I & IE
UTK

Simulation

Statistical Models in Simulation

Dr. Xueping Li
University of Tennessee

Based on
Banks, Carson, Nelson & Nicol
Discrete-Event System Simulation

Purpose & Overview

- The world the model-builder sees is probabilistic rather than deterministic.
  - Some statistical model might well describe the variations.
- An appropriate model can be developed by sampling the phenomenon of interest:
  - Select a known distribution through educated guesses
  - Make estimate of the parameter(s)
  - Test for goodness of fit
- We will:
  - Review several important probability distributions
  - Present some typical application of these models

Review of Terminology and Concepts

- In this section, we will review the following concepts:
  - Discrete random variables
  - Continuous random variables
  - Cumulative distribution function
  - Expectation

Discrete Random Variables

[Probability Review]

- X is a discrete random variable if the number of possible values of X is finite, or countably infinite.
- Example: Consider jobs arriving at a job shop.
    - Let X be the number of jobs arriving each week at a job shop.
    - R_x = possible values of X (range space of X) = $\{0,1,2,\ldots\}$
    - $p(x_i)$ = probability the random variable is $x_i = P(X = x_i)$
  - $p(x_i)$, $i = 1,2, \ldots$ must satisfy:

    1. $p(x_i) \geq 0, \text{ for all } i$

    2. $\sum_{i=1}^{\infty} p(x_i) = 1$

  - The collection of pairs $[x_i, p(x_i)]$, $i = 1,2,\ldots$, is called the probability distribution of X, and $p(x_i)$ is called the probability mass function (pmf) of X.

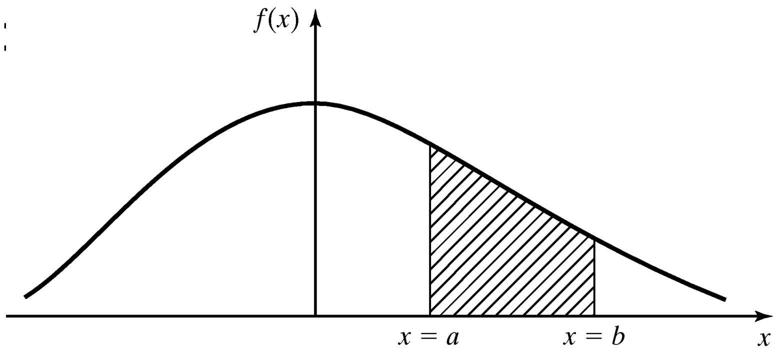
Continuous Random Variables [Probability Review]

- X is a continuous random variable if its range space R_x is an interval or a collection of intervals.
- The probability that X lies in the interval *[a,b]* is given by:

$$P(a \le X \le b) = \int_a^b f(x)dx$$

- *f(x)*, denoted as the pdf of *X*, satisfies:

1. $f(x) \ge 0$, for all x in R_X
2. $\int_{R_X} f(x)dx = 1$
3. $f(x) = 0$, if x is not in R_X



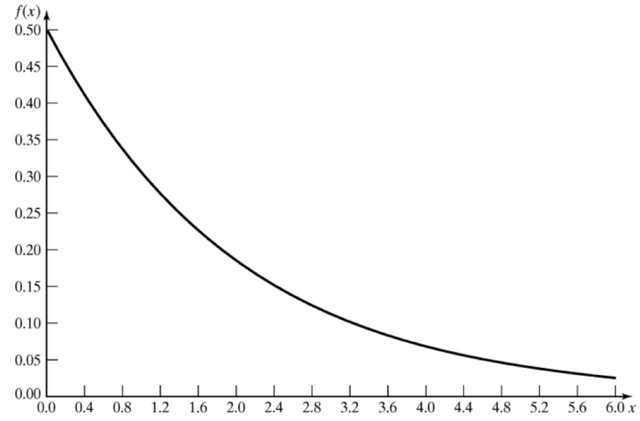
- Properties

1. $P(X = x_0) = 0$, because $\int_{x_0}^{x_0} f(x)dx = 0$
2. $P(a \le X \le b) = P(a \prec X \le b) = P(a \le X \prec b) = P(a \prec X \prec b)$

Continuous Random Variables [Probability Review]

- Example: Life of an inspection device is given by *X*, a continuous random variable with pdf:

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$



  - *X* has an exponential distribution with mean 2 years
  - Probability that the device's life is between 2 and 3 years is:

$$P(2 \leq x \leq 3) = \frac{1}{2}\int_2^3 e^{-x/2}dx = 0.14$$

Cumulative Distribution Function [Probability Review]

- Cumulative Distribution Function (cdf) is denoted by *F(x)*, where *F(x) = P(X <= x)*
  - If *X* is discrete, then $$F(x) = \sum_{\substack{\text{all} \\ x_i \le x}} p(x_i)$$
  - If *X* is continuous, then $$F(x) = \int_{-\infty}^{x} f(t)\,dt$$
- Properties

  1. F is nondecreasing function. If $a \prec b$, then $F(a) \le F(b)$
  2. $\lim_{x\to\infty} F(x) = 1$
  3. $\lim_{x\to-\infty} F(x) = 0$

- All probability question about *X* can be answered in terms of the cdf, e.g.:

$$P(a \prec X \le b) = F(b) - F(a), \text{ for all } a \prec b$$

Cumulative Distribution Function [Probability Review]

- Example: An inspection device has cdf:

$$F(x) = \frac{1}{2}\int_0^x e^{-t/2}dt = 1 - e^{-x/2}$$

  - The probability that the device lasts for less than 2 years:

$$P(0 \le X \le 2) = F(2) - F(0) = F(2) = 1 - e^{-1} = 0.632$$

  - The probability that it lasts between 2 and 3 years:

$$P(2 \le X \le 3) = F(3) - F(2) = (1 - e^{-(3/2)}) - (1 - e^{-1}) = 0.145$$

Expectation

- The expected value of *X* is denoted by *E(X)*
  - If *X* is discrete $$E(x) = \sum_{\text{all } i} x_i p(x_i)$$
  - If *X* is continuous $$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$
  - a.k.a the mean, m, or the 1st moment of *X*
  - A measure of the central tendency
- The variance of *X* is denoted by *V(X)* or *var(X)* or σ^2
  - Definition: $V(X) = E[(X - E[X]^2]$
  - Also, $V(X) = E(X^2) - [E(x)]^2$
  - A measure of the spread or variation of the possible values of X around the mean
- The standard deviation of *X* is denoted by σ
  - Definition: Square root of *V(X)*
  - Expressed in the same units as the mean

Expectations

- Example: The mean of life of the previous inspection device is:

$$E(X)=\frac{1}{2}\int_0^{\infty} xe^{-x/2}dx = -xe^{-x/2}\Big|_0^{\infty}+\int_0^{\infty} e^{-x/2}dx = 2$$

- To compute variance of X, we first compute $E(X^2)$:

$$E(X^2)=\frac{1}{2}\int_0^{\infty} x^2e^{-x/2}dx = -x^2e^{-x/2}\Big|_0^{\infty}+\int_0^{\infty} e^{-x/2}dx = 8$$

- Hence, the variance and standard deviation of the device's life are:

$$V(X)=8-2^2=4$$

$$\sigma=\sqrt{V(X)}=2$$

Useful Statistical Models

- In this section, statistical models appropriate to some application areas are presented. The areas include:
  - Queueing systems
  - Inventory and supply-chain systems
  - Reliability and maintainability
  - Limited data

Queueing Systems [Useful Models]

- In a queueing system, interarrival and service-time patterns can be probablistic .
- Sample statistical models for interarrival or service time distribution:
  - Exponential distribution: if service times are completely random
  - Normal distribution: fairly constant but with some random variability (either positive or negative)
  - Truncated normal distribution: similar to normal distribution but with restricted value.
  - Gamma and Weibull distribution: more general than exponential (involving location of the modes of pdf’s and the shapes of tails.)

Inventory and supply chain [Useful Models]

- In realistic inventory and supply-chain systems, there are at least three random variables:
  - The number of units demanded per order or per time period
  - The time between demands
  - The lead time
- Sample statistical models for lead time distribution:
  - Gamma
- Sample statistical models for demand distribution:
  - Poisson: simple and extensively tabulated.
  - Negative binomial distribution: longer tail than Poisson (more large demands).
  - Geometric: special case of negative binomial given at least one demand has occurred.

Reliability and maintainability [Useful Models]

- Time to failure (TTF)
  - Exponential: failures are random
  - Gamma: for standby redundancy where each component has an exponential TTF
  - Weibull: failure is due to the most serious of a large number of defects in a system of components
  - Normal: failures are due to wear

Other areas

- For cases with limited data, some useful distributions are:
  - Uniform, triangular and beta
- Other distribution: Bernoulli, binomial and hyperexponential.

Discrete Distributions

- Discrete random variables are used to describe random phenomena in which only integer values can occur.
- In this section, we will learn about:
  - Bernoulli trials and Bernoulli distribution
  - Binomial distribution
  - Geometric and negative binomial distribution
  - Poisson distribution

Bernoulli Trials and Bernoulli Distribution [Discrete Dist'n]

- Bernoulli Trials:
  - Consider an experiment consisting of n trials, each can be a success or a failure.
    - Let X_j = 1 if the jth experiment is a success
    - and X_j = 0 if the jth experiment is a failure
  - The Bernoulli distribution (one trial):

$$p_j(x_j) = p(x_j) = \begin{cases} p, & x_j = 1, j = 1,2,\ldots,n \\ 1-p=q, & x_j = 0, j = 1,2,\ldots,n \\ 0, & \text{otherwise} \end{cases}$$

  - where $E(X_j) = p$ and $V(X_j) = p(1-p) = pq$
- Bernoulli process:
  - The n Bernoulli trials where trails are independent:

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)\, p_2(x_2) \ \ldots \ p_n(x_n)$$

Binomial Distribution

[Discrete Dist'n]

- The number of successes in *n* Bernoulli trials, *X*, has a binomial distribution.

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

The number of outcomes having the required number of successes and failures

Probability that there are x successes and (n-x) failures

  - The mean, *E(x)* = *p* + *p* + … + *p* = *n*p*
  - The variance, *V(X)* = *pq* + *pq* + … + *pq* = *n*pq*

Geometric & Negative Binomial Distribution

[Discrete Dist'n]

- Geometric distribution
  - The number of Bernoulli trials, *X*, to achieve the 1st success:

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

  - *E(x) = 1/p, and V(X) = q/p*^{2}

- Negative binomial distribution
  - The number of Bernoulli trials, X, until the *k*th success
  - If Y is a negative binomial distribution with parameters p and k, then:

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

  - *E(Y) = k/p, and V(X) = kq/p*^{2}

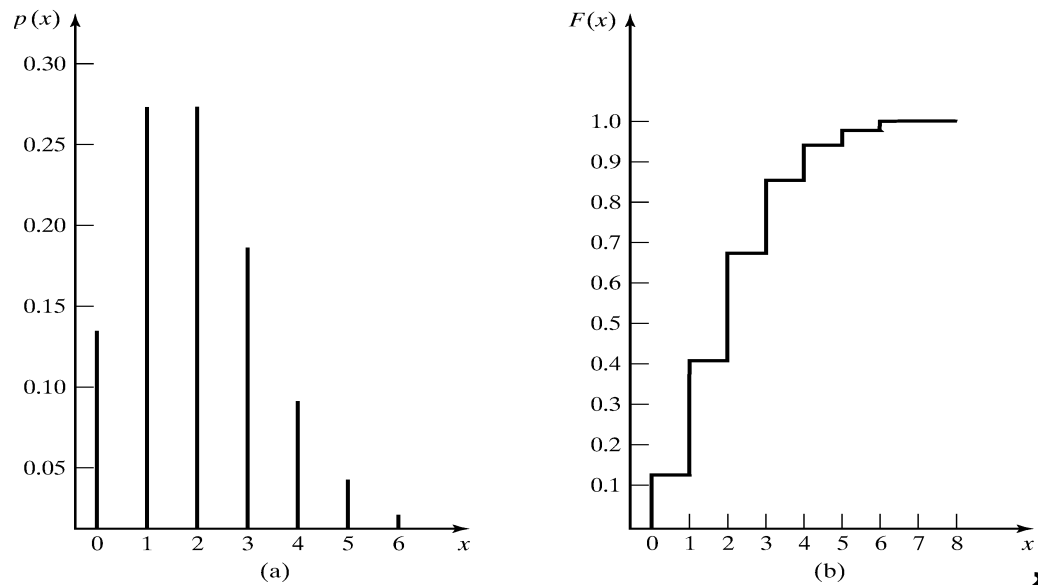
Poisson Distribution [Discrete Dist'n]

- Poisson distribution describes many random processes quite well and is mathematically quite simple.
  - where $\alpha > 0$, pdf and cdf are:

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

$$F(x) = \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^{i}}{i!}$$

  - $E(X) = \alpha = V(X)$



Poisson Distribution [Discrete Dist'n]

- Example: A computer repair person is "beeped" each time there is a call for service. The number of beeps per hour ~ Poisson(α = 2 per hour).
  - The probability of three beeps in the next hour:

    $p(3) = e^{-2}2^3/3! = 0.18$

    also, $p(3) = F(3) - F(2) = 0.857-0.677=0.18$

  - The probability of two or more beeps in a 1-hour period:

    $$\begin{aligned} p(2 \text{ or more}) &= 1 - p(0) - p(1) \\ &= 1 - F(1) \\ &= 0.594 \end{aligned}$$

Continuous Distributions

- Continuous random variables can be used to describe random phenomena in which the variable can take on any value in some interval.
- In this section, the distributions studied are:
  - Uniform
  - Exponential
  - Normal
  - Weibull
  - Lognormal

Uniform Distribution [Continuous Dist'n]

- A random variable X is uniformly distributed on the interval (a,b), $U(a,b)$, if its pdf and cdf are:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases} \qquad F(x) = \begin{cases} 0, & x \prec a \\ \dfrac{x-a}{b-a}, & a \le x \prec b \\ 1, & x \ge b \end{cases}$$

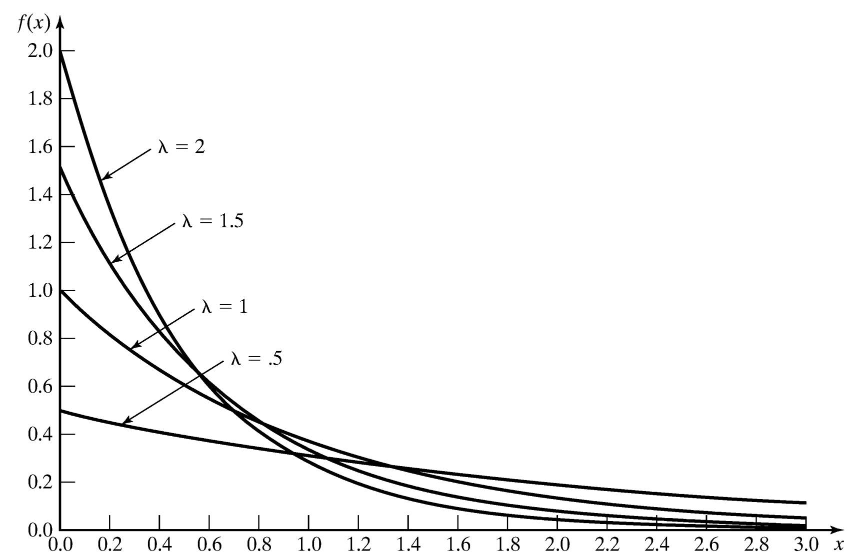
- Properties
  - $P(x_1 < X < x_2)$ is proportional to the length of the interval [$F(x_2) - F(x_1) = (x_2-x_1)/(b-a)$]
  - $E(X) = (a+b)/2$ $\qquad$ $V(X) = (b-a)^2/12$
- U(0,1) provides the means to generate random numbers, from which random variates can be generated.

Exponential Distribution [Continuous Dist'n]

- A random variable X is exponentially distributed with parameter $\lambda > 0$ if its pdf and cdf are:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases} \qquad F(x) = \begin{cases} 0, & x \prec 0 \\ \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

  - $E(X) = 1/\lambda$ $\qquad V(X) = 1/\lambda^2$
  - Used to model interarrival times when arrivals are completely random, and to model service times that are highly variable
  - For several different exponential pdf's (see figure), the value of intercept on the vertical axis is λ, and all pdf's eventually intersect.



Exponential Distribution [Continuous Dist'n]

- Memoryless property
  - For all s and t greater or equal to 0:

$$P(X > s+t \mid X > s) = P(X > t)$$

  - Example: A lamp ~ exp(λ = 1/3 per hour), hence, on average, 1 failure per 3 hours.
    - The probability that the lamp lasts longer than its mean life is: $P(X > 3) = 1-(1-e^{-3/3}) = e^{-1} = 0.368$
    - The probability that the lamp lasts between 2 to 3 hours is:

      $P(2 <= X <= 3) = F(3) – F(2) = 0.145$
    - The probability that it lasts for another hour given it is operating for 2.5 hours:

      $P(X > 3.5 \mid X > 2.5) = P(X > 1) = e^{-1/3} = 0.717$

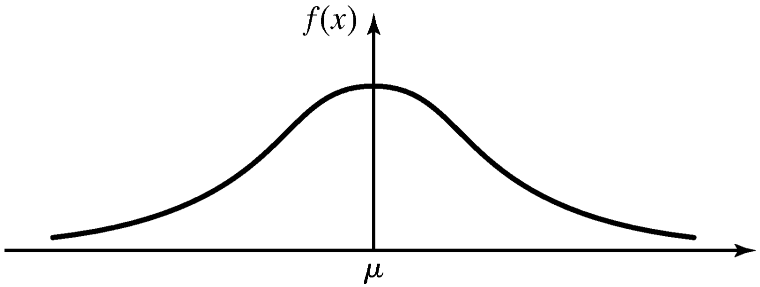
Normal Distribution

[Continuous Dist'n]

- A random variable X is normally distributed has the pdf:

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

  - Mean: $-\infty \prec \mu \prec \infty$
  - Variance: $\sigma^2 \succ 0$
  - Denoted as $X \sim N(\mu, \sigma^2)$



- Special properties:
  - $\lim_{x\to-\infty} f(x) = 0, \text{ and } \lim_{x\to\infty} f(x) = 0.$
  - $f(\mu-x)=f(\mu+x)$; the pdf is symmetric about μ.
  - The maximum value of the pdf occurs at $x = \mu$; the mean and mode are equal.

Normal Distribution

[Continuous Dist'n]

- Evaluating the distribution:
  - Use numerical methods (no closed form)
  - Independent of μ and σ, using the standard normal distribution:

    $Z \sim N(0,1)$
  - Transformation of variables: let Z = (X - μ) / σ,

$$F(x) = P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right)$$

$$= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= \int_{-\infty}^{(x-\mu)/\sigma} \phi(z) dz = \Phi\left(\tfrac{x-\mu}{\sigma}\right) \quad \text{, where} \quad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

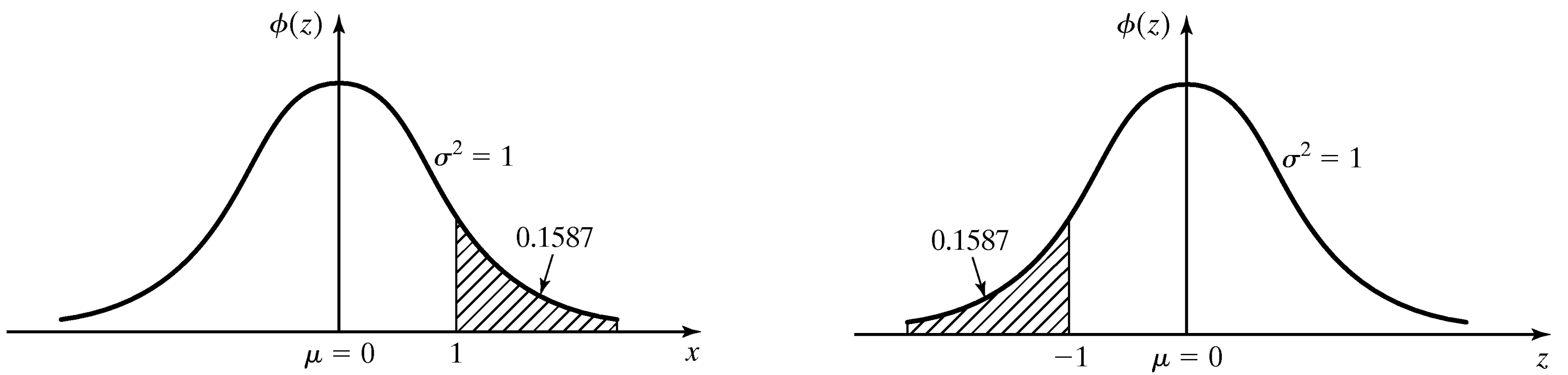
Normal Distribution

[Continuous Dist'n]

- Example: The time required to load an oceangoing vessel, *X*, is distributed as *N(12,4)*
  - The probability that the vessel is loaded in less than 10 hours:

$$F(10) = \Phi\left(\frac{10-12}{2}\right) = \Phi(-1) = 0.1587$$

    - Using the symmetry property, $\Phi(1)$ is the complement of Φ (-1)



(a) (b)

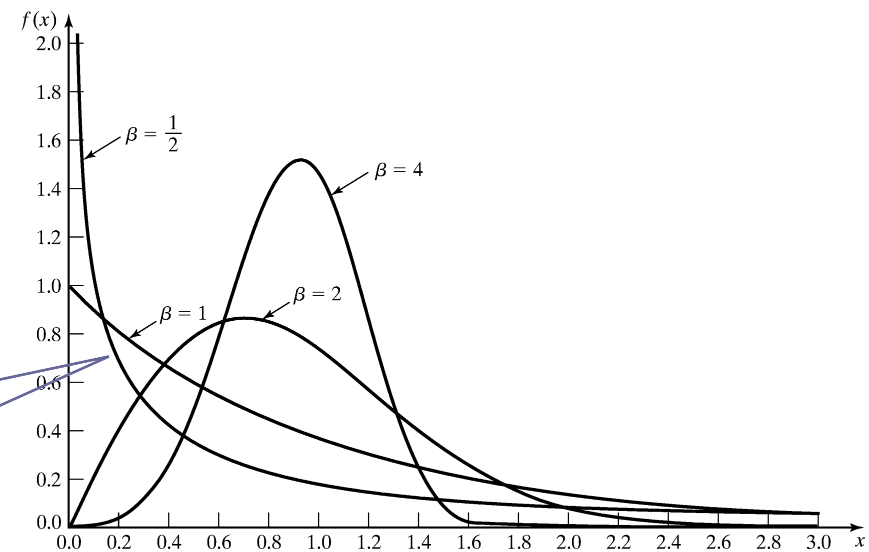
Weibull Distribution

[Continuous Dist’n]

- A random variable *X* has a Weibull distribution if its pdf has the form:

$$f(x) = \begin{cases} \dfrac{\beta}{\alpha}\left(\dfrac{x-\nu}{\alpha}\right)^{\beta-1} \exp\left[-\left(\dfrac{x-\nu}{\alpha}\right)^{\beta}\right], & x \geq \nu \\ 0, & \text{otherwise} \end{cases}$$

- 3 parameters:
  - Location parameter: υ, $(-\infty \prec \nu \prec \infty)$
  - Scale parameter: β, $(\beta > 0)$
  - Shape parameter. α, (> 0)
- Example: $\upsilon = 0$ and $\alpha = 1$:

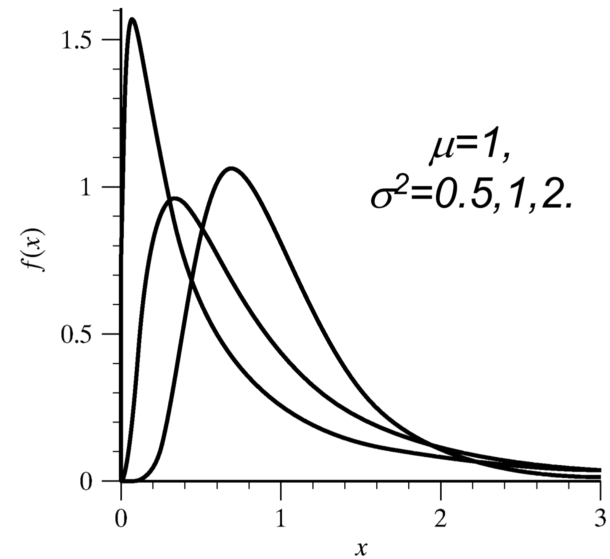


When $\beta = 1$,
$X \sim exp(\lambda = 1/\alpha)$

Lognormal Distribution [Continuous Dist'n]

- A random variable X has a lognormal distribution if its pdf has the form:

$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\dfrac{(\ln x - \mu)^2}{2\sigma^2} \right], & x \succ 0 \\ 0, & \text{otherwise} \end{cases}$$



  - Mean $E(X) = e^{\mu+\sigma^2/2}$
  - Variance $V(X) = e^{2\mu+\sigma^2/2} (e^{\sigma^2} - 1)$

- Relationship with normal distribution
  - When $Y \sim N(\mu, \sigma^2)$, then $X = e^Y \sim$ lognormal(μ, σ^2)
  - Parameters μ and σ^2 are not the mean and variance of the lognormal

Poisson Distribution

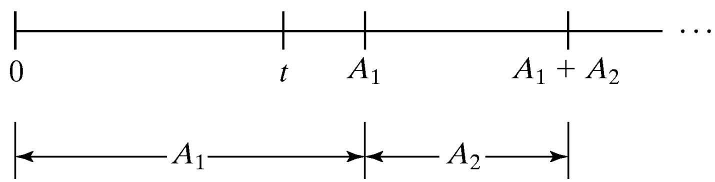
- Definition: *N(t)* is a counting function that represents the number of events occurred in *[0,t].*
- A counting process {*N(t), t>=0*} is a Poisson process with mean rate λ if:
  - Arrivals occur one at a time
  - {*N(t), t>=0*} has stationary increments
  - {*N(t), t>=0*} has independent increments
- Properties

$$P[N(t)=n]=\frac{e^{-\lambda t}(\lambda t)^n}{n!}, \qquad \text{for } t\geq 0 \text{ and } n=0,1,2,\ldots$$

  - Equal mean and variance: $E[N(t)] = V[N(t)] = \lambda t$
  - Stationary increment: The number of arrivals in time *s* to *t* is also Poisson-distributed with mean $\lambda(t\text{-}s)$

Interarrival Times [Poisson Dist'n]

- Consider the interarrival times of a Possion process ($A_1, A_2, \ldots$), where A_i is the elapsed time between arrival i and arrival $i+1$



  - The 1st arrival occurs after time t iff there are no arrivals in the interval [0,t], hence:

$$P\{A_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

$$P\{A_1 <= t\} = 1 - e^{-\lambda t} \qquad \text{[cdf of exp}(\lambda)\text{]}$$

  - Interarrival times, $A_1, A_2, \ldots$, are exponentially distributed and independent with mean $1/\lambda$

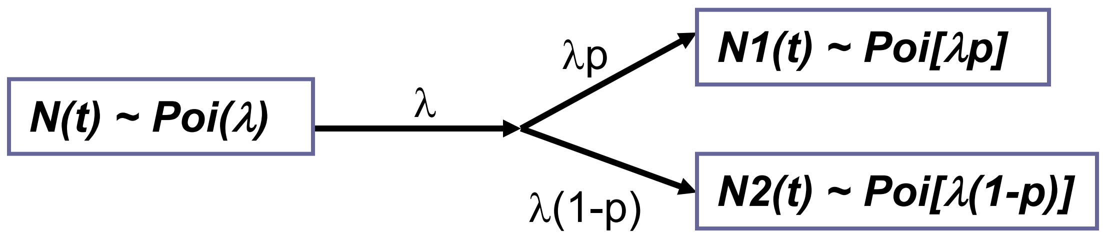
Arrival counts $\sim Poi(\lambda)$ ⟺ Interarrival time $\sim Exp(1/\lambda)$

Stationary & Independent

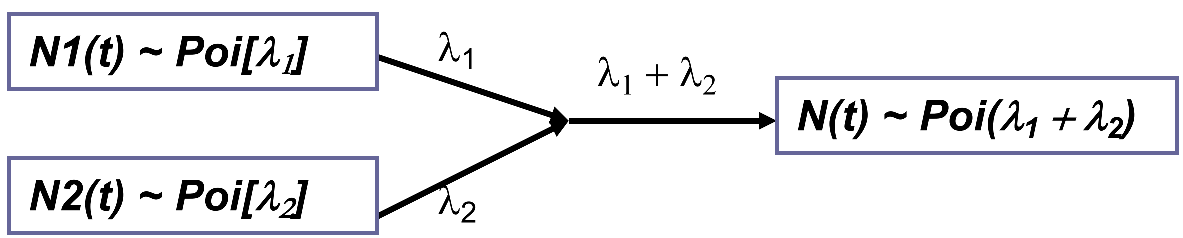
Memoryless

Splitting and Pooling [Poisson Dist'n]

- Splitting:
  - Suppose each event of a Poisson process can be classified as Type I, with probability *p* and Type II, with probability *1-p.*
  - $N(t) = N1(t) + N2(t)$, where $N1(t)$ and $N2(t)$ are both Poisson processes with rates λp and $\lambda(1-p)$



- Pooling:
  - Suppose two Poisson processes are pooled together
  - $N1(t) + N2(t) = N(t)$, where $N(t)$ is a Poisson processes with rates $\lambda_1 + \lambda_2$



Nonstationary Poisson Process (NSPP)

- Poisson Process without the stationary increments, characterized by $\lambda(t)$, the arrival rate at time t.
- The expected number of arrivals by time t, $\Lambda(t)$:

$$\Lambda(t) = \int_0^t \lambda(s)ds$$

- Relating stationary Poisson process $n(t)$ *with rate* $\lambda=1$ and NSPP $N(t)$ with rate $\lambda(t)$:
  - Let arrival times of a stationary process with rate $\lambda = 1$ be $t_1, t_2, \ldots$, and arrival times of a NSPP with rate $\lambda(t)$ be $T_1, T_2, \ldots$, we know:

$$t_i = \Lambda(T_i)$$

$$T_i = \Lambda^{-1}(t_i)$$

Nonstationary Poisson Process (NSPP)

- Example: Suppose arrivals to a Post Office have rates 2 per minute from 8 am until 12 pm, and then 0.5 per minute until 4 pm.
- Let t = 0 correspond to 8 am, NSPP *N(t)* has rate function:

$$\lambda(t) = \begin{cases} 2, & 0 \le t \prec 4 \\ 0.5, & 4 \le t \prec 8 \end{cases}$$

Expected number of arrivals by time t:

$$\Lambda(t) = \begin{cases} 2t, & 0 \le t \prec 4 \\ \int_0^4 2ds + \int_4^t 0.5ds = \frac{t}{2} + 6, & 4 \le t \prec 8 \end{cases}$$

- Hence, the probability distribution of the number of arrivals between 11 am and 2 pm.

$$\begin{aligned} P[N(6) - N(3) = k] &= P[N(\Lambda(6)) - N(\Lambda(3)) = k] \\ &= P[N(9) - N(6) = k] \\ &= e^{(9-6)}(9-6)^k/k! \quad = e^3(3)^k/k! \end{aligned}$$

Empirical Distributions [Poisson Dist'n]

- A distribution whose parameters are the observed values in a sample of data.
  - May be used when it is impossible or unnecessary to establish that a random variable has any particular parametric distribution.
  - Advantage: no assumption beyond the observed values in the sample.
  - Disadvantage: sample might not cover the entire range of possible values.

Summary

- The world that the simulation analyst sees is probabilistic, not deterministic.
- We have:
  - Reviewed several important probability distributions.
  - Showed applications of the probability distributions in a simulation context.
- Important task in simulation modeling is the collection and analysis of input data, e.g., hypothesize a distributional form for the input data. Reader should know:
  - Difference between discrete, continuous, and empirical distributions.
  - Poisson process and its properties.