
  1. Events happen one at a time
  2. Distribution of number of events within an interval (e.g., rage of t to t+s) just depends on the interval length (i.e., s) not the starting point .
  3. Distribution of number of events during non-overlapping time intervals are independent random variables. Thus, a large or small number of arrivals in one time interval has no effect on the number of arrivals in subsequent time intervals.

Statistical models in simulation

4. Several useful continuous distributions :

- *Uniform distribution* (Banks et al. 2004, p146)

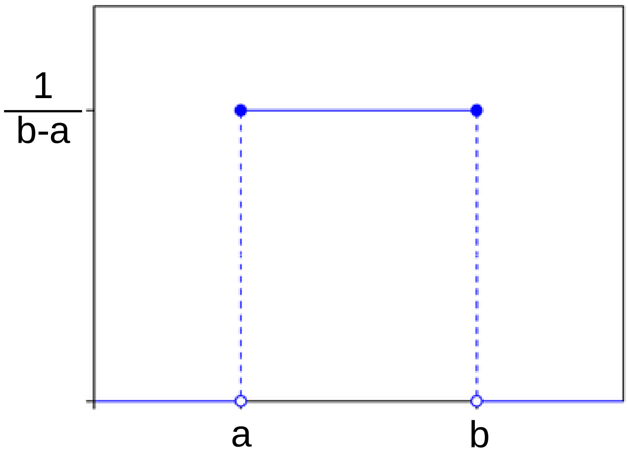
Uniform distribution is used for a random variable X uniformly distributed on the interval (a, b):

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

$$\begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a, b) \\ 1 & \text{for } x \ge b \end{cases}$$

mean (μ): $\frac{1}{2}(a+b)$

Variance (σ^2): $\frac{1}{12}(b-a)^2$



Use uniform distribution for estimating duration of activities when activity experts just can estimate a duration range (with a minimum of a and a maximum of b) with no priority determined within the interval.

Statistical models in simulation

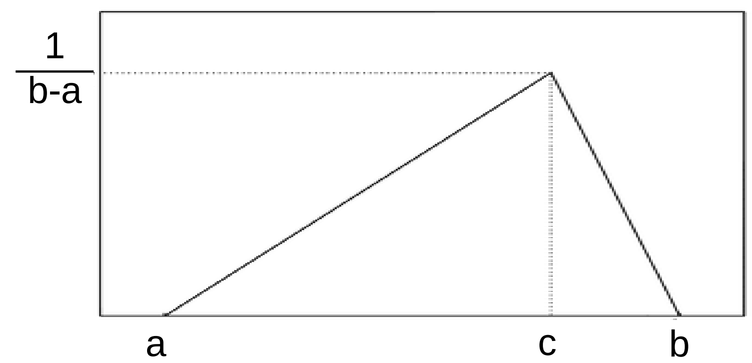
4. Several useful continuous distributions :

- *Triangular distribution* (Banks et al. 2004, p167)

Triangular distribution is used for a random variable X distributed on the interval (a, b) while it is more dense at the value of c (within the interval):

$$f(x|a,b,c) = \begin{cases} 0 & \text{for } x < a, \\ \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a \le x \le c, \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c < x \le b, \\ 0 & \text{for } b < x, \end{cases}$$

$$F(x|\ a,\ b,\ \ c) = \begin{cases} 0 & \text{for } x < a, \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{for } a \le x \le c, \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{for } c < x \le b, \\ 1 & \text{for } b < x. \end{cases}$$



Mean(μ): $\dfrac{a+b+c}{3}$

Variance(σ^2): $\dfrac{a^2+b^2+c^2-ab-ac-bc}{18}$

Use triangular distribution duration for estimating duration of activities when activity experts give a duration range (with a minimum of a and a maximum of b) for the activity duration and a most likely value within the range (distribution mode of c).

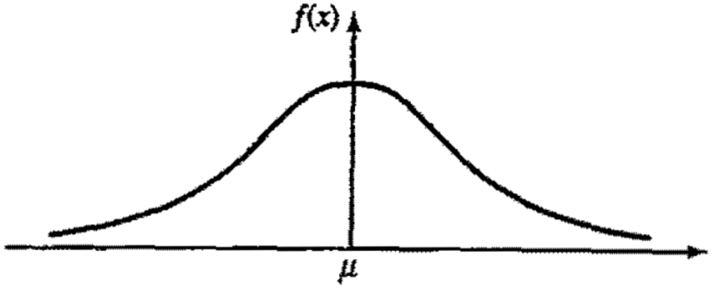
Statistical models in simulation

4. Several useful continuous distributions :

- *Normal (or Gaussian) distribution* (Banks et al. 2004, p152)

  A random variable X with mean $-\infty<\mu<+\infty$ and variance $\sigma^2>0$ has a normal distribution if it has the pdf:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],\quad -\infty<x<\infty$$



Physical quantities which are expected to be the sum of many independent processes/ elements (such as engineering tolerances and measurement errors) often have a distribution very close to the normal. In construction operations when an activity duration is *equipment driven*, usually we can estimate normal distribution for activity duration.

Normal distribution is also known as *natural distribution*. In the natural and social sciences normal distribution is used for real-valued random variables whose distributions are not known. (Casella and Berger 2001)

Statistical models in simulation

4. Several useful continuous distributions :

- *Exponential distribution* (Banks et al. 2004, p152)

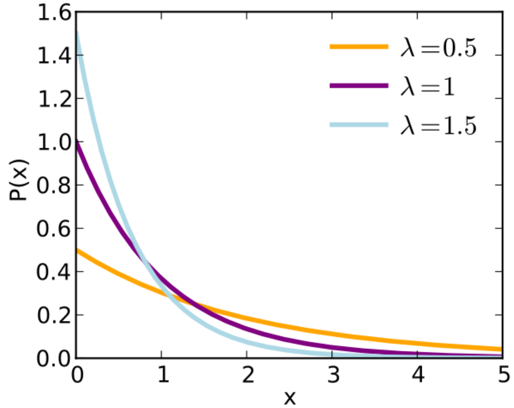
  Exponential distribution is the probability distribution that describes the duration time between events in a Poisson process. The distribution formula is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$E(X) = 1/\lambda$$

$$Var(X) = 1/\lambda^2$$

$$F(x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



In simulation models exponential distribution usually is used for modeling time between (entity) arrival events specially when these events are randomly distributed and externally originated. Examples for exponentially distributed time between events can be time between new projects introduced to the market or time between equipment breaks.

Question: What is the relation between λ as the parameter in Poisson distribution and in Exponential distribution?

Statistical models in simulation

4. Several useful continuous distributions :

- *Beta distribution* (Banks et al. 2004, p164)

A random variable y is beta-distributed with parameters α>0 and β> 0 if its pdf is given by:

$$f(y) = \begin{cases} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$E[Y] = \frac{\alpha}{\alpha+\beta}$$

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

$$V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

With special values of α and β Beta distribution behaves similar to other distributions:

- If X ~ Beta(α, 1) then −ln(X) ~ Exponential(α)

- Beta(1, 1) ~ Uniform(0, 1)

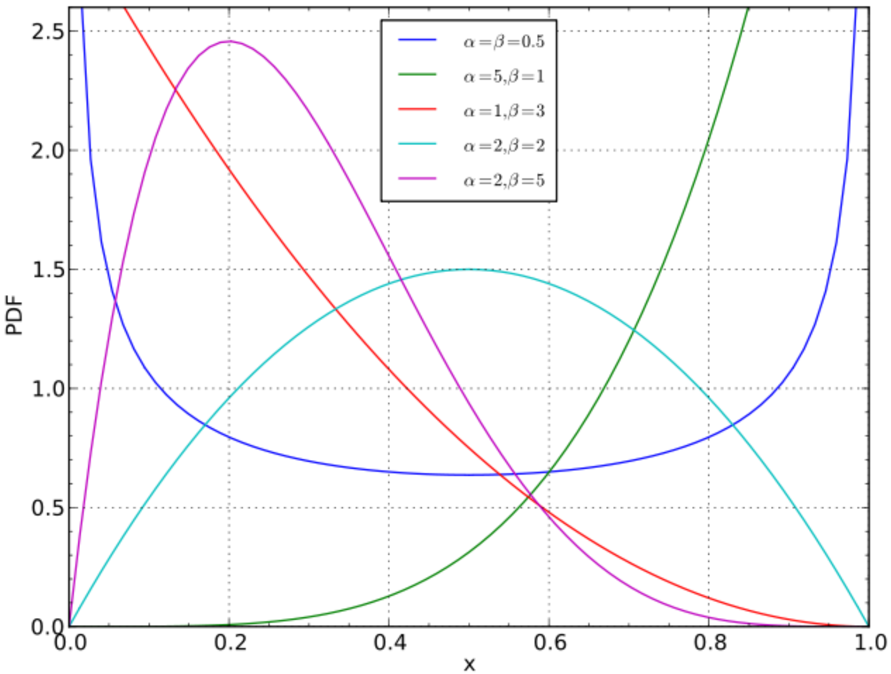
- For large number of α and β Beta distribution approximates Normal distribution

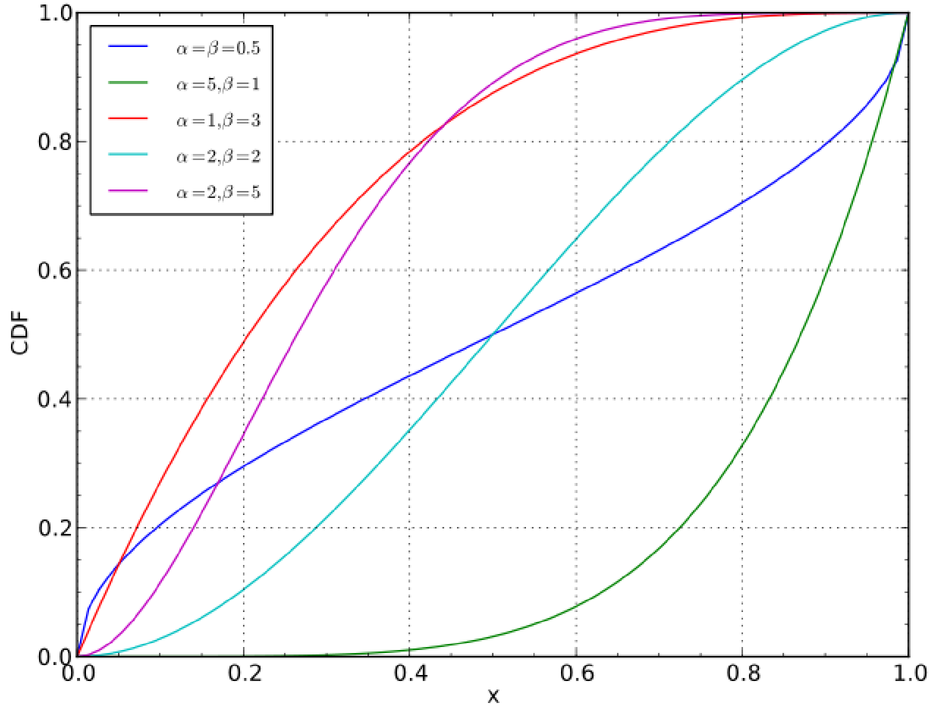
Beta distribution is a very flexible distribution in terms of fitting data with different trends:

Statistical models in simulation

4. Several useful continuous distributions :

- *Beta distribution* (cont'd)





Use Beta randomly distributed variable to fit historical data with strange shape/ behavior!

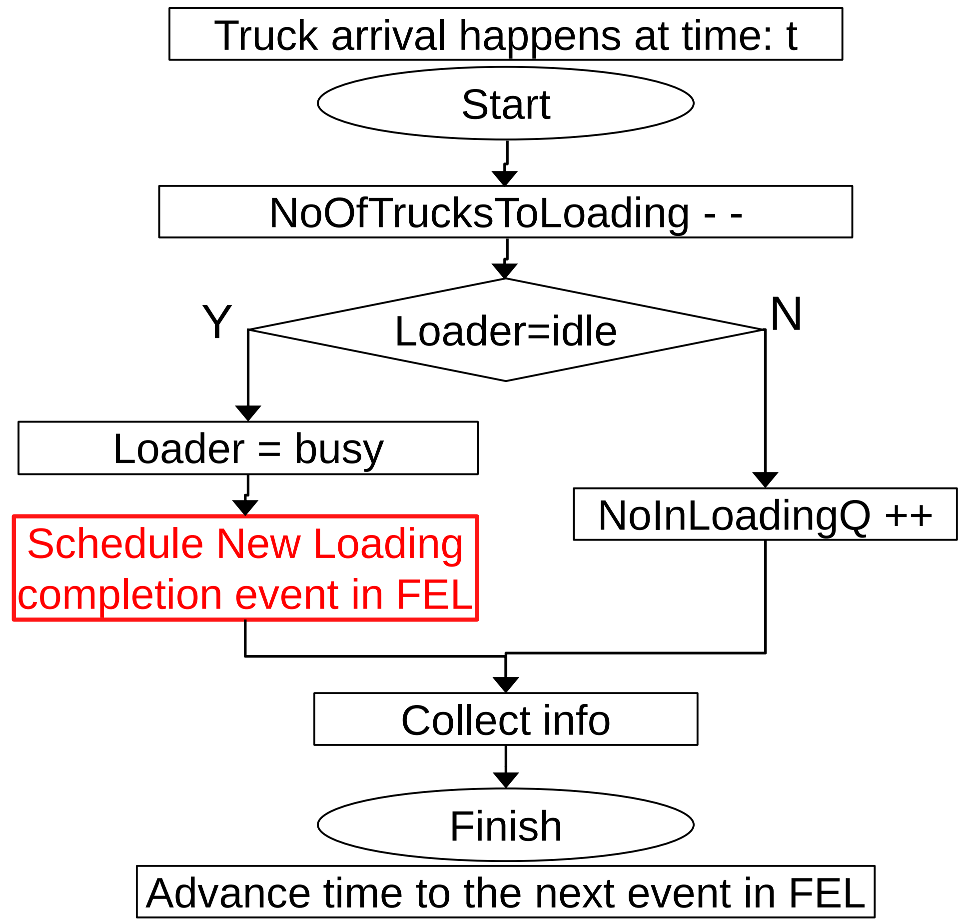
Random variable generation

- **Question:** Why do we need to generate random variables?

Random variable generation

Question: Why do we need to generate random variables?

- Remember the way that we update FEL using *event occurrence procedure*.
- Using activity duration and scheduling activity completion events in FEL is our main tool for adding new events into FEL and keep our simulation engine going!
- **Example:** Truck arrival procedure:



Random variable generation

- For deterministic activities with fixed durations scheduling new activity completion is quite straight forward: add the activity duration to the current time and schedule new activity completion in FEL. This is exactly what we did in our hand simulation!
- But for stochastic aspects of our system like stochastic activities, results and other system specification with randomly distributed distributions what are we going to do? How can we make sure that values we create in our simulation program follow specific distribution functions they are distributed in accordance with.
- Before start generating random *variables* distributed under different statistical distributions lets start with generation of the simplest and most prevalent random variable generation; generation of random numbers or uniform random variables between 0 and 1.

Random variable generation

- **Random *Number* Generation** (Banks et al. 2004, chapter 7)**:**
  - In general, random numbers generation is generation of random variables uniformly distributed between 0 and 1:

    $$x \sim U[0,1]$$

  - This is what you are going to receive when you refer to random number generating functions available in computer programming, calculator and widely used general applications like M.S. Excel.
  - Random numbers Vs Pseudo random numbers:
    - What these programs create are not real random numbers, but *Pseudo random numbers.* These *Pseudo random numbers* are generated through algorithms which try to generate numbers following random number behaviour. So when we talk about random number generation in simulation, we actually mean Pseudo random number generation!

    Question: How can we really generate real random numbers?

Random variable generation

- **Random *Number* Generation:**
  - Properties of random numbers:

Density function: $$\text{function } f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$$

Expected value: $$E[X] = \int_0^1 x dx = \frac{x^2}{2}\Big|_0^1 = 1/2$$

Variance: $$\mathrm{Var}[X] = E[X^2] - (E[X])^2 = 1/3 - 1/4 = 1/12$$

So any pseudo random number generating algorithm is expected to generate numbers with real random number properties!

There are some other properties of random variables which are tested for commercial and highly loaded pseudo random number generating programs. You are encouraged to study them (Banks et al. 2004, Section 7.4), but we avoid more discussion here!

Random variable generation

- **Random *Number* Generation:**
  - Linear congruential generation (LCG) technique *(Banks et al. 2004, Section 7.3.1)*:
    - This method initially proposed by Lehmer [1951].
    - It is one of the most efficient methods and is the base for many current methods for random number generation.

The generator produces a sequence of integers, X_1, X_2, ... between zero and m-1 by following a recursive relationship:

$$X_{i+1} = (aX_i + c) \bmod m$$

Where:

$i = 0, 1, 2, 3, \ldots$

$m > 0$ (modulus)

$0 < a < m$ (multiplier)

$0 \leq c < m$ (increment)

$0 < X_0 < m$ (seed or start value)

At ith recursion, pseudo random number Ri is created through: $R_i = X_i / m$

Random variable generation

- **Random *Number* Generation:**
  - Linear congruential generation (LCG) technique:

Example: With :X_0 = 27, a = 17, c = 43 and m = 100 following LCG method we have:

X_1 = (17 * **27** + 43) mod 100 = 2 => R_1 = 2 / 100 = 0.02

X_2 = (17* 2 + 43) mod 100 = 77 => R_2 = 77/ 100 = 0.77

X_3 = (17 * 77 + 43) mod 100 = 52 => R_3 = 52 / 100 = 0.52

X_4 = (17 * 52 + 43) mod 100 = **27** => R_4 = 27 / 100 = 0.27

From now on these numbers are repeated, no new number will be generated and therefore numbers generated will be dependent and not valid to be used as random variables!

In class practice 2

With $a = 13$, $c = 0$ and $m = 64$ follow LCG method for creating pseudo random numbers. First time with $X_0 = 4$ and second time with $X_0 = 2$.

Random variable generation

- **Random *Number* Generation:**
  - Linear congruential generation (LCG) technique:

    Some points regarding random number generation:
    - At the best case scenario period length of generated random number numbers is m.
    - In simulation model programs we usually need to generate many random numbers, so m should be big enough and X_0, a and c should be picked as reasonable values which can generate the length we require. There are many research efforts done to be able to find proper parameters which behave like random number and have long period!
    - Some guidelines for picking parameters are: Select m as a big number of power of 2 e.g., 2^64 (what will be the maximum period length?); X_0 to be prime to m; c and m must be relatively prime; have (a – 1) divisible by all prime factors of m; have (a – 1) a product of 4 if m is a product of 4.

Parameters in some famouse random number genetors based on LCG

Source	***m***	***(multiplier) a***	***(increment) c***
Numerical Recipes	2^32	1664525	1013904223
Borland C/C++	2^32	22695477	1
glibc (used by GCC)[5]	2^31	1103515245	12345
ANSI C: Watcom, Digital Mars, CodeWarrior, IBM VisualAge C/C++ [6]	2^31	1103515245	12345
Borland Delphi, Virtual Pascal	2^32	134775813	1
Microsoft Visual/Quick C/C++	2^32	214013 (343FD16)	2531011 (269EC316)
Microsoft Visual Basic (6 and earlier)[7]	2^24	1140671485 (43FD43FD16)	12820163 (C39EC316)
RtlUniform from Native API[8]	2^31-1	2147483629 (7FFFFFED16)	2147483587 (7FFFFFC316)
Apple CarbonLib, C++11's minstd_rand0[9]	2^31-1	16807	0
C++11's minstd_rand[9]	2^31-1	48271	0
MMIX by Donald Knuth	2^64	6.36414E+18	1.4427E+18
Newlib	2^64	6.36414E+18	1
VAX's MTH$RANDOM,[10] old versions of glibc	2^32	69069	1
Java's java.util.Random, glibc [ld]rand48[_r]()	2^48	25214903917	11

Random variable generation

- **Random *Number* Generation:**
  - Linear congruential generation (LCG) technique :
    - LCG random number generators usually come with supporting set of data containing different start values or seeds. Since period length of the random number generator is very large (or never ending), to make the generator create different number in every new run of random number generation, it just needs to move from one seed to the other for every new run.
    - There are other methods for creating random numbers (e.g., MidSquare and Multiplicative Congruential Method), Almost all of them have recursive function technique to generate random numbers; though with different formulations. LCG is one of the most efficient and commonly used method.
    - Our main purpose here was to introduce basic concept of random number generation. More deeply studies on random number generation methods might be the case with the computer science or mathematical departments!

Random variable generation

- **Random *Variable* Generation- basis** (Banks et al. 2004, chapter 8)
  - There are different methods for random variable generations. However, most of them require *random numbers* as their inputs.
  - *Main methods we briefly discuss in this lecture are:*
    - Inverse transform technique
    - *Acceptance-rejection Technique*
    - *Direct Transformation for the Normal distribution*

Random variable generation

- **Random *Variable* Generation- *Inverse transform technique*** (Banks et al. 2004, chapter 8)

  Generating random variables using inverse transform function is the most straight forward method for distribution functions with known inverse CDF.

  - ***Exponential distribution:***
    - In a normal form of exponential distribution our input to the distribution functions is exponentially distributed values and the outputs are probability density (from pdf) and probability (from cdf):

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

λ means the rate of events happening over course of time and 1/λ is the average time between two consecutive events. $E(X) = 1/\lambda$

Random variable generation

- **Random *Variable* Generation- *Inverse transform technique***
  - ***Exponential distribution (cont'd):***
    - However, the goal here is to develop a procedure for generating values X1, X2, X3, ... that have an: exponential distribution.
    - One step-by step procedure for the inverse-transform technique, illustrated by the exponential distribution, consists of the following steps:

Step 1. Compute the cdf of the desired random variable X.

For the exponential distribution, the cdf is $F(x) = 1 - e^{-\lambda x}$; $x \geq 0$.

***Step 2*.** Set $F(X) = R$ on the range of X.

For the exponential distribution, it becomes $1 - e^{-\lambda x} = R$ on the range $x \geq 0$.

X is a random variable (with the exponential distribution in this case), so $1 - e^{-\lambda x}$ is also a random variable, here called R. As it is proved, R has a uniform distribution over the interval of [0, 1].

Random variable generation

- **Random *Variable* Generation- *Inverse transform technique***
  - ***Exponential distribution (cont'd):***

    Step 3. Solve the equation *F(X)* = *R* for *X* in terms of *R*.

    For the exponential distribution, the solution proceeds as follows:

$$1-e^{-\lambda X}=R$$

$$e^{-\lambda X}=1-R$$

$$-\lambda X=\ln(1-R)$$

$$X=-\frac{1}{\lambda}\ln(1-R)$$

This equation is called a random-variate generator for the exponential distribution which is $X=F^{-1}(R)$.

Step 4. Generate (as needed) uniform random numbers R1, R2, R3, … and compute the desired random variates X1, X2, X3, … by using $X_i=F^{-1}(R_i)$. For the exponential case, $F^{-1}(R) = (-1/\lambda) \ln(1 - R)$ and equation will be: $X_i= (-1/\lambda) \ln(1 - R_i)$ *Since R is a random number with uniform[0,1] distribution, 1-R is also a random number of uniform[0,1] and the function can be simplified as:* $\boldsymbol{X_i= (-1/\lambda) \ln(R_i)}$

Random variable generation

- **Random *Variable* Generation- *Inverse transform technique***
  - ***Exponential distribution (cont'd):***

    Example: Time between new projects introduced to the market has an exponential distribution with the average of 10 days. Random numbers generated for randomly generating time between intervals are: 0.921, 0.324, 0.489

    Random-variate generator for this function is: *:* $X = (-10)\ln(R)$

$$X1 = (-10)\ln(0.921) = 0.82 \text{ days}$$

$$X1 = (-10)\ln(0.324) = 11.27 \text{ days}$$

$$X1 = (-10)\ln(0.489) = 7.15 \text{ days}$$

In class practice 3

We are going to simulate a structural steel construction where in average 20 steel elements are randomly arrived to the site from fabrication shop on daily basis. There is also an average 2.1 crane breakdown in week (= 5 working days and 10 working hours each day). Both time between steel arrival and time between crane break downs have exponential distributions.

We are going to use a Linear congruential generation (LCG) method with m of 1024, a of 64, c of 101 and X0 of 31.

You are going to use your random number generator to randomly generate first time of steel arrival to the site and the first crane break down!

In class practice 4

According to the data received from the welding foreman, welding duration for each cubic inch is a random variable with a triangular distribution with minimum duration (a) of 2 minutes and maximum duration (b) of 5 minutes and a most likely value (c) of 3 minutes. We know CDF of triangular distribution is:

$$F(x|\ a, b,\ c) = \begin{cases} 0 & \text{for } x < a, \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{for } a \le x \le c, \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{for } c < x \le b, \\ 1 & \text{for } b < x. \end{cases}$$

What would random-variant generator for welding one cubic inch look like?

Random variable generation

- **Random *Variable* Generation- *Inverse transform technique***

  With pursuing inverse transform technique we are going to have following random-variant generator for the following distributions:

  - ***Uniform distribution:***

$$X = a + (b - a)\,U$$

  - ***Triangular distribution:***

$$\begin{cases} X = a + \sqrt{U(b-a)(c-a)} & \text{for } 0 < U < F(c) \\ X = b - \sqrt{(1-U)(b-a)(b-c)} & \text{for } F(c) \le U < 1 \end{cases}$$

Random variable generation

- **Random *Variable* Generation- *Inverse transform technique***
  - ***Discrete distributions:*** All discrete distributions can be generated via the inverse-transform technique through a table-lookup procedure.

    Example: A chance of number of client's safety inspector visits our construction site is as in below:

$$p(0) = P(X = 0) = 0.50$$
$$p(1) = P(X = 1) = 0.30$$
$$p(2) = P(X = 2) = 0.20$$

For randomly generating number of safety visits during the day we can prepare its distribution table as:

x	$p(x)$	$F(x)$
0	0.50	0.50
1	0.30	0.80
2	0.20	1.00

Random variable generation

- **Random *Variable* Generation- *Inverse transform technique***

  Example (cont'd): By generating a random number we are going to look up the cumulative values on the table, the discrete random value will be determined as follows:

  Generated value = xi if $F(x_{i-1}) < R_j \leq F(x_i)$ for $i>0$

  Generated value = x1 if $R_j \leq F(x_1)$

  For example for the received random numbers of 0.42, 0.75 and 0.87 we have:

  $R=0.42 \leq F(x=0)=0.5$ =>generated x = 0

  $F(x=0)=0.5 < R = 0.75 \leq F(x=1)=0.8$ => generated x = 1

Random variable generation

- **Random *Variable* Generation- *Acceptance-rejection technique*** (Banks et al. 2004, chapter 8.2)

  Acceptance-rejection technique is used for generating random variables of a distribution by using random variables generated from a related distribution; when generating random variables from second distribution is easier (computationally or analytically).

  - ***Uniform distribution:***

  Random variables of a uniform distribution can be generated by using random numbers as input and following formula: $X = a + (b - a) U$

  This means after generation of a random number we need to do several mathematical operations.

  Suppose a random variable $X \sim U[0.25, 1]$. A substitute method for generating random number is:

  - Generate a random number of u (between 0 and 1)

  - If $u \geq 0.25$ accept u as a random variable of $X \sim U[0.25, 1]$

  -If $U < 0.25$ reject this sample and generate a new random number!

  This method is very similar or a kind of Monte-Carlo simulation technique!

Random variable generation

- **Random *Variable* Generation- *Acceptance-rejection technique***
  - ***Poisson distribution:***
    - Poisson distribution counts occurrence of events within the unite interval where time between two consecutive events are exponentially distributed.
    - Creating exponentially distributed numbers are possible through inverse CDF.
    - For generating Poisson numbers in acceptance-rejection technique we are going to consecutively generate exponential values (with the same λ parameter value of Poisson distribution) and add exponential values received until summation of exponentially generated values exceeds the time interval for the first time.
    - Our Poisson number will be number of total sampled exponential values minus one!

Generated Poisson value = n if $A1+A2+ \dots +An \leq 1 < A1+A2+ \dots +An + A(n+1)$ where $Ai \sim$ exponential distribution with the same λ parameter value as Poisson

As you might have noticed acceptance-rejection technique is not a very efficient method!

Random variable generation

- **Random *Variable* Generation- *Acceptance-rejection technique***
  - ***Poisson distribution:***

Example: Number of site broken down equipment per week follows a Poisson distribution with an average number of 5. Randomly generate number of broken down equipment for the first week. For random numbers presented in the table we have:

Exponential random-variant generator: $X_i = (-1/5)\ ln(R_i)$

i	Ri	Xi	Cumulative Sum
1	0.905	0.020	0.020
2	0.131	0.407	0.427
3	0.756	0.056	0.483
4	0.718	0.066	0.549
5	0.318	0.229	0.778
6	0.155	0.373	1.150

The random Poisson number is: **5**

In class practice 5

Time between arrival of structural steel elements on site has an exponential distribution with the mean of 4 hours. Every day site opens from 7 am to 7 pm. Use following random numbers to randomly generate number of steel elements arrive on site on *daily basis* for 3 days.

	Ri
1	0.930
2	0.604
3	0.026
4	0.921
5	0.633
6	0.325
7	0.156
8	0.169
9	0.906
10	0.837
11	0.663
12	0.482
13	0.271
14	0.150
15	0.072
16	0.401
17	0.374
18	0.870
19	0.299

Random variable generation

- **Random *Variable* Generation- *Direct Transformation for the Normal distribution technique*** (Banks et al. 2004, chapter 8.3.1)

CDF inverse of normal distribution is not known, so we can not create use CDF inverse method for Normal distribution

- There is specific method developed in 1958 by Box and Muller for generating random numbers with normal standard distribution (refer to Banks et al. 2004, chapter 8.3.1 for the proof):

$$Z_1 = (-2\ln R_1)^{1/2}\cos(2\pi R_2)$$
$$Z_2 = (-2\ln R_1)^{1/2}\sin(2\pi R_2)$$

- In this method we are going to receive two normal *standard* variable by using 2 random numbers!
- Then Normal standard variables Z~N(0, 1) can easily be transformed to Normal variables of X~N(μ, σ) through: $X = \mu + Z\sigma$

Random variable generation

- **Random *Variable* Generation- *Direct Transformation for the Normal distribution technique***
- ***Example:*** Duration of loading activity has a normal distribution with the mean of 2 minutes and standard deviation of 20 seconds. Using randomly generated numbers of 0.901 and 0.347 for *sampling* loading durations we have:

*Z1 =(-2*LN(R1))^(1/2)* COS(2 * PI() *R2) = -0.261*

*Z2= =(-2*LN(R1))^(1/2)* SIN(2 * PI() * R2) = 0.374*

X1= 2 + 0.261 x 20 / 60 = 2.12 minute

X2= 2 - 0.374 x 20 / 60 = 1.91 minute

Home assignment 4

Assume an earthmoving example with 2 trucks (truck 1, and truck 2) and 2 loaders (loader 1 and loader 2) in the system where on a truck arrival to the loading site, if two loaders are idle, truck refers to the loader 1 for loading, if one loader is busy and the other one is idle, truck refers to the idle loader. In this operation only one dumping site is available. Job site works from 7am to 7 pm. With the following assumption hand simulate the system for 1 hour; calculate the average length of the queue and the waiting time in the dumping site, average idle time of each loader and the productivity of the system.

- Time between first system arrivals (morning arrivals) has an exponential distribution with the average of 1 minute.
- Loading: Normal (mean=12 minutes, SD=2 minutes)
- Trip to the dumping site: Normal (mean=2 minutes, SD=0.3 minute)
- Dumping 5 minutes
- Trip from dumping: Exponential distribution with the average of 1 minute.
- For generating random numbers use LCG method with m=2^20, a=511, C=1, X_0=1000

(Due in 2 weeks)

Reference

- Banks, J., Carson, J.S., Nelson, B.L. And Nicol D.M. (2004) “Discrete event simulation” Prentice Hall, ISBN: 0131446797., chapters 5, 7 and 8.
- Banks, J. (1998) "HANDBOOK OF SIMULATION, Principles, Methodology, Advances, Applications, and Practice” John Wiley and Sons, Toronto, Canada. ISBN 0-471-13403-1.
- Casella, G. and Berger, R. L. (2001). Statistical Inference (2nd ed.). Duxbury. ISBN 0-534-24312-6.

Thank you!