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

Probability (Review)

- Is a *measure of chance*
- **Laplace's Classical Definition:** The Probability of an event *A* is defined a-priori, without actual experimentation as

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}}$$

  provided all these outcomes are *equally likely*.
- **Relative Frequency Definition**: The **probability of an event** A is defined as

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

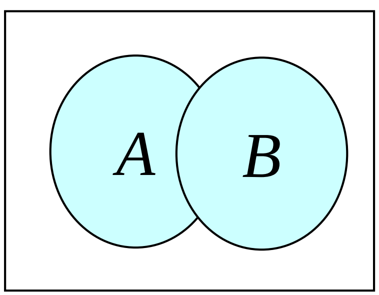
  where n_A is the number of occurrences of A and n is the total number of trials

Probabilty (continued ..)

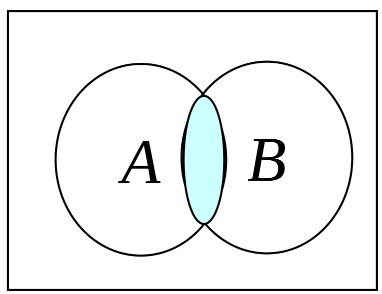
- The ***axiomatic approach*** to probability, due to Kolmogorov developed through a set of axioms
- For any Experiment *E*, has a set *S or* Ω of all possible outcomes called **sample space**

$$\Omega = \{ \xi_1, \xi_2, \cdots, \xi_k, \cdots \}$$

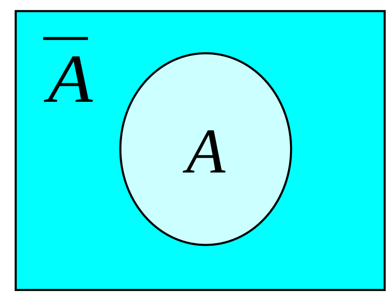
- Ω has **subsets** *{A, B, C, …..}* called events*.* If $A \cap B = \phi$, the empty set, then *A* and *B* are said to be ***mutually exclusive events***.



$A \cup B$



$A \cap B$



$\overline{A}$

Axioms of Probability (Review)

- For any event *A*, we assign a number *P*(*A*), called the probability of the event *A*. This number satisfies the following three conditions that act the axioms of probability.

$$\text{(i)} \quad P(A) \geq 0 \quad \text{(Probability is a nonnegative number)}$$

$$\text{(ii)} \quad P(\Omega) = 1 \quad \text{(Probability of the whole set is unity)}$$

$$\text{(iii)} \quad \text{If } A \cap B = \phi, \quad \text{then} \quad P(A \cup B) = P(A) + P(B).$$

(Note that (iii) states that if *A* and *B* are mutually exclusive (M.E.) events, the probability of their union is the sum of their probabilities.)

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

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

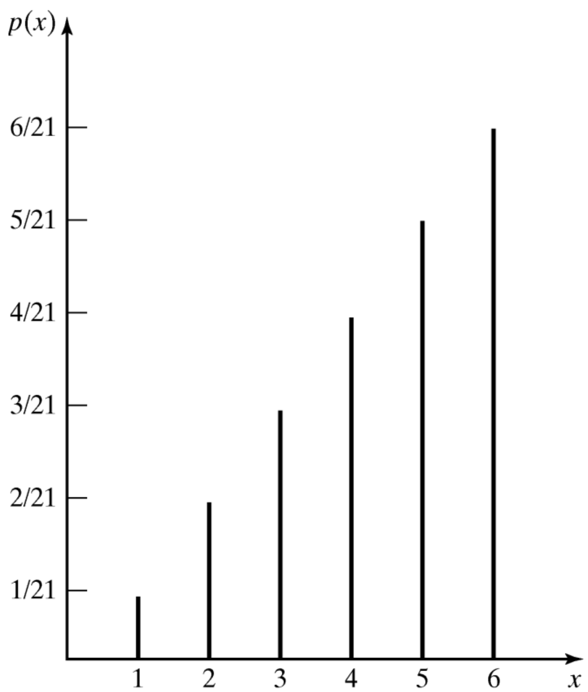
    2. $\sum_{i=1}^{\infty} p(x_i) = 1$

  - 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.

Example

- Consider the experiment of tossing a single die. Define *X* as the number of spots on the up face of the die after a toss.
- R_X={1,2,3,4,5,6}
- Assume the die is **loaded (weighted)** so that the probability that a given face lands up is **proportional to the number of spots** showing

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



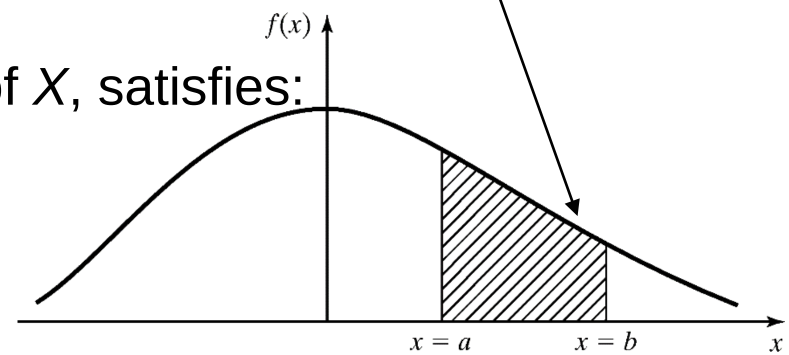
- What if all the faces are equally likely??

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:

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

shown as shaded area



- *f(x)*, probability density function (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

f(x) is called probability density function

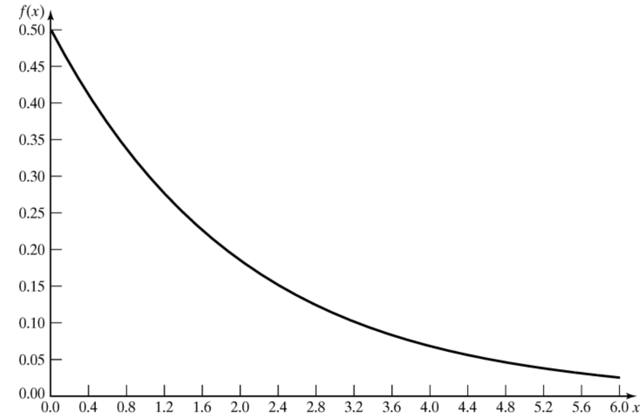
- Properties

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

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 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)*, measures the probability that the random variable $X \le x$, i.e., $F(x) = P(X \le x)$
  - 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 $a < b$, then $F(a) \le F(b)$
  2. $\lim_{x\to\infty} F(x) = 1$
  3. $\lim_{x\to-\infty} F(x) = 0$

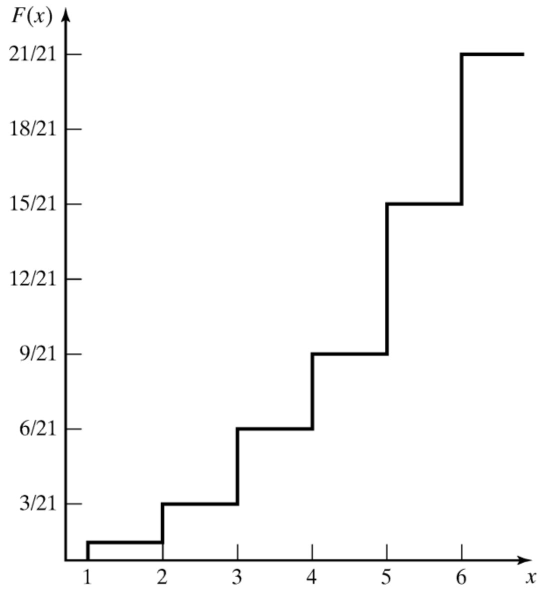
- All probability questions about X can be answered in terms of the cdf, e.g.:

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

CDF

- Consider the loaded die example

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



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)=μ*
  - If *X* is discrete

$$E(X) = \sum_{\text{all } i} x_i p(x_i)$$

  - If *X* is continuous

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

  - Expected value is also known as the mean (μ), or the 1st moment of *X*
  - A measure of the central tendency
- *E(Xⁿ), n ≥1* is called ***nth* moment of *X*** ***(very useful for analysis)***
  - If X is discrete

$$E(X^n) = \sum_{\text{all } i} x_i^{\,n} p(x_i)$$

  - If X is continuous

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx$$

Measures of Dispersion

- The *variance of X* is denoted by $V(X)$ or $var(X)$ or σ^2
  - Definition: $V(X) = E[(X - E[X])^2] = E[(X - \mu)^2]$
  - Also, $V(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$
  - A measure of the spread or variation of the possible values of X around the mean μ
- The *standard deviation of X* is denoted by σ
  - Definition: square root of $V(X)$ *i.,e*

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

  - Expressed in the same units as the mean

Example Discrete RV

- The mean and variance of loaded die is:

$$E(X)=1\left(\frac{1}{21}\right)+2\left(\frac{2}{21}\right)+\cdots+6\left(\frac{6}{21}\right)=\left(\frac{91}{21}\right)=4.33$$

To compute Variance, first compute Second moment

$$E\left(X^2\right)=1^2\left(\frac{1}{21}\right)+2^2\left(\frac{2}{21}\right)+\cdots+6^2\left(\frac{6}{21}\right)=21$$

Then, Variance of X is :

$$V(X)=E\left(X^2\right)-E^2(X)=21-18.78=2.22$$

Standard Deviation of X is :

$$\sigma=\sqrt{V(X)}=\sqrt{2.22}=1.49$$

Example Continuous RV

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

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

- To compute variance of *X*, we first compute *E(X²)*:

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

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

$$V(X) = 8 - 2^2 = 4 \text{ years}^2$$

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

Mode

- In the discrete RV case, the *mode* is the value of the random variable that occurs most frequently
- In the continuous RV case, the *mode* is the value at which the pdf is maximized
- Mode might not be unique
- If the modal value occurs at two values of the random variable, it is said to **bi-modal**

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, inter-arrival and service-time patterns can be probabilistic (for more queueing examples, see Chapter 2).
- Sample statistical models for inter-arrival 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 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 (time between the placing of an order for stocking the inventory system 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.

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 distributions: *Bernoulli, binomial and hyper-exponential.*

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.
    - Let $X_j = 1$ if the j^{th} trial is a success with probability p
    - and $X_j = 0$ if the j^{th} trial is a failure

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

  - For one trial, it is called the Bernoulli distribution 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 with parameters *n* and *p*: ***B(n,p)***.

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

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

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

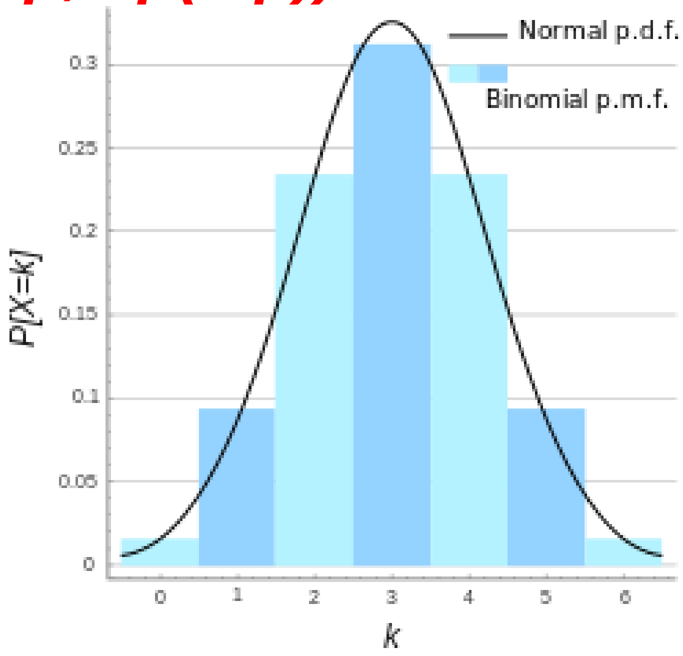
  - Easy approach is to consider the binomial distribution X as a sum of n independent Bernoulli Random variables *(X=X1+X2+…+Xn)*
  - The mean, *E(X) = p + p + … + p = n*p*
  - The variance, *V(X) = pq + pq + … + pq = n*pq*

Other "shape" measures

- **Skewness:** A measure of asymmetry of a distribution around it's mean. Defined as **third** standardized moment
  - $\gamma_1 = E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$
- **Kurtosis:** A measure of peakedness in a distribution around it's mean. Defined as **fourth** standardized moment
  - $\text{Kurt}[X] = E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}$

Binomial Distribution (contd ..)

- If *n* is large enough, then the skew (asymmetry) of the distribution is not too great. In this case a reasonable approximation to ***B(n, p)*** is given by the normal distribution ***N(np, np(1-p))***



Binomial Distribution Example

- A production process produces memory chips on the average at 2% non conforming. Everyday, a sample size of 50 is taken from the process. If the sample contains more than two non conforming chips, the process will be stopped. What is the probability that the process is stopped by the sampling scheme?

$$p(x) = \begin{cases} \binom{50}{x}(0.02)^x(0.98)^{50-x}, & x = 0,1,2,\cdots,50 \\ 0, & \text{otherwise} \end{cases}$$

$$P(X > 2) = 1 - P(X \le 2)$$

$$P(X \le 2) = \sum_{x=0}^{2}\binom{50}{x}(0.02)^x(0.98)^{50-x} = 0.92$$

The Probability the production is stopped = (1-0.92) = 0.08

The mean number of non-conforming chips in a random sample of 50 is:

E(X)=np = 50(0.02)=1
V(X)=npq= 0.98

Geometric & Negative Binomial Distribution

[Discrete Dist'n]

- Geometric distribution (Used frequently in data networks)
  - The number of Bernoulli trials, *X*, to achieve the 1st success:

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

  - *E(x) = 1/p, and V(X) = q/p²*

- 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/p²*
  - *Y is the sum of k independent geometric RVs*

Example

- 40% of the assembled ink-jet printers are rejected at the inspection station. Find the probability that the first acceptable ink-jet printer is the third one inspected. Considering each inspection as a Bernoulli trial with *q=0.4 and p=0.6,*

  $$p(3) = 0.4^2(0.6) = 0.096$$

  Thus, in only about 10% of the cases is the first acceptable printer is the third one from any arbitrary starting point
- What is the probability that the third printer inspected is the second acceptable printer?

  Use Negative Binomial Distribution with *y=3 and k=2*

$$p(3) = \binom{3-1}{2-1} 0.4^{3-2} (0.6)^2 = 0.288$$

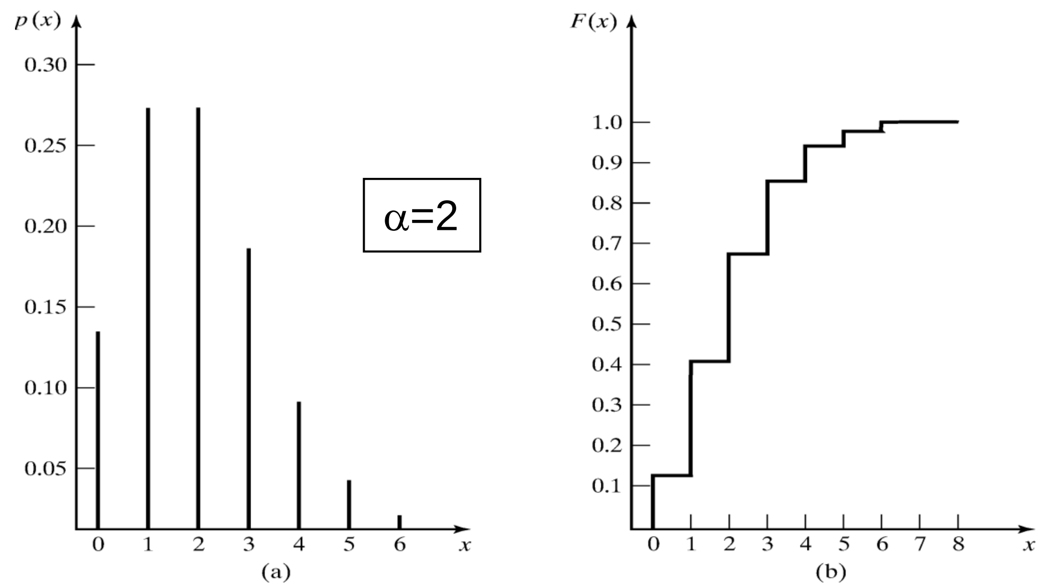
Poisson Distribution [Discrete Dist'n]

- Poisson distribution describes many random processes quite well and is mathematically quite simple. The pmf and cdf are:

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

where $\alpha > 0$

- ***E(X) = α = 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

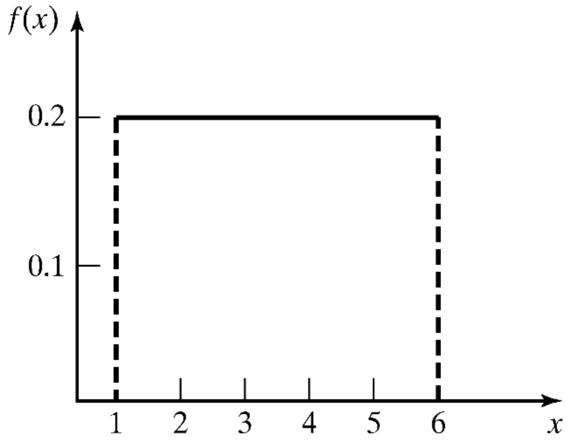
Uniform Distribution

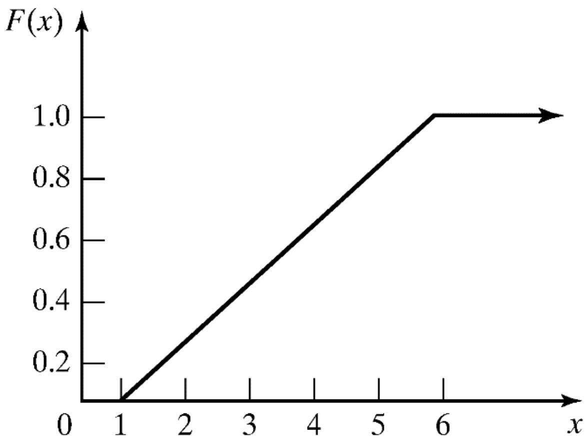
[Continuous Dist'n]

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

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

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





Example with *a = 1* and *b = 6*

Uniform Distribution

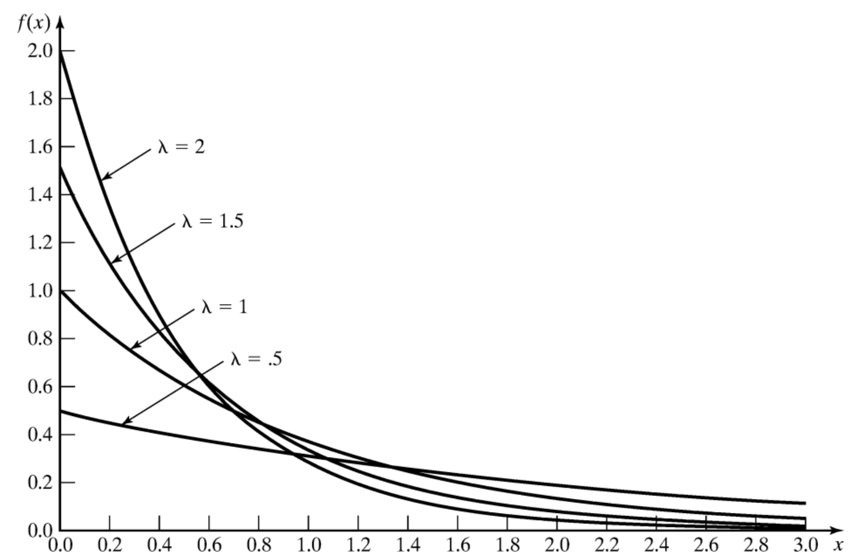
- Properties
  - $P(x_1 \le X < x_2)$ is proportional to the length of the interval $[F(x_2) - F(x_1) = (x_2-x_1)/(b-a)]$
  - $E(X) = (a+b)/2$ $V(X) = (b-a)^2/12$
- *U(0,1) provides the means to generate random numbers, from which random variates can be generated.*
- ***Example:*** In a warehouse simulation, a call comes to a forklift operator about every 4 minutes. With such a limited data, it is assumed that time between calls is uniformly distributed with a mean of 4 minutes with ($a=0$ and $b=8$)

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 Example

- Example: A lamp life (in thousands of hours) is exponentially distributed with failure rate ($\lambda = 1/3$), hence, on average, 1 failure per 3000 hours.
  - The probability that the lamp lasts longer than its “mean life” is:

    $P(X > 3) = 1-(1-e^{-3/3}) = e^{-1} = 0.368$

    This is independent of λ. That is, the probability that an exponential random variable is greater than it’s mean is 0.368 for any λ

  - The probability that the lamp lasts between 2000 to 3000 hours is:

    $P(2 \le X \le 3) = F(3) - F(2) = 0.145$

Memoryless Property of Exponential Distribution

- Memoryless property is one of the important properties of exponential distribution
  - For all $s \geq 0$ and $t \geq 0$:

    $$P(X > s+t \mid X > s) = P(X > t) = P(X>s+t)/P(X>s) = e^{-\lambda t}$$

  - Let X represent the life of a component and is exponentially distributed. Then, the above equation states that the probability that the component lives for at least $s+t$ hours, given that it survived s hours is the same as the probability that it lives for at least t hours. That is, the component doesn't remember that it has been already in use for a time *s*. ***A used component is as good as new!!!***
  - **<u>Light bulb example:</u>** The probability that it lasts for another 1000 hours given it is operating for 2500 hours is the same as the new bulb will have a life greater than 1000 hours

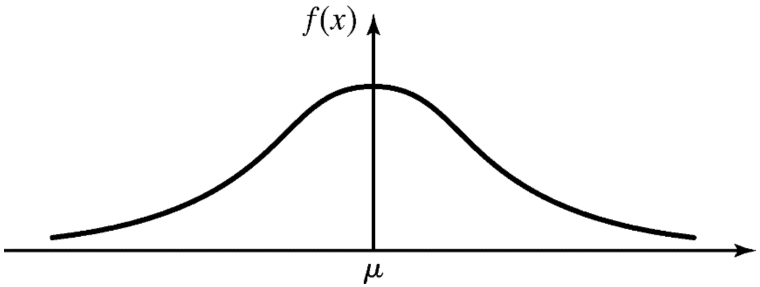
    $$P(X > 3.5 \mid X > 2.5) = P(X > 1) = e^{-1/3} = 0.717$$

Normal Distribution [Continuous Dist'n]

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

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

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



- Special properties:
  - $\lim_{x \to -\infty} f(x) = 0$, 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.

CDF of Normal Distribution

- The CDF of Normal distribution is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] \mathrm{d}t,$$

- It is not possible to evaluate this in closed form
- Numerical methods can be used but it would be necessary to evaluate the integral for each pair (μ, σ^2).
- A transformation of variable allows the evaluation to be independent of μ and σ.

Normal Distribution [Continuous Dist'n]

- Evaluating the distribution:
  - Independent of μ and σ, using the standard normal distribution:

    Z ~ N(0,1)
  - Transformation of variables: let Z = (X - *μ*) / *σ*,

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

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

$$= \int_{-\infty}^{(x-\mu)/\sigma} \phi(z) dz = \Phi(\tfrac{x-\mu}{\sigma})$$

where $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ is very well tabulated.

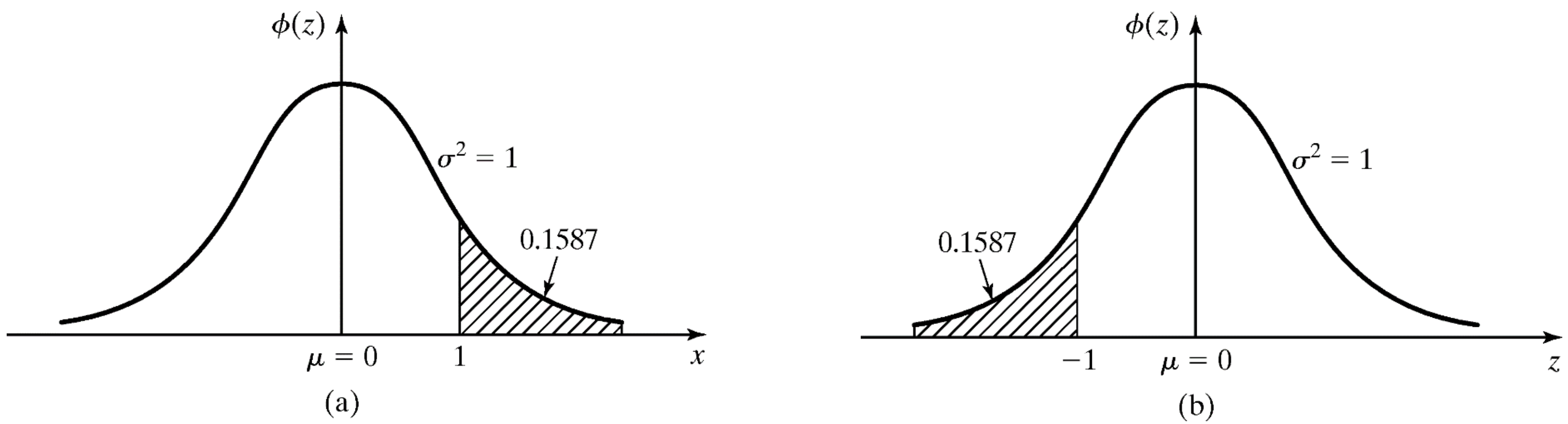
Normal Distribution

[Continuous Dist'n]

- Example: The time required to load an ocean going 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), i.e., Φ (-1) = 1- $\Phi(1)$



Example

- The time to pass through a queue to begin self-service at a cafeteria is found to be *N(15,9).* The probability that an arriving customer waits between 14 and 17 minutes is:

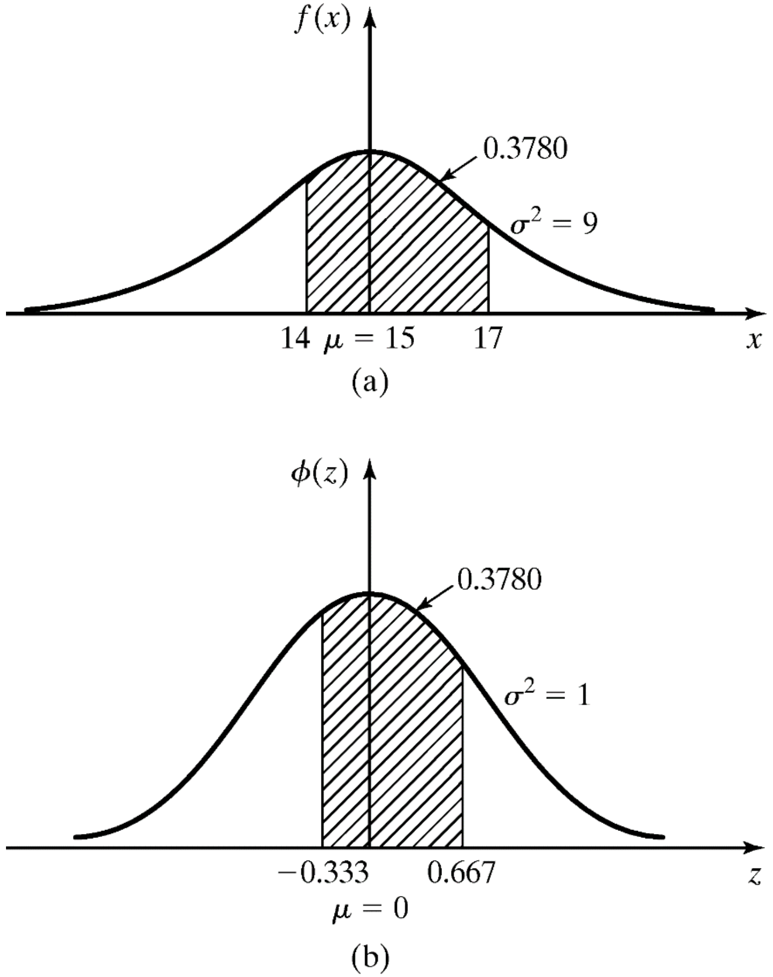
$$P(14 \le X \le 17) = F(17) - F(14)$$

$$= \phi((17-15)/3) - \phi((14-15)/3)$$

$$= \phi(0.667) - \phi(-0.333) = 0.3780$$

Transformation

- Transformation of pdf for the queue example is shown here



Weibull Distribution

[Continuous Dist'n]

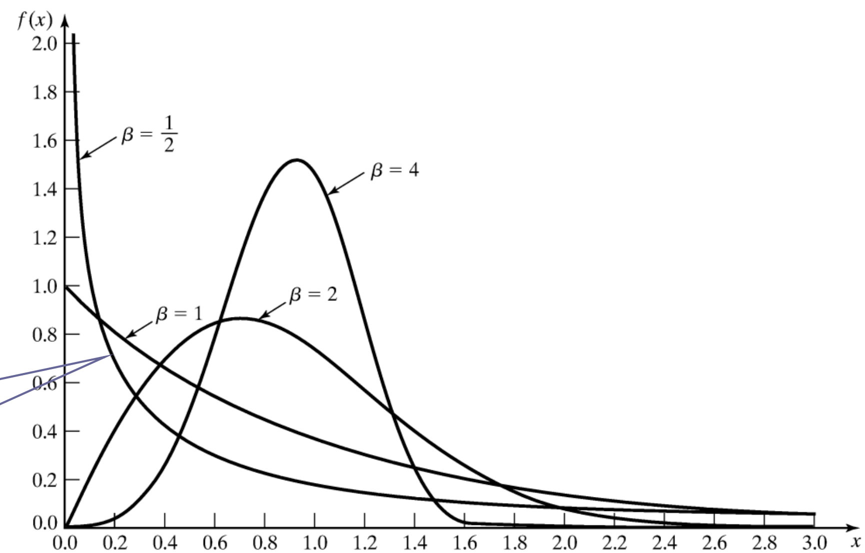
- A random variable *X* has a Weibull distribution if its pdf has the form:

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

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

Exponential Distribution

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



Mean and Variance of Weibull

- The mean and variance of Weibull is given by

$$E(X) = \nu + \alpha\Gamma\left(\frac{1}{\beta} + 1\right)$$

$$V(X) = \alpha^2\left[\Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right)\right]^2\right]$$

where $\Gamma(.)$ is a Gamma function defined as $\Gamma(\beta) = \int_0^\infty x^{\beta-1}e^{-x}dx$

If β is an integer, $\Gamma(\beta) = (\beta - 1)!$

- The CDF is given by

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

Example

- The time it takes for an aircraft to land and clear the runway at a major international airport has a Weilbull distribution with ν=1.35 minutes, β=0.5, α=0.04 minute. Find the probability that an incoming aircraft will take more than 1.5 minute to land and clear the runway.

$$\boldsymbol{P}(\boldsymbol{X} > 1.5) = 1 - \boldsymbol{P}(\boldsymbol{X} \le 1.5)$$

$$\boldsymbol{P}(\boldsymbol{X} \le 1.5) = \boldsymbol{F}(1.5) = 1 - \exp\left[-\left(\frac{1.5-1.34}{0.04}\right)^{0.5}\right] = 0.865$$

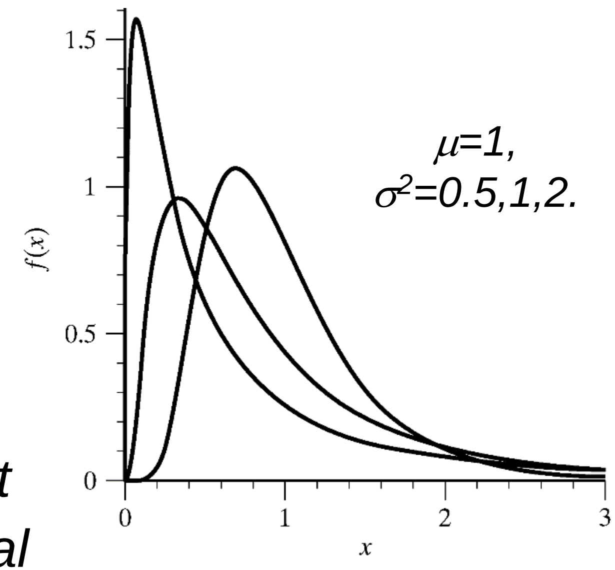
Therefore, the probability that an aircraft will require more than 1.5 minutes to land and clear runway is

$$\mathbf{1 - 0.865 = 0.135}$$

Lognormal Distribution [Continuous Dist’n]

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

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



  - Mean $E(X) = e^{\mu+\sigma^2/2}$
  - Variance $V(X) = e^{2\mu+\sigma^2/2} (e^{\sigma^2} - 1)$
  - *Note that **parameters μ and σ^2** are not the mean and variance of the lognormal*

- Relationship with normal distribution
  - When $Y \sim N(\mu, \sigma^2)$, then $X = e^Y \sim \text{lognormal}(\mu, \sigma^2)$

Lognormal (continued ..)

If the mean and variance of lognormal are known to be

μ_{L} and σ_L^2 respectively, then the parameters μ and σ^2 is given by:

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

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

- Example: The rate of return on a volatile investment is modeled as lognormal with mean 20% (=μ_L) and standard deviation 5% (=σ_L^2). What are the *parameters* for lognormal?
  - μ = 2.9654; σ^2=0.06

Poisson Distribution

- Definition: $N(t), t \geq 0$ is a counting function that represents the number of events occurred in $[0,t]$.
  - *e.g.,* arrival of jobs, e-mails to a server, boats to a dock, calls to a call center
- 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**: The distribution of number of arrivals between t and $t+s$ depends only on the length of interval s and not on starting point t. Arrivals are completely random without rush or slack periods.
  - **$\{N(t), t \geq 0\}$ has independent increments**: The number of arrivals during non-overlapping time intervals are independent random variables.

Properties of Poisson

- Properties

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

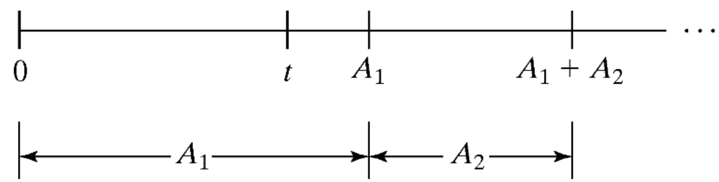
  - Equal mean and variance: $E[N(t)] = V[N(t)] = \lambda t$
  - Stationary increment: For any s and *t,* such that *s* < *t,* the number of arrivals in time *s* to *t* is also Poisson-distributed with mean $\lambda(t-s)$

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

$$\text{and } E\left[N(t)-N(s)\right]=\lambda(t-s)=V\left[N(t)-N(s)\right]$$

Interarrival Times [Poisson Dist'n]

- Consider the inter-arrival 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:

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

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

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

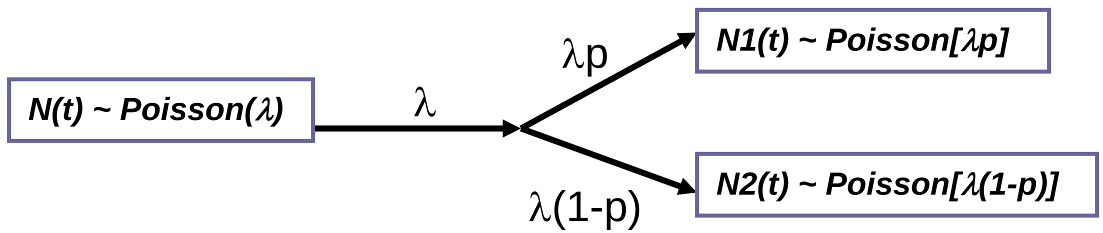
Arrival counts ~ *Poisson*(λ)	⟺	Inter-arrival time ~ *Exp*($1/\lambda$)
Stationary & Independent		Memoryless

Example

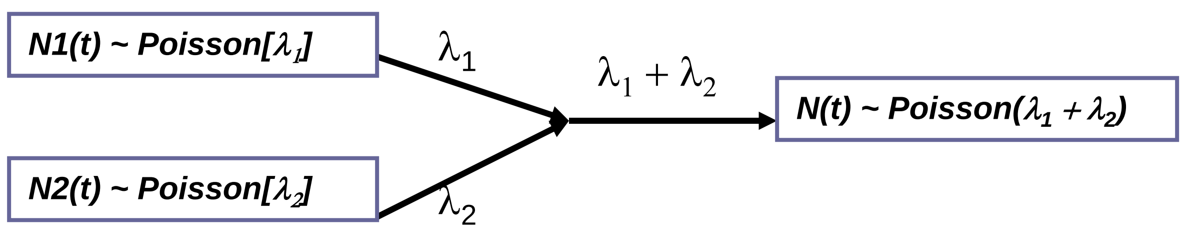
- The jobs at a machine shop arrive according to a Poisson process with a mean of $\lambda = 2$ jobs per hour. Therefore, the inter-arrival times are distributed exponentially with the expected time between arrivals being $E(A)=1/\lambda=0.5$ hour

Properties of Poisson Process [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 $1\text{-}p$.
  - $N(t) = N_1(t) + N_2(t)$, where $N_1(t)$ and $N_2(t)$ are both Poisson processes with rates λp and $\lambda(1\text{-}p)$



- Pooling:
  - Suppose two Poisson processes are pooled together
  - $N_1(t) + N_2(t) = N(t)$, where $N(t)$ is a Poisson processes with rates $\lambda_1 + \lambda_2$



Example

- Suppose jobs arrive at a shop with a Poisson process of rate λ. Suppose further that each arrival is marked “high priority” with probability 1/3 (Type I event) and “low priority” with probability 2/3 (Type II event). Then $N_1(t)$ and $N_2(t)$ will be Poisson with rates $\lambda/3$ and $2\lambda/3$.

Non-stationary Poisson Process (NSPP)

- Poisson Process without the stationary increments, characterized by $\lambda(t)$, the arrival rate at time t. (Drop assumption 2 of Poisson process, stationary increments)
- 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:

    $t_i = \Lambda(T_i)$ *[Expected # of arrivals]*

    $T_i = \Lambda^{-1}(t_i)$
  - An NSPP can be transformed into a stationary Poisson process with arrival rate 1 and vice versa.

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:

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

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, corresponds to times 3 and 6 respectively.

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

Empirical Distributions

- 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.

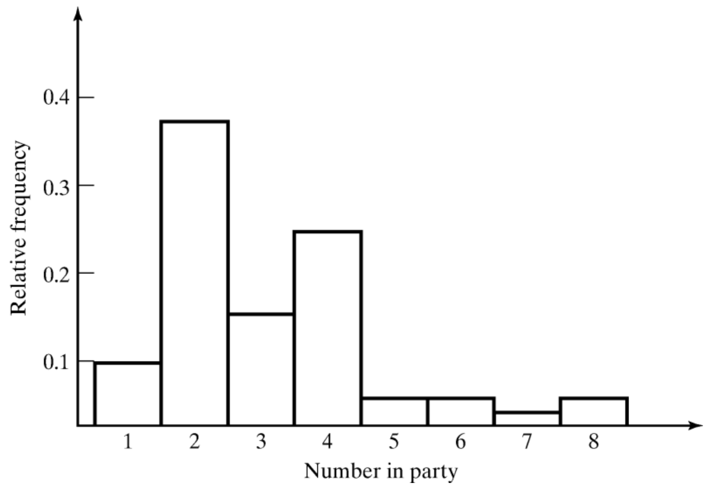
Empirical Example - Discrete

- Customers at a local restaurant arrive at lunch time in groups of eight from one to eight persons. The number of persons per party in the last 300 groups has been observed. The results are summarized in Table 5.3. A histogram of the data is plotted and a CDF is constructed. The CDF is called the empirical distribution

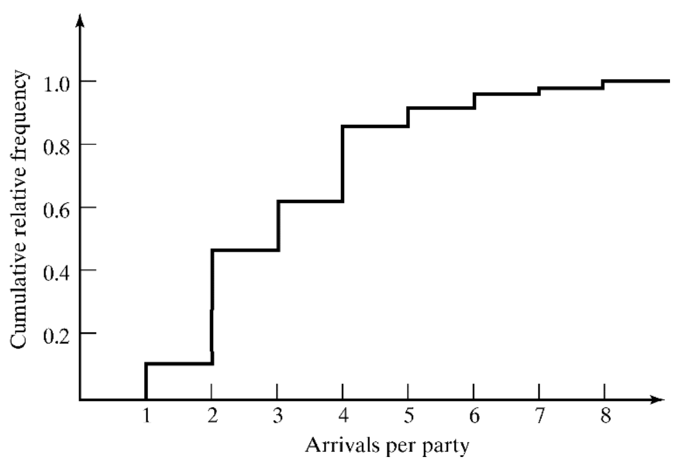
Table 5.3 Arrivals per Party Distribution

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 Example – Discrete (contd ..)



Histogram



CDF

Empirical Example - Continuous

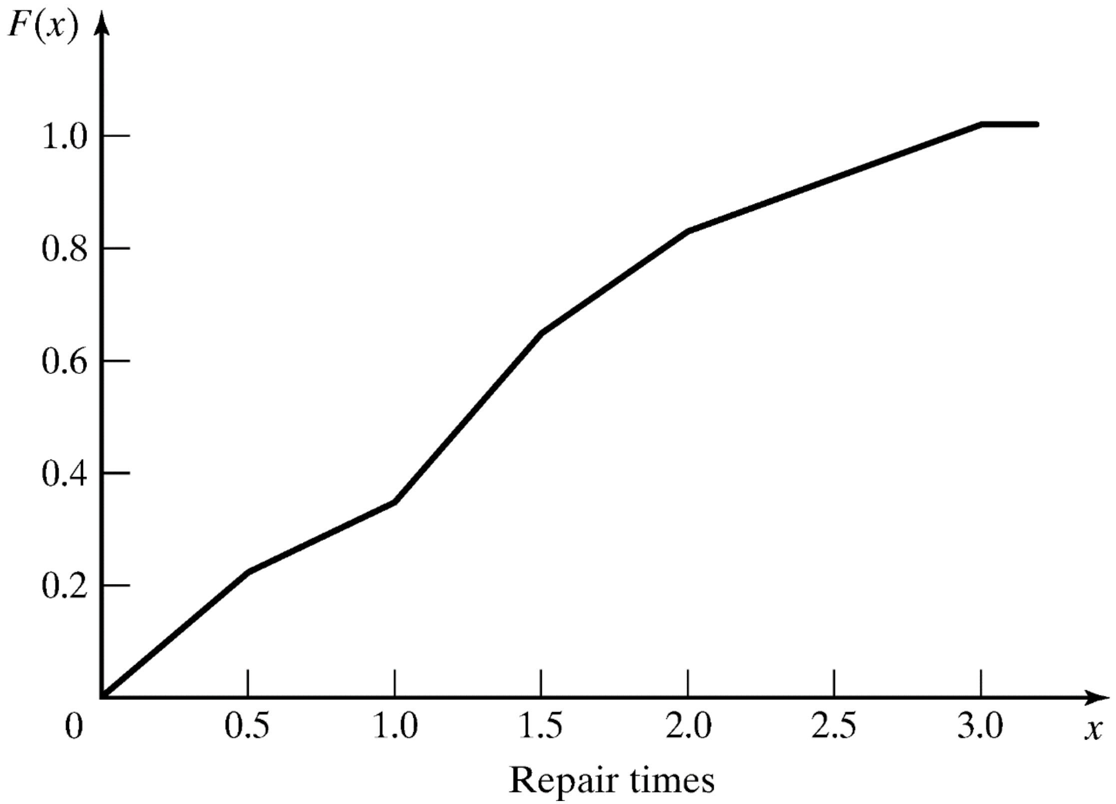
- The time required to repair a conveyor system that has suffered a failure has been collected for the last 100 instances; the results are shown in Table 5.4. There were 21 instances in which the repair took between 0 and 0.5 hour, and so on. The empirical cdf is shown in Figure 5.29. A piecewise linear curve is formed by the connection of the points of the form *[x,F(x)].* The points are connected by a straight line. The first connected pair is (0, 0) and (0.5, 0.21); then the points (0.5, 0.21) and (1.0, 0.33) are connected; and so on. More detail on this method is provided in Chapter 8

Empirical Example – Continuous (contd ..)

Table 5.4 Repair Times for Conveyor

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

Empirical Example – Continuous (contd ..)



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.