Fair Use Notice

- The material used in this presentation i.e., pictures/graphs/text, etc. is solely intended for educational/teaching purpose, offered free of cost to the students for use under special circumstances of Online Education due to COVID-19 Lockdown situation and may include copyrighted material - the use of which may not have been specifically authorized by Copyright Owners. It's application constitutes Fair Use of any such copyrighted material as provided in globally accepted law of many countries. The contents of presentations are intended only for the attendees of the class being conducted by the presenter.

Chapter 5
Statistical Models in Simulation

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

- In this chapter:
  - 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

- 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) = $\{0,1,2,\dots\}$
  - $p(x_i)$ = probability the random variable is $x_i = P(X = x_i)$
  - $p(x_i)$, $i = 1,2, \dots$ must satisfy:

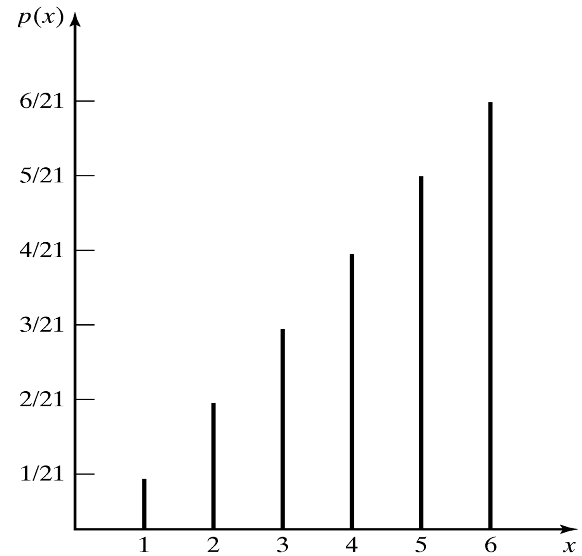
    1. $p(x_i) \geq 0, \text{ for all } i$
    2. $\sum_{i=1}^{\infty} p(x_i) = 1$

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

Discrete Random Variables

- Example: Assume the die is loaded so that the probability that a given face lands up is proportional to the number of spots showing.

x_i	1	2	3	4	5	6
$P(x_i)$	1/21	2/21	3/21	4/21	5/21	6/21



  - $p(x_i)$, $i = 1,2, \ldots$ must satisfy:

    1. $p(x_i) \geq 0$, for all i
    2. $\sum_{i=1}^{\infty} p(x_i) = 1$

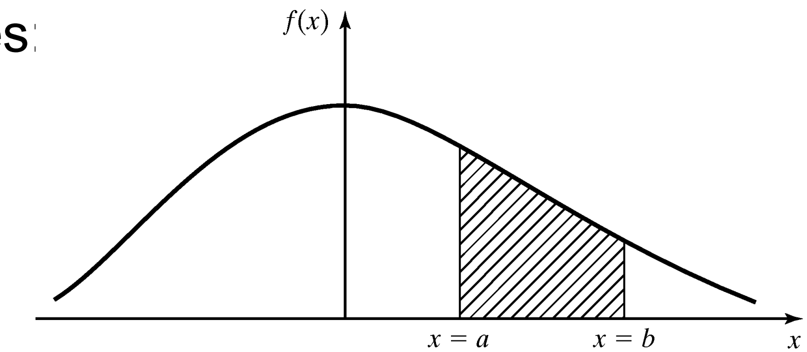
Continuous Random Variables [Probability Review]

- 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

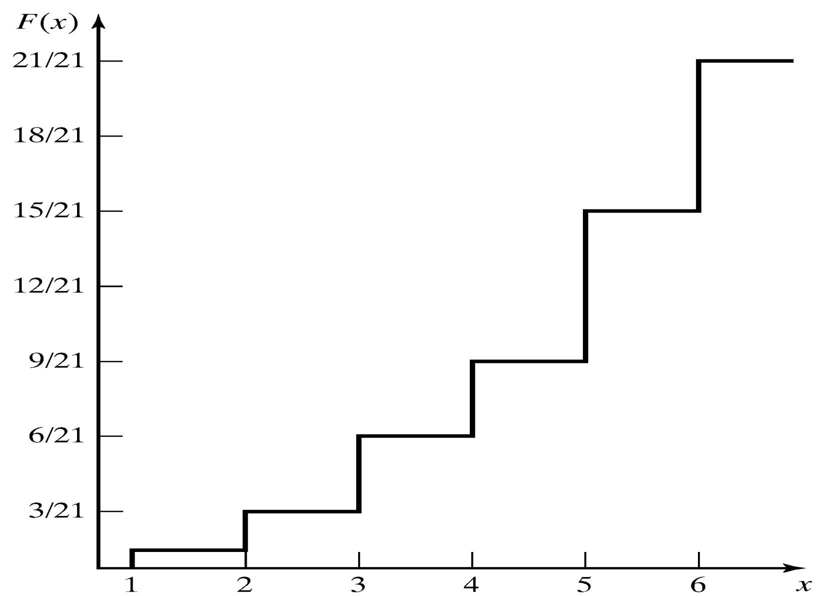
1. $P(X = x_0) = 0$, because $\int_{x_0}^{x_0} f(x)dx = 0$
2. $P(a \le X \le b) = P(a \prec X \le b) = P(a \le X \prec b) = P(a \prec X \prec b)$

Continuous Random Variables

- Example: The die-tossing experiment described in last example has a cdf given as follows:

x	(-∞,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,∞)
F(x)	0	1/21	3/21	6/21	10/21	15/21	21/21

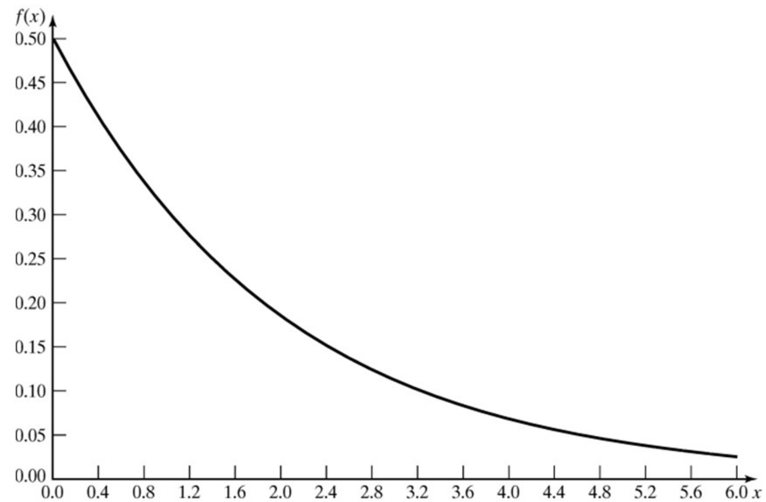
  - [a,b) = {a ≤ x ≤ b}



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

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

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:

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

- 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.

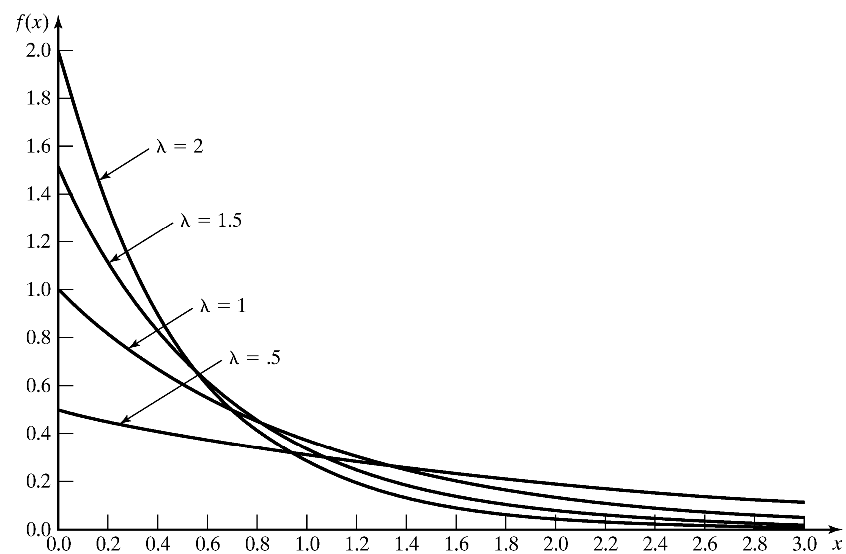
Exponential Distribution [Continuous Dist'n]

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

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

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

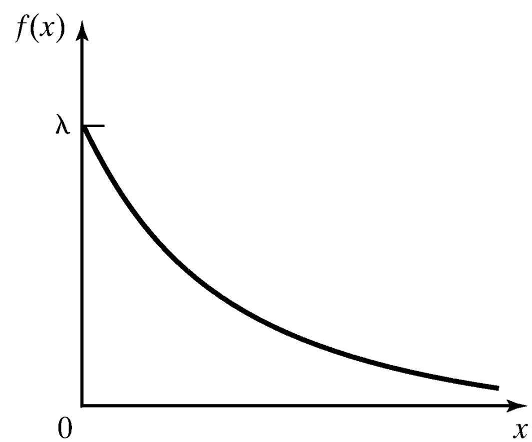
  - $E(X) = 1/\lambda$ $V(X) = 1/\lambda^2$
  - Used to model interarrival times when arrivals are completely random, and to model service times that are highly variable
  - For several different exponential pdf's (see figure), the value of intercept on the vertical axis is λ, and all pdf's eventually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

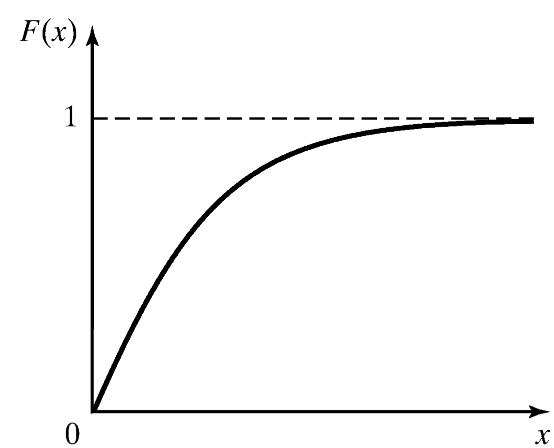


Exponential Distribution [Probability Review]

- Model times between events
  - Times between arrivals
  - Times between failures
  - Times to repair
  - Service Times
- A random variable X is said to be exponentially distributed with parameter if its PDF is given by

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





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

Gamma Distribution [Probability Review]

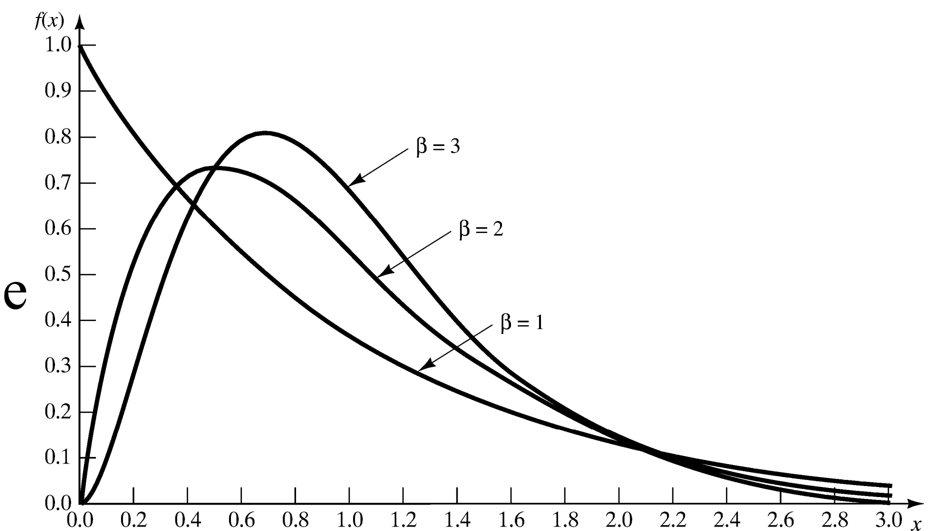
- A function used in defining the gamma distribution is the gamma function, which is defined for all $\beta > 0$ as

$$\Gamma(\beta) = \int_0^{\infty} x^{\beta-1} e^{-x} dx$$

- A random variable X is gamma distributed with parameters β and θ if its PDF is given by

$$f(x) = \begin{cases} \dfrac{\beta\theta}{\Gamma(\beta)} (\beta\theta x)^{\beta\text{-}1} \mathrm{e}^{\text{-}\beta\theta x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$E(X) = \frac{1}{\theta} \qquad V(X) = \frac{1}{\beta\theta^2}$$

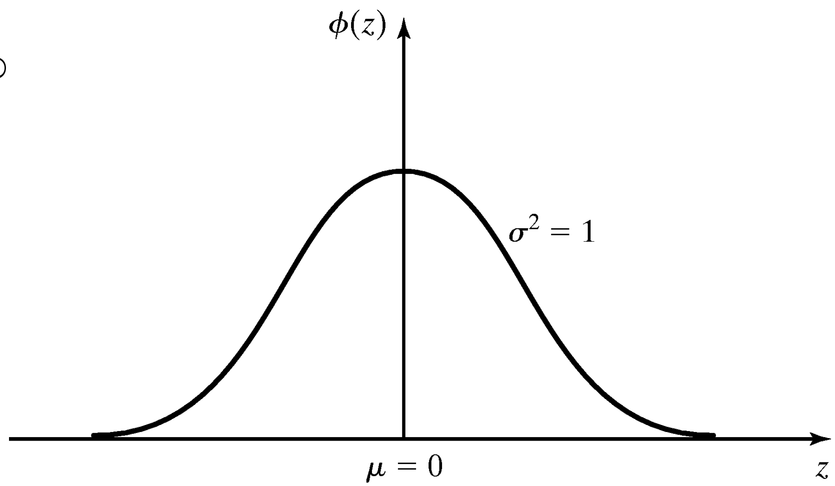


Normal Distribution

[Continuous Dist'n]

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

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



  - Mean: $-\infty < \mu < \infty$
  - Variance: $\sigma^2 > 0$
  - 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 ~ N(0,1)
  - Transformation of variables: let Z = (X - μ) / σ,

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

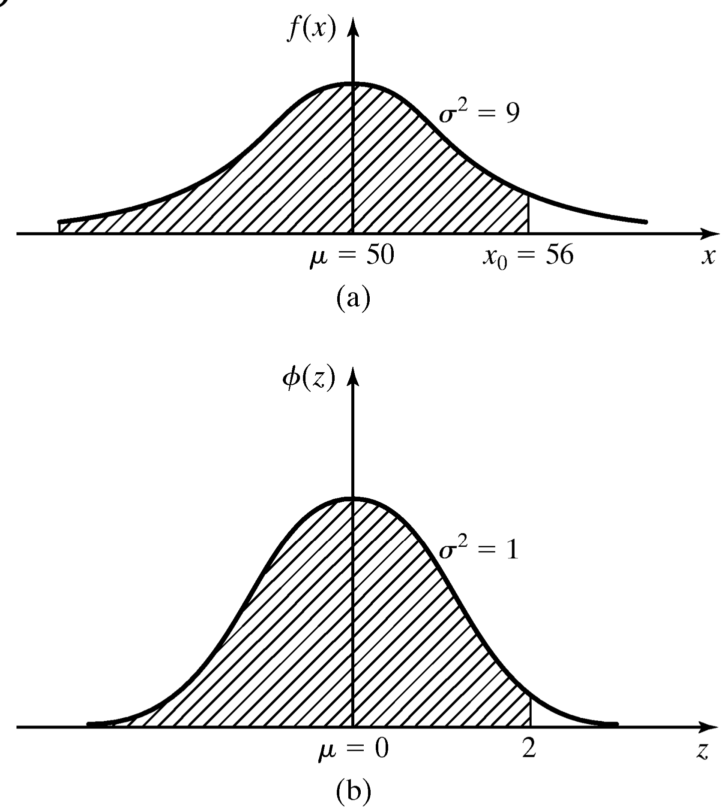
$$= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

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

Normal Distribution [Probability Review]

- Example: Suppose that *X ~ N (50, 9).*

$$F(56) = \Phi\left(\frac{56-50}{3}\right) = \Phi(2) = 0.9772$$



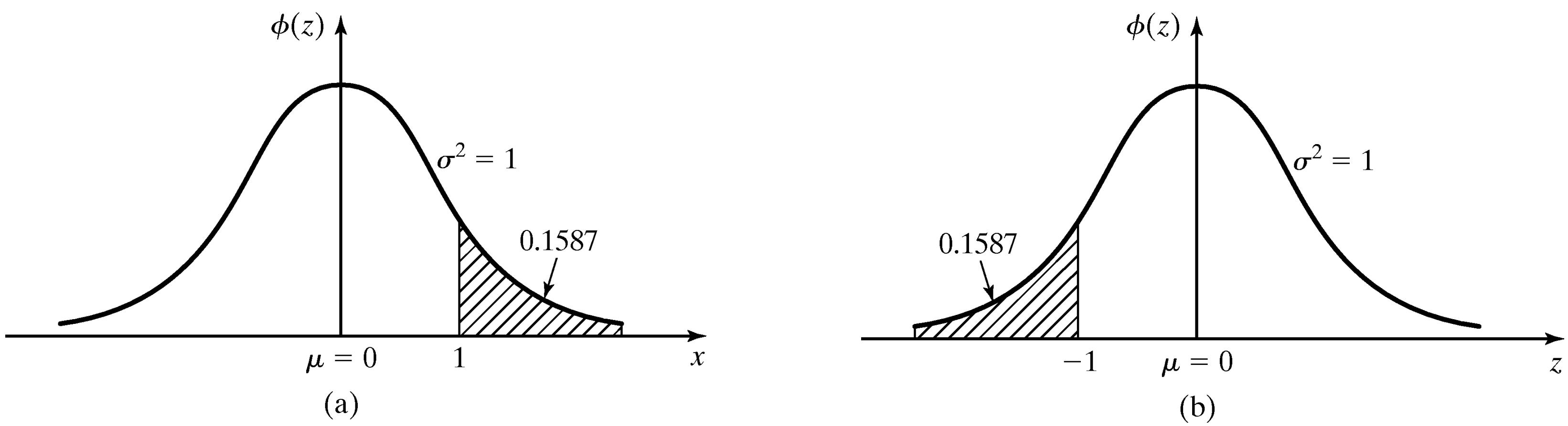
Normal Distribution

[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:

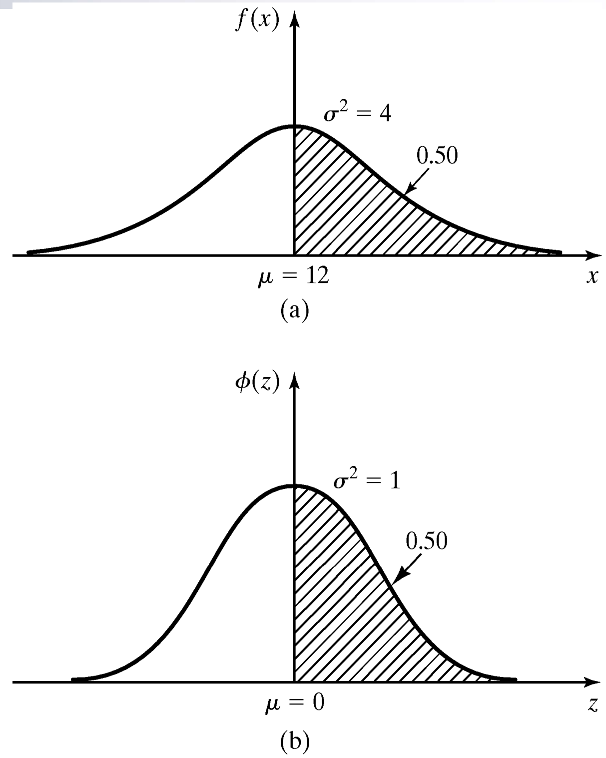
$$F(10) = \Phi\left(\frac{10-12}{2}\right) = \Phi(-1) = 0.1587$$

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



Normal Distribution [Probability Review]

- Example: The time in hours required to load a ship, X, is distributed as $N(12, 4)$. The probability that 12 or more hours will be required to load the ship is:



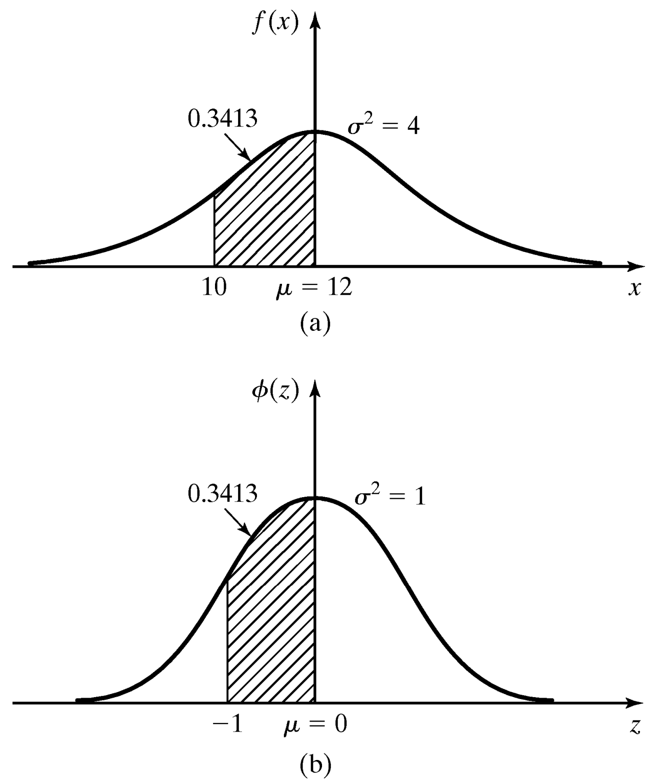
$$P(X > 12) = 1 - F(12) = 1 - 0.50 = 0.50$$

(The shaded portions in both figures)

Normal Distribution [Probability Review]

- Example (cont.):

  The probability that between 10 and 12 hours will be required to load a ship is given by



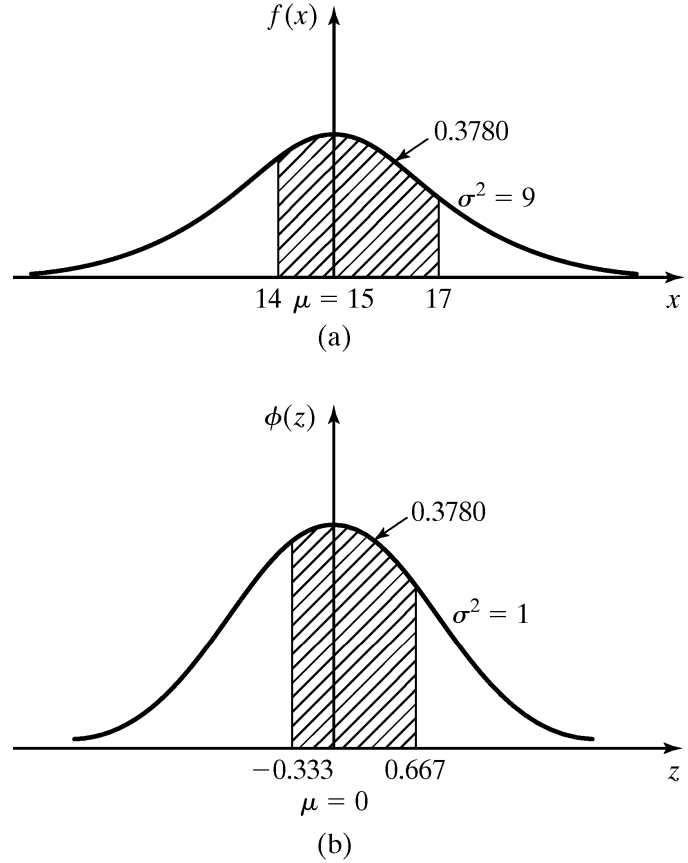
$$P(10 \le X \le 12) = F(12) - F(10) = 0.5000 - 0.1587 = 0.3413$$

The area is shown in shaded portions of the figure

Normal Distribution [Probability Review]

- Example: The time to pass through a queue is *N(15, 9)*. The probability that an arriving customer waits between 14 and 17 minutes is:

Type equation here.



$P(14 \le X \le 17) = F(17) - F(14) =$

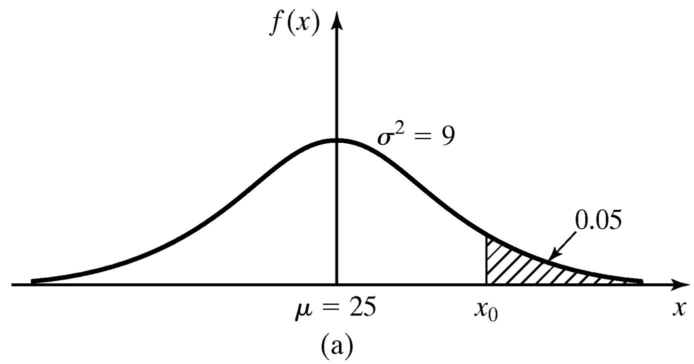
$$\Phi\left(\frac{17-15}{3}\right) - \Phi\left(\frac{14-15}{3}\right) = \Phi(0.667) - \Phi(-0.333) = 0.7476 - 0.3696 = 0.3780$$

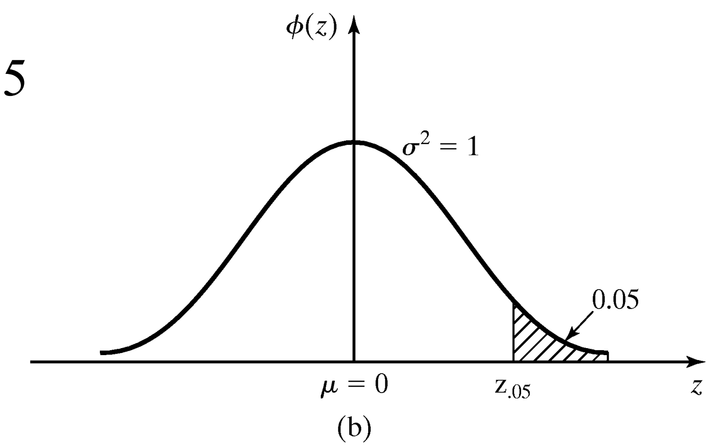
$\varphi(-.333) = 1 - \varphi(0.333) = 1 - 0.6304 = 0.3696$

Normal Distribution [Probability Review]

- Example: Lead-time demand, *X*, for an item is *N(25, 9)*.

  Compute the value for lead-time that will be exceeded only *5%* of time.





$$P(X > x_0) = P(Z > \frac{x_0 - 25}{3}) = 1 - \Phi(\frac{x_0 - 25}{3}) = 0.05$$

$$\Phi(1.645) = 0.95.$$

$$\Phi\left(\frac{x_0 - 25}{3}\right) = 0.95$$

$$\frac{x_0 - 25}{3} = 1.645$$

$$x_0 = 29.935$$

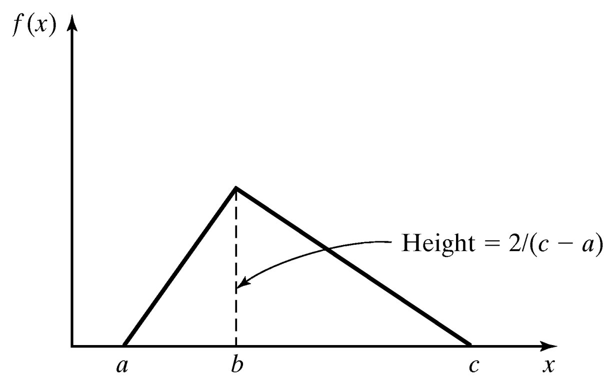
Triangular Distribution [Probability Review]

- A random variable X has a triangular distribution if its PDF is given by

$$f(x) = \begin{cases} \dfrac{2(x-a)}{(b-a)(c-a)}, & a \le x \le b \\[2ex] \dfrac{2(c-x)}{(c-b)(c-a)}, & b < x \le c \\[2ex] 0, & \text{elsewhere} \end{cases}$$

where $a \le b \le c$.

$$E(X) = \frac{a+b+c}{3}$$



Triangular Distribution

- A burger franchise planning a new outlet in Auckland wants to determine the probability the new outlet will have weekly sales of less than $2000. If the weekly sales are less than this the outlet is unlikely to cover its costs.
- So, they wish to calculate $\Pr(X < 2000)$ They use a triangular distribution to model the future weekly sales with a minimum value of *a*=$1000, and maximum value of *b*=$6000 and a peak value of *c*=$3000.

Triangular Distribution

Area corresponding to $\Pr(X < 2000)$

The area of a triangle is $area = \frac{1}{2} base \times height$. In this case $base = 2000 - 1000 = 1000$

the height is given by $f(2000)$ using the probability density function formula. Since 2000 is between *a*=1000 and *c*=3000 we have:

$$height = f(2000) = \frac{2(2000 - a)}{(b - a)(c - a)} = \frac{2 \times (2000 - 1000)}{(6000 - 1000) \times (3000 - 1000)} = 0.0002$$

So $area = \frac{1}{2} \times 1000 \times 0.0002 = 0.1$ so $\Pr(X < 2000) = 0.1$ (or 10%)

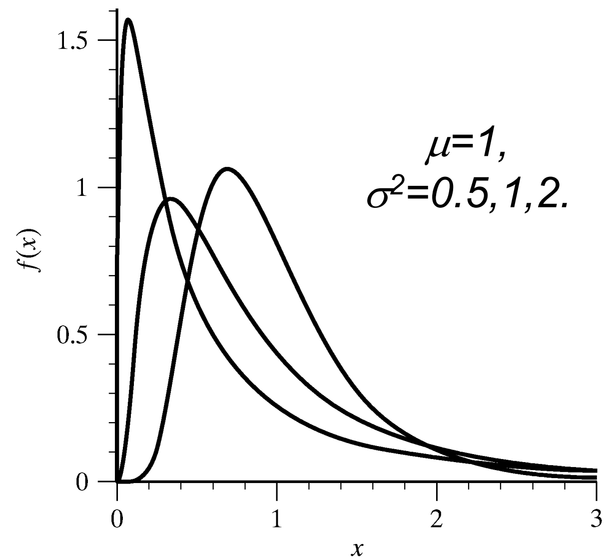
Lognormal Distribution [Continuous Dist'n]

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

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

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

median $= e^{\mu}$ mode $= e^{\mu-\sigma^2}$



- 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

Lognormal Distribution

[Continuous Dist'n]

- If the mean and variance of the lognormal are known to be μ_L and σ_L^2, respectively,
- then the parameters μ and σ^2 are given by

$$\mu = \ln\left(\frac{\mu_L^2}{\sqrt{\mu_L^2 + \sigma_L^2}}\right)$$

$$\sigma^2 = \ln\left(\frac{\mu_L^2 + \sigma_L^2}{\mu_L^2}\right)$$

Lognormal Distribution [Continuous Dist'n]

- The rate of return on a volatile investment is modeled as having a lognormal distribution with mean 20% and standard deviation 5%. Compute the parameters for the lognormal distribution.

$$\mu = \ln\left(\frac{20^2}{\sqrt{20^2 + 5^2}}\right) \doteq 2.9654$$

$$\sigma^2 = \ln\left(\frac{20^2 + 5^2}{20^2}\right) \doteq 0.06$$

Beta Distribution

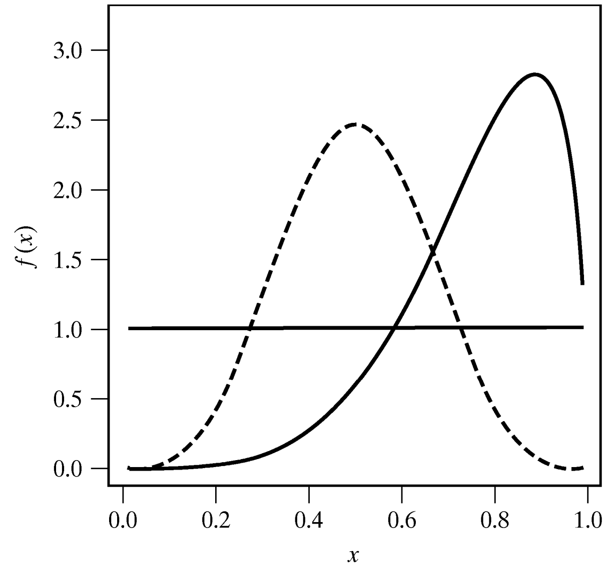
[Probability Review]

- A random variable *X* is beta-distributed with parameters $\beta_1 > 0$ and $\beta_2 > 0$ if its PDF is given by

$$f(x) = \begin{cases} \dfrac{x^{\beta_1 - 1}(1 - x)^{\beta_2 - 1}}{B(\beta_1, \beta_2)}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where

$$B(\beta_1, \beta_2) = {\Gamma(\beta_1)\Gamma(\beta_2)} \big/ {\Gamma(\beta_1 + \beta_2)}$$



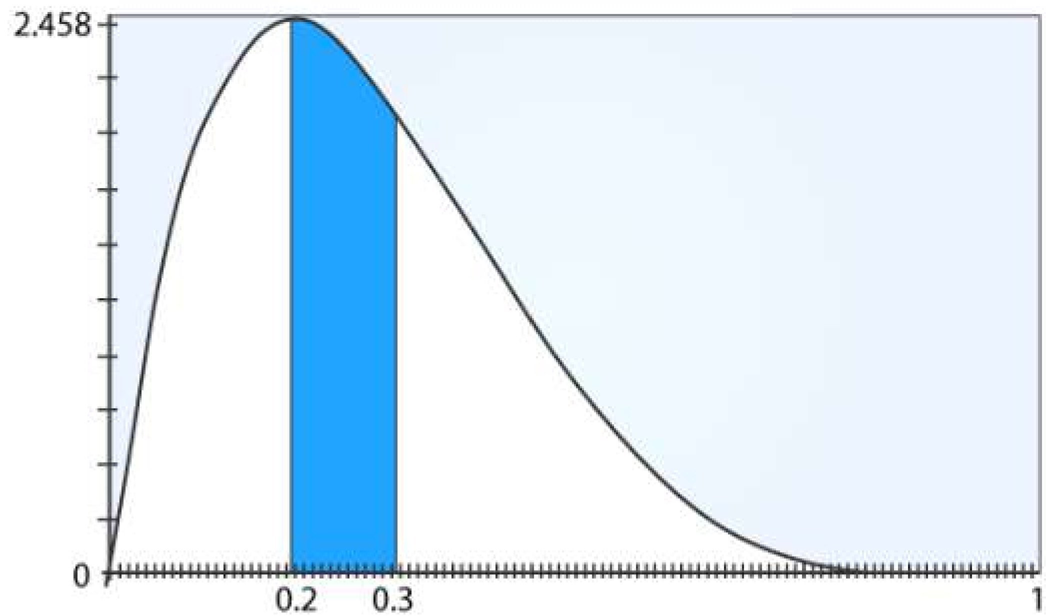
Beta Distribution

Problem: Suppose, if in a basket there are balls which are defective with a Beta distribution of α=2 and β=5. Compute the probability of defective balls in the basket from 20% to 30%.

$$P(x) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha,\beta)$$

$$P(0.2 \leq x \leq 0.3) = \sum_{0.2}^{0.3} x^{2-1}(1-x)^{5-1}/B(2,5)$$

$=0.235185$



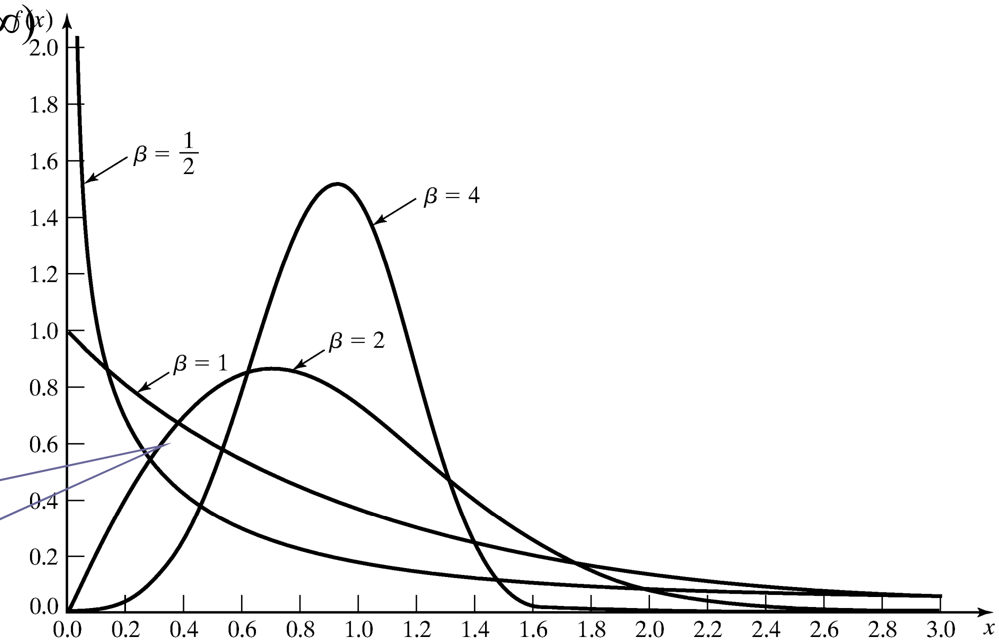
Weibull Distribution

[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, $(-\infty < \nu < \infty)$
  - Scale parameter: β, $(\beta > 0)$
  - Shape parameter. α, (> 0)
- Example: $\upsilon = 0$ and $\alpha = 1$:



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

Weibull Distribution

[Continuous Dist'n]

- The time it takes for an aircraft to land and clear the runway at a major international airport has a Weibull distribution with v = 1.34 minutes, β = 0.5, and α= 0.04 minute. Find the probability that an incoming airplane will take more than 1.5 minutes to land and clear the runway.

$$P(X \le 1.5) = F(1.5)$$

$$= 1 - \exp\left[-\left(\frac{1.5 - 1.34}{0.04}\right)^{0.5}\right]$$

$$= 1 - e^{-2} = 1 - 0.135 = 0.865$$

- Therefore, the probability that an aircraft will require more than 1.5 minutes to land and clear the runway is 0.135.

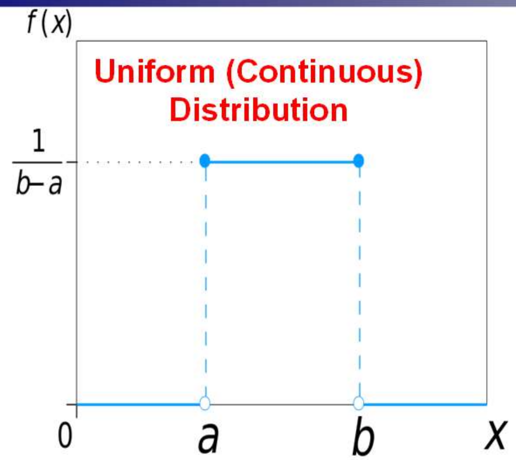
Continuous Distributions

- For events that are highly variable (interarrival times) or instantaneous occurrences (failure of a light bulb) → exponential distribution
- Sum of independent exponential distributions → Gamma distribution. Extremely flexible, used to model non-negative variables
- The sum of *k* independent random variables → normal distribution
- The product of *k* independent random variables → lognormal distribution
- Bounded random variables → beta distribution

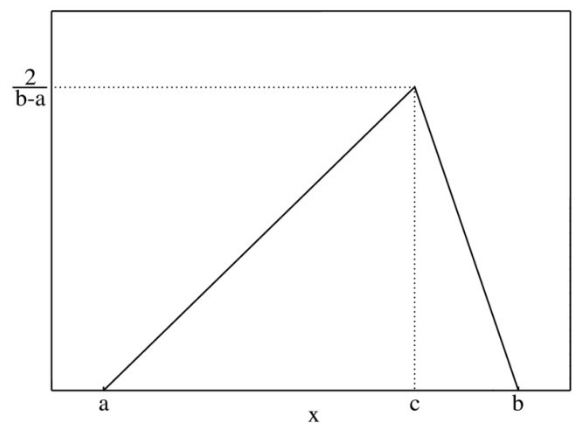
Continuous Distributions

- Weibull distribution can be thought of as a stretched exponential distribution. That's why it has a longer tail.
- Uniform distribution → complete uncertainty
- Triangular distribution → maximum and minimum are known
- When no theoretical distribution seems appropriate → Empirical data

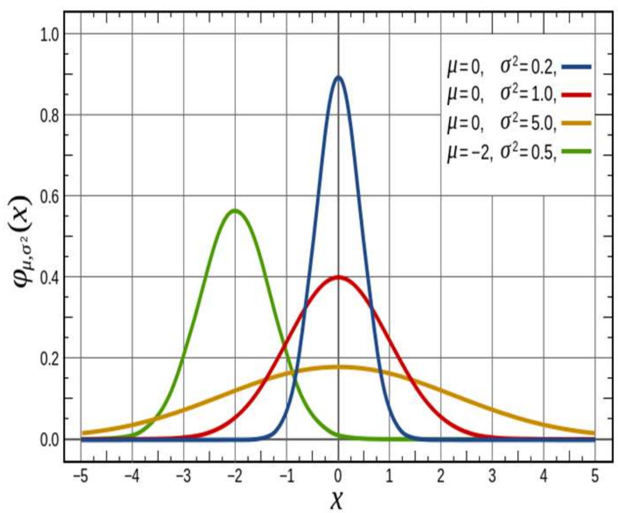
Continuous Probability Distributions



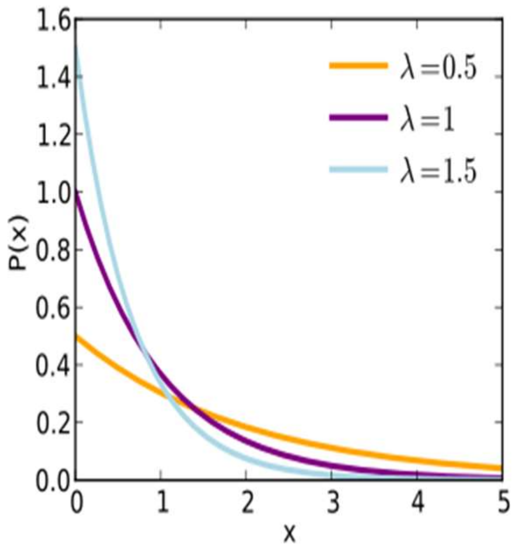
Triangular Distribution



Normal Distribution

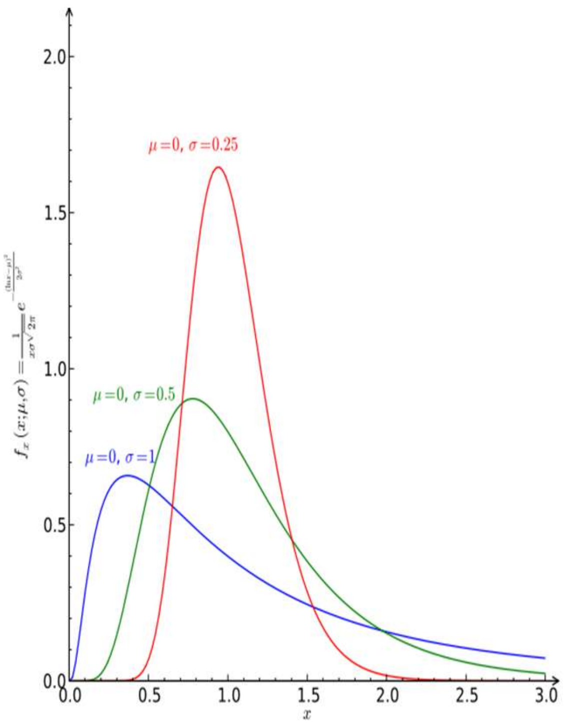


Exponential Distribution

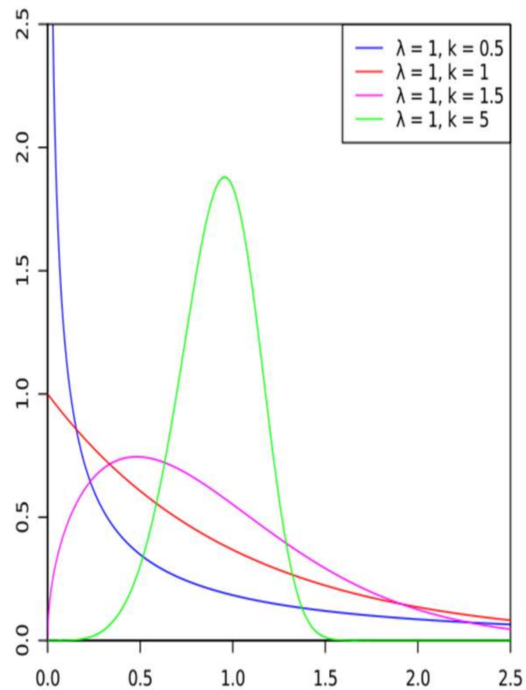


Continuous Probability Distributions

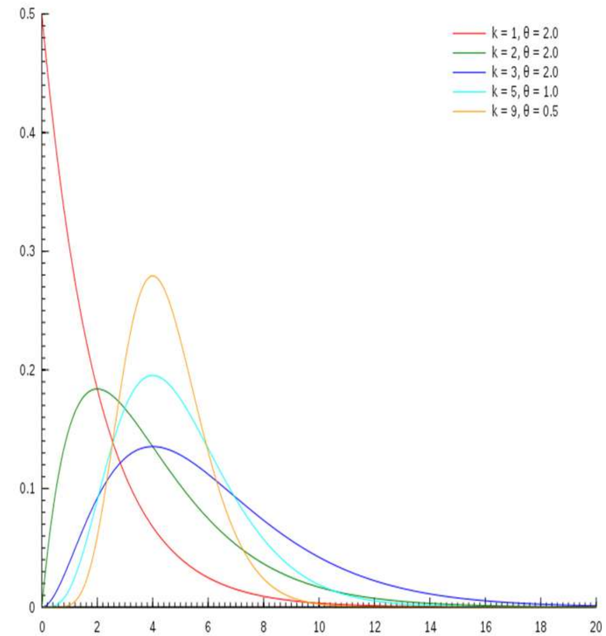
Lognormal Distribution



Weibull Distribution



Gamma Distribution

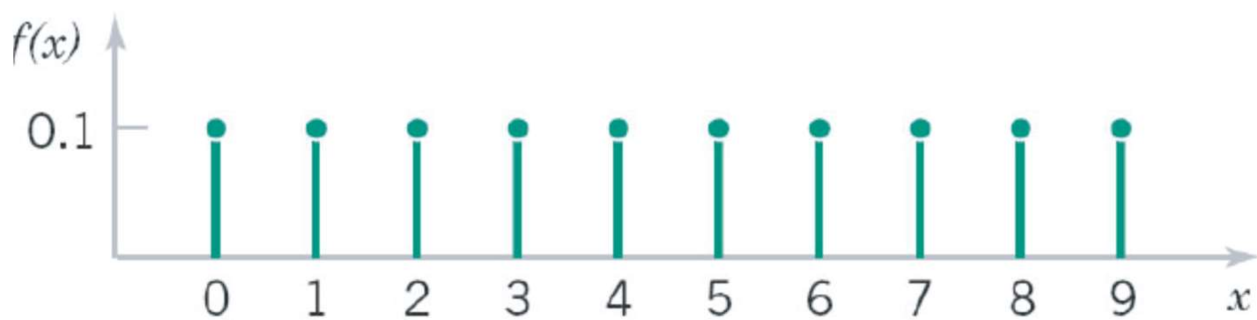


Discrete Distributions

- 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

Uniform Distribution

- A random variable X has a discrete uniform distribution if each of the n values in its range, say $x_1, x_2, \ldots\ldots\ldots\ x_n$, has equal probability. Then, $f(x_i) = 1/n$
- Let X represent a random variable taking on the possible values of {0; 1; 2; 3; 4; 5; 6; 7; 8; 9}, and each possible value has equal probability.
- This is a discrete uniform distribution and the probability for each of the 10 possible value is $P(X = x_i) = f(x_i) = 1/10 = 0.10$



Bernoulli Trials and Bernoulli Distribution

- 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, $E(x) = p + p + \ldots + p = n{*}p$
  - The variance, $V(X) = pq + pq + \ldots + pq = n{*}pq$

Binomial Distribution

- Suppose 40% of a very large population of registered voters favor candidate Obama. A random sample of n = 5 voters will be selected, and X, the number favoring Obama out of 5, is to be observed.
- $p = P(\text{success}) = 0.40$
- $1 - p = P(\text{failure}) = 0.60$

Binomial Distribution

- What is the probability of getting no one who favors Obama, i.e. what is P(X = 0)?
- P(X = 0) = (0.6)(0.6)(0.6)(0.6)(0.6)

  No No No No No

  $= (0.6)^5$

  $= 0.07776$

What is the probability of getting 1 person who favors Obama?

P(X = 1) $= \binom{5}{1} (0.4)^1 (0.6)^4 = 0.25920$

Binomial Distribution

- What is the probability of getting 2 person who favors Obama? P(X = 2) =?

$$\binom{5}{2} = \frac{5!}{2!\,3!} = 10 \begin{cases} \text{Y Y N N N} & (0.4)(0.4)(0.6)(0.6)(0.6) \\ \text{Y N Y N N} & (0.4)(0.6)(0.4)(0.6)(0.6) \\ \vdots & \vdots \\ \text{N N N Y Y} & (0.6)(0.6)(0.6)(0.4)(0.4) \end{cases}$$

10 configurations

$$P(X = 2) = \binom{5}{2}(0.4)^2(0.6)^3 = 10 \cdot (0.4)^2(0.6^3) = 0.34560$$

Geometric 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/p²*

- Let X be a geometric random variable with p = 0.25. What is the probability that X = 4 (i.e. that the first success occurs on the 4th trial)?
- ANS: For X to be equal to 4, we must have had 3 failures, and then a success.

Negative Binomial Distribution

- 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:

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

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

Poisson Distribution

- 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 $\lambda(t-s)$

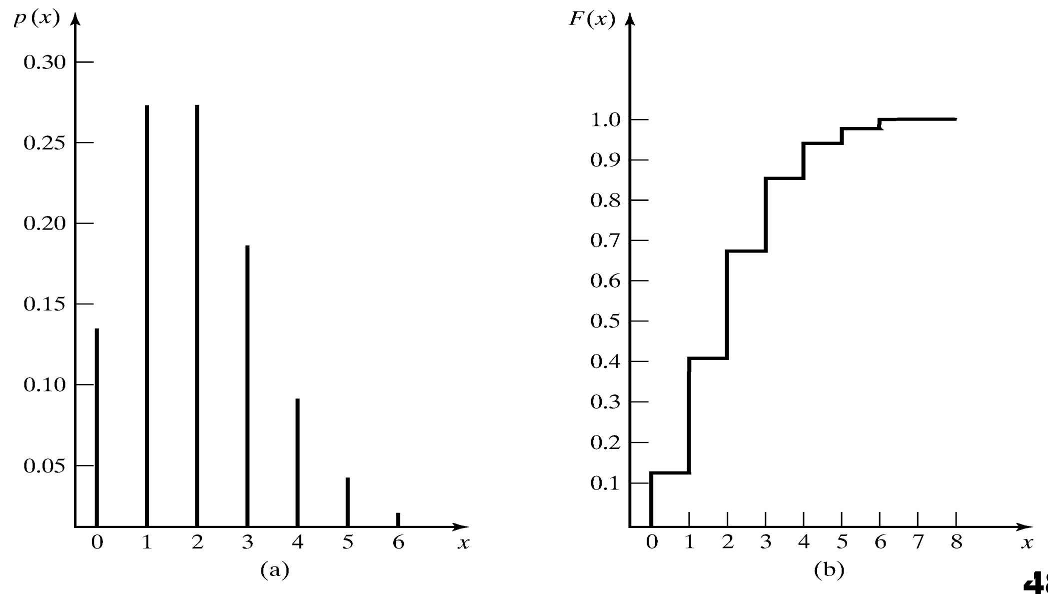
Poisson Distribution

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

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

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

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



Poisson Distribution

[Discrete Dist'n]

If the random variable X follows a Poisson distribution with mean 3.4, find $P(X=6)$.

This can be written more quickly as: if $X \sim Po(3.4)$ find $P(X=6)$.

Now $$P(X=6) = \frac{e^{-\lambda}\lambda^{6}}{6!}$$

$$= \frac{e^{-3.4}(3.4)^{6}}{6!}$$

= 0.072

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:

    $p(2 \text{ or more}) = 1 – p(0) – p(1)$

    $= 1 – F(1)$

    $= 0.594$

Poisson Distribution

The number of industrial injuries per working week in a particular factory is known to follow a Poisson distribution with mean 0.5.

Find the probability that

(a) in a particular week there will be:

(i) less than 2 accidents,

(ii) more than 2 accidents;

(b) in a three week period there will be no accidents.

Let A be 'the number of accidents in one week', so $A \sim P_0(0.5)$.

Poisson Distribution

$$P(A<2) = P(A=0)+P(A=1)$$

$$= e^{-0.5} + \frac{e^{-0.5}\times 0.5}{1!}$$

$$= \frac{3}{2}e^{-0.5}$$

$$\simeq 0.9098.$$

(ii) $\quad P(A>2) = 1-P(A\leq 2)$

Poisson Distribution [Discrete Dist'n]

$$1 - [P(A=0) + P(A=1) + P(A=2)]$$

$$= 1 - \left[e^{-0.5} + e^{-0.5} 0.5 + \frac{e^{-0.5}(0.5)^2}{2!} \right]$$

$$= 1 - e^{-0.5}(1 + 0.5 + 0.125)$$

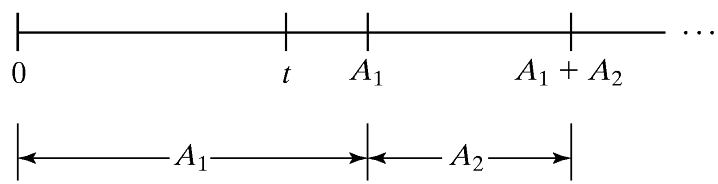
$$= 1 - 1.625\, e^{-0.5}$$

$$\approx 0.0144.$$

(b) $P\,(0 \text{ in } 3 \text{ weeks}) = \left(e^{-0.5}\right)^3 \approx 0.223.$

Interarrival Times [Poisson Dist'n]

- The experiment results in outcomes that can be classified as successes or failures.
- 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:

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

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

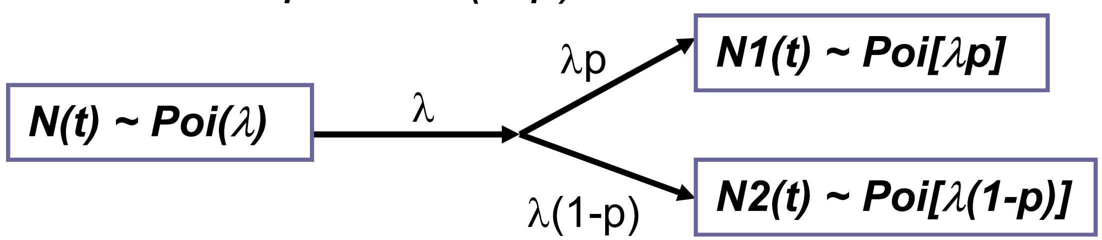
  - Interarrival times, $A_1, A_2, \ldots$, are exponentially distributed and independent with mean $1/\lambda$

Arrival counts ~ $Poi(\lambda)$ ⇔ Interarrival time ~ $Exp(1/\lambda)$

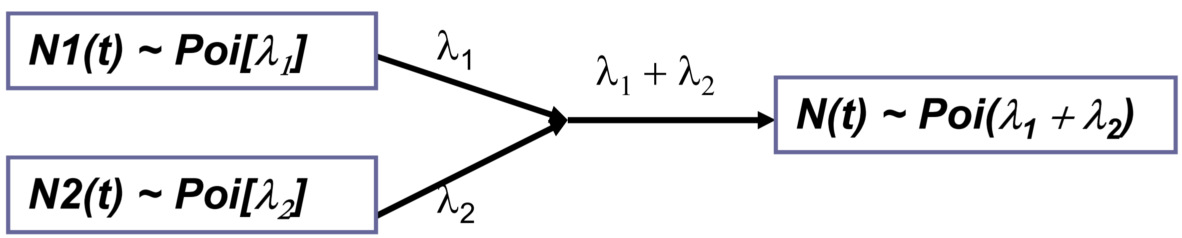
Stationary & Independent | Memoryless

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-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)

[Poisson Dist'n]

- 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:

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

Expected number of arrivals by time t:

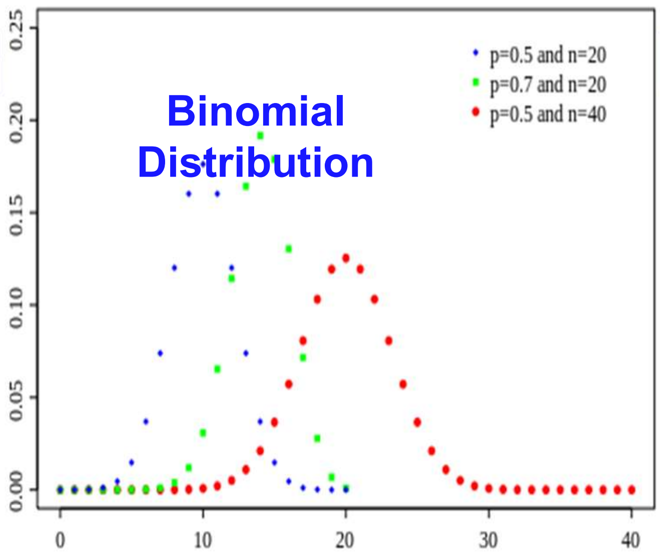
$$\Lambda(t) = \begin{cases} 2t, & 0 \le t < 4 \\ \int_0^4 2ds + \int_4^t 0.5ds = \frac{t}{2} + 6, & 4 \le t < 8 \end{cases}$$

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

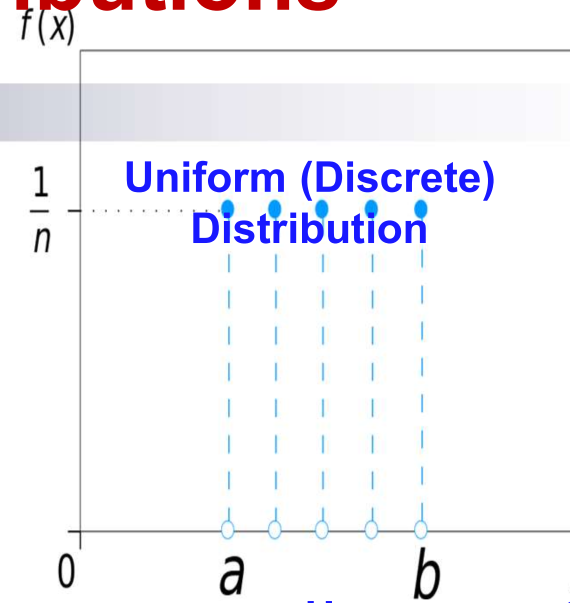
$$\begin{aligned} P[N(6) - N(3) = k] &= P[N(\Lambda(6)) - N(\Lambda(3)) = k] \\ &= P[N(9) - N(6) = k] \\ &= e^{(9-6)}(9-6)^k/k! \quad = e^3(3)^k/k! \end{aligned}$$

Discrete Probability Distributions

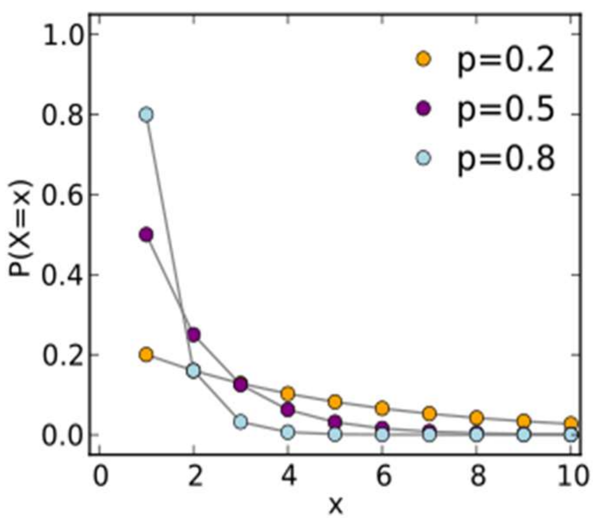
Binomial Distribution



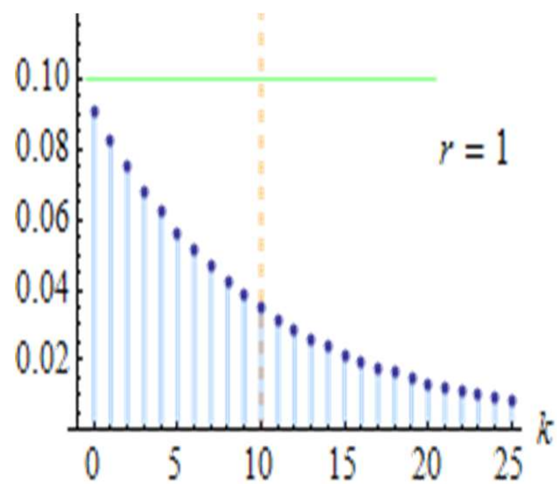
Uniform (Discrete) Distribution



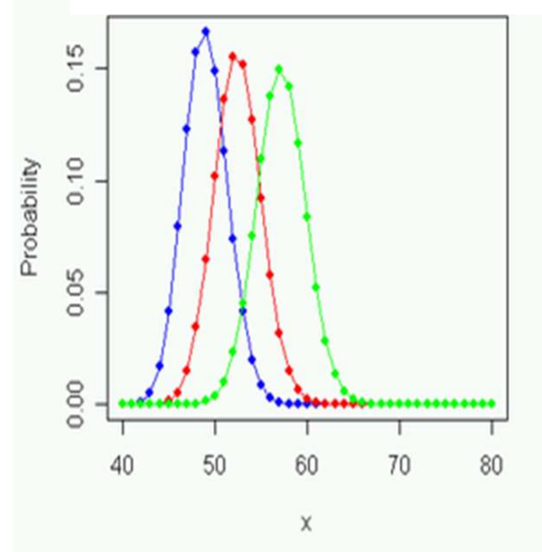
Geometric



Negative Binomial Distribution



Hypergeometric Distribution



Discrete Distributions

- If an experiment only has two possible outcomes → Binomial distribution (ex: packet successfully received or not)
- If we need to count the number of trials until the first success → geometric distribution
- If we need to count the number of trials until the *kth* success, $k = 1, 2, \ldots$ →negative binomial distribution.
- Negative binomial distribution can thought of as a sum of independent geometric distributions
- Ex: What is the probability that the third inspected product at a manufacturing plant is the second one accepted.
- If we need to count the number of event occurrences within a period of time → Poisson distribution (ex: number of calls in an hour)

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.

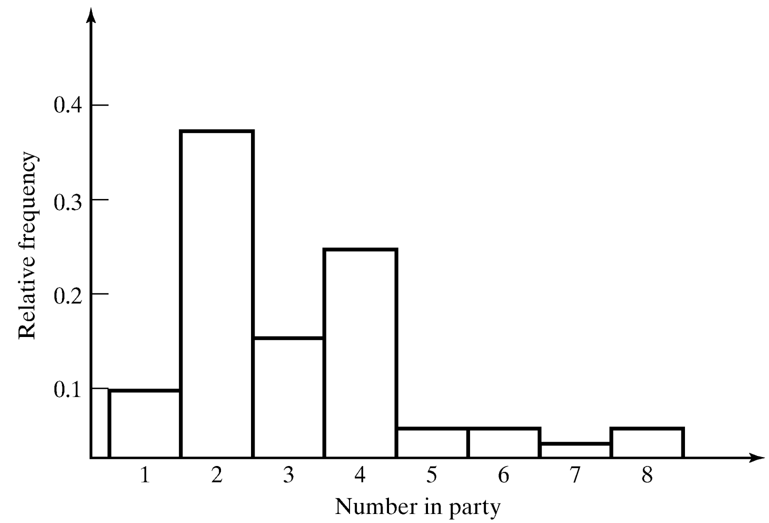
Empirical Distributions

- Example:
- Customers arrive at lunchtime in groups of from one to eight persons.
- The number of persons per party in the last 300 groups has been observed.
- The results are summarized in a table.
- The histogram of the data is also included.

Empirical Distributions (cont.)

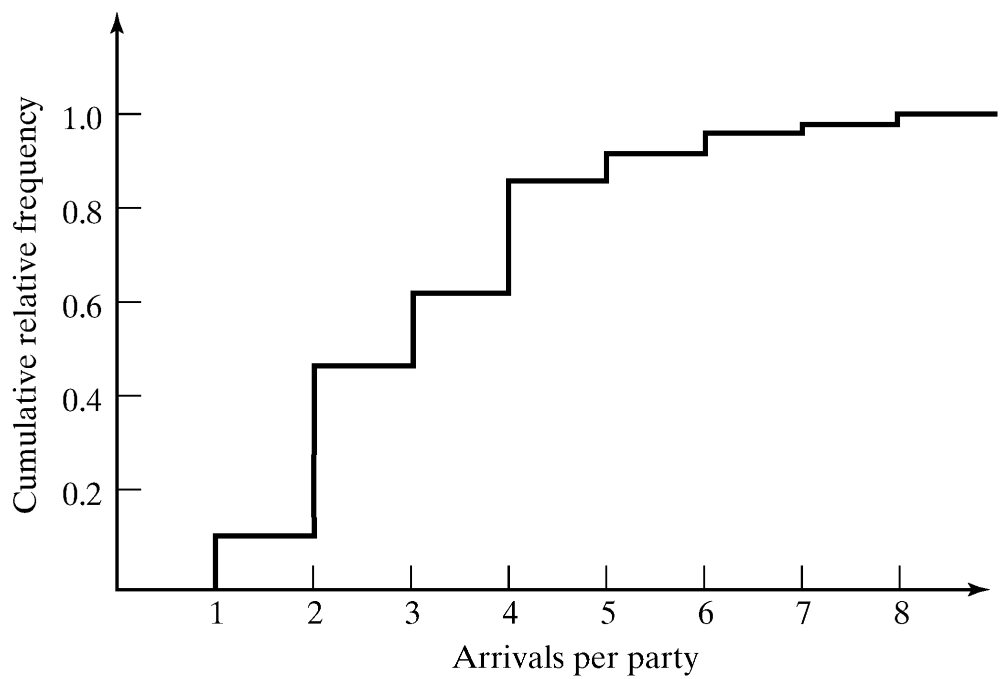
[Probability Review]

Arrivals per Party	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



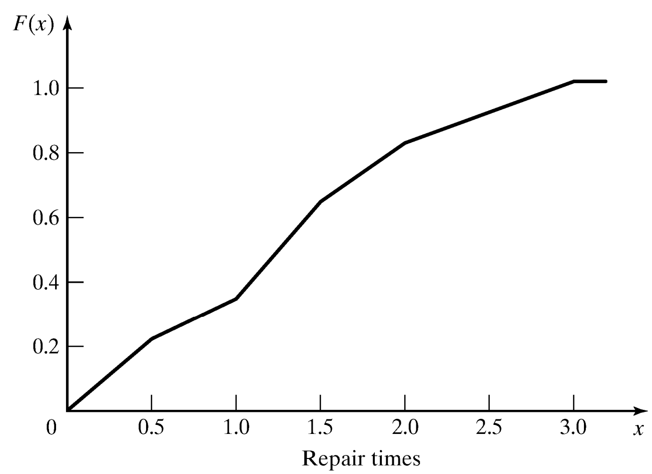
Empirical Distributions (cont.) [Probability Review]

- The CDF in the figure is called the empirical distribution of the given data.



Empirical Distributions

- Example:
- The time required to repair a system that has suffered a failure has been collected for the last 100 instances.
- The empirical CDF is shown in the figure



Empirical Distributions (cont.)

Intervals (Hours)	Frequency	Relative Frequency	Cumulative Frequency
0<x<0.5	21	0.21	0.21
0.5<x<1.0	12	0.12	0.33
1.0<x<1.5	29	0.29	0.62
1.5<x<2.0	19	0.19	0.81
2.0<x<2.5	8	0.08	0.89
2.5<x<3.0	11	0.11	1.00

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

- 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 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 (to satisfy demands)
- 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
  - Lognormal distribution: time to failure for some types of components

Limited Data

- If the data obtained is limited or incomplete, there are usually three distributions that are used:

Uniform Distribution	Triangular Distribution	Beta Distribution
• Inter-arrival or service times are known to be random • No other information is known about the distribution	• When assumptions can be made about the maximum, minimum and modal values of the distribution	• Provides a variety of distribution forms on the unit interval • These distributions can be shifted to any other interval

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. Reader should know:
  - Difference between discrete, continuous, and empirical distributions.
  - Poisson process and its properties.