MCA Sem. I Core

# Mathematical Techniques for Computer Applications(MCAC 103) L 7

## Continuous Distributions

24 Feb 2021

Vasudha Bhatnagar
Department of Computer Science,University of Delhi
Delhi, India.

# Outline

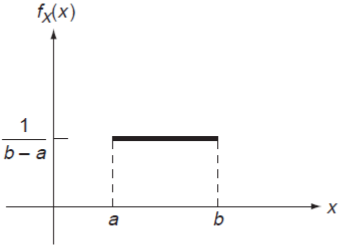
1. **Uniform Distribution**
2. **Exponential Distribution**
3. **Normal Distribution**
4. **Practice Exercises**

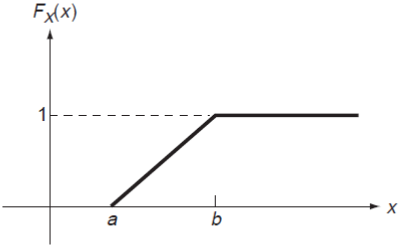
# Uniform Distribution

1. **Simplest statistical distribution**
2. **Random variable takes values in a continuous interval and all values between two boundaries are equally probable**
3. **pdf:** $f(x) = \frac{1}{b-a}$, **for** $a \leq x \leq b, 0$ **otherwise**
4. $\mu = \frac{a+b}{2}$ **and** $\sigma^2 = \frac{(b-a)^2}{12}$
5. **Discrete version also common in real world ( Digits of** $\pi$ **)**

Real World Use cases

1. Useful in case where nothing is known about the variable, except the fact that it lies between two bounds, a minimum and a maximum.
2. Random number generator that produces Uniform random variables
3. Duration of a flight might be considered uniformly distributed over a certain time interval (Continuous)
4. Time required by a student to travel from home to college is uniformly distributed between min-max minutes.




**The Uniform distribution with a = 0 and b = 1 is called Standard Uniform distribution.**

**Uniform Property:**

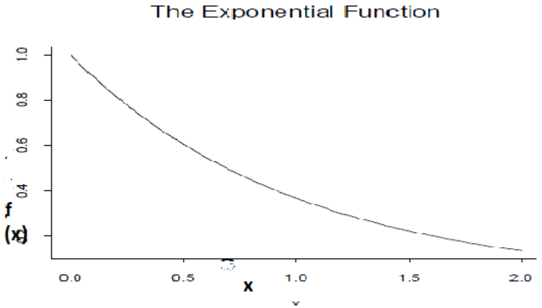
For any $h > 0$ and $X \in [t, t+h]$ , the probability $p(t < X \leq t+h) = \int_t^{t+h} dx/(b-a)$

# Exponential Distribution

1. **Waiting time distribution - predicts the amount of time elapsed till the event occurs**
2. **Often used to model time: waiting time, interarrival time, hardware lifetime, failure time etc.**
3. **Continuous analog of Geometric distribution**
4. **pdf:** $f(x) = \lambda e^{-\lambda x}$ **for** $x \geq 0, 0$ **otherwise**
5. $\mu = 1/\lambda$ **and** $\sigma^2 = 1/\lambda^2$

Real World Use cases

1. Length, in minutes, of long distance telephone calls
2. Amount of money customers spend in one trip to the supermarket
3. Calculations of product reliability



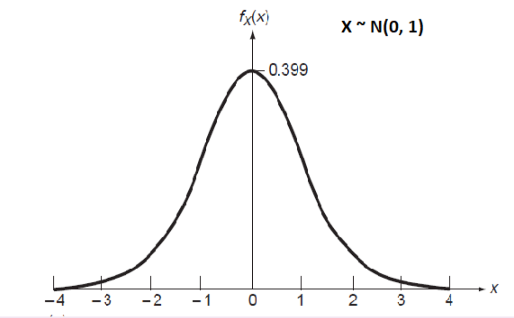

**Memoryless Property:** $P(X > x + t | X > t) = P(X > x)$ **for all** $x, t > 0$

Let X represent the length of time (measured in some unit) that a certain component works before failing. Consider the probability that an item that is working till t units, will continue to work for atleast x more units of time. This means that the component works for $x + t$ units of time. I.e. we are interested in $P(X > x + t | X > t)$

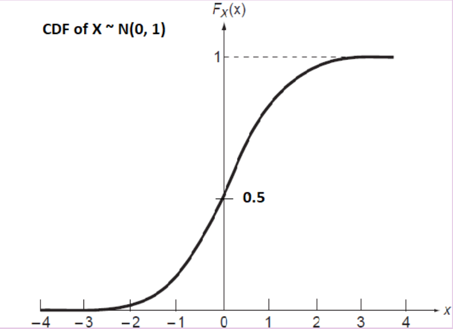
**The fact of having waited for t minutes gets "forgotten," and it does not affect the future waiting time**

# Normal Distribution

1. **Most important probability distribution in theory as well as in application**
2. **Good model for several physical variables like weight, height, temperature, voltage, pollution level etc.and Errors**
3. **pdf: if $X \sim N(\mu, \sigma^2)$, then**
   $f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp - \frac{(x-\mu)^2}{2\sigma^2}, -\infty < x < \infty$



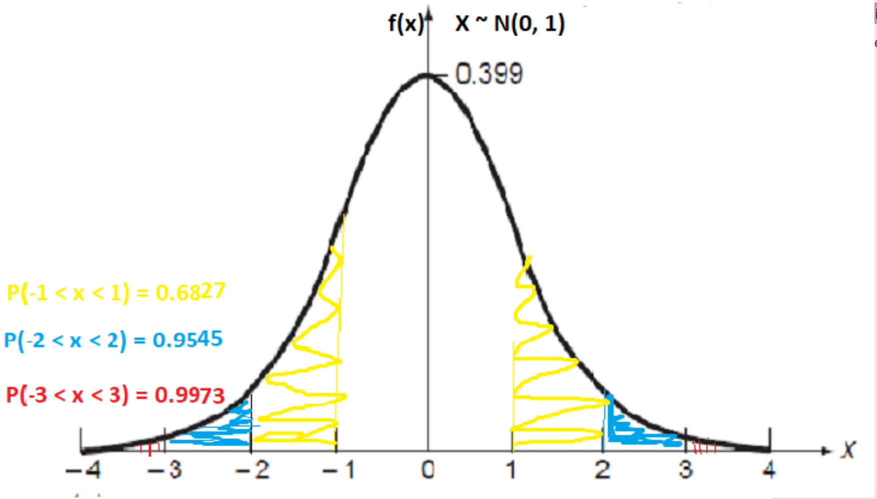

Density is known as the bell-shaped curve, symmetric and centered at $\mu$ , its spread being controlled by $\sigma^2$




$$F_X(x) = P(X < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} exp \frac{x^2}{2}$$

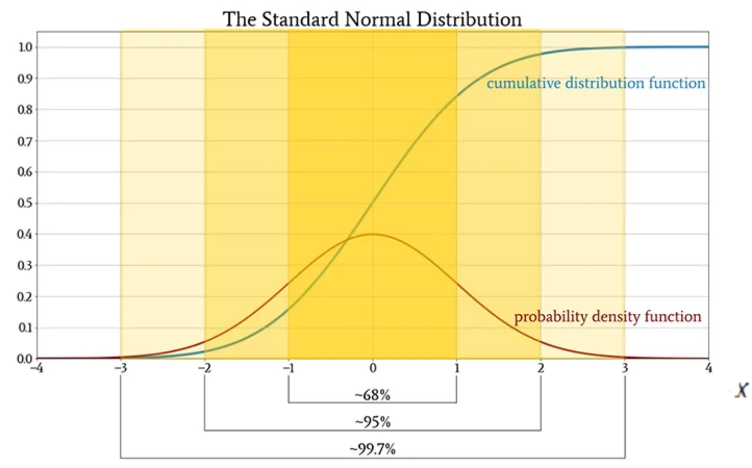
1. E(X) = $\mu$ and Var(X) = $\sigma^2$
2. Knowledge of its mean and variance completely characterizes a normal distribution.
3. $\mu$ and $\sigma$ are called location and scale parameters
4. Mean = Median = Mode

1. Density is symmetric about $\mu$ and attains its maximum value of $\frac{1}{\sqrt{2\pi}\sigma} \sim 0.399/\sigma$ at $x = \mu$
2. If X is normal with mean $\mu$ and variance $\sigma^2$, then $Y = \alpha X + \beta$ is normal with mean $\alpha\mu + \beta$ and variance $\alpha^2\sigma^2$
3. Sum of independent normal rvs is also a normal rv.




Many statistical tests are based on the assumption of normality.

# The complete picture



# Central Limit Theorem

**The central limit theorem states that as the sample size (N) becomes large,**

1. **The sampling distribution of the mean becomes approximately normal regardless of the distribution of the original variable.**
2. **The sampling distribution of the mean is centered at the population mean ($\mu$) of the original variable.**
3. **The standard deviation of the sampling distribution of the mean approaches $\sigma/\sqrt{N}$.**

# Central Limit Theorem

**The central limit theorem states that as the sample size (N) becomes large,**

1. **The sampling distribution of the mean becomes approximately normal regardless of the distribution of the original variable.**
2. **The sampling distribution of the mean is centered at the population mean ($\mu$) of the original variable.**
3. **The standard deviation of the sampling distribution of the mean approaches $\sigma/\sqrt{N}$.**

Formally,
Let $X_1, X_2, \ldots, X_n$ be a sequence of mutually independent, identically distributed random variables with means $\mu$ and variances $\sigma^2$. Further, let $S_n = \Sigma_{i=1}^{n} X_i$, and the normalized random variable $Z$ be defined as $\frac{S_n - n\mu}{\sqrt{n}\sigma}$. Then the probability distribution function of $Z$ , $F_Z(z)$, converges to N(0, 1) as $n \to \infty$.

## Example Applications

1. A disk has free space of 330 megabytes. Is it likely to be sufficient for 300 independent images, if each image has expected size of 1 megabyte with a standard deviation of 0.5 megabytes?
2. You wait for an elevator, whose capacity is 2000 Kg. The elevator comes with ten adult passengers. Suppose your own weight is 60 Kg, and you heard that human weights are normally distributed with the mean of 65 Kg and the standard deviation of 10 Kg. Would you board this elevator or wait for the next one?
3. 

## Practice Exercises

1. It is known that disks produced by a certain company will be defective with probability .01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective. What proportion of packages is returned? If someone buys three packages, what is the probability that exactly one of them will be returned?

# Practice Exercises

1. It is known that disks produced by a certain company will be defective with probability .01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective. What proportion of packages is returned? If someone buys three packages, what is the probability that exactly one of them will be returned? (0.005, 0.015)
2. A communications system consists of n components, each of which will, independently, function with probability p. The total system will be able to operate effectively if at least one-half of its components function.
(a) For what values of p is a 5-component system more likely to operate effectively than a 3-component system?
(b) In general, when is a $2k + 1$ component system better than a $2k - 1$ component system?

## Practice Exercises

1. It is known that disks produced by a certain company will be defective with probability .01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective. What proportion of packages is returned? If someone buys three packages, what is the probability that exactly one of them will be returned? (0.005, 0.015)
2. A communications system consists of n components, each of which will, independently, function with probability p. The total system will be able to operate effectively if at least one-half of its components function.
(a) For what values of p is a 5-component system more likely to operate effectively than a 3-component system?
(b) In general, when is a $2k+1$ component system better than a $2k-1$ component system? ($p \geq 1/2, p \geq 1/2$)
3. Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week.

## Practice Exercises

1. It is known that disks produced by a certain company will be defective with probability .01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective. What proportion of packages is returned? If someone buys three packages, what is the probability that exactly one of them will be returned? (0.005, 0.015)
2. A communications system consists of n components, each of which will, independently, function with probability p. The total system will be able to operate effectively if at least one-half of its components function.
(a) For what values of p is a 5-component system more likely to operate effectively than a 3-component system?
(b) In general, when is a $2k+1$ component system better than a $2k-1$ component system? ($p \geq 1/2, p \geq 1/2$)
3. Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week. (0.95)

1. At a party n people put their hats in the center of a room, where the hats are mixed together. Each person then randomly chooses a hat. If X denotes the number of people who select their own hat then. Show that $E(X) = 1$.
2. If X is a normal random variable with mean 3 and variance 16, find (a) $P(X < 11)$ (b) $P(X > -1)$ (c) $P(2 < X < 7)$

1. At a party n people put their hats in the center of a room, where the hats are mixed together. Each person then randomly chooses a hat. If X denotes the number of people who select their own hat then. Show that $E(X) = 1$.
2. If X is a normal random variable with mean 3 and variance 16, find (a) $P(X < 11)$ (b) $P(X > -1)$ (c) $P(2 < X < 7)$ (0.9772, 0.8413, 0.4400)
3. Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches. Assuming that the precipitation totals for the next 2 years are independent,
   (a) find the probability that the total precipitation during the next 2 years will exceed 25 inches.
   (b) find the probability that next year's precipitation will exceed that of the following year by more than 3 inches.

1. At a party n people put their hats in the center of a room, where the hats are mixed together. Each person then randomly chooses a hat. If X denotes the number of people who select their own hat then. Show that $E(X) = 1$.
2. If X is a normal random variable with mean 3 and variance 16, find (a) $P(X < 11)$ (b) $P(X > -1)$ (c) $P(2 < X < 7)$ (0.9772, 0.8413, 0.4400)
3. Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches. Assuming that the precipitation totals for the next 2 years are independent,
   (a) find the probability that the total precipitation during the next 2 years will exceed 25 inches.
   (b) find the probability that next year's precipitation will exceed that of the following year by more than 3 inches. ( 0.4240, 0.2469)

1. Since it is more economical to limit long-distance telephone calls to three minutes or less, the PDF of X – the duration in minutes of long-distance calls – may be of the form

$$F_X(x) = \begin{cases} (1 - e^{-x/3}) & \text{for } 0 \leq x < 3 \\ (1 - \frac{e^{-x/3}}{2}) & \text{for } x \geq 3 \\ 0 & \text{elsewhere} \end{cases}$$

Determine the probability that X is (a) more than two minutes and (b) between two and six minutes.

1. Since it is more economical to limit long-distance telephone calls to three minutes or less, the PDF of X – the duration in minutes of long-distance calls – may be of the form

$$F_X(x) = \begin{cases} (1 - e^{-x/3}) & \text{for } 0 \leq x < 3 \\ (1 - \frac{e^{-x/3}}{2}) & \text{for } x \geq 3 \\ 0 & \text{elsewhere} \end{cases}$$

Determine the probability that X is (a) more than two minutes and (b) between two and six minutes. ($e^{-2/3}$, $(e^{-2/3} - e^{-2}/2)$)

1. A driver is eagerly eyeing a precious parking space some distance down the street. There are five cars in front of the driver, each of which having a probability 0.2 of taking the space. What is the probability that the car immediately ahead will enter the parking space?

1. A driver is eagerly eyeing a precious parking space some distance down the street. There are five cars in front of the driver, each of which having a probability 0.2 of taking the space. What is the probability that the car immediately ahead will enter the parking space? (0.82)
2. A curbside parking facility has a capacity for three cars. Determine the probability that it will be full within 10 minutes. It is estimated that 6 cars will pass this parking space within the timespan and, on average, 80% of all cars will want to park there.

1. A driver is eagerly eyeing a precious parking space some distance down the street. There are five cars in front of the driver, each of which having a probability 0.2 of taking the space. What is the probability that the car immediately ahead will enter the parking space? (0.82)
2. A curbside parking facility has a capacity for three cars. Determine the probability that it will be full within 10 minutes. It is estimated that 6 cars will pass this parking space within the timespan and, on average, 80% of all cars will want to park there. (0.983)
3. Income levels are classified as low, medium, and high in a study of incomes of a given population. If, on average, 10% of the population belongs to the low-income group and 20%belongs to the high-income group, what is the probability that, of the 10 persons studied, 3 will be in the low-income group and the remaining 7 will be in the medium-income group? What is the marginal distribution of the number of persons (out of 10) at the low-income level?

1. A driver is eagerly eyeing a precious parking space some distance down the street. There are five cars in front of the driver, each of which having a probability 0.2 of taking the space. What is the probability that the car immediately ahead will enter the parking space? (0.82)
2. A curbside parking facility has a capacity for three cars. Determine the probability that it will be full within 10 minutes. It is estimated that 6 cars will pass this parking space within the timespan and, on average, 80% of all cars will want to park there. (0.983)
3. Income levels are classified as low, medium, and high in a study of incomes of a given population. If, on average, 10% of the population belongs to the low-income group and 20%belongs to the high-income group, what is the probability that, of the 10 persons studied, 3 will be in the low-income group and the remaining 7 will be in the medium-income group? What is the marginal distribution of the number of persons (out of 10) at the low-income level? (0.983)