



SHARIF
UNIVERSITY OF
TECHNOLOGY



Computer Simulation

Dr. Bardia Safaei

Chapter Three: Statistical Models



Purpose and Overview (1)



- What is the goal of studying statistical models?
 - Many phenomenon in real-world are rather **probabilistic** rather than being **deterministic**
 - Therefore, it is necessary to **search for good statistical models** to describe the **alterations** of the intended probabilistic phenomenon correctly with good precision
- Based on sampling the intended phenomena and the results, an appropriate model can be developed by:
 - Select a known distribution through educated **guesses**
 - Make **estimate of the parameter(s)** for the guessed distribution function(s)
 - Test for goodness of fit to select the best candidate

Purpose and Overview (2)



- Note: It is possible to have multiple distribution functions describing a single phenomena
 - It is our job to compare the outcomes of the functions to the real-world data
 - This is what we call as **reconciliation**
 - Afterwards, the most suited distribution function will be selected
- In this chapter, we are going to study:
 - Different distribution functions
 - Their specifications
 - Applications that they could be applied to

Random Variables (1)



- A random variable is a parameter of a specific phenomena which is measurable but its value is random
- Discrete random variables
 - If X could be assigned with only a set of countable values
 - Example: tossing a coin, status of a production line which is working or is failed
- Continuous random variables
 - If X could be assigned with any values within a finite or infinite range
 - Example: Received Signal Strength Index (RSSI) in wireless communications (in dB), or strength of wind in power systems

Random Variables (2)



- Now consider the number of jobs arriving at a job shop
 - Let X be the number of jobs arriving each week at a the shop
 - R_X is all of the possible values of X , which could be $\in \{0, 1, 2, \dots\}$
 - R_X is called as support or sample space
 - It is possible to have no requests, while we could have more
 - The probability that the random variable X equals to x_i is denoted by $P(x_i) = P(X = x_i)$
 - Note: Probability of having event x_i for any i is greater or equal to zero:

$$P(x_i) \geq 0, \text{ for any } i$$

- Note: Sum of all probability values for all the events x_i (any i) would be equal to 1:

$$\sum_{i=1}^{\infty} P(x_i) = 1$$

Random Variables (3)



- Couples of $[x_i, P(x_i)]$ for all of the i values, is called the **probability distribution** of random variable X
 - $P(x_i)$ is known as Probability Mass Function (PMF) of random variable X
 - PMF gives us the probability that X is equal to x_i
- Reminding some important statistical knowledge
- Random experiment
 - A function which is replicable, but gives us different random outcomes every time we iterate it
 - Notify that all the outcomes are known priorly, but we don't know which one would come up after the experiment
 - Example: tossing a coin, throwing a dice, distant to the target in archery



Random Variables (4)



■ Sample space:

- Set of all the possible outcome values for a random experiment
- It is indicated with Ω
 - Example: In tossing a coin we have $\Omega = \{H, T\}$, or in archery, the distance to the red target has $\Omega = [0, r)$
 - Since in the later, x_i can get any numbers and we have an uncountable set, the random variable is continuous

■ Event:

- Consider a countable sample space
- Any subset of this sample space, creates an event
 - Example: In tossing a coin for two times, $A = \{TH, TT\}$ is an event representing an observation of tails in the first round of throwing
- If $A \subset \Omega \rightarrow \bar{A}(\text{Complement Event}) \subset \Omega$
 - $\Omega (\text{Definite Event}) \subset \Omega \rightarrow \emptyset (\text{Impossible Event}) \subset \Omega$



Random Variables (5)



■ Event space:

- A set of all the subsets of our sample space (Ω)
- Denoted with F
 - $\emptyset \in F$, and $\Omega \in F$
 - If A is member of F , then its complement (\bar{A}) is also member of F
 - If A_1, A_2, \dots are sequence of independent events ($\in F$), then their union will be also member of F

$$\bigcup_{i=1}^{\infty} A_i \in F$$

■ Inconsistent events:

- If A , and B are two events of random variable X , they are called inconsistent if their intersection is null

$$\bigcap_{i=1}^{\infty} A_i = \emptyset$$



Random Variables (6)



■ Inconsistent events (Cont.):

- In this case, only one event (A , or B) could be occurred
- Sequence of A_1, A_2, \dots are called mutually exclusive if only one of them could occur at a time instance:

$$A_i \cap A_j = \emptyset$$

- If the sequence of events A_1, A_2, \dots have the following property, then we call them monotone

$$A_1 \subset A_2 \subset A_3 \dots$$

■ Probability function (formal definition):

- Any function who makes a correspondence between event space (F) to real numbers (R)



Random Variables (7)



■ Probability function (Cont.):

□ Any probability function must guarantee the three following conditions:

■ Probability value for every event is a non-negative number

□ $P(A) \geq 0$

■ The probability of the sample space equals to 1

□ $P(\Omega) = 1$

■ Based on the definition of mutual exclusive events, we have the following relation

□ $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$



Important Statistical Theorems

- Probability of event \emptyset is zero: $P(\emptyset) = 0$
- Probability of the uniting two mutual exclusive events is equal to adding their probability individually:

$$P(A \cup B) = P(A) + P(B)$$

- Probability of complement of A, equals to subtracting probability of A from 1: $P(\bar{A}) = 1 - P(A)$
- If event A is subset of event B, probability value of A is less than probability value of B: *If $A \subseteq B$ then $P(A) \leq P(B)$*
- If C is an event created by uniting A, and B, then probability of C equals to:

$$P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Final Marks About Random Variable (1)



- With the help of random variables, we can assign a real number for every event of the sample space
- Since events occur randomly, it is obvious that we consider random values for the every event
 - This is why we call them random variables
- Nevertheless, we could say that random variables are somehow a function
 - Example: In case of throwing a symmetric coin for two times, the sample space would be $\Omega = \{HH, HT, TH, TT\}$
 - If we consider our random variable X , as the number of heads:
$$X(\{HT\}) = X(\{TH\}) = 1, X(\{TT\}) = 0, X(\{HH\}) = 2$$
 - As you can see, we have assigned a real number for every member of the sample space



Final Marks About Random Variable (2)



■ Now, let's determine the probabilities for different values of our random variable

- We want to calculate $P(X = x_i)$
- Recall: x_i is member of sample space, and it could get those values we discussed in last slide $\in \{0,1,2\}$

$$P(X = 1) = P(\{HT, TH\}) = \frac{2}{4} = 0.5$$

$$P(X = 2) = P(\{HH\}) = \frac{1}{4} = 0.25$$

$$P(X = 4) = P(X < 0) = P(1 < X < 2) = P(\emptyset) = 0$$

■ Formal definition of a random variable:

- Consider a random experiment with Ω as its sample space, F as its event space, and A as its simple events
- Random variable is a function which assigns a real value to every event, and is denoted with $X(A) \in R$

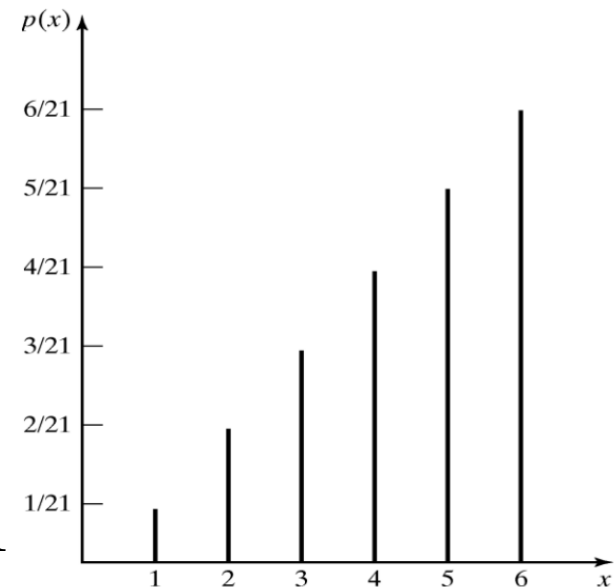




Example

- Assume a dice is loaded so that the probability of having a given face landed up is proportional to the number of spots showing
 - What is our random variable? The face that lands up
 - So, every face will be assigned to $\Omega \in \{1, 2, \dots, 6\}$
- In the following chart and table, the probability values are depicted for different values of X

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



- As you can see:

$$P(X = x_i) \geq 0 \text{ for all } i \quad \text{and} \quad \sum_{i=1}^6 P(x_i) = 1$$

Continuous Random Variables (1)



- First we discuss a little bit more about support (R_x)
 - Support (or range space) is a set of values which a random variable could get with a non-negative probability
 - Typically, the probability function is only indicated for values that are supported by R_x
 - For values out of R_x , probability function gives zero
- If X could be assigned with **any** values within a finite or infinite range, it is a continuous random variable
 - The range space R_x of a continuous variable is an **interval** or a **collection of intervals**
- Lifetime of a lamp bulb is an instance of continuous random variables



Continuous Random Variables (2)



- Probability of X having values in $[a,b]$ interval is calculated with the following:

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

- In this equation, $f(x)$ is the probability density function (PDF) for continuous random variable X
- Is it possible to calculate the probability in a specific point (y) (not an interval)?
 - First, we need to get familiar with another concept
- In statistics, F_X is known as the cumulative distribution function (CDF), and it indicates $F_X(y) = P(X \leq y)$

Continuous Random Variables (3)



- CDF is a continuous function:

$$F_X(y) = F_X(y^+) = F_X(y^-)$$

- In order to calculate $P(X = y)$:

$$P(X = y) = F_X(y^+) - F_X(y^-) = 0$$

- Therefore, the probability value for continuous random variables in a single point of the support set would be **zero**

- Any other option? This is where $f(x)$ comes in, and represents the derivative of the CDF → Does it really help?
- If X is a random variable with strict continuous CDF (F_X), a function f exists:

$$F_X(y) = \int_{-\infty}^y f(t) dt \rightarrow f_X(y) = F_X'(y)$$

Specifications of PDF (1)



- For all of the values of R_x , the PDF is equal or greater than zero: $f_X(x_i) \geq 0$, for all $x_i \in R_x$

- Integration of $f(x)$ for values over the support is 1:

$$\int_{R_x} f(x) d_x = 1$$

- Value of $f(x)$ for all of other values out of R_x is zero:

$$f_X(x_i) = 0, \text{ if } x_i \text{ is not in } R_x$$

- Based on what we have discussed, even with using the PDF, the probability of a continuous random variable in a specific point is zero:

$$P(X = x_0) = P(x_0 \leq X \leq x_0) = \int_{x_0}^{x_0} f(x) d_x = 0$$

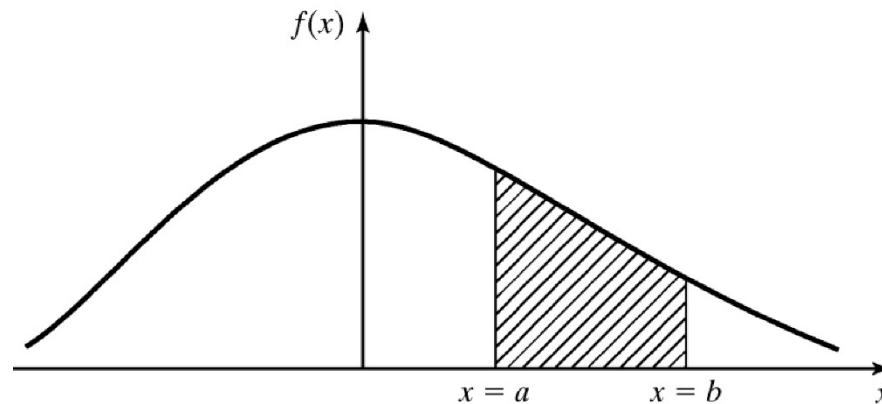
Specifications of PDF (2)



- Note: probability of X between a , and b does not depend on openness or closeness of the boundaries

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

- According to the principles of integration
 - In order to obtain the probability of continuous variable X in $[a,b]$, we can use the area under the PDF curve



Example



- Life of an inspection device is given by X , a continuous random variable with a PDF depicted in the following:

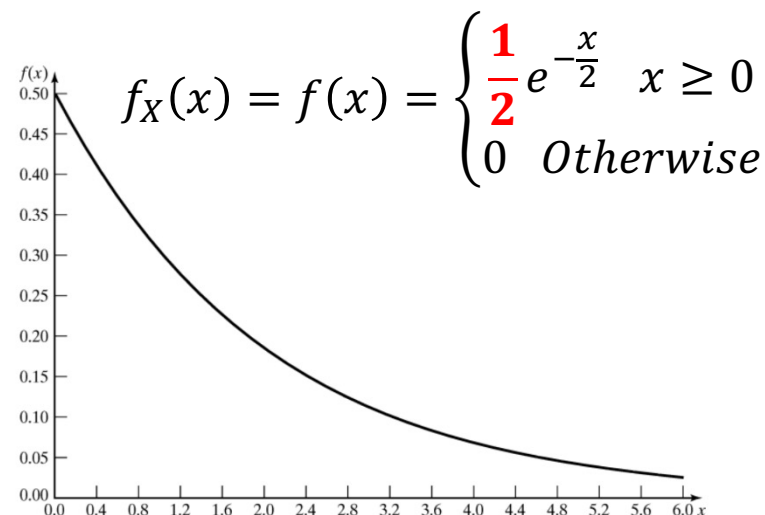
- As we will discuss later, this random variable has an exponential distribution

- Based on its PDF, it has a mean lifetime of 2 years

- This number could be obtained by reversing the value of λ

- We will discuss about the specifications of the distribution functions later in this chapter

- Calculate the probability of a lifetime between 2 to 3 years?





Cumulative Distribution Function (1)

- Cumulative Distribution Function (CDF) is used whenever $P(X \leq x_i)$ is required
- As mentioned before, CDF of a random variable X is denoted by $F_X(x_i)$
- If X is discrete, then:
$$F_X(x) = \sum_{\substack{\text{all} \\ x_i \leq x}} P(x_i)$$
- If X is continuous, then
$$F_X(x) = \int_{-\infty}^x f(t)dt$$
- Most important specifications of CDF:
 - $F_X(x)$ is a strict ascending function: $a < b \rightarrow F_X(a) < F_X(b)$
 - Since $F_X(x_i) = P(X \leq x_i)$, and it adds up all of the probabilities of values less than x_i , it is obvious that with increasing x_i , value of $F_X(x_i)$ gets bigger and gets closer to 1

Cumulative Distribution Function (2)



■ Boundries of $F_X(x_i)$

□ Assume that R_x could range between $-\infty$, and $+\infty$ (any other values could be considered)

□ $F_X(-\infty) = P(X \leq -\infty) = 0 \rightarrow \lim_{x \rightarrow -\infty} F_X(x) = 0$

■ Because x cannot get values less than the lower boundary of R_x

□ $F_X(+\infty) = P(X \leq +\infty) = 1 \rightarrow \lim_{x \rightarrow +\infty} F_X(x) = 1$

■ Based on the principles of probability

□ Therefore, F_X could range between $[0,1]$

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

□ Example:

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

Example (1)

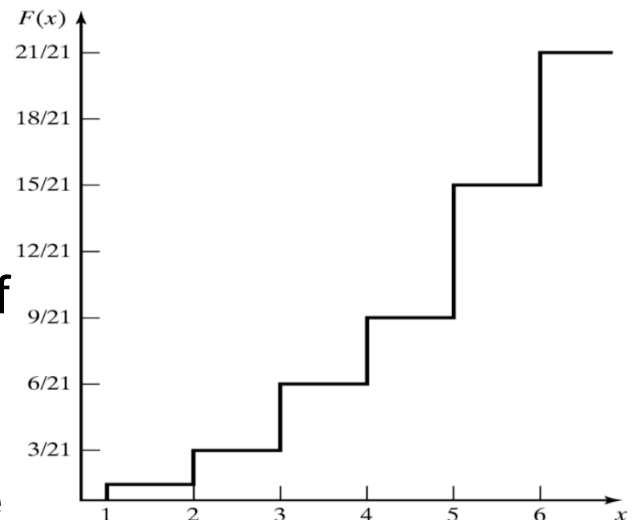


- Recall our dice-tossing example in previous slides
- As you can see, the table of values are different from what we have seen for PDF

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

- Notes:

- Since values less than 1 are not part of R_x , value of F_X would be zero
- On the other hand, for values greater than the maximum bound, F_X would be equal to 1
- Similar to PDF, the data in the table could be represented as a chart





Example (2)

- In our previously mentioned inspection device, the PDF of the lifetime was defined as an exponential distribution:

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

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

- First obtain CDF based on our PDF

$$F_X(x) = \int_{-\infty}^{+\infty} f(x) dx = \int_0^x \frac{1}{2} e^{-\frac{t}{2}} dt = 1 - e^{-\frac{x}{2}}$$

$$\rightarrow P(0 \leq X \leq 2) = F_X(2) - F_X(0) = 1 - e^{-\frac{2}{2}} = 0.63$$

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

$$P(2 \leq X \leq 3) = F_X(3) - F_X(2) = (1 - e^{-\frac{3}{2}}) - (1 - e^{-\frac{2}{2}}) = 0.14$$



- Statistical indexes are parameters or criteria, which are used to compare at least two statistical populations
- Indexes are generally grouped into three categories
 - Measures of central tendency
 - Mode, median, mean (also known as expected value), quantile (quartile, decile, percentile)
 - Measures of dispersion
 - Variance, and standard deviation
 - Measures of relative dispersion
 - Coefficient of variation, coefficient of skewness, and coefficient of Kurtosis



Measures of central tendency

- Any index, which determines the center of the population and is surrounded by other observations
- Expected value is one the most important central tendency measures
 - Assume random variable X
 - The expected value for X is denoted by $E(X)$, and is calculated as follows:
 - If X is discrete: $E(x) = \sum_{\text{all } i} x_i p(x_i)$ **Probability Mass Function**
 - If X is continuous: $E(x) = \int_{-\infty}^{\infty} x f(x) dx$ **Probability Distribution Function**
- $E(X)$ is only defined for values residing in R_x
 - $-\infty$, and $+\infty$ should be replaced with the boundaries if needed
 - Because, $f(x)$ gives zero outside of R_x



Expected Value (1)



- $E(X)$ talks about the values which may come in future
 - It may be considered identical to average (or mean) of the sampled data, which has happened until now
- $E(X)$ could be also represented with the first order momentum
 - Momentum in statistics is a quantitative value or a measure for determining the shape of a probability function or the distribution of random variables
 - In physics, momentum or torque is also used to define the specifications of an entity (mass, center of gravity, ...)
 - It is useful to exploit momentum to describe different aspects of a random variable
 - This is done with different orders of momentum over R_x

Expected Value (2)



- The more complicated the distribution function gets, more number of momentum orders are required for understanding its specifications
- The n^{th} order momentum for a real function $f(x)$ around c is defined as follows:

$$\mu_n = \int_{-\infty}^{+\infty} (x - c)^n f(x) dx$$

- If $c = \bar{x}$: it is called as the central momentum
- Typically, c is set as zero, when μ is called raw momentum (or the momentum around zero)
 - In this case, if $n=1$: $\mu_1 = \int_{-\infty}^{+\infty} xf(x)dx = E(X)$
 - This is why we say expected value is the 1st order momentum of the random variable X

Measures of Dispersion



- Based on the relation between $E(X)$, and momentum
 - The expected value tries to evaluate the **focus** of the intended population around the intended point
- But, there is a more important statistical criterion
 - Variance, which is denoted by $V(X)$, $VAR(X)$, σ^2
- While $E(X)$, and mean are important measures of tendency for evaluation of two populations around them, but we may face with two identical values for each population
 - In such cases, $E(X)$ cannot be used for comparison
 - We need to use measures of dispersion
 - They are used to compare the **dispersion** of populations around the center, and select the desired set of data

Variance and Standard Deviation



■ Variance

- One of the most important measures of dispersion
 - Indicates the deviation of observations (x_1, x_2, \dots) around the mean value

$$V(X) = \sigma^2 = \frac{\sum (x_i - \text{mean})^2}{n} = E \left[(x - E(X))^2 \right]$$

$$V(X) = \sigma^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n} \right)^2 = E(X^2) - E(X)^2$$

■ Standard Deviation

- Assume that the unit of observations is meters (m)
 - Hence, the unit of $V(X)$ will be m^2
 - It is complicated to do the compare the dispersion between populations → to solve this issue, we use the square root of $V(X)$
 - It is called the standard deviation, which gives us a measure of dispersion with same unit as our data (and its mean)




Example (1)



- The expected value (mean) of lifetime of the previous inspection device is calculated as:

$$f_X(x) = f(x) = \begin{cases} \frac{1}{2} e^{-\frac{x}{2}} & x \geq 0 \\ 0 & \text{Otherwise} \end{cases} \quad E(X) = \int_0^{+\infty} x \times \frac{1}{2} e^{-\frac{x}{2}} dx = \frac{1}{2} \int_0^{+\infty} x e^{-\frac{x}{2}} dx$$
$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Partial Integration  $E(X) = -x e^{-\frac{x}{2}} [0, +\infty] + \int_0^{+\infty} e^{-\frac{x}{2}} dx = 2$

- As you can remember, we mentioned that in exponential distribution, the mean value could be obtained by inverting the rate (λ)

- Here, $\lambda = \frac{1}{2} \rightarrow E(X) = \text{mean} = 2$





Example (2)

- The variance, and standard deviation is calculated as follows:

$$E(X^2) = \frac{1}{2} \int_0^{+\infty} x^2 e^{-\frac{x}{2}} dx = -x^2 e^{-\frac{x}{2}} [0, +\infty] + \int_0^{+\infty} e^{-\frac{x}{2}} dx = 8$$

- Recall: $V(X)$ can be calculated as $E(X^2) - E(X)^2$
 - $8 - 2^2 = 4$
- The standard deviation is the square root of $V(X)$
 - $\sigma = 2$

Useful Statistical Models



- In the rest of this chapter, the well-known statistical distributions for discrete and continuous random variables are introduced
 - Their specifications will be discussed
- Prior to this, a number of 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 (1)



- In a queueing systems, **interarrival** and **service-time** patterns can be probabilistic
 - For more queueing examples, see Chapter 2
- Sample statistical models for interarrival or service time distribution:
 - Exponential distribution
 - If service times are completely random
 - In other words, it is unclear when the customers arrive or how long their service will take
 - Normal distribution
 - If there isn't any difference between the probability of different values
 - Fairly constant but with some random variability (either positive or negative)





- Sample statistical models for interarrival or service time distribution (Cont.):
 - Truncated normal distribution
 - Similar to normal distribution but with **restricted boundaries**
 - The values of the random variable are bounded
 - In normal distribution, R_x could range from $-\infty$ to $+\infty$, but we may have an entity where its values are only positive (or negative)
 - Gamma and Weibull distributions
 - More general than exponential
 - Their mathematical presentation has more parameters for tuning and adjusting to different variables
 - Different values of parameters results in different distribution shapes
 - They are more complicated, but gives us more freedom to apply to different applications



Inventory and Supply Chain (1)



SHARIF
UNIVERSITY OF
TECHNOLOGY

- In realistic inventory and supply-chain systems, there are at least three random variables:
 - The number of units demanded per order or per time period
 - The time between demands
 - The lead time
- Sample statistical models for lead time distribution:
 - Gamma
- Sample statistical models for demand distribution:
 - Poisson:
 - Since it is a discrete distribution, it is used for modeling number of demands in a period, or number of units per order
 - It is simple with only a single parameter for tuning with the sampled data



Inventory and Supply Chain (2)



SHARIF
UNIVERSITY OF
TECHNOLOGY

- Sample statistical models for demand distribution (Cont.):
 - Poisson:
 - Its CDF is available in form of tables
 - Even without tables, due to simplicity of calculations with its equation, the probability values could be simply obtained
 - Negative binomial distribution:
 - Creates longer tail than Poisson
 - Useful for more large demands
 - Geometric:
 - A special case of negative binomial given **at least one demand has been received**
- Based on the specifications of the intended random variable, suitable distribution is selected





- Reliability is the conditional probability of correct functioning of a device at t , while it was correctly working at the beginning of our analysis ($t = 0$)
 - Denoted with $R(t)$, and reported for $(0, t]$ interval
- Time to failure (TTF) is one of the most important parameters for reliability analysis
 - Indicates the time between two consecutive failures
- If failures occur randomly, TTF is modeled with exponential distribution
- If standby redundancy is considered for different parts of the system with **each** having exponential TTF
 - TTF of the **entire system** is modeled with Gamma



Reliability and Maintainability



SHARIF
UNIVERSITY OF
TECHNOLOGY

- If failure is due to a large number of defects in a system of components
 - The Weibull distribution could be used
- If failures occur due to wear
 - The normal distribution will be used





- For cases with limited data, some useful distributions are:
 - Uniform, triangular and beta
- Other distributions include Bernoulli, binomial and hyper-exponential
- The important issue is to select the best-fit distribution based on the collected data
 - To model the intended variable with high precision
 - In case we realized the distribution does not act well, change it or try to tune its parameters
 - More information in Chapter 8
 - This skill could be improved with more practice and using your previous experience

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
- Note: Many distributions are constructed based on a basic model
 - Example: Tossing a coin and observing H/T is a random variable with Bernoulli distribution
 - If we iterate this experiment and define a variable which indicates number of H/Ts in n iterations \rightarrow a variable with binomial distribution



Bernoulli Trials and Bernoulli Distribution (1)



■ Bernoulli Trial

- An experiment with only 2 outcomes: success or failure
- If we define X as our Bernoulli random variable (outcome of the Bernoulli trial):
 - Let $X = 1$ if the experiment is a success
 - and $X = 0$ if the experiment is a failure
 - The Bernoulli distribution:
 - Probabilities in j -th trial:
$$P(x_j) = \begin{cases} p, & x_j = 1, j = 1, 2, 3, \dots, n \\ 1 - p = q, & x_j = 0, j = 1, 2, 3, \dots, n \\ 0, & \text{Otherwise} \end{cases}$$
- Example:
 - Tossing a homogenous coin, observing $H \rightarrow p=1/2$, and $q=1/2$
 - We may have an unusual coin with other probability values, but we still have Bernoulli trial
 - Throwing a dice 2 times, observing same values $\rightarrow p=1/6$, and $q=5/6$

Bernoulli Trials and Bernoulli Distribution (2)



- Expected value for a random variable with Bernoulli distribution:
 - $E(X) = \sum_{i=0,1} x_i \times P(x_i) = 1 \times p + 0 \times (1 - p) = p$
- Variance value for a random variable with Bernoulli distribution:
 - $V(X) = E(X^2) - E(X)^2 = p - p^2 = p \times (1 - p) = p \times q$
- Bernoulli process:
 - If we repeat n Bernoulli trials, where trials are independent, we call it a Bernoulli process
 - To obtain the probability of a sequence of Bernoulli events:
 - $P(x_1, x_2, \dots, x_n) = P(x_1) \times P(x_2) \times \dots \times P(x_n)$

Binomial Distribution (1)



- Consider a **single** Bernoulli trial with a constant success probability p
 - If we assume the **number of successes** in **n trials** as a random variable X , it will have Binomial distribution
 - Note: The number of trials n is constant
- The probability function of X is defined as the following:

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

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

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



Binomial Distribution (2)



- In the PMF equation, it is obvious that X cannot get values more than n
 - Because we have repeated the trials for n times
 - So, for $x > n$, the value of $P(x)=0$
- Expected value for random variable X is obtained according to the following:

$$E(X) = \sum_{\text{for all } i} x_i P(x_i) = np$$

- Variance could be also calculated based on the previous equations

$$V(X) = n \times p q$$



Binomial Distribution Example (1)

- Assume that in the process of manufacturing ICs, there is a 20% probability of having a defected chip
 - Every day, a group of 50 ICs are sampled randomly
 - If they observe more than 2 broken ICs, the production process will be shut down
- Calculate the probability of shut down in this manufacturing plant?
 - First we need to determine the random variable
 - Since every IC could be produced correctly or with defect, each instance of production is a Bernoulli trial
 - Since the **number of failed productions** is important to us (a number of failed Bernoulli trials), we are facing with a binomial distribution
 - Since we want to count the **number of defected ICs**, we are facing a **discrete** variable



Binomial Distribution Example (2)



- Therefore, we have 50 Bernoulli trials in this example
 - Assume that $p=0.02$ is indicating the probability of failure, and $q=0.98$ is representing the correct IC
 - To calculate the probability of x failures in 50 trials:

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

- We want to obtain probability of shutdown (having more than two IC defects)

$$P(X > 2) = 1 - P(X \leq 2) = 1 - \sum_{x=0}^2 \binom{50}{x} (0.02)^x (0.98)^{50-x} = 1 - [(0.98)^{50} + 50(0.02)(0.98)^{49} + 1225(0.02)^2(0.98)^{48}] = 1 - 0.92 = \mathbf{8\%}$$

Geometric Distribution (1)



- In binomial distribution, the random variable is the **number of successes** in **n repetitions** of Bernoulli trials with a constant success probability p
- Now assume we modify this a little bit:
 - **How many trials must be taken to reach the first success?**
 - Number of trials will be our random variable with a geometric distribution
 - Total number of trials is known and is constantly n
- Anytime we need to model the number of trials to reach the first success we use this distribution
 - Example: In a computer network, the probability of having the first collision after a number of packet receptions in a hub is modeled with geometric distribution



Geometric Distribution (2)



- The probability function of geometric distribution is represented as the following:

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

- p is indicating the first success occurring on the xth trial, and after x-1 trials, which were failed
 - q is the fail probability
- In order to obtain $E(X)$, and $V(X)$, use the equations below:

$$E(X) = 1/p$$

$$V(X) = q/p^2$$

Negative Binomial Distribution (1)



- It is used to model the number of Bernoulli trials required (X) to reach k^{th} success (or failure)
 - In other words, how many times we should repeat the trials to have k successes (or failures)
 - Probability function for random variable X with parameters p, and k is defined as follows:

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

- Values of expected value, and variance for random variable X is calculated with:

$$E(X) = \frac{k}{p} \quad V(X) = kq/p^2$$





Example

- 1) It has been reported that 40% of the ink-jet printers are returned after production in the quality-check section
 - In what chance, having the first approved printer is the 3rd printer?
 - Answer: we want to have the first success after 3 failures
 - **Geometric distribution** with $p=0.6$, and $q=0.4$
- 2) If we desire to calculate the probability of approving the 2nd printer on our 3rd inspection
 - **Negative binomial distribution** with $p=0.4$, and $k=2$

$$P(X = 3) = \binom{3-1}{2-1} \times (0.4)^{3-2} (0.6)^2 = \binom{2}{1} \times 0.4 \times (0.6)^2 = 0.288 = \sim 30\%$$



Poisson Distribution (1)



- Similar to binomial distribution, the number of successes is important in Poisson distribution
 - **But, in a temporal or spatial unit**
- In other words, if in random experiments, the events are dependent on temporal or spatial interval, the variable could be modeled with Poisson
- Examples:
 - Number of customers entering a market during 8:00-9:00 A.M.
 - Number of pizza orders during a working day
 - Number of packet drops in a router gateway in a week
 - Number of vehicles passing a specific 1Km of a highway



Poisson Distribution (2)



- A random variable with Poisson distribution must meet the three following conditions:
 - Events occurring in different intervals are independent
 - This is also known as **memoryless** aspect of Poisson distribution
 - When an event occurs, it does not have any effect on the probability of the next event
 - Probability of having more than one event at the same time is zero
 - Probability of an event in an interval is proportional to the size of the interval
- When to use Poisson?
 - Modeling events, which their **average interarrivals are known**, but the **exact instance of occurrence is unknown**, and is **random**
 - So, the events will be stochastic



Poisson Distribution Function (1)



- Recall: Poisson is similar to binomial distribution, where the number of successes is proportional to the number of trials n
 - Therefore, the probability of success P_n will be also dependent on n (n could be assumed to be our interval 😊)
 - What is the relation between P_n and n ? $n \times P_n = \alpha$
 - Where α is indicating the number of events occurred in an interval
 - Note: If P_n is constant \rightarrow number of success has **binomial distribution**
 - Accordingly, let's obtain the Poisson distribution function
 - Consider X to be the number of successes
 - Based on binomial distribution, we have:

$$P(X = x) = \underbrace{\binom{n}{x}}_A \underbrace{\left(\frac{\alpha}{n}\right)^x}_{\text{dashed box}} \underbrace{\left(1 - \frac{\alpha}{n}\right)^{n-x}}_{\text{dashed box}} \quad \text{B} \quad \boxed{\text{I}}$$

Poisson Distribution Function (2)



- Consider part **A** of the equation

- If n moves towards ∞ :

$$\lim_{n \rightarrow \infty} \frac{n!}{x! (n-x)!} \times \left(\frac{\alpha}{n}\right)^x = \frac{\alpha^x}{x!} \lim_{n \rightarrow \infty} \frac{n!}{(n-x)! \times n^x} = \frac{\alpha^x}{x!} \quad \text{II}$$

- Consider part **B** of the equation

- If n moves towards ∞ :

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\alpha}{n}\right)^{n-x} = e^{-\alpha} \quad \text{III}$$

- According to I, II, and III:

$$P(X = x) = e^{-\alpha} \frac{\alpha^x}{x!} \quad x = 0, 1, 2, \dots$$

Poisson Distribution Function (3)



- PDF of the Poisson distribution is defined as follows:

$$P(x) = \begin{cases} \frac{e^{-\alpha} \alpha^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

- Where $X=x$ indicates the number of successes (observations) in a Poisson experiment
- It is called that X has a Poisson distribution with parameter α
- α could be obtained by $\alpha = \lambda t$
 - Where λ is representing the **rate of occurrence** for the Poisson events

$$\lambda = \frac{\text{Number of Events}}{\text{Time Unit}} \rightarrow \alpha = \frac{\text{Number of Events}}{\text{Time Unit}} \times t$$

Time period which you want to analyze
or observe the random variable



Example



- An insurance company signs 3 contracts per week
 - What is the probability of signing at least one contract in one week?
 - Answer: Our random variable X is the number of contracts in an interval
 - So, Poisson is one candidate
 - But, does X support three conditions of Poisson distribution?
 - Only one contracts could be signed at a time ☺
 - Signing a contract is independent from other contracts ☺
 - More time indicates more contracts to be signed ☺
- First determine the $\alpha = \lambda t = \frac{3 \text{ contracts}}{1 \text{ week}} \times 1 \text{ week} = 3$
- $P(X > 0) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = 95\%$

Yes



Poisson Distribution Function (4)

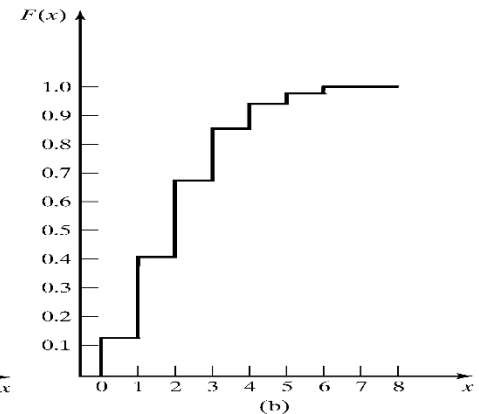
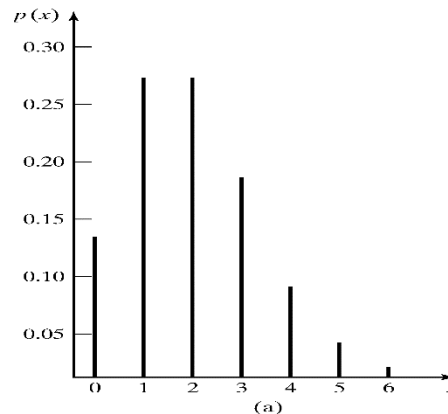


- The expected value, and variance of a random variable with Poisson distribution is obtained by the following:

$$E(X) = V(X) = \lambda t = \alpha$$

- The CDF of this distribution is calculated by: $F(x) = \sum_{i=0}^x \frac{e^{-\alpha} \alpha^i}{i!}$

- For $\alpha = 2$, the PDF, and CDF graphs for different success values (i in the above) are depicted



- Continuous random variables can be used to describe random phenomena in which the variable can take **any values in an interval**
 - The distributions that are used to model such variables are continuous distributions
- In this section, the distributions that will be studied are:
 - Uniform
 - Exponential
 - Normal
 - Gamma
 - Erlang
 - Weibull
 - Lognormal
 - Triangular
 - Beta

Uniform Distribution (1)



- This distribution could be used for both discrete and continuous variables
 - This depends on the support and the type of our intended variable
 - While discrete uniform variables can take values $S = \{1, 2, 3, \dots, n\}$, continuous uniform variables can take any values in (a, b) interval
- Why they are called uniform?
 - Because their probability function/density function gives identical values for any value in the support interval
 - Recall: we use probability function for discrete variables, and density function for continuous variables



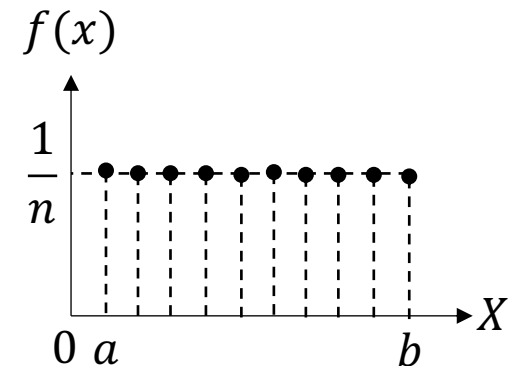
Uniform Distribution (2)



- If X is a discrete uniform variable with $S = \{1, 2, 3, \dots, n\}$, the probability function of this variable is denoted with:

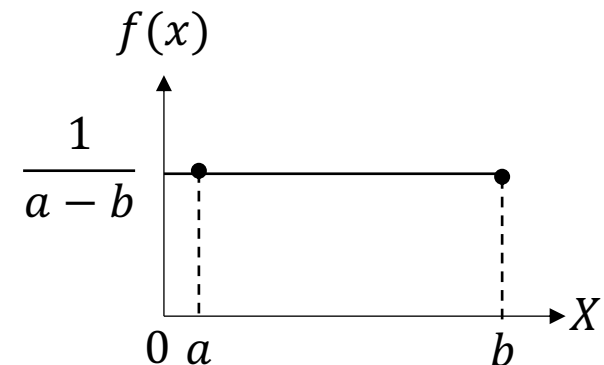
$$f_X(x) = P(X = x) = \frac{1}{n}, \quad x \in S$$

- As you can see, all the values are $1/n$



- If X is a continuous uniform variable with $S = (a, b)$, the PDF of this variable is denoted with:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$



Uniform Distribution (3)



- Value of $E(X)$ in the uniform distribution is an instance between the two end points of the support: $E(X) = \frac{a+b}{2}$
- The variance is obtained by: $V(X) = \frac{(b-a)^2}{12}$
- Since the variable is continuous, we use integration to obtain CDF from $f_X(x)$:

$$F_X(x) = P(X \leq x) = \int_a^x f_X(x) dx = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}, a \leq x \leq b$$

- It is obvious that for values less than a , the CDF gives zero, and for values greater than b , it gives 1
 - So, the uniform CDF is defined as:

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

Example



- Assume an athlete, who tries to shoot a 2 meter long line
 - What is the probability of having a distance less than 0.2m to the center of this line?
- Answer:
 - Our random variable here is the distance to the center of the line
 - So, X has a uniform distribution in $(0,2)$
 - The center of the line stands at 1 meter
 - Therefore, the distance to the center of the line would be $|X-1|$

$$P(|X - 1| \leq 0.2) = P(-0.2 \leq X - 1 \leq 0.2) = P(0.8 \leq X \leq 1.2) = F_X(1.2) - F_X(0.8)$$

$$F_X(x) = \frac{x - a}{b - a}, a = 0, \text{ and } b = 2$$

$$\rightarrow P = \frac{1.2 - 0}{2 - 0} - \frac{0.8 - 0}{2 - 0} = 0.6 - 0.4 = 0.2$$

Exponential Distribution (1)

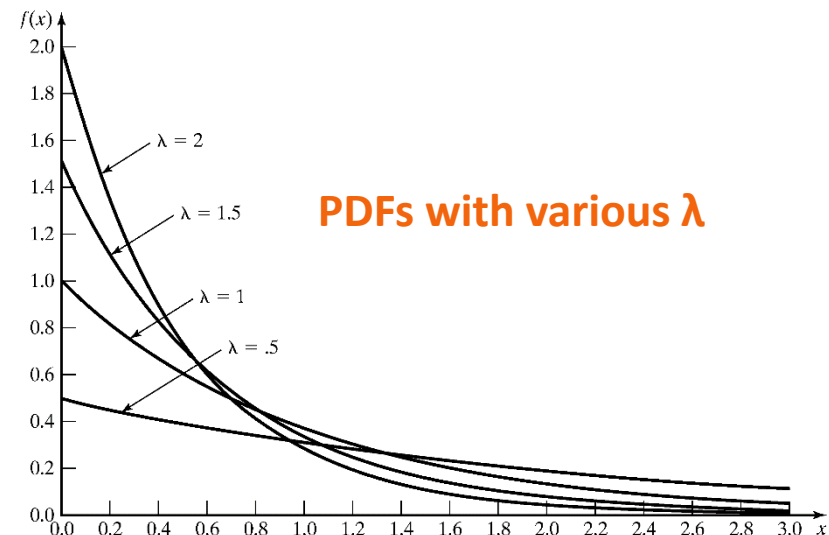


- If X is a continuous random variable with two conditions, it follows the exponential distribution:

- It has a **non-negative** real support
- Its probability density function is:
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

- We say X has an exponential distribution with parameter λ

- Used to model interarrival times when arrivals are completely **random**, and to model service times that are **highly variable**



Exponential Distribution (2)



- In Poisson experiments, the number of events (success or failure) in a specific interval is important
 - But, if we define X as the **time to the occurrence of events** (success or failure), then it follows the exponential distribution



- In Poisson, λ indicates the number of events in a time unit
 - Therefore, the time between two events could be obtained by $1/\lambda$
 - Indicated with θ
- The PDF of exponential distribution could be also defined as:

$$f_X(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}, x \geq 0$$

Exponential Distribution (3)

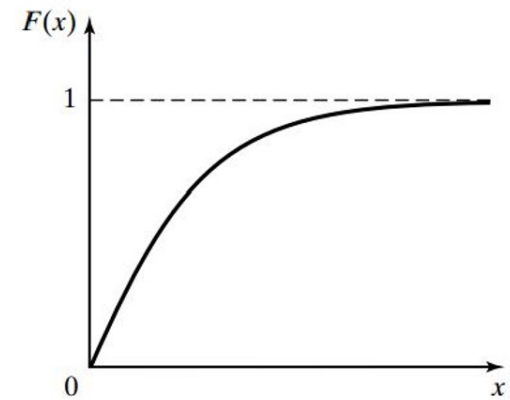


- Examples for this distribution:
 - The time a customer waits in a bank to make a deposit
 - The length of a fabric to be inspected for finding a tear
 - Life-time of ICs (or anything else)
- Exponential distribution has also the **memoryless** property
- Expected value, and the variance in exponential distribution are calculated with the following equations:

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

- The CDF of this distribution:

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



Example



- On average, there is 1 crack in every 2 meters of a 6m grinder
 - If we define X as the length of the grinder, which does not include cracks, what is the probability that he grinder does not have any cracks?
- Answer:
 - First, we are talking about an **interval**, the first candidates for X would be **Poisson**, and **exponential**
 - Based on the problem, if we consider that the first crack stays on zero point of the grinder, and the second occurs at a length longer than 6 meters → There is an interval → exponential 😊

$$\left. \begin{aligned} P(X \geq 6) &= 1 - P(X < 6) = 1 - F_X(6) = 1 - (1 - e^{-\lambda \times 6}) \\ \lambda &= \frac{1}{2} = 0.5 \end{aligned} \right\} \begin{aligned} P(X \geq 6) &= e^{-3} \\ &= 0.0498 \end{aligned}$$

Gamma Distribution (1)



- Gamma distribution has two parameters in its structure
 - Therefore, is one of the important distributions
- The random variable X has a Gamma distribution with parameters β , and θ and the following density function:

$$f_X(x) = \begin{cases} \frac{\beta\theta}{\Gamma(\beta)} (\beta\theta x)^{\beta-1} e^{-\beta\theta x}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

- Where β is the shape parameter (due to its effect on the shape of the PDF graph), and θ is the scaling parameter
 - With tuning these two parameters, we could better adjust the model with the sampled data
- **What is $\Gamma(\beta)$?**

Gamma Distribution (2)



- The Gamma function for all positive values of β is defined as the following:

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

- If we use partial integration: $\Gamma(\beta) = (\beta - 1)\Gamma(\beta - 1)$
- For integer values of $\beta > 0$, the value of the Gamma function could be obtained with $\Gamma(\beta) = (\beta - 1)!$
- The value of expected value, and variance are calculated with:

$$E(X) = 1/\theta \qquad V(X) = 1/\beta\theta^2$$

- In future chapters, we will discuss that for estimating proper values of parameters in Gamma distribution, we could estimate θ based on $E(X)$, and β based on $V(X)$

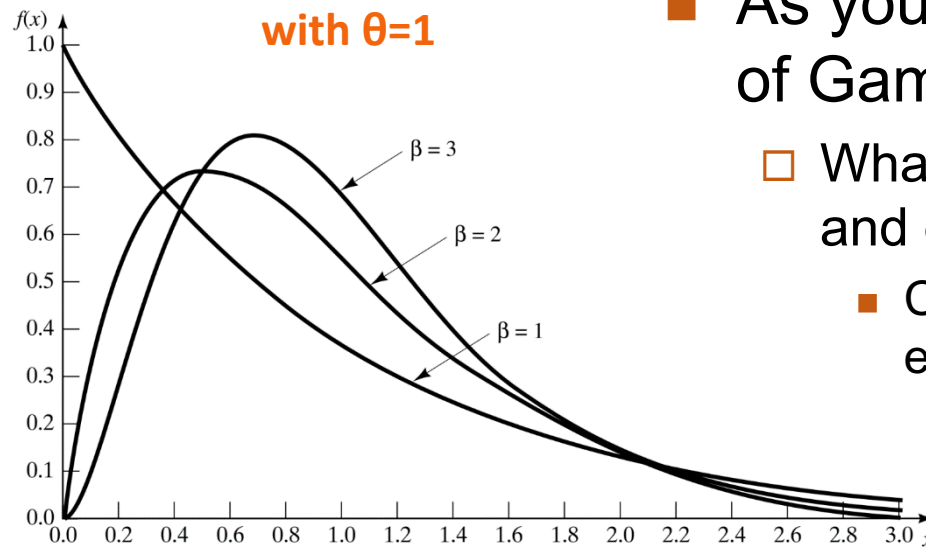
Gamma Distribution (3)



- The CDF for this distribution is determined as follows:

$$F_X(x) = \begin{cases} 1 - \int_x^{\infty} \frac{\beta\theta}{\Gamma(\beta)} (\beta\theta t)^{\beta-1} e^{-\beta\theta t} dt, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

PDFs for several
Gamma distributions
with $\theta=1$



- As you can see, for $\beta = 1$, the PDF of Gamma is similar to exponential

- What is the relation between Gamma and exponential?

- Consider X as the sum of β independent exponential variables with parameter $\beta\theta$

$$X = X_1 + X_2 + \dots + X_\beta$$

- X will support Gamma with θ , and β parameters

Erlang Distribution (1)



- The PDF of Gamma distribution is often referred to as the Erlang distribution of order (or number of phases) k , when $\beta=k$, is an integer
 - Therefore, Erlang is a more specific form of Gamma
- Erlang was a Danish scientist, who was an expert in telecommunications
- Erlang distribution and its relation with exponential distribution could be explained as follows:
 - A customer wants to get service from a telecommunication company
 - His/her request must pass through k stations
 - The next customer cannot apply for service until all k stations have served for the previous customer
 - Each station has an exponentially distributed service time (parameter $k\theta$)



Erlang Distribution (2)



- The CDF of the Erlang distribution is obtained by:

$$F_X(x) = \begin{cases} 1 - \sum_{i=0}^{k-1} \frac{e^{-k\theta x} (k\theta x)^i}{i!}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

- Expected value, and variance of random variable X with Erlang distribution is defined as follows:

$$E(X) = 1/\theta \qquad V(X) = 1/k\theta^2$$

- We have k exponentially distributed variables with $\lambda = k\theta$
- And expected value is a linear function:

$$E(X + Y) = E(X) + E(Y)$$
$$E(\alpha X) = \alpha E(X)$$

$$\Rightarrow E\left(\sum_{i=0}^{k-1} X_i\right) = k \times \left(\frac{1}{k\theta}\right)$$

Example (1)



- A college professor of electrical engineering is leaving home for the summer, but would like to have a light burning at all times to discourage burglars
 - The professor installs a device that will hold two light bulbs
 - The device will switch the current to the second bulb if the first bulb fails
 - The box in which the light bulbs are packaged says:
 - “Average life 1000 hours, exponentially distributed”
 - The professor will be gone 90 days (2160 hours)
- What is the probability that a light will be burning when the summer is over and the professor returns?



Example (2)



■ Answer:

□ The probability of a device working for more than x hours could be obtained by $P(X > x) = R(X > x) = 1 - F(X \leq x)$

■ $R(X)$ is indicating the reliability **somehow**

■ $F(X)$ is representing the probability of failure, which the device is working less than x hours

□ Easier to calculate based on CDF $F_X(X)$

□ In this example, the total operating time of the system is calculated by adding the operation time of two independent light bulbs ($X_S = X_1 + X_2$)

■ So, $k=2$

□ According to the problem, $\lambda = \frac{1}{1000} \rightarrow k\theta = \frac{1}{1000} \rightarrow \theta = \frac{1}{2000}$

$$F(X \leq 2160) = F_X(2160) = 1 - e^{-2.16} \sum_{i=0}^1 \frac{(2.16)^i}{i!} = 0.636 = \sim 64\% \rightarrow R(X) = \sim 36\%$$



Normal Distribution (1)



- It is also known as the natural distribution, bell graph Gaussian distribution, and Gauss-Laplace distribution
- Normal distribution is one the most important distributions available in statistics
 - Why? Because, many of the fluctuations in values given by natural, and physical events surrounding us in the environment are centered around a specific value
 - These values are highly adjusted with the outcomes of this model
- Many applications are supported by Normal distribution:
 - Height of people
 - Dimension of a specific production manufactured in a factory
 - Errors in measurements
 - Blood pressure



Normal Distribution (2)

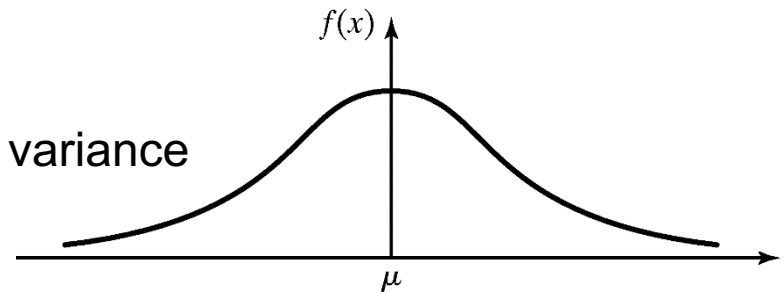


- It is indicated with $N(\mu, \sigma^2)$, and has two parameters:

- ☐ The mean value μ
- ☐ The standard deviation σ

- Note: For presentation, we use the variance

- PDF of the normal distribution:



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

- Special properties:

- ☐ $\lim_{x \rightarrow -\infty} f_X(x) = \lim_{x \rightarrow +\infty} f_X(x) = 0$
- ☐ PDF is symmetric about μ : $f_X(\mu - x) = f_X(\mu + x)$
- ☐ The maximum value of the pdf occurs at $x = \mu$
 - The mean and mode are equal

Normal Distribution (3)



- Evaluation of CDF is complex
- There are two major approaches for this issue:
 - Use numerical methods (no closed form)
 - Independence of μ and σ ($\mu=0$, $\sigma=1$) using the **standard normal distribution**: $Z \sim N(0,1)$
 - Transformation of variables

Normal Distribution (4)



■ Transforming the variables

- For random variable $X \sim N(\mu, \sigma^2)$, we recommend using a modified version of CDF, which transforms X to Z

$$F_X(x) = P(X < x) = P\left(\frac{X - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = P\left(\mathbf{Z} < \frac{x - \mu}{\sigma}\right) = \phi\left(\frac{x - \mu}{\sigma}\right)$$

- The PDF, and CDF of this distribution are defined as follows:

$$\varphi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < +\infty$$

$$\Phi_Z(z) = P(Z < z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

- To simplify the calculation of probability values for the standard normal distribution, look-up tables are used:

Z	0.00	0.01	0.02	...	0.05	...	0.09
0.0	<div>↓</div>						
0.1							
...							
1.5	<div>→</div>				0.9394		
...					$\Phi(Z < 1.55)$		
2.9							

Example (1)



- The time required to load an oceangoing vessel, X , is distributed as $\sim N(12, 4)$
 - Calculate the probability that the vessel is loaded in less than 10 hours?

- Answer:

- $\mu = 12$, and $\sigma = 2$, we are seeking $F_X(10)$

- Transforming X to Z : $z = \frac{10-12}{2} = -1$

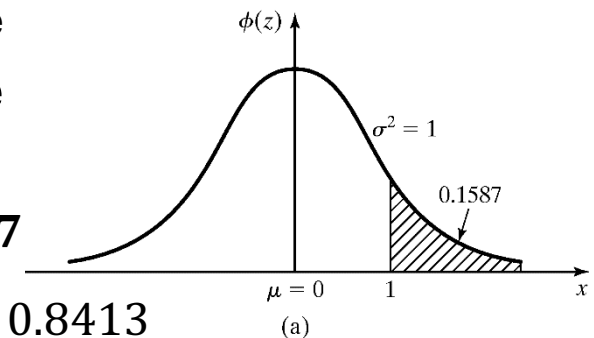
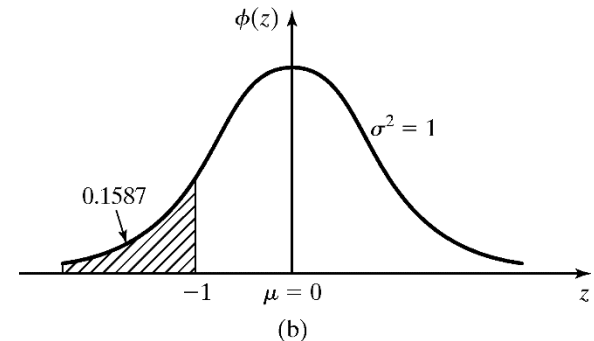
- We must calculate $\Phi_Z(-1) = P(Z < -1)$

- Negative values are not indicated in the table

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

$$P(Z < -1) = P(Z > 1) = 1 - P(Z < 1) = \mathbf{0.1587}$$

- From table (Page 540): $P(Z < 1) = \Phi_Z(1) = 0.8413$



Example (2)



- What is the probability that the vessel is loaded in 12 hours or more?

- Answer:

□ $\mu = 12$, and $\sigma = 2$, we are seeking $P(X \geq 12) = 1 - \boxed{P(X < 12)}$

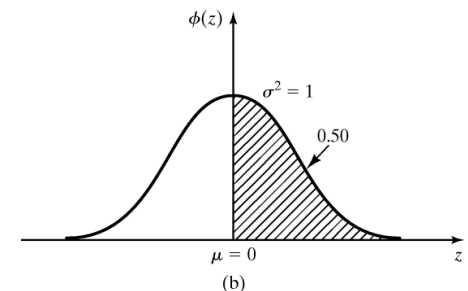
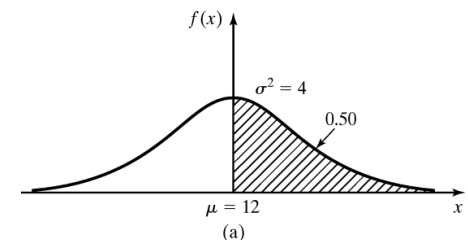
□ Transforming X to Z : $z = \frac{12-12}{2} = 0$

- According to the table, $F_X(12) = \Phi_Z(0) = 0.5$

$$P(X \geq 12) = 1 - \Phi_Z(0) = \mathbf{0.5}$$

□ This could be obtained without calculation:

- The mean value was 12
- We also wanted to evaluate values $X \geq 12$
- Due to the symmetric nature of normal graph, probability of having values higher or lower than the mean are the same = 50%



Example (3)

■ What is the probability $P(10 \leq X \leq 12)$?

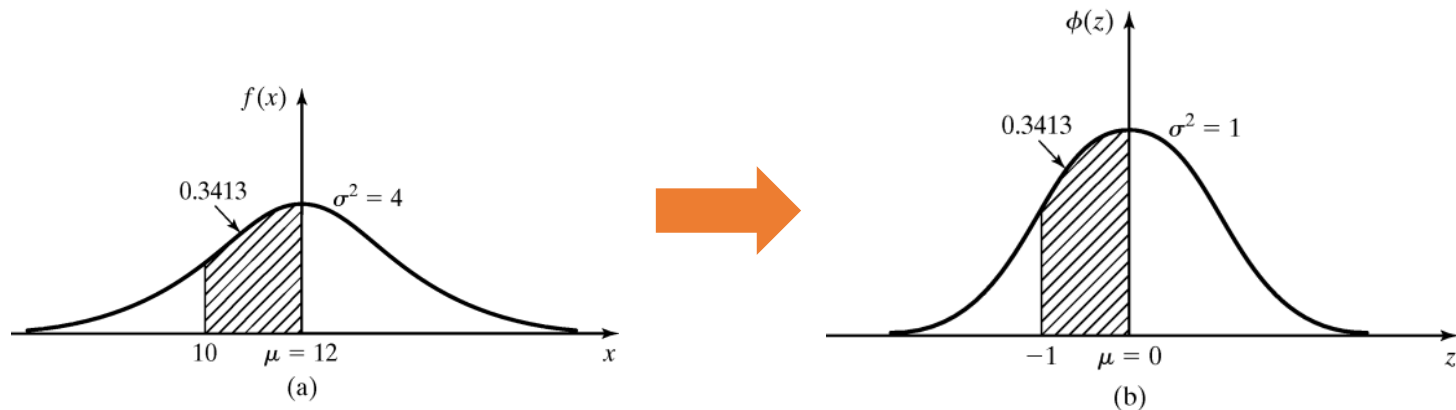
■ Answer:

□ $\mu = 12$, and $\sigma = 2$, we are seeking $P(10 \leq X \leq 12) = F_X(12) - F_X(10)$

□ Transforming X to Z : $z = \frac{12-12}{2} = 0$, and $z = \frac{10-12}{2} = -1$

■ Based on what we calculated previously

$$P(10 \leq X \leq 12) = \Phi_Z(0) - \Phi_Z(-1) = 0.3413$$



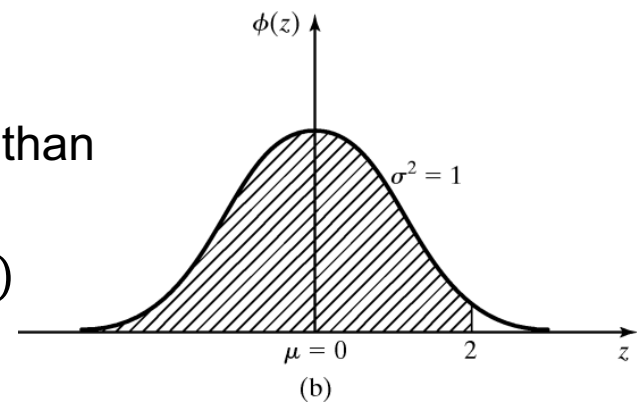
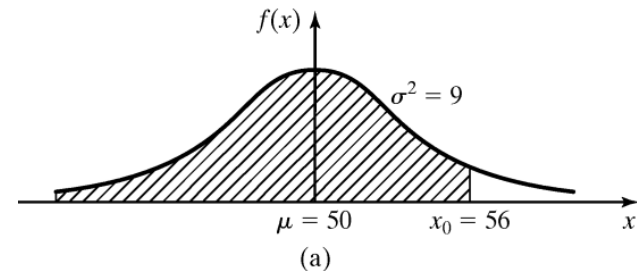


Example (4)

- Suppose a random variable $X \sim N(50, 9)$
 - Calculate $F_X(56)$
- Answer:
 - $\mu = 50$, and $\sigma = 3$
 - Transforming X to Z : $z = \frac{56-50}{3} = 2$

$$P(Z < 2) = \Phi_Z(2) = \mathbf{0.9772}$$

- With 98% chance, X gets values less than 56
- As you can see, the area under $\Phi_Z(2)$ has been nearly fully scratched



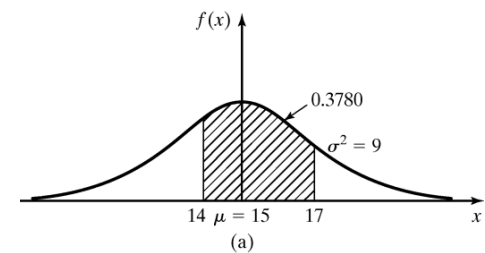
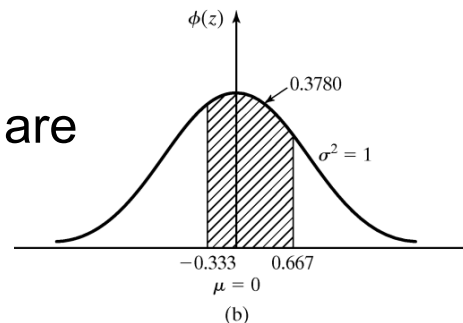
Example (5)



- The time required to pass a queue is modeled with $N(15,9)$
 - Calculate the probability that a customer spends 14 to 17 minutes in this queue
 - $X \sim N(15,9) \rightarrow \mu = 15$, and $\sigma = 3$
 - Transforming X to Z : $z = \frac{17-15}{3} = \frac{2}{3}$, and $z = \frac{14-15}{3} = -\frac{1}{3}$

$$P(14 \leq X \leq 17) = \Phi_Z(0.667) - \Phi_Z(-0.333) = \Phi_Z(0.667) - 1 + \Phi_Z(0.333) = 0.7476 - 0.3696 = \mathbf{0.3780}$$

- What is the point?
 - Obtaining the values, which are not explicitly indicated in the table, such as $\Phi_Z(0.667)$



Estimating Values Not Indicated in the Table



- There is no value corresponding to $\Phi_Z(0.667)$
 - But we have values for $\Phi_Z(0.66)$, and $\Phi_Z(0.67)$ in the table
- In order to obtain $\Phi_Z(0.667)$ we use a simple interpolation:

z_α	0.05	0.06	0.07	0.08	0.09	z_α
0.0	0.519 94	0.523 92	0.527 90	0.531 88	0.535 86	0.0
0.1	0.559 62	0.563 56	0.567 49	0.571 42	0.575 34	0.1
0.2	0.598 71	0.602 57	0.606 42	0.610 26	0.614 09	0.2
0.3	0.636 83	0.640 58	0.644 31	0.648 03	0.651 73	0.3
0.4	0.673 64	0.677 24	0.680 82	0.684 38	0.687 93	0.4
0.5	0.708 84	0.712 26	0.715 66	0.719 04	0.722 40	0.5
0.6	0.742 15	0.745 37	0.748 57	0.751 75	0.754 90	0.6
0.7	0.773 37	0.776 37	0.779 35	0.782 30	0.785 23	0.7
0.8	0.802 34	0.805 10	0.807 85	0.810 57	0.813 27	0.8
0.9	0.824 94	0.831 47	0.833 97	0.836 46	0.838 91	0.9
1.0	0.853 14	0.855 43	0.857 69	0.859 93	0.862 14	1.0
1.1	0.874 93	0.876 97	0.879 00	0.881 00	0.882 97	1.1
1.2	0.894 35	0.896 16	0.897 96	0.899 73	0.901 47	1.2
1.3	0.911 49	0.913 08	0.914 65	0.916 21	0.917 73	1.3
1.4	0.926 47	0.927 85	0.929 22	0.930 56	0.931 89	1.4
1.5	0.939 43	0.940 62	0.941 79	0.942 95	0.944 08	1.5
1.6	0.950 53	0.951 54	0.952 54	0.953 52	0.954 48	1.6
1.7	0.959 94	0.960 80	0.961 64	0.962 46	0.963 27	1.7
1.8	0.967 84	0.968 56	0.969 26	0.969 95	0.970 62	1.8
1.9	0.974 41	0.975 00	0.975 58	0.976 15	0.976 70	1.9
2.0	0.979 82	0.980 30	0.980 77	0.981 24	0.981 69	2.0
2.1						2.1
2.2						2.2
2.3						2.3
2.4						2.4
2.5						2.5
2.6						2.6
2.7	0.997 02	0.997 11	0.997 20	0.997 28	0.997 36	2.7
2.8	0.997 81	0.997 88	0.997 95	0.998 01	0.998 07	2.8
2.9	0.998 41	0.998 46	0.998 51	0.998 56	0.998 61	2.9
3.0	0.998 86	0.998 89	0.998 93	0.998 97	0.999 00	3.0
3.1	0.999 18	0.999 21	0.999 24	0.999 26	0.999 29	3.1
3.2	0.999 42	0.999 44	0.999 46	0.999 48	0.999 50	3.2
3.3	0.999 60	0.999 61	0.999 62	0.999 64	0.999 65	3.3
3.4	0.999 72	0.999 73	0.999 74	0.999 75	0.999 76	3.4
3.5	0.999 81	0.999 81	0.999 82	0.999 83	0.999 83	3.5
3.6	0.999 87	0.999 87	0.999 88	0.999 88	0.999 89	3.6
3.7	0.999 91	0.999 92	0.999 92	0.999 92	0.999 92	3.7
3.8	0.999 94	0.999 94	0.999 95	0.999 95	0.999 95	3.8
3.9	0.999 96	0.999 96	0.999 96	0.999 97	0.999 97	3.9

$$\Phi_Z(0.667) = \Phi_Z(0.66) + 7 \times \left(\frac{\Phi_Z(0.67) - \Phi_Z(0.66)}{10} \right) = 0.7476$$

- We have a 2 decimal precision in the table, but 3 decimal is required
 - The gap between values should be divided by 10 units

Example (6)



- Assume that the lead time for serving a request is following $N(25,9)$
 - Compute the lead time ($X = x$) that will be exceeded only in 5% of the time?

- Note: In this example, we have the probability, but x is unknown

$$P(X \geq x) = 1 - P(X < x) = 1 - F_X(x) = 1 - \Phi\left(\frac{x - 25}{3}\right) = 0.05$$

- Now we refer to the table and search for the z that gives $\Phi(z)=0.95$

- Similar to previous example, you cannot find any z which supports above

- Find two z_1 , and z_2 who cover 0.95

- $\Phi(z_1 = 1.64) = 0.94950$, and $\Phi(z_2 = 1.65) = 0.95053$

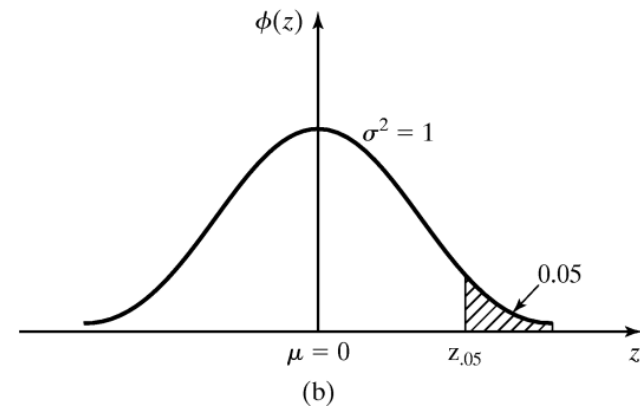
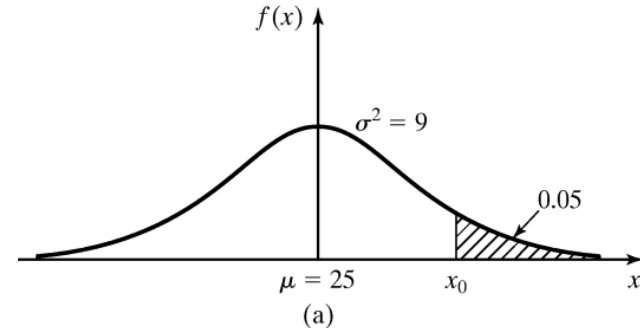
- Use **mean** value to obtain $\Phi(z) = 0.95 \rightarrow z = \frac{x-25}{3} = \frac{1.64+1.65}{2} = 1.645$

- According to the above, $x = 29.935 \rightarrow$ In 5% of the time, customers wait more than 29.935 minutes

Example (7)



- Checkout the indicated values in the graphs
 - Both, the standard deviation, and the mean
 - Worthy to emphasis that the area under the curve is identical in both graphs
 - So, the probability will be the same



Weibull Distribution (1)



- This distribution plays an important role in many applications, especially in analyzing reliable, and fault tolerant techniques
- If random variable X with continuous, and non-negative support, follows the indicated PDF, it could be modeled with Weibull distribution:
$$f_X(x) = \begin{cases} \frac{\beta}{\alpha} \left(\frac{x-\nu}{\alpha} \right)^{\beta-1} \exp \left[- \left(\frac{x-\nu}{\alpha} \right)^{\beta} \right], & x \geq \nu \\ 0, & \text{otherwise} \end{cases}$$
- It has 3 parameters:
 - Location parameter: ν
 - Shape parameter: β , ($\beta > 0$)
 - Scale parameter: α , ($\alpha > 0$)
- It could be well-tuned with different data sets with higher degree of freedom

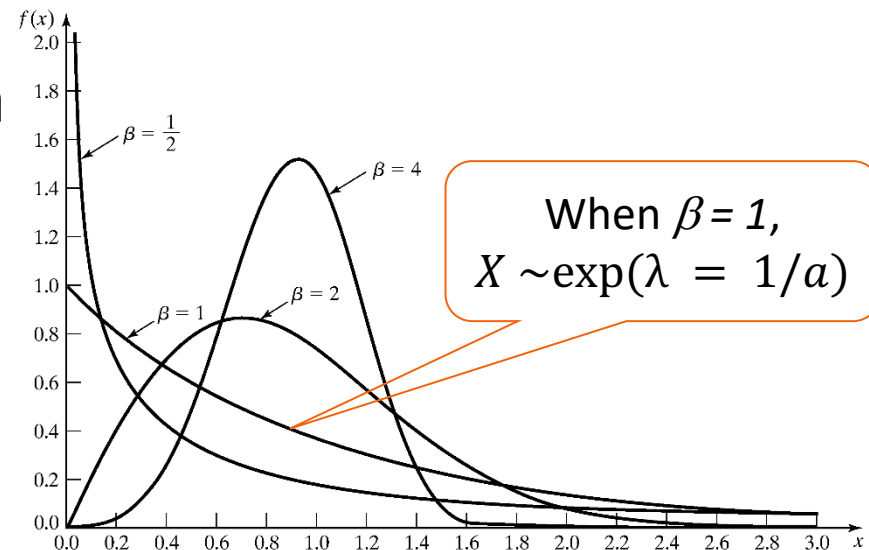
Weibull Distribution (2)



■ Example:

- Illustrated graph has been depicted with $v = 0$, and $\alpha = 1$
- As you can see, with tuning parameters of the PDF, different graphs with completely different features could be obtained
 - Highly flexible

- While the Weibull distribution could be adjusted with various sampled data, it is **computationally complicated**



Relation Between Weibull and Other Distributions



- As indicated in the previous slide, if $v = 0$, and $\beta = 1$, Weibull distribution turns into exponential:

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

- If $\alpha = 2$, the random variable X follows the Rayleigh distribution with the following PDF:

$$f_X(x) = \begin{cases} \frac{\beta}{2} \left(\frac{x - v}{2} \right)^{\beta-1} \exp \left[- \left(\frac{x - v}{2} \right)^{\beta} \right], & x \geq v \\ 0, & \text{otherwise} \end{cases}$$

- This version of the distribution is widely used in modeling multipath fading in wireless communications, electromagnetic radiation, wind speed, etc.

Weibull Distribution (3)



- Expected value and variance of the Weibull distribution are calculated via the following equations:

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

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

- Recall: $\Gamma(x)$ is the Gamma function that we have previously discussed
- The CDF is calculated as follows:

$$F_X(x) = \begin{cases} 1 - e^{-\left(\frac{x-v}{\alpha}\right)^\beta}, & x \geq v \\ 0, & \text{otherwise} \end{cases}$$

Example



- Assume the TTF of a device has a Weibull distribution with $v = 0$, $\alpha = 200H$ and $\beta = 1/3$

- Calculate the average TTF

$$\left. \begin{aligned} E(X) &= \cancel{v}^0 + \alpha \Gamma\left(\frac{1}{\beta} + 1\right) = 200 \times \Gamma(3 + 1) = 200\Gamma(4) \\ \Gamma(4) &= (4 - 1)! = 6 \end{aligned} \right\} E(X) = 1200H$$

- What is the probability of failure before 2000H?

$$\left. \begin{aligned} P(X < 2000) &= F_X(X = x) \\ x &= 2000H \geq v = 0 \end{aligned} \right\} P(X < 2000) = 1 - e^{-\left(\frac{2000-0}{200}\right)^{\frac{1}{3}}} = 1 - e^{-\frac{2}{15}} = 0.884$$

- **With 89% chance, the device will be failed in less than 2000 hours**





Log-Normal Distribution (1)

- It is used to model variables, which are changing in a cumulative fashion
 - Number of reviews regarding a gadget in Amazon
 - Amount of rainfall
 - Checkers play time
- There is a tight relation between log-normal and normal distributions
 - If there exists a random variable $Y \sim N(\mu, \sigma^2)$, a random variable X is also available, which $X = e^Y$
 - Then, X is called to support log-normal distribution with parameters μ, σ , and is denoted as $X \sim \text{Lognormal}(\mu, \sigma^2)$
 - Note: Parameters μ, σ^2 are not indicating the mean and variance of X , but they correspond to Y , and used **only for presentation**



Log-Normal Distribution (2)

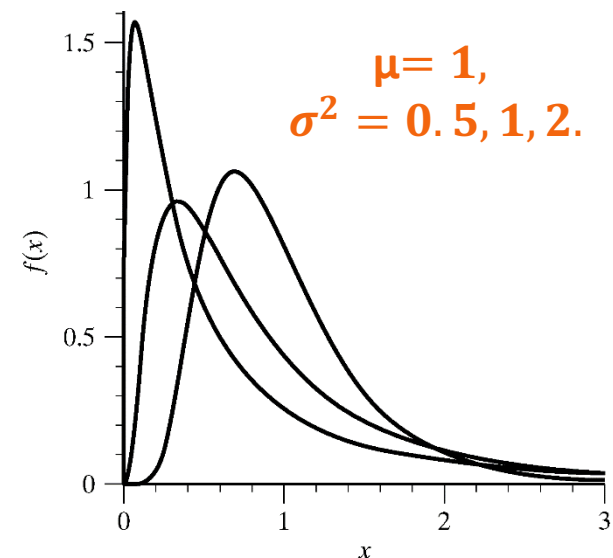


- The PDF of lognormal:
$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma x}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

- Expected value and variance:
$$E(X) = e^{\mu + \frac{\sigma^2}{2}} \quad V(X) = e^{2\mu + \frac{\sigma^2}{2}} (e^{\sigma^2} - 1)$$

- With tuning the parameters of this distribution, we could fit the model with various data samples

- This is in contrast with the normal distribution, which the structure of the PDF diagram does not change
 - It only gets shifted to sides (**constant σ , and variable μ**) or expansion/contraction (**variable σ , and constant μ**)



Triangular Distribution (1)

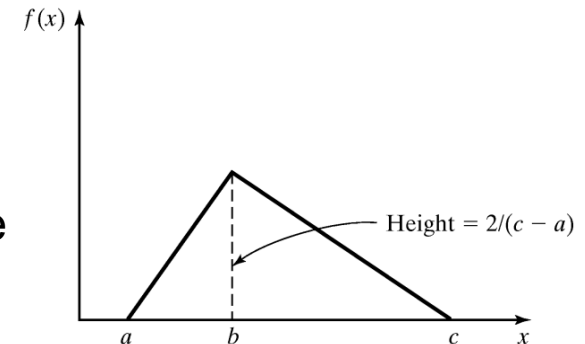


- A distribution used to model **fuzzy variables**

- In contrast with binary, variables can get more than two values
- These values are not necessarily numbers
 - They can be descriptive
 - Example: Energy level of a harvesting battery could be high, medium, or low
- With Fuzzy logic, the system decisions could be made in a similar way to human decision makings

- It has three parameters: $a \leq b \leq c$

- Lower bound a
- The mode b
 - The event with most frequency of occurrence
- Upper bound c



Triangular Distribution (2)



- PDF of the random variable X:

$$f_X(x) = \begin{cases} \text{Ascending linear function} \leftarrow \frac{2(x-a)}{(b-a)(c-a)}, & a \leq x \leq b \\ \frac{2(c-x)}{(c-b)(c-a)}, & b < x \leq c \\ \text{Descending linear function} \leftarrow 0, & \text{otherwise} \end{cases}$$

- To calculate the expected value for variable X, use the following equation:

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

- Mode has a more pivotal role than $E(X)$ in triangular distribution

□ It is calculated as: $Mode = b = 3E(X) - (a + c)$

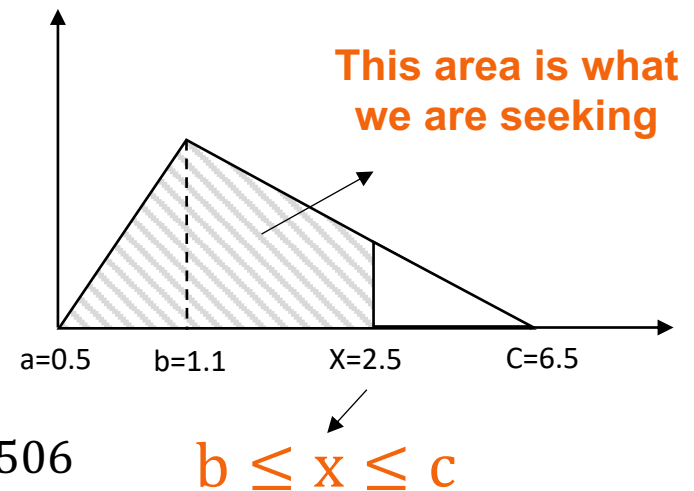


Triangular Distribution (3)



- CDF for the triangular distribution:
- Example: An embedded processor executes the tasks with a triangularly distributed make-span $\sim T(0.5\text{ms}, 1.1\text{ms}, 6.5\text{ms})$
 - What is the probability of executing a task in less than 2.5ms?

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ \frac{(x-a)^2}{(b-a)(c-a)}, & a < x \leq b \\ 1 - \frac{(c-x)^2}{(c-b)(c-a)}, & b < x \leq c \\ 1, & x > c \end{cases}$$

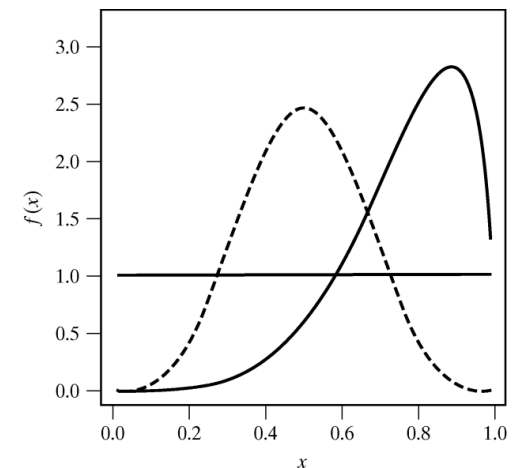


$$P(X < 2.5) = 1 - \frac{(6.5 - 2.5)^2}{(6.5 - 1.1)(6.5 - 0.5)} = 0.506$$

Beta Distribution (1)



- In cases, where random variable X ranges between 0, and 1, it **could be** modeled with Beta distribution
 - This is not a rule
 - Depends on the application and type of data
 - For instance, if the probability of all values are identical, uniform is preferred
- It has two shape parameters $0 < \beta_1, \beta_2$
- PDF of a Beta distributed variable X is defined as:
$$f_X(x) = \begin{cases} \frac{x^{\beta_1-1}(1-x)^{\beta_2-1}}{B(\beta_1, \beta_2)}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$
- B is the Beta function: $B(\beta_1, \beta_2) = \frac{\Gamma(\beta_1)\Gamma(\beta_2)}{\Gamma(\beta_1 + \beta_2)}$



Beta Distribution (2)



- What if the support is ($a \neq 0, b \neq 1$)?
 - We define a transformed random variable Y:

$$Y = a + (b - a)X$$

$$R_X = (0,1)$$

$$R_Y = (a, b)$$

- In order to obtain $E(X)$, and $V(X)$ we use the following:

$$E(X) = \frac{\beta_1}{\beta_1 + \beta_2} \quad V(X) = \frac{\beta_1 \beta_2}{(\beta_1 + \beta_2)^2 (\beta_1 + \beta_2 + 1)}$$

- Obtain the expected value and variance equations for random variable Y



Poisson Process (1)



- In many applications, a counting function $N(t)$ could be considered, which represents the number of events occurred in $[0, t]$
 - $N(t)$ is the number of observations $\in \{0, 1, 2, \dots\}$
 - $t=0$ indicates the instance, where our observations has started
- A counting process $\{N(t), t \geq 0\}$ is a **Poisson process** with mean rate λ if:
 - In every instance of time, only one arrival could occur
 - The counting function has **stationary increments**
 - Meaning that the number of arrivals in $[t, t+\Delta t]$ is only dependent on the length of interval (Δt) not the start of the interval (t)
 - The counting function has **independent increments**
 - Number of arrivals in two non-overlapping intervals are independent



Poisson Process (2)



- In a Poisson process, the probability of having n arrivals in $[0, t]$ interval is calculated as:

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

- Recall: We could have used α instead of λt
 - λ indicates the rate of arrivals in a time unit
 - α indicates the number of arrivals in a time interval
- $E(X)$, and $V(X)$ are identical in Poisson process:

$$E(N(t)) = V(N(t)) = \alpha = \lambda t$$

- The above equations are valid for cases where we want to analyze the process in an interval starting from $t=0$
 - What if we want to analyze a Poisson process in $[t, t + \Delta t]$, $t \neq 0$?



Poisson Process (3)



- If it is desired to consider the number of arrivals in an interval $[t, t + \Delta t]$, $t \neq 0$
 - According to the stationary incremental property of Poisson process, in order to obtain $P[N(t + \Delta t) - N(t)]$, we must use $\lambda \Delta t$:

$$P[N(t + \Delta t) - N(t)] = \frac{e^{-\lambda \Delta t} (\lambda \Delta t)^n}{n!}, t \geq 0, n = 0, 1, 2, \dots$$

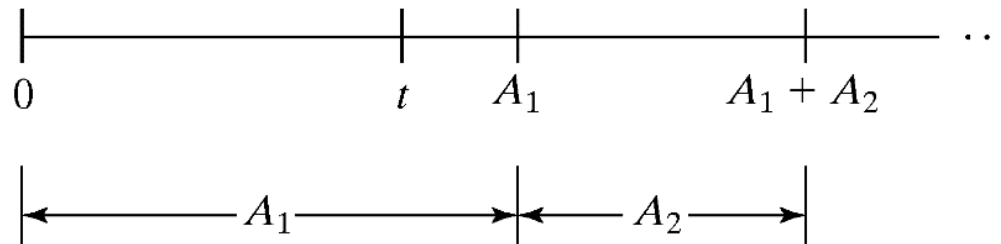
- Similarly, the $E(X)$, and $V(X)$, are calculated as:

$$E(N(t + \Delta t) - N(t)) = V(N(t + \Delta t) - N(t)) = \lambda \Delta t$$

Poisson Process (4)



- Now consider a Poisson process with a sequence of arrivals occurring on a temporal basis
 - 1st arrival happens on A_1 , 2nd on $A_1 + A_2$, 3rd on $A_1 + A_2 + A_3$, etc.
 - Arrivals are depicted below:



- Where A_i is the elapsed time between arrival $i-1$ and arrival i
 - We are talking about **interarrival times** here
- What should we do to calculate the probability that the 1st arrival occurs before $t \rightarrow P(A_1 \leq t)$

Poisson Process (5)



- Assume that no arrivals occur between 0, and t

□ First arrival occurs after t $\rightarrow A_1 > t$, and $N(t) = 0$

$$P(A_1 > t) = P(N(t) = 0) = P(n = 0) = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \xrightarrow[n! \rightarrow 1]{n \rightarrow 1} e^{-\lambda t}$$
$$\Rightarrow P(A_1 > t) = e^{-\lambda t}$$

- In order to calculate the probability that the 1st arrival occurs in $[0, t]$, the following equation is used:

$$P(A_1 \leq t) = 1 - P(A_1 > t) = 1 - e^{-\lambda t}$$

- This equation indicates that the interarrival times in a Poisson process are exponential with parameter λ

Arrival counts \longleftrightarrow Interarrival time
 $\sim \text{Poi}(\lambda)$ $\sim \text{Exp}(1/\lambda)$

Stationary & Independent

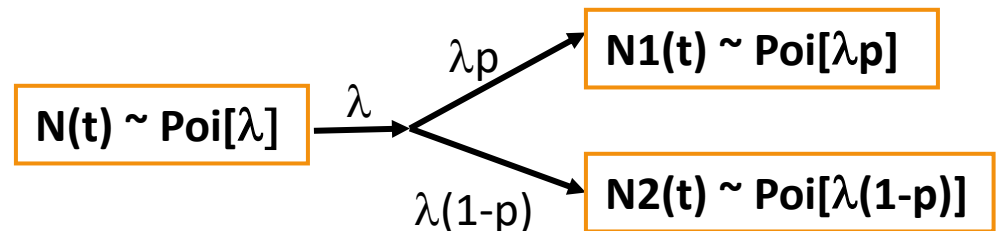
Memoryless



Splitting and Pooling (1)



- A Poisson process could be generated with **merging** two or more smaller Poisson process, or it could be **divided** into two or more processes
- Assume a Poisson process $\{N(t), t \geq 0\}$ with rate λ
- Splitting:
 - Suppose every event of this process can be classified as Type I, with probability p and Type II, with probability $1-p$
 - During $[0, t]$, $N_1(t)$ and $N_2(t)$ indicate the number of events corresponding to both of these types, where $N(t) = N_1(t) + N_2(t)$
 - It could be proven $N_1(t)$, and $N_2(t)$ are both Poisson processes with rates λp , and $\lambda(1 - p)$

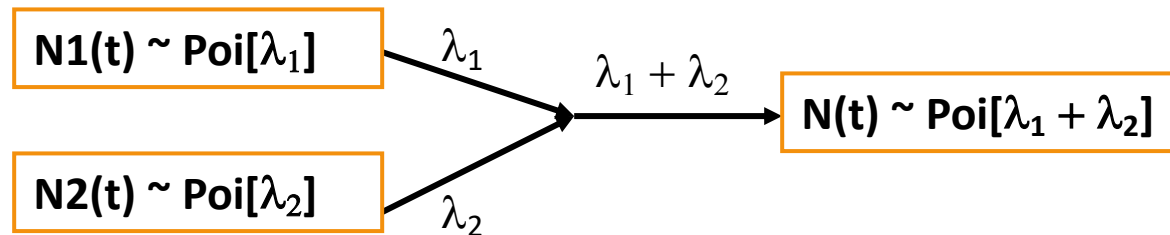


Splitting and Pooling (2)



■ Pooling:

- Suppose two Poisson processes $N_1(t)$, and $N_2(t)$ with parameters λ_1 , and λ_2 are pooled together
- With merging these two processes, a bigger Poisson process $N(t)$ is created, where $N_1(t) + N_2(t) = N(t)$
- The rate of the new process is $\lambda_1 + \lambda_2$



Splitting and Pooling (3)



■ Example 1:

- Jobs arriving at a shop $N(t)$ are in accordance with a Poisson process having rate λ
- Suppose every arrival is marked “**high priority**” (type I) with probability $1/3$ and “**low priority**” (type II) with probability $2/3$
- $N_1(t)$ and $N_2(t)$ are indicating the number of jobs, which both follow the Poisson process, with rates $\lambda/3$ and $2\lambda/3$, respectively

■ Example 2:

- A Poisson arrival stream with $\lambda_1 = 10$ arrivals per hour is combined (or pooled) with a Poisson arrival stream with $\lambda_2 = 17$ arrivals per hour
- The combined process is a Poisson process with $\lambda = 27$ arrivals per hour

Nonstationary Poisson Process (NSPP)



- Recall: Poisson processes have 3 conditions
- A Poisson process without the stationary increments (condition 2) is called as NSPP
 - NSPP are used for systems where the rate of arrival changes with time
 - So, NSPP are characterized by $\lambda(t)$, the arrival rate at time t
- Depending on the observation interval, λ would be different
- A number of use-cases for NSPP
 - Number of calls during a work-day
 - Number of pizza deliveries

Nonstationary Poisson Process (NSPP)



- One of the most important performance measurement parameters in NSPP is the expected number of arrivals by time t
 - It is denoted with $\Lambda(t)$:

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

- To be useful as an arrival-rate function, $\lambda(t)$ must be nonnegative and integrable
- For a SPP with rate λ , we have $\Lambda(t) = \lambda t$ as expected

Empirical Distributions (1)



- In many situations, the gathered and sampled data could not be described with known distributions
 - They cannot be expressed with previously mentioned mathematical models
 - What should we do?
 - Use **graphics** and **tables** to represent the data as it is
 - These are called empirical distribution
- May be used when it is **impossible** or **unnecessary** to conclude 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 (2)



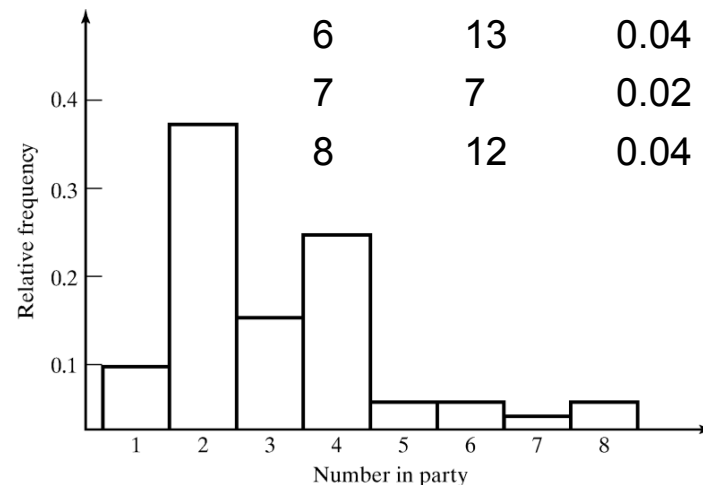
■ Example

- Assume that the students are entering the dining area in groups of 1-8 members
- We have monitored 300 groups, and counted number of members in every group
- Results are depicted in the histogram

■ This chart is somehow the PDF for random variable X

- Where X is the number of students in a group

Arrivals per Party	Frequenc y	Relative Frequenc y	Cumulati ve Relative Frequenc y
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 (3)



- Based on the histogram, there isn't any corresponding distributions matching the data
 - We have many upward and downward fluctuations
 - This amount of alterations cannot be described with known mathematical distributions
 - There may exist few distributions with **partial similarities**
 - We can use them as an estimate to simplify our modeling and calculations
 - But they will not produce an appropriate match to the data
- Shall we introduce a new distribution?
 - It is not rational
 - Because, there may not exist any other similar phenomena to be modeled with the same distribution
 - It is time consuming and complex

Stick with the table 😊



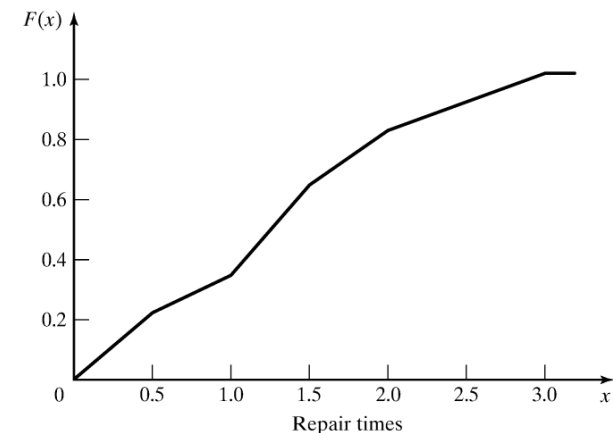
Empirical Distributions (4)



■ Let's talk about another example:

- The time required to repair a system, which has suffered a failure has been collected for the last 100 instances
- The empirical CDF for this random variable is depicted in the following

Intervals (Hours)	Frequency	Relative Frequency	Cumulativ e 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



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 the input data
 - Example: Hypothesize a distributional form for the input data
- Reader should know:
 - Difference between discrete, continuous, and empirical distributions
 - Poisson process and its properties

