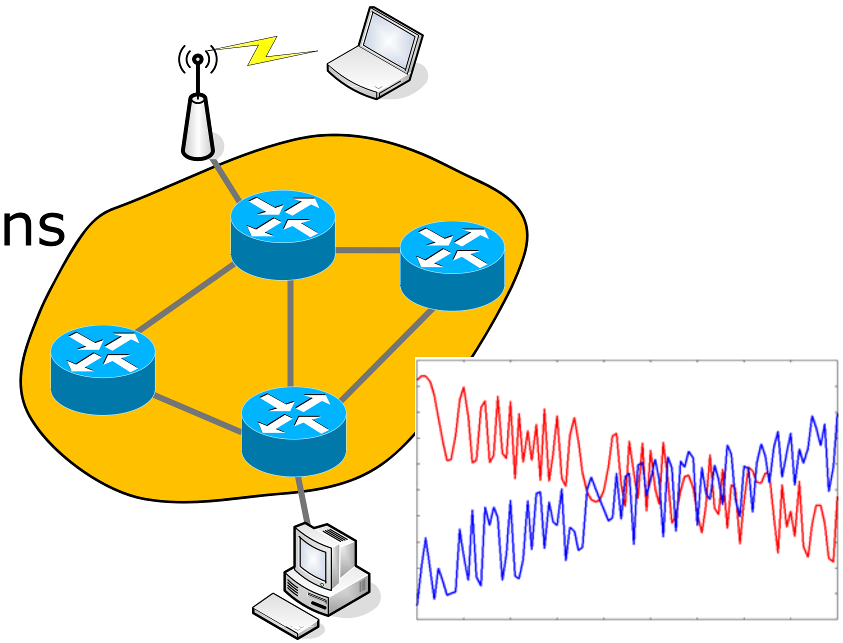
Chapter 5

Statistical Models in Simulations



Contents

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 parameters
  - Test for goodness of fit

- In this chapter:
  - Review several important probability distributions
  - Present some typical application of these models

Basic Probability Theory Concepts

Review of Terminology and Concepts

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

Discrete Random Variables

- X is a discrete random variable if the number of possible values of X is finite, or countable infinite.
- Example: Consider packets arriving at a router.
  - Let X be the number of packets arriving each second at a router.

    R_X = possible values of X (range space of X) = $\{0,1,2,\ldots\}$

    $p(x_i)$ = probability the random variable X is x_i , $p(x_i) = P(X = x_i)$
  - $p(x_i),\ i = 1,2, \ldots$ 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

- 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:

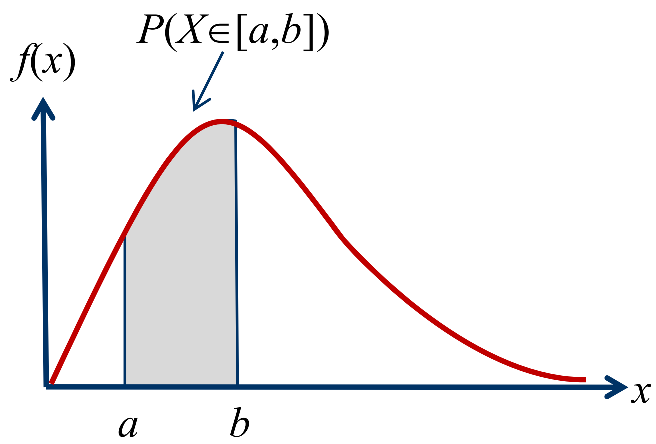
$$P(a \le X \le b) = \int_a^b f(x)dx$$

- $f(x)$ is called the **probability density function** (PDF) of X, and satisfies:

$$1.\ f(x) \ge 0,\ \text{for all } x \text{ in } R_X$$

$$2.\ \int_{R_X} f(x)dx = 1$$

$$3.\ f(x) = 0,\ \text{ if } x \text{ is not in } R_X$$



- Properties

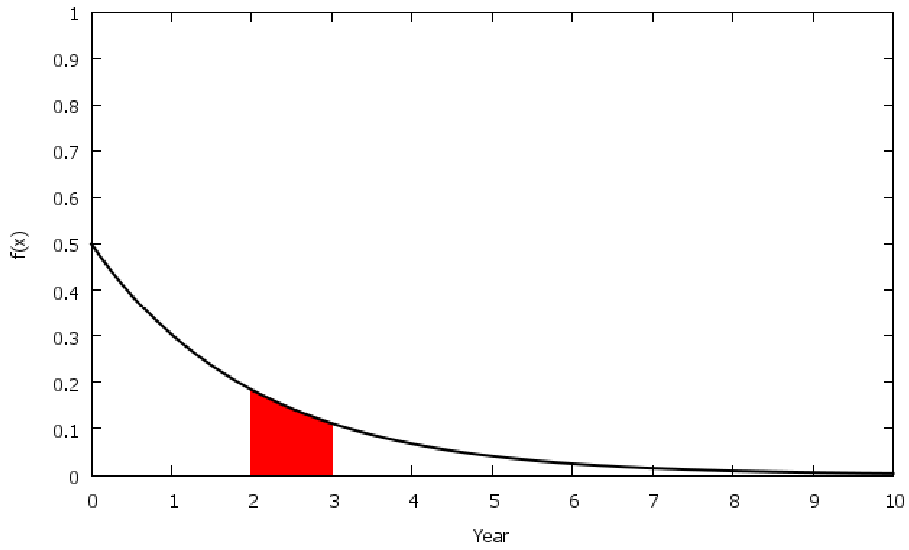
$$1.\ P(X = x_0) = 0,\ \text{because} \int_{x_0}^{x_0} f(x)dx = 0$$

$$2.\ P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$

Continuous Random Variables

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

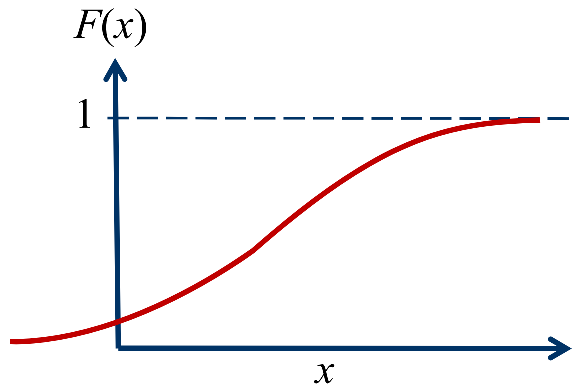
$$P(2 \leq x \leq 3) = \frac{1}{2} \int_2^3 e^{-x/2} dx = 0.145$$

Cumulative Distribution Function

- Cumulative Distribution Function (**CDF**) is denoted by $F(x)$, where

$$F(x) = P(X \le x)$$

  - If X is discrete, then $$F(x) = \sum_{x_i \le x} p(x_i)$$
  - If X is continuous, then $$F(x) = \int_{-\infty}^{x} f(t)dt$$



- Properties

$$1.\ F \text{ is nondecreasing function. If } a \le b, \text{ then } F(a) \le F(b)$$

$$2.\ \lim_{x \to \infty} F(x) = 1$$

$$3.\ \lim_{x \to -\infty} F(x) = 0$$

- All probability questions about X can be answered in terms of the CDF:

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

Cumulative Distribution Function

- Example: The 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:

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

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

$$P(2 \leq X \leq 3) = F(3) - F(2) = \left(1 - e^{-\frac{3}{2}}\right) - \left(1 - e^{-1}\right) = 0.145$$

Expected value

- The expected value of X is denoted by $E(X)$

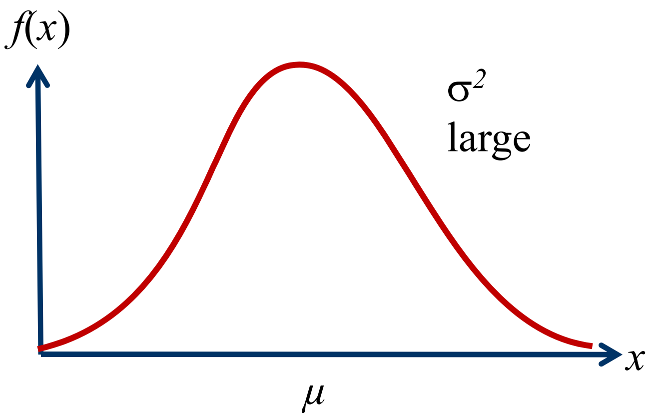
  - If X is discrete $$E(X) = \sum_{\text{all } i} x_i p(x_i)$$

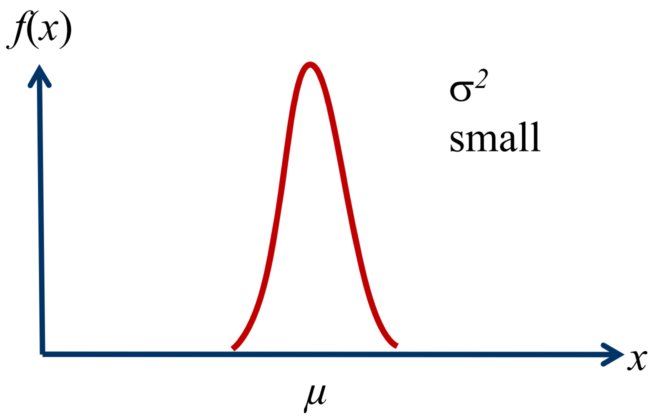
  - If X is continuous $$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

  - a.k.a the mean, m, μ, or the 1st moment of X
  - A measure of the central tendency

Variance

- 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





Standard deviation

- The standard deviation (SD) of X is denoted by σ
  - Definition: $\sigma = \sqrt{V(x)}$
  - The standard deviation is **expressed** in the **same units** as the **mean**
  - Interprete σ always together with the mean
- Attention:
  - The standard deviation of two different data sets may be difficult to compare

Expected value and variance: Example

- 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 the 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:

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

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

Expected value and variance: Example

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

Partial Integration

$$\int u(x)v'(x)dx = u(x)v(x) - \int u'(x)v(x)dx$$

Set

$$u(x) = x$$

$$v'(x) = e^{-x/2}$$

$$\Rightarrow$$

$$u'(x) = 1$$

$$v(x) = -2e^{-x/2}$$

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

Mean and variance of sums

- If $x_1, x_2, \ldots, x_k$ are k random variables and if $a_1, a_2, \ldots, a_k$ are k constants, then

$$E(a_1x_1+a_2x_2+\ldots+a_kx_k) = a_1E(x_1)+a_2E(x_2)+\ldots+a_kE(x_k)$$

- For independent variables

$$\mathrm{Var}(a_1x_1 + a_2x_2 + \cdots + a_kx_k) = a_1^2\,\mathrm{Var}(x_1) + a_2^2\,\mathrm{Var}(x_2) + \ldots + a_k^2\,\mathrm{Var}(x_k)$$

Coefficient of variation

- The ratio of the standard deviation to the mean is called coefficient of variation (C.O.V.)
  - Dimensionless
  - Normalized measure of dispersion

$$C.O.V = \frac{\text{standard deviation}}{\text{mean}} = \frac{\sigma}{\mu} \qquad , \mu > 0$$

  - Can be used to compare different datasets, instead the standard deviation.

Covariance

- Given two random variables x and y with μ_x and μ_y, their covariance is defined as

$$\text{Cov}(x, y) = \sigma^2_{xy} = E[(x-\mu_x)(y-\mu_y)] = E(xy) - E(x)\,E(y)$$

  - $\text{Cov}(x, y)$ measures the dependency of x and y, i.e., how x and y vary together.

- For independent variables, the covariance is **zero**, since

$$E(xy) = E(x)E(y)$$

Correlation coefficient

- The normalized value of covariance is called the correlation coefficient or simply correlation

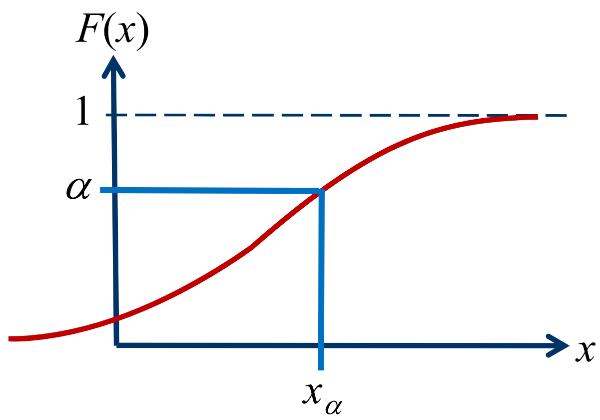
$$\text{Correlation}(x, y) = \rho_{x,y} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}$$

- The correlation lies between -1 and +1

Quantile

- The x value at which the CDF takes a value α is called the **α-quantile** or **100α-percentile**. It is denoted by x_α.

$$P(X \le x_\alpha) = F(x_\alpha) = \alpha \quad , \ \alpha \in [0,1]$$

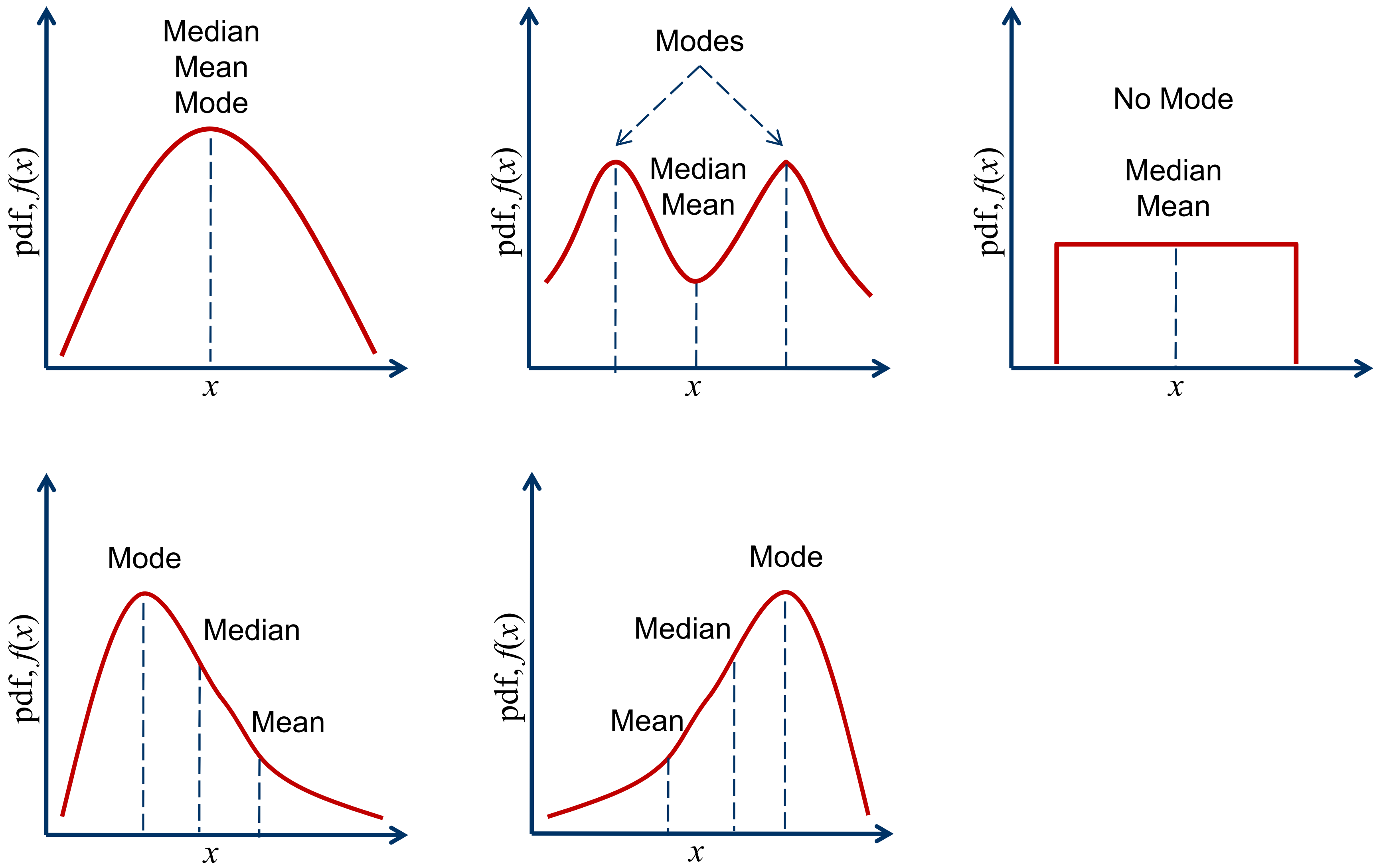


- Relationship:
  - The **median** is the **50-percentile** or **0.5-quantile**

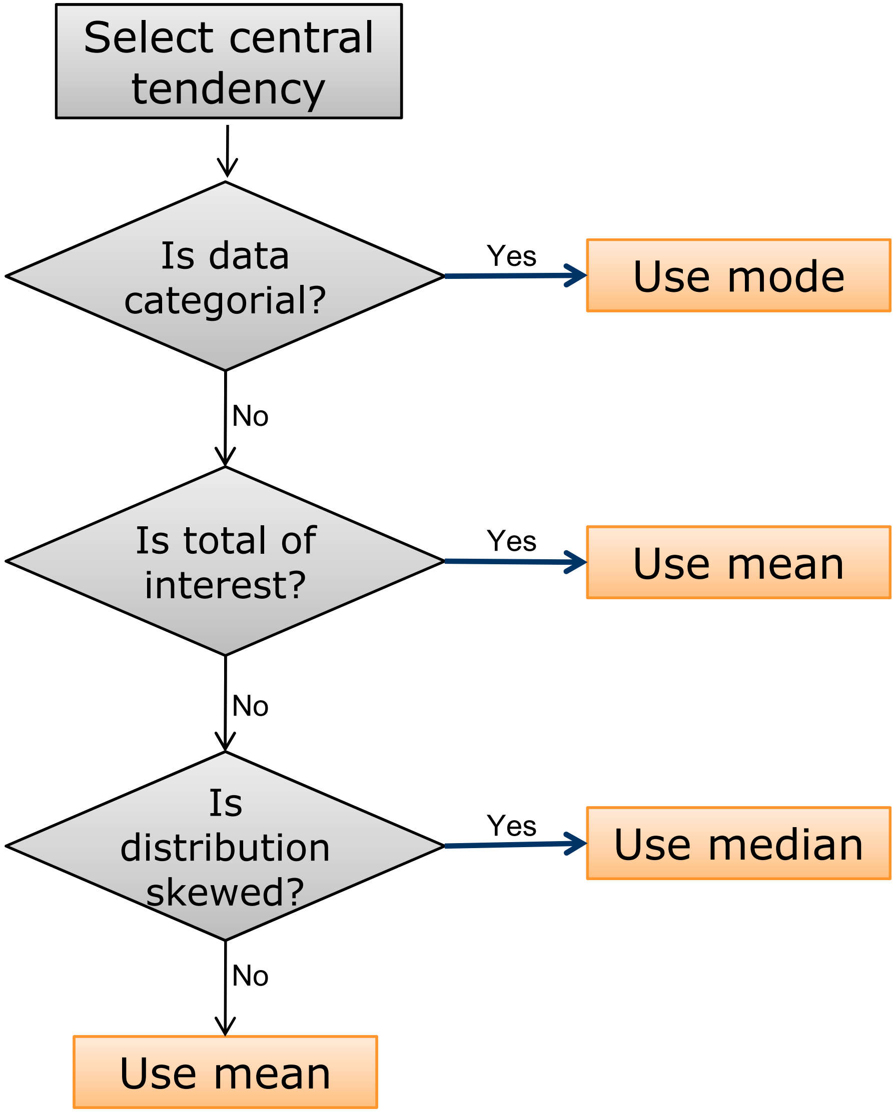
Mean, median, and mode

- Three different indices for the central tendency of a distribution:
  - Mean: $$E(X) = \mu = \sum_{i=1}^{n} p_i x_i = \int_{-\infty}^{\infty} x \cdot f(x) dx$$
  - Median: The 0.5-quantile, i.e., the x_i for that half of the values are smaller and the other half is larger.
  - Mode: The most likely value, i.e., the x_i that has the highest probabiliy p_i or the x at which the PDF is maximum.

Mean, median, and mode



Selecting among mean, median, and mode



Relationship between simulation and probability theory

Central limit theorem

- Let Z_n be the random variable

$$Z_n = \frac{X(n) - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

- and $F_n(z)$ be the distribution function of Z_n for a sample size of n, i.e., $F_n(z)=P(Z_n \le z)$, then

$$F_n(z) \xrightarrow[n \to \infty]{} \Theta(z)$$

- where $\Theta(z)$ is normal distribution with $\mu=0$ and $\sigma^2=1$

$$\Theta(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-y^2}{2}} dy \quad \text{for} -\infty < z < \infty$$

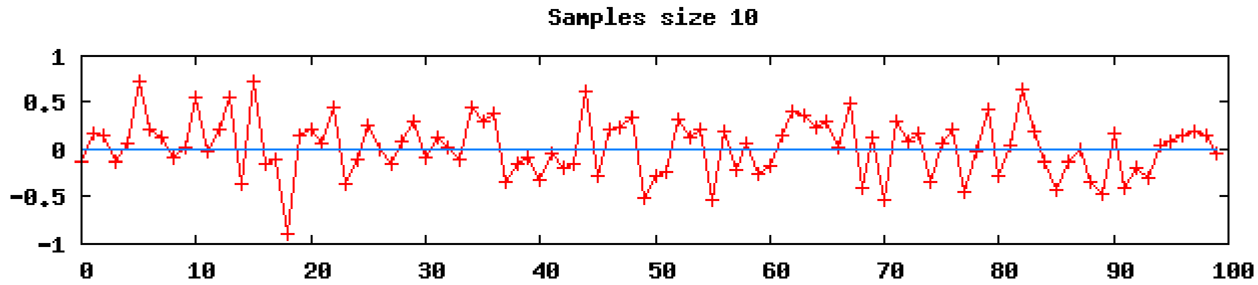
Strong law of large numbers

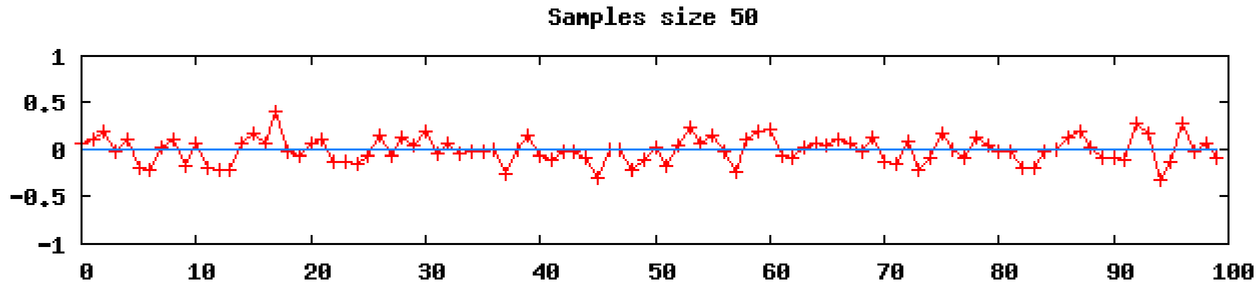
- Let $X_1, X_2, \ldots, X_n$ be IID random variables with mean μ.

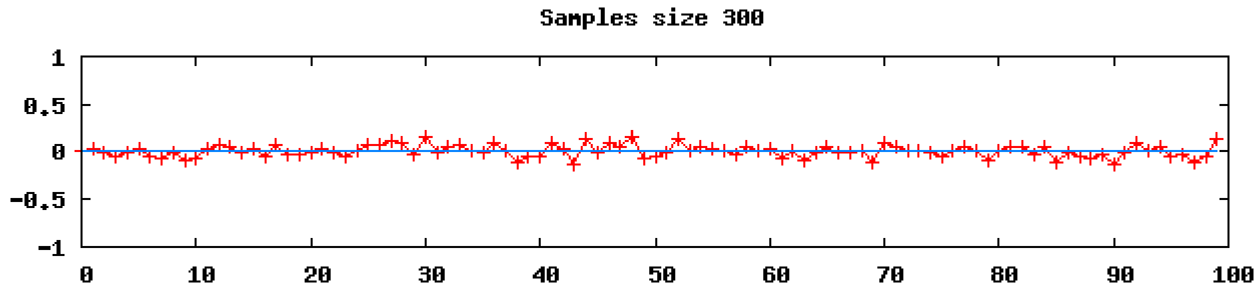
$$\overline{X}(n) \xrightarrow[n \to \infty]{} \mu \quad \text{with probability } 1$$

Sample mean

Strong law of large numbers







Discrete Distributions

Discrete Distributions

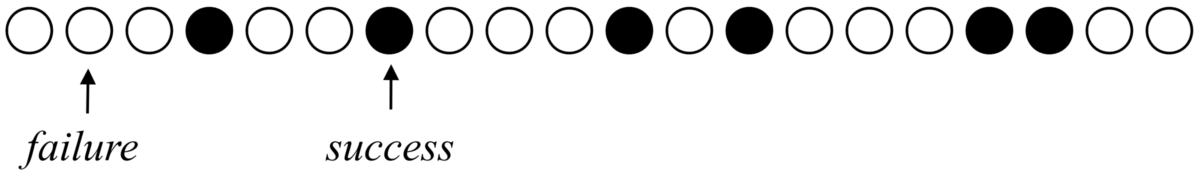
- Discrete random variables are used to describe random phenomena in which only integer values can occur.

- 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

- Bernoulli trials:
  - Consider an experiment consisting of n trials, each can be a success or a failure.

$$X_j = \begin{cases} 1 & \text{if the j-th experiment is a success} \\ 0 & \text{if the j-th experiment is a failure} \end{cases}$$



  - The Bernoulli distribution (one trial):

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

  - where $E(X_j) = p$ and $V(X_j) = p(1\text{-}p) = pq$

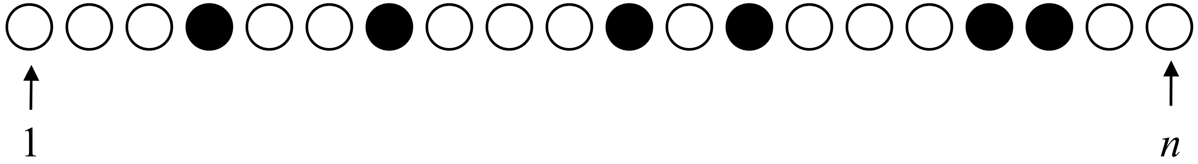
Bernoulli Trials and Bernoulli Distribution

- Bernoulli process:
  - n Bernoulli trials where trials 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

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



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

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

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

- The mean, $E(x) = p + p + \ldots + p = n \times p$
- The variance, $V(X) = pq + pq + \ldots + pq = n \times pq$

Geometric Distribution

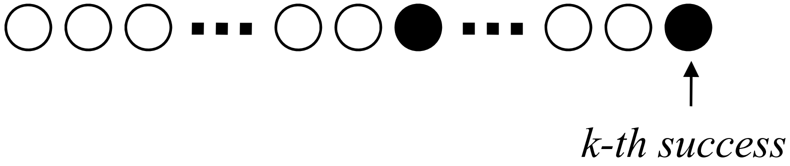
- Geometric distribution
  - The **number** of Bernoulli trials, X, to achieve the 1st success:

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

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

Negative Binomial Distribution

- Negative binomial distribution
  - The number of Bernoulli trials, X, until the k-th success



  - If X is a negative binomial distribution with parameters p and k, then:

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

$$p(x) = \underbrace{\binom{x-1}{k-1} q^{x-k} p^{k-1}}_{(k\text{-}1)\text{ successes}} \cdot \underbrace{p}_{k-\text{th success}}$$

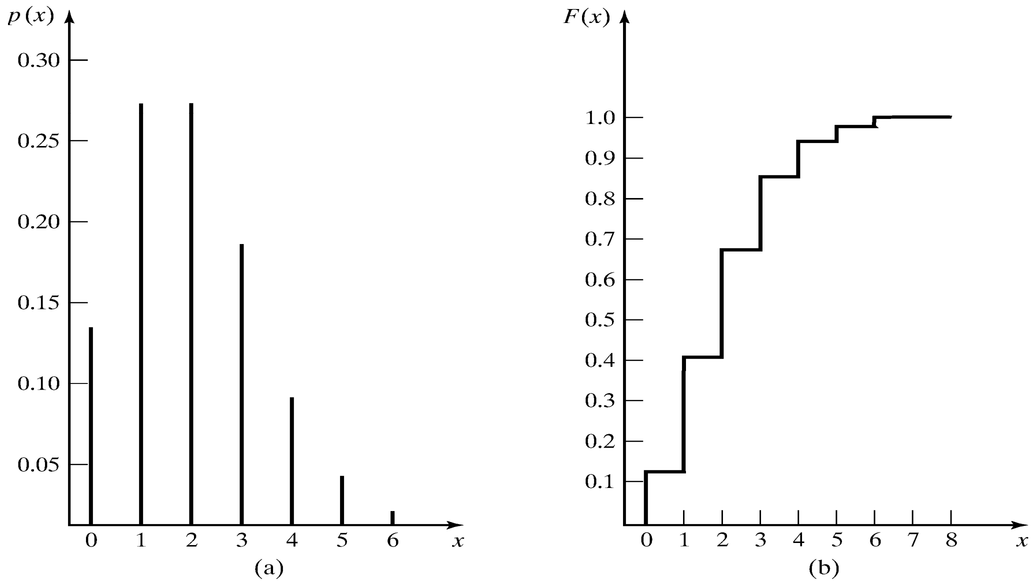
  - $E(X) = k/p$, and $V(X) = kq/p^2$

Poisson Distribution

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

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

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



Poisson Distribution

- Example: A computer repair person is "beeped" each time there is a call for service. The number of beeps per hour ~ Poisson($\alpha = 2$ per hour).

  - The probability of three beeps in the next hour:

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

also, $$p(3) = F(3) - F(2) = 0.857\text{-}0.677\text{=}0.18$$

  - The probability of two or more beeps in an 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 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
  - Weibull
  - Normal
  - Lognormal

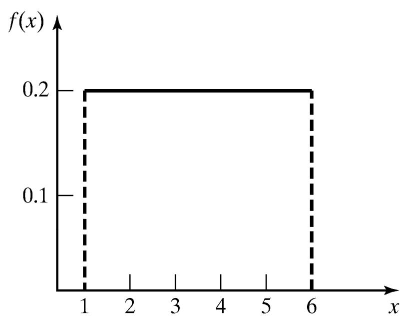
Uniform Distribution

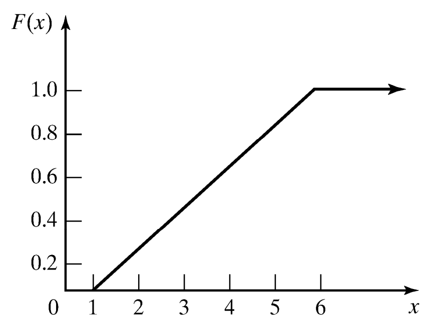
- A random variable X is uniformly distributed on the interval (a, b), $U(a, b)$, if its PDF and CDF are:

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

- 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.



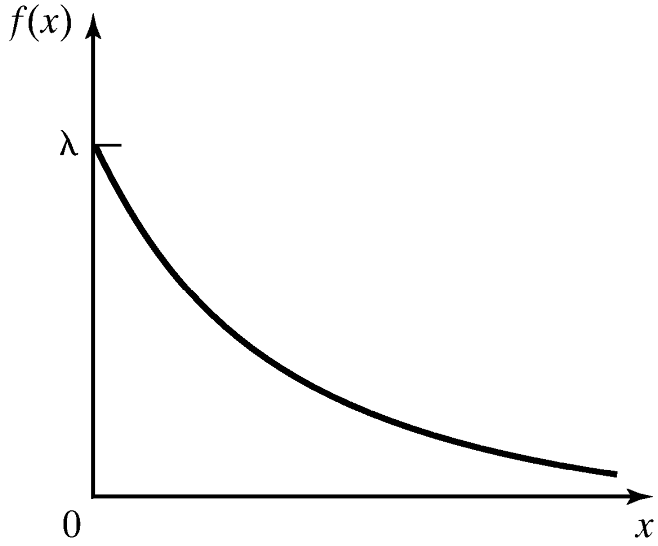


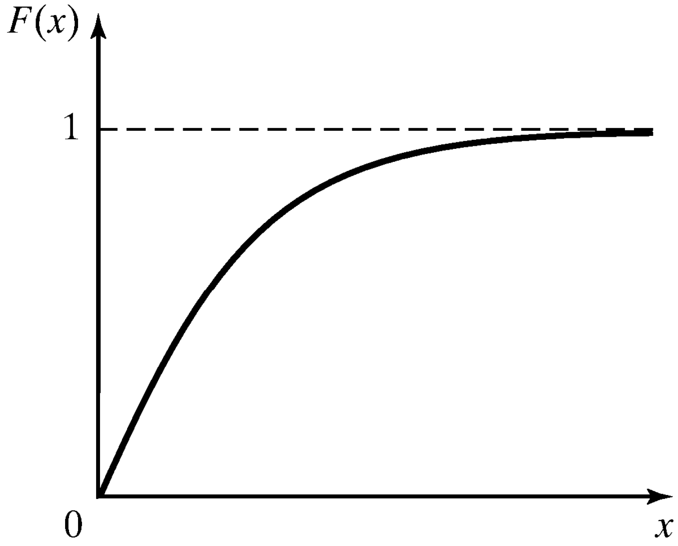
Exponential Distribution

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

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

  - $E(X) = 1/\lambda$ $\quad$ $V(X) = 1/\lambda 2$

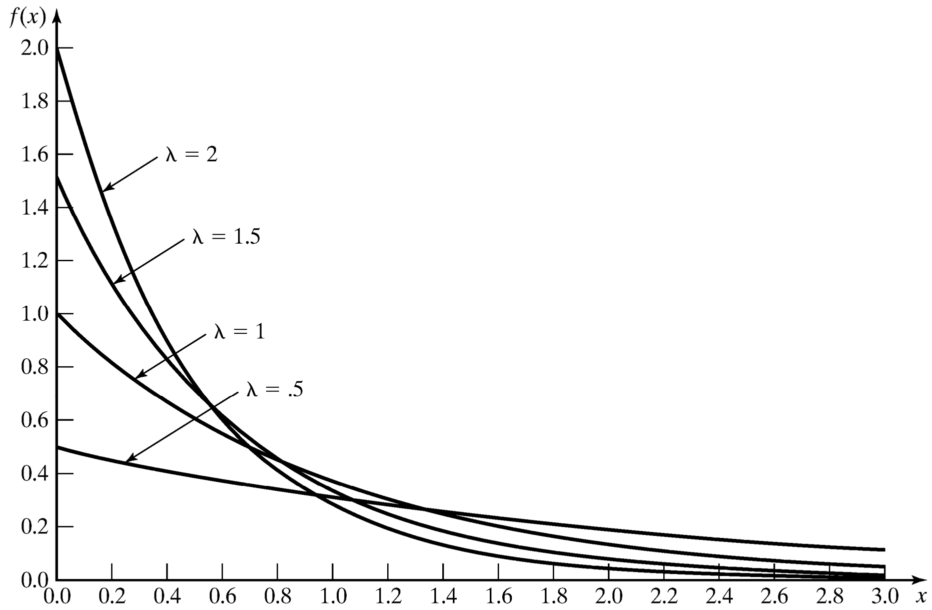




Exponential Distribution

- 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

- 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 $\sim\exp(\lambda = 1/3 \text{ per hour})$, hence, on average, 1 failure per 3 hours.
    - The probability that the lamp lasts longer than its mean life is: $P(X > 3) = 1 - P(X < 3) = 1 - (1 - e^{-3/3}) = e^{-1} = 0.368$
    - The probability that the lamp lasts between 2 to 3 hours is:
      $P(2 \leq X \leq 3) = F(3) - F(2) = 0.145$
    - The probability that it lasts for another hour given it is operating for 2.5 hours:
      $P(X > 3.5 \mid X > 2.5) = P(X > 1) = e^{-1/3} = 0.717$

Exponential Distribution

- Memoryless property

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

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}$$

$$= e^{-\lambda t}$$

$$= P(X > t)$$

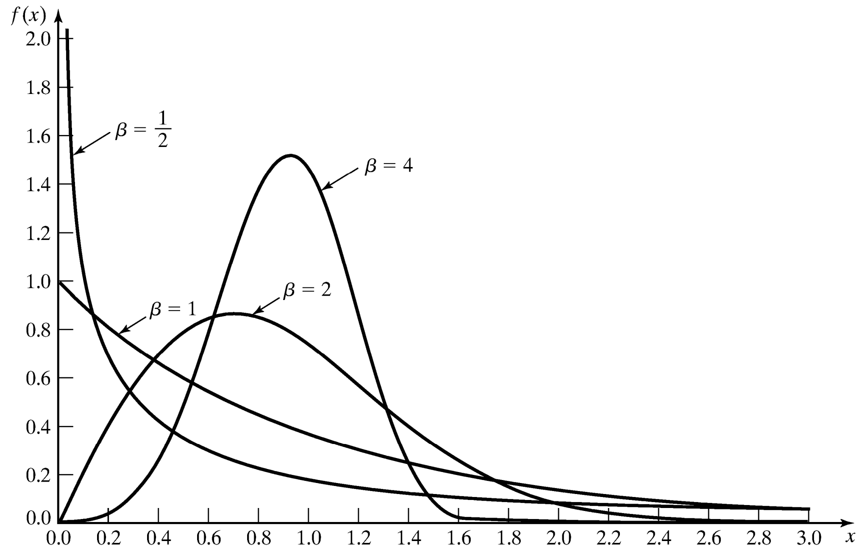
Weibull Distribution

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

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

- 3 parameters:
  - Location parameter: υ, $(-\infty<\nu<\infty)$
  - Scale parameter: β, $(\beta>0)$
  - Shape parameter: α, (>0)

- Example: $\upsilon=0$ and $\alpha=1$:



Weibull Distribution

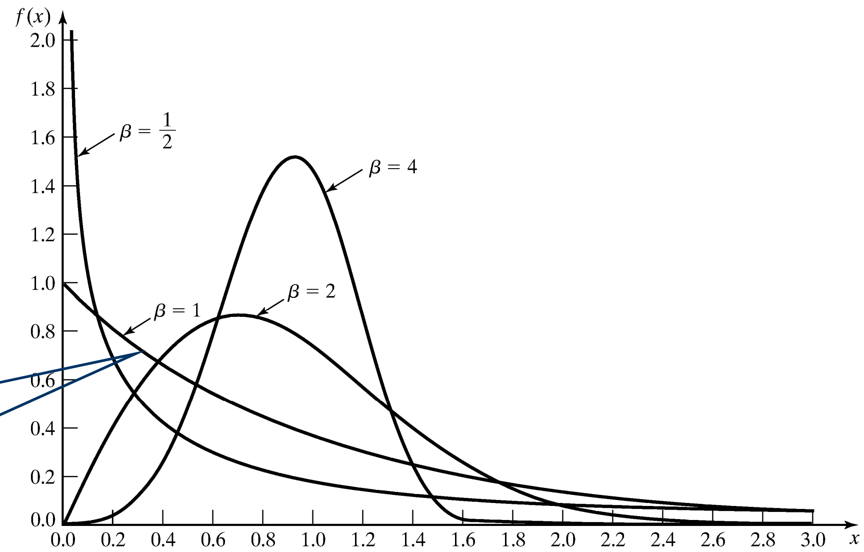
- Weibull Distribution

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

- For $\beta = 1$, $\upsilon=0$

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

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

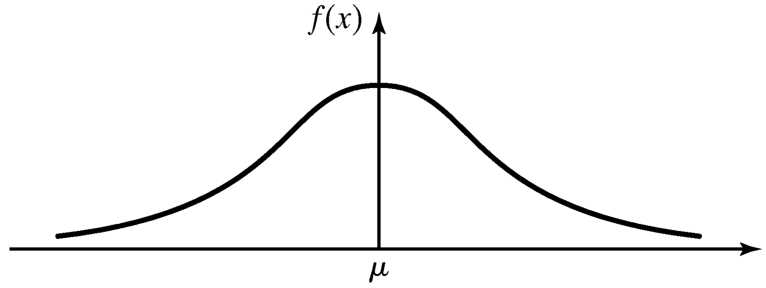


Normal Distribution

- A random variable X is normally distributed if it has the PDF:

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

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



- 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

- 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 $Z = \frac{X - \mu}{\sigma}$,

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

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

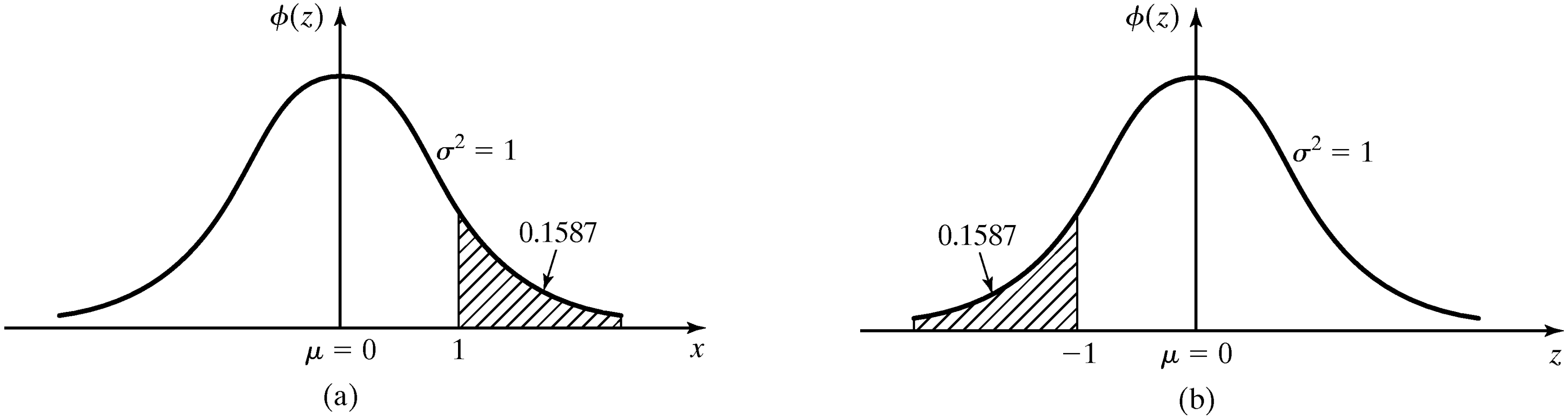
$$= \int_{-\infty}^{(x-\mu)/\sigma} \phi(z) dz = \Phi\left(\tfrac{x-\mu}{\sigma}\right) \quad \text{, where } \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Normal Distribution

- Example: The time required to load an oceangoing vessel, X, is distributed as $N(12,4)$, μ=12, σ=2
  - The probability that the vessel is loaded in less than 10 hours:

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

  - Using the symmetry property, $\Phi(1)$ is the complement of $\Phi(-1)$, i.e., $\Phi(-x) = 1-\Phi(x)$



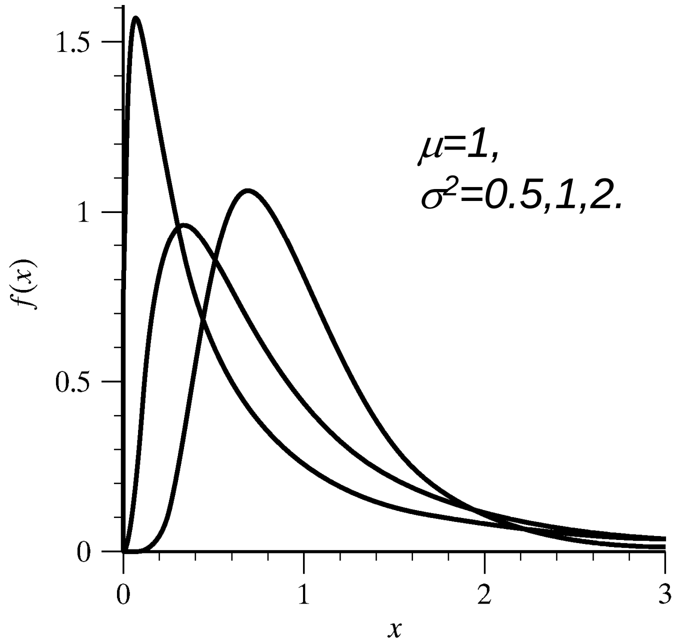
Normal Distribution

- Why is the normal distribution important?
  - The most commonly used distribution in data analysis
  - The sum of n independent normal variates is a normal variate.
  - The sum of a large number of independent observations from any distribution has a normal distribution.

Lognormal Distribution

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

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



  - 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 $X = e^Y \sim \text{lognormal}(\mu, \sigma^2)$
  - Parameters μ and σ^2 are not the mean and variance of the lognormal random variable X

Poisson Process

Poisson Process

- Definition: $N(t)$ is a counting function that represents the number of events occurred in $[0,t]$.

- A counting process $\{N(t), t \geq 0\}$ is a Poisson process with mean rate λ if:
  - Arrivals occur one at a time
  - $\{N(t), t \geq 0\}$ has stationary increments
    - Number of arrivals in $[t, t+s]$ depends only on s, not on starting point t
    - Arrivals are completely random

  - $\{N(t), t \geq 0\}$ has independent increments
    - Number of arrivals during non-overlapping time intervals are independent
    - Future arrivals occur completely random

Poisson Process

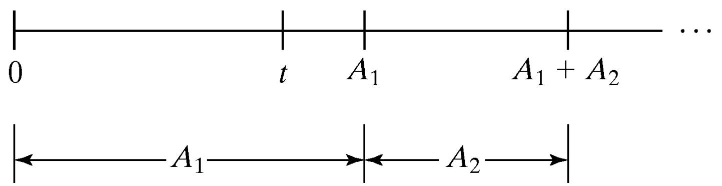
- Properties

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

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

Poisson Process: Interarrival Times

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

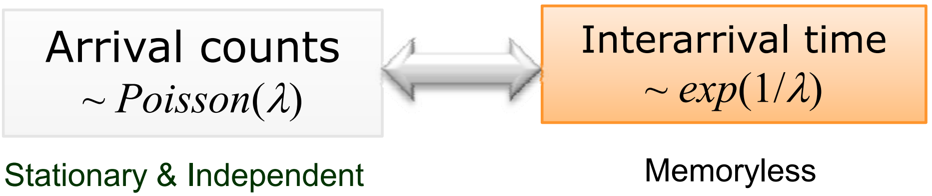


- The 1st arrival occurs after time t iff there are no arrivals in the interval $[0, t]$, hence:

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

$$P(A_1 \leq t) = 1 - P(A_1 > t) = 1 - e^{-\lambda t}$$ [CDF of exp(λ)]

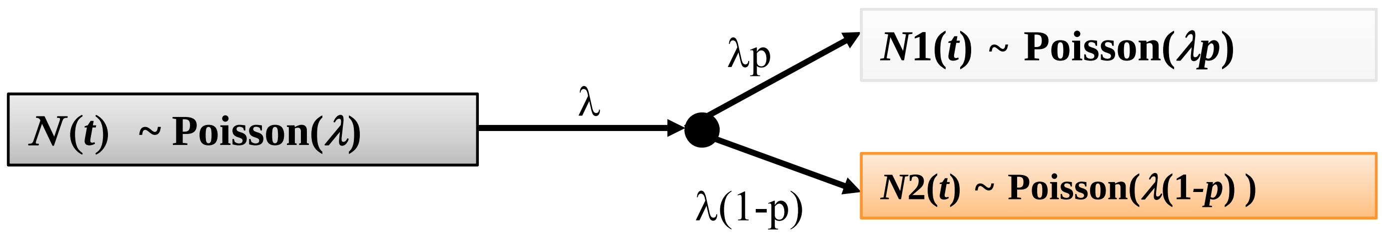
- Interarrival times, $A_1, A_2, \ldots$, are exponentially distributed and independent with mean $1/\lambda$



Poisson Process: Splitting and Pooling

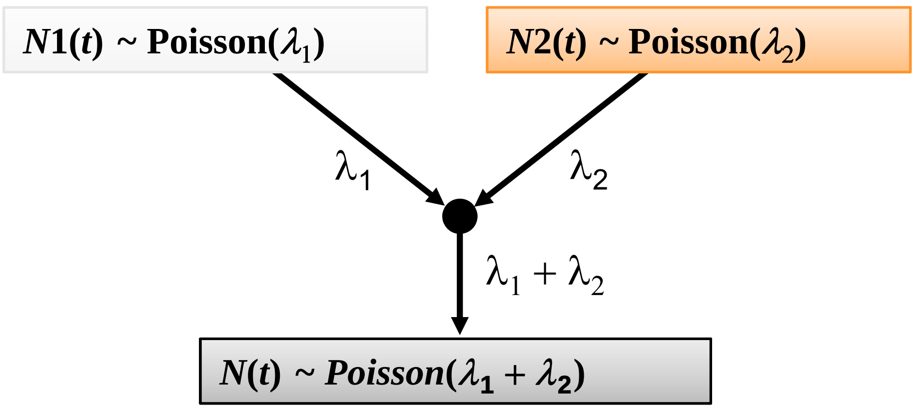
- Splitting:
  - Suppose each event of a Poisson process can be classified as Type I, with probability p and Type II, with probability $1\text{-}p$.

  - $N(t) = N1(t) + N2(t)$, where $N1(t)$ and $N2(t)$ are both Poisson processes with rates λp and $\lambda(1\text{-}p)$



Poisson Process: Splitting and Pooling

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



$$P(N_1 + N_2 = n) = \sum_{j=0}^{n} P(N_1 = j)P(N_2 = n - j)$$

$$= \sum_{j=0}^{n} \frac{(\lambda_1 t)^j}{j!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^{n-j}}{(n-j)!} e^{-\lambda_2 t}$$

$$= e^{-\lambda_1 t} e^{-\lambda_2 t} \sum_{j=0}^{n} \frac{(\lambda_1 t)^j}{j!} \frac{(\lambda_2 t)^{n-j}}{(n-j)!}$$

$$= e^{-(\lambda_1 + \lambda_2)t} t^n \sum_{j=0}^{n} \frac{\lambda_1^j}{j!} \frac{\lambda_2^{n-j}}{(n-j)!}$$

$$= e^{-(\lambda_1 + \lambda_2)t} \frac{t^n}{n!} \sum_{j=0}^{n} n! \frac{\lambda_1^j}{j!} \frac{\lambda_2^{n-j}}{(n-j)!}$$

$$= e^{-(\lambda_1 + \lambda_2)t} \frac{t^n}{n!} \sum_{j=0}^{n} \binom{n}{j} \lambda_1^j \lambda_2^{n-j}$$

$$= e^{-(\lambda_1 + \lambda_2)t} \frac{t^n}{n!} (\lambda_1 + \lambda_2)^n$$

Empirical Distributions

Empirical Distributions

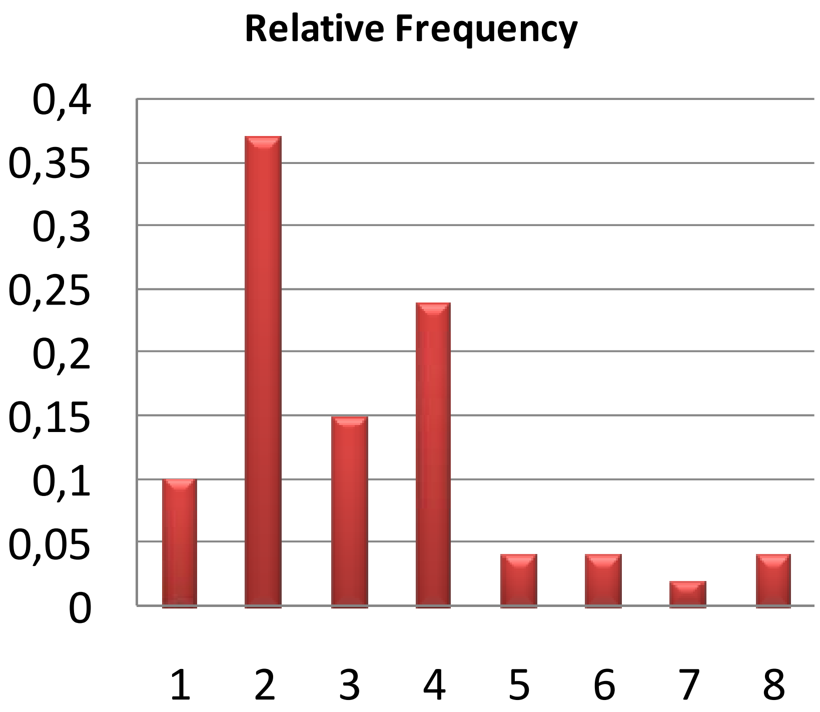
- A distribution whose parameters are the observed values in a sample of data.
  - May be used when it is impossible or unnecessary to establish that a random variable has any particular parametric distribution.
  - Advantage: no assumption beyond the observed values in the sample.
  - Disadvantage: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

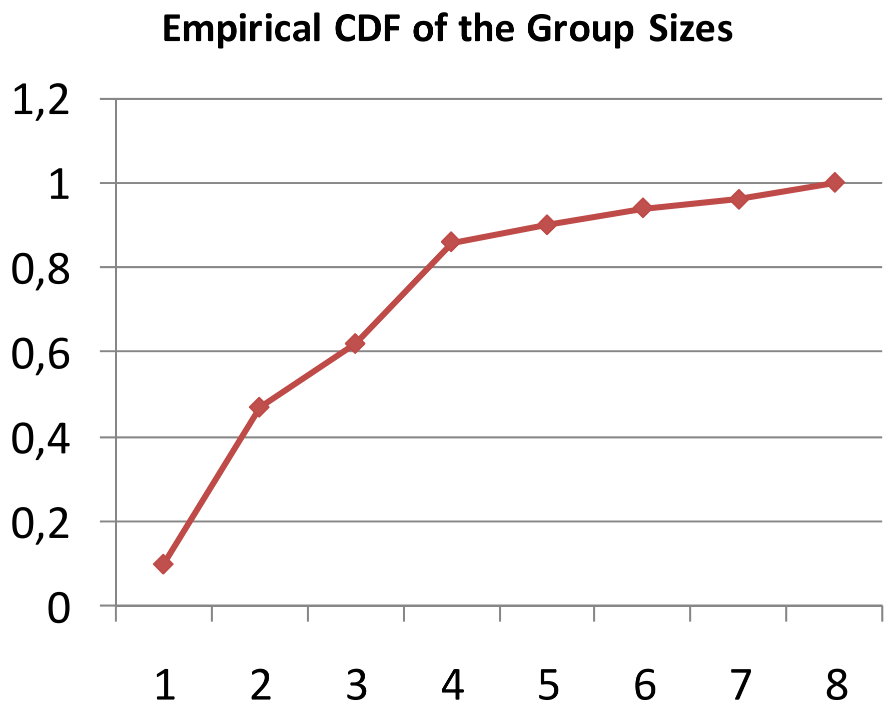
Empirical Distributions: Example

- Customers arrive in groups from 1 to 8 persons
- Observation of the last 300 groups has been reported
- Summary in the table below

Group Size	Frequency	Relative Frequency	Cumulative Relative Frequency
1	30	0.10	0.10
2	110	0.37	0.47
3	45	0.15	0.62
4	71	0.24	0.86
5	12	0.04	0.90
6	13	0.04	0.94
7	7	0.02	0.96
8	12	0.04	1.00

Empirical Distributions: Example





Useful Statistical Models

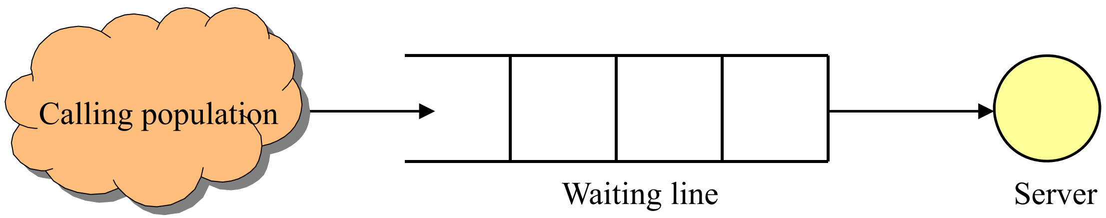
Useful Statistical Models

- 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

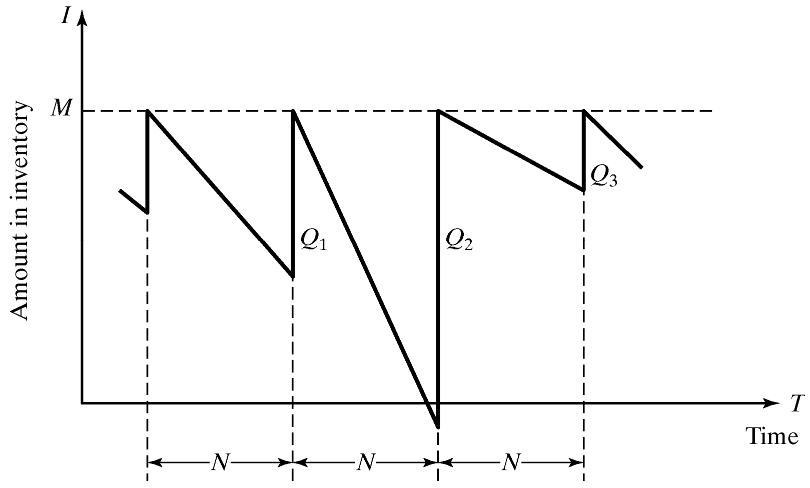
Useful models: Queueing Systems

- In a queueing system, interarrival and service-time patterns can be probabilistic.
- 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 values.
  - Gamma and Weibull distributions: more general than exponential (involving location of the modes of PDF’s and the shapes of tails.)



Useful models: Inventory and supply chain

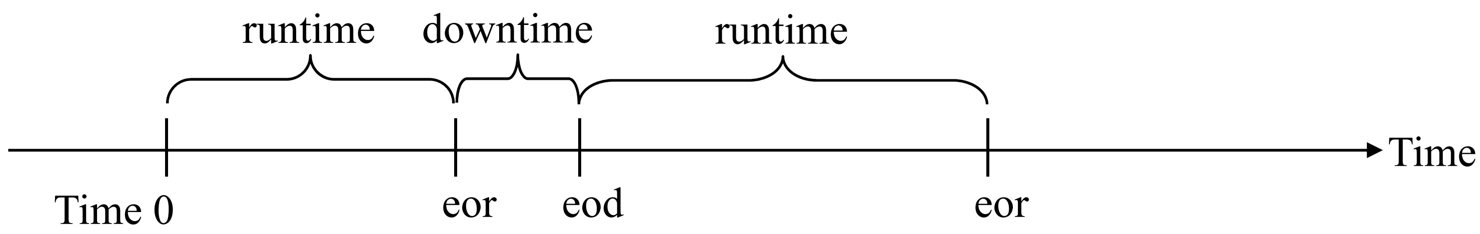
- In realistic inventory and supply-chain systems, there are at least three random variables:
  - The number of units demanded per order or per time period
  - The time between demands
  - The lead time = Time between placing an order and the receipt of that order



- 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.

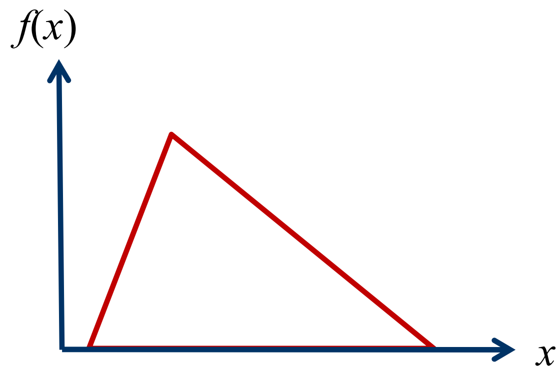
Useful models: Reliability and maintainability

- 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



Useful models: Other areas

- For cases with limited data, some useful distributions are:
  - Uniform
  - Triangular
  - Beta



- Other distribution:
  - Bernoulli
  - Binomial
  - Hyperexponential

Summary

- The world that the simulation analyst sees is probabilistic, not deterministic.
- In this chapter:
  - 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.
- Student should know:
  - Difference between discrete, continuous, and empirical distributions.
  - Poisson process and its properties.