BASIC PROBABILITY CONCEPTS

1.1 **Sample Space**

The set of all possible outcomes of a statistical experiment is called the sample space and is represented by the symbol S.

1.1.1 Events

An event is a subset of a sample space.

1.1.2 Complement of event

The complement of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement, of A by the symbol A'.

1.1.3 Intersection

The intersection of two events A and B, denoted by the symbol **A∩B**, is the event containing all elements that are common to A and B.

1.1.4 Mutually Exclusive

Two events A and B are mutually exclusive, or disjoint, if **A∩B=φ**, i.e. if A and B have no elements in common.

1.1.5 Union

The union of the two events A and B, denoted by the symbol **AUB**, is the event containing all the elements that belong to A or B or both

Exercise

1. An experiment involves tossing a pair of dice, 1 green and 1 red and recording the numbers that come up. If **x** equals the outcome on the green die and **y** the outcome on the red die, describe the sample space S
   a) by listing the elements {x,y)
   b) by using the rule method.
2. An experiment, consists of tossing a die and then flipping a coin once if the number on the: die is even. If the number on the die is odd, the coin is flipped twice. Using the notation 4H, for example, to denote the outcome that the die comes up 4 and then the coin comes up heads, and 3HT to denote the outcome that the die comes up 3 followed by a head and then a tail on the coin, construct a tree diagram to show the 18 elements of the sample space S.
3. Four students are selected at random from a chemistry class and classified as male or female. List the elements of the sample space S_1 using the letter M for "male" and F for "female." Define a second sample space S_2 where the elements represent the number of females selected.
4. An engineering firm is hired to determine if certain waterways in Kano State are safe for fishing. Samples are taken from three rivers.
   a) List the elements of a sample space S, using the letters F for "safe to fish" and N for "not safe to fish."

b) List the elements of S corresponding to event E that at least two of the rivers are safe for fishing.
c) Define an event that has as its elements the points {FFF, NFF, FFN, NFN}.

5. Exercise and diet are being studied as possible substitutes for medication to lower blood pressure. Three groups of subjects will be used to study the effect of exercise. Group one is sedentary while group two walks and group three swims for 1 hour a day. Half of each of the three exercise groups will be on a salt-free diet. An additional group of subjects will not exercise nor restrict their salt, but will take the standard medication. Use Z for sedentary, W for walker, S for swimmer, Y for salt, N for no salt, M for medication, and F for medication free.
   a) Show all of the elements of the sample space S.
   b) Given that A is the set of non-medicated subjects and B is the set of walkers, list the elements of **AUB**.
   c) List the elements of **A∩B**.

1.2 Probability of an Event

To find the probability of an event .4, we sum all the probabilities assigned to the sample points in A. This sum is called the probability of A and is denoted by P(A)

Definition: The probability of an event A is the sum of the weight of all sample joint in A. Therefore,

$$0 < P(A) < 1, \qquad P(\emptyset) = 0, \qquad and\ P(S) = 1$$

Furthermore if A_1, A_2, A_3,… is a sequence of mutually exclusive events then

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

Again, if the sample space S or an experiment contains N elements, all of which are equally likely to occur, we assign a probability equal to $1/N$ to each of the N points. The probability of any event **A** containing **n** of these N sample points is then the ratio of the number of elements in A to the number of elements in S.

Definition: If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is

$$P(A) = \frac{n}{N}$$

Exercise

6. A box contains 500 envelopes of which 75 contain N100 in cash, 150 contain N25, and 275 contain N10. An envelope may be purchased for N25. What is the sample space for the different amounts of money? Assign probabilities to the sample points and then find the probability that the first envelope purchased contains less than N100.

7. An automobile manufacturer is concerned about a possible recall of its best-selling four-door sedan. If there were a recall, there is 0.25 probability that a defect is in the brake system, 0.18 in the transmission, 0.17 in the fuel system, and 0.40 in some other area.
   a) What is the probability that the defect is the brakes or the fuelling system if the probability of defects in both systems simultaneously is 0.15?
   b) What is the probability that there are no defects in either the brakes or the fuelling system?
8. If each coded item in a catalogue begins with 3 distinct letters followed by 4 distinct nonzero digits, find the probability of randomly selecting one of these coded items with the first letter a vowel and the last digit even.
9. A pair of fair dice is tossed. Find the probability of getting
   a) a total of 8
   b) at most a total of 5
10. Two cards are drawn in succession from a deck without replacement. What is the probability that both cards are greater than 2 and less than 8?

1.3 Conditional Probability

The probability of an event B occurring when it is known that some event A has occurred is called a conditional probability and is denoted by $P(B/A)$. The symbol $P(B/A)$ is usually read "**the probability that B occurs given that A occurs" or simply "the probability of B, given A.**"

Definition: The conditional probability of B, given A, denoted by $P(B/A)$is defined by

$$P(B/A) = \frac{P(A \cap B)}{P(A)} \; provided \; P(A) > 0$$

Exercise

11. The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$; the probability that it arrives on time is $P(A) = 0.82$; and the probability that it departs and arrives on time is $P(A \cap B) = 0.78$.
    a) Find the probability that a plane arrives on time given that it departed on time,
    b) departed on time given that it has arrived on time
12. The probability that the head of a household is home when a telemarketing representative calls is 0.4. Given that the head of the house is home, the probability that goods will be bought from the company is 0.3. Find the probability that the head of the house is home and goods being bought from the company.
13. The probability that a doctor correctly diagnoses a particular illness is 0.7. Given that the doctor makes an incorrect diagnosis, the probability that the patient enters a law suit is 0.9. What is the probability that the doctor makes an incorrect diagnosis and the patient sues?

1.4 Bayes' Rule

Theorem: If the event $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S $P(B_i) \neq 0 \; for \; i = 1, 2, 3, \ldots, k$ then for any event A of S,

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P\,(B_i) P(A/B_i)$$

$$P(B_k / A) = \frac{P(B_k \cap A)}{\sum_{i=1}^{n} P(B_i \cap A)} = \frac{P(B_k) P(A/B_k)}{\sum_{i=1}^{n} P(B_i) P(A/B_i)}$$

Exercise

14. Suppose that the four inspectors at a film factory are supposed to stamp the expiration date on each package of film at the end of the assembly line. John, who stamps 20% of the packages, fails to stamp the expiration date once in every 200 packages: Tom, who stamps 60% of the packages, fails to stamp the expiration date once in every 100 packages; Jeff, who stamps 15% of the packages, fails to stamp the expiration date once in every 90 packages; and Pat, who stamps 5% of the packages, fails to stamp the expiration date once in every 200 packages. If a customer complains that her package of film does not show the expiration date, what is the probability that it was inspected by John?
15. A regional telephone company operates three identical relay stations at different locations. During a one year period, the number of malfunctions reported by each station and the causes are shown below.

Station	**A**	**B**	**C**
Problems with electricity supplied	2	1	1
Computer malfunction	4	3	2
Malfunctioning electrical equipment	5	4	2
Caused by other human error	7	7	5

Suppose that a malfunction was reported and it was found to be caused by other human errors. What is the probability that it came from station C?

16. A paint-store chain produces and sells latex and semi-gloss paint. Based on long-range sales, the probability that a customer will purchase latex paint is 0.75. Of those that purchase latex paint, 60% also purchase rollers. But only 30% of semi-gloss paint buyers purchase rollers. A randomly selected buyer purchases a roller and a can of paint. What is the probability that the paint is latex?
17. A truth serum has the property that 90% of the guilty suspects are properly judged while, of course, 10% of guilty suspects are improperly found innocent. On the other hand, innocent suspects are misjudged 1% of the time. If the suspect was selected from a group of suspects of which only 5% have ever committed a crime, and the serum indicates that he is guilty, what is the probability that he is innocent?

18. A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products respectively. The "defect rate:" is different for the three procedures as follows:

$$P(D/P_1) = 0.30 \quad P(D/P_2) = 0.20 \quad and \quad P(D/P_3) = 0.50$$

Where $P(D/P_j)$ is the probability of a defective product, given plan j. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

Definition: A random variable is a function that associates a. real number with each element in the sample space

Definition: If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a discrete sample space.

Definition: If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a continuous sample space.

A random variable is called a ***discrete random variable*** if its set of possible outcomes is countable. And when a random variable can take on values on a continuous scale, it, is called a ***continuous random valuable***.

2.1 Discrete Probability Distributions

A discrete random variable assumes each of its values with a certain probability.

Definition: The set of ordered pairs $(x, f(x))$ is ***a probability function, probability mass function, or probability distribution*** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0,$
2. $\sum_x f(x) = 1,$
3. $P(X = x) = f(x)$

2.1.1 Cumulative distribution

There are many problems where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x. Writing $F(x) = P(X \leq x)$ for every real number x, we define $F(x)$ to be the cumulative distribution function of the random variable X.

Definition: The c**umulative distribution function $F(x)$ of a discrete random variable** X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \qquad for - 00 < x < 00$$

2.2 Continuous Probability Distributions

A continuous random variable has a probability of zero of assuming exactly any of its values. Consequently, its probability distribution cannot be given in tabular form.

Definition: The: function $f(x)$ is a **probability density function** for the continuous random variable A, defined over the set of real number R, if

1. $f(x) \geq 0$ for all $x \varepsilon R$
2. $\int_{-\alpha}^{\alpha} f(x)dx = 1$
3. $P(a < x < b) = \int_{a}^{b} f(x)dx$

2.2.1 Cumulative distribution

Definition: The **cumulative distribution function $F(x)$ of a continuous random variable** A with density function $f(x)$ is

$$F(x) = P(X \leq a) = \int_{-\infty}^{\infty} f(t)dt \qquad for - 00 < x < \infty$$

Exercise

19. The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \\ 0 & elsewhere \end{cases}$$

Find the probability that over the period of one year, a family runs their vacuum cleaner

a) Less than 120 hours
b) Between 50 and 100 hours

20. A coin is flipped until 3 heads in succession occur. List only those elements of the sample space that require 6 or less tosses. Is this a discrete sample space? Explain.

21. A continuous random variable X that can assume values between $x = 2$ and $x = 5$ has a density function given by $f(x) = 2(1 + x)/27$. Find

a) $P(X < 4)$
b) $P(3 \leq X < 4)$

22. The time to failure in hours of an important piece of electronic equipment used in a manufactured DVD player has the density function

$$f(x) = \begin{cases} \frac{1}{2000} e^{-(t/_{2000})} & x \geq 0 \\ 0, & x < 0 \end{cases}$$

a) Find F(x)

b) Determine the probability that the component (and thus the DVD player) lasts more than 1000 hours before the component needs to be replaced.
c) Determine the probability that the component fails before 2000 hours.

23. Suppose it is known from large amounts of historical data that X, the number of cars that arrive at a specific intersection during a 20 second time period, is characterized by the following discrete probability function

$$f(x) = e^{-6}\frac{6^x}{x!}, \qquad x = 0,1,2,\ldots$$

a) Find the probability that in a specific 20-second time period, more than 8 cars arrive at intersection.
b) Find the probability that only two cars arrive.

MATHEMATICAL EXPECTATION

3.1 Mean of a Random Variable

Whereas the probability mass function or the probability density function provides complete information about the probabilistic properties of a random variable, it is often useful to employ some summary measures of these properties. One of the most basic summary measures is the **expectation or mean of a random variable**, which is denoted by **E(X)** and represents an "average" value of the random variable. Two random variables with the same expected value can be thought of as having the same average value, although their probability mass functions or probability density functions may be quite different.

Definition: Let X be a random variable with probability distribution $f(x)$. The **mean or expected value of X** is, If X is discrete

$$\mu = E(X) = \sum_x x\, f(x)$$

And if X is continuous,

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

Theorem 3.1: Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(x)$ is, **if X is discrete,**

$$\mu_g(X) = E[g(X)] = \sum_x g(x)f(x)$$

And if X is continuous

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

We shall now extend our concept of mathematical expectation to the case of two random variables X and Y with joint probability distribution

Definition: Let $X\ and\ Y$ be random variables with joint probability distribution $f(x.y)$. The mean or expected value of the random variable $f(X,Y)$ is, **if X and Y are discrete**

$$\mu_g(X,Y) = E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y)$$

And if X and Y are continuous,

$$\mu_g(X) = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x.y)f(x,y)dxdy$$

Exercise

24. The probability distribution of the discrete random variable X is

$$f(x) = \binom{3}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{3-x} \qquad x = 0,1,2,3.$$

Find the mean of X.

25. Two tire-quality experts examine stacks of tires and assign quality ratings to each tire on a 3-point scale. Let X denote the grade given by expert A and Y denote the grade given by B. The following table gives the joint distribution for X and Y.

$f(x,y)$	y = 1	2	3
x = 1	0.10	0.05	0.02
2	0.10	0.35	0.05
3	0.03	0.10	0.20

Find $\mu_x\ and\ \mu_y$

26. Suppose that X and Y have the following joint probability function:

$f(x,y)$	x = 1	2
y = 1	0.10	0.15
2	0.20	0.30
5	0.10	0.15

a) Find the expected value of $g(X,Y) = XY^2$
b) Find$\mu_x\ and\ \mu_y$

27. A continuous random variable X has the density function

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

Find the expected value of $g(X) = e^{2X/3}$

28. A coin is biased so that a head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

3.2 Variance and Covariance of Random Variables

Definition: Let X be a random variance of X is, **if X is discrete**

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

And if X is continuous

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

The positive square root of the variance, $\boldsymbol{\sigma}$ is called the **standard deviation of X.**

An alternative and preferred formula for finding σ^2, which often simplifies the calculations, is stated in the following theorem.

Theorem 3.2.1: The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2$$

We shall now extend our concept of the variance of a random variable X to also include random variables related to X. For the random variable f(X), the variance will be denoted by$\sigma^2{}_{g(X)}$, and is calculated by means of the following theorem.

Theorem 3.2.2: Let X be a random variable with probability distribution f(x). The variance of the random variable $g(X)$ is, **if X is discrete**

$$\sigma^2{}_{g(X)} = E\left\{\left[g(X) - \mu_{g(X)}\right]^2\right\} = \sum_x \left[g(X) - \mu_{g(X)}\right]^2 f(x)$$

And if X is continuous,

$$\sigma^2{}_{g(X)} = E\left\{\left[g(X) - \mu_{g(X)}\right]^2\right\} = \int_{-\infty}^{\infty} \left[g(X) - \mu_{g(X)}\right]^2 f(x)dx$$

If $g(X,Y) = (X-\mu_X)(Y-\mu_Y) where\ \mu_X = E(X) and\ \mu_Y = E(Y)$. Theorem 3.2.1 yields an expected value called the covariance of X and Y, which we denote by $\boldsymbol{\sigma_{XY}\ or\ Cov\ (X,Y)}$

Definition: Let X and Y be random variables with joint probability distribution f(x, y). The covariance of X and Y is, ***if X and Y are discrete***

$$\sigma_{XY}\ or\ Cov\ (X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sum_{XY} (x-\mu_x)(y-\mu_y)\, f(x,y)$$

And if X and Y arc continuous,

$$\sigma_{XY} \ or \ Cov\ (X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)f(x,y)dxdy$$

The alternative and preferred formula for $\sigma_{XY} \ or \ \ (X,Y)$ is stated below

Theorem 3.2.3: The covariance of two random variables X and Y with means ux and uY, respectively, is given by

$$\sigma_{XY} \ or \ Cov\ (X,Y) = E(XY) - \mu_X\mu_Y$$

Exercise

29. Let X be a random variable with the following probability distribution:

x	-2	3	5
$f(x)$	0.3	0.2	0.5

Find the standard deviation of X.

30. The random variable X, representing the number of errors per 100 lines of software code, has the following probability distribution:

x	2	3	4	5	6
$f(x)$	0.01	0.25	0.4	0.3	0.04

Using Theorem 3.2.1, find the variance of X.

31. Let X denote the number of times a certain numerical control machine will malfunction: 1, 2, or 3 times on any given day. Let Y denote the number of times a technician is called on an emergency call. Their joint probability distribution is given as

$f(x,y)$		x		
		1	2	3
y	1	0.05	0.05	0.1
	2	0.05	0.1	0.35
	3	0	0.2	0.1

Find the covariance of the random variables X and Y

3.3 Chebyshev's Theorem

Theorem 3.3.1: The probability that any random variable X will assume a value within k standard deviation of the mean is at least $1-\frac{1}{k^2}$. That is,

$$P(p-k\sigma < X < \mu + k\sigma) > 1 - \frac{1}{k^2}$$

Exercise

32. Seventy new jobs are opening up at an automobile manufacturing plant, but 1000 applicants show up for the 70 positions. To select the best 70 from among the applicants, the company gives a test that covers mechanical skill, manual dexterity, and mathematical ability. The mean grade on this test turns out to be 60, and the scores have a standard deviation 6. Can a

person who has an 84 score count on getting one of the jobs? [Hint Use Chebyshev's theorem] Assume that the distribution is symmetric about the mean.

33. An electrical firm manufactures a 100-watt light bulb, which, according to specifications written on the package, has a mean life of 900 hours with a standard deviation of 50 hours. At most, what percentage of the bulbs fail to last even 700 hours? Assume that the distribution is symmetric about the mean

SOME DISCRETE PROBABILITY DISTRIBUTIONS

4.1 Discrete Uniform Distribution

The simplest of all discrete probability distributions is one where the random variable assumes each of its values with an equal probability. Such a probability distribution is called ***a discrete uniform distribution***.

If the random variable X assumes the values $x_1, x_2, \dots, x_k$, with equal probaDistribution bilities, them the discrete unil'orni distribution is given by

$$f(x;k) = \frac{1}{k} \qquad x = x_1, x_2, \dots, x_k$$

We have used the notation $f(x;k)$ instead of $f(x)$ to indicate that the uniform distribution depends on the parameter k.

Theorem: The mane and variance of the discrete uniform distribution $f(x;k)$ are

$$\mu = \frac{1}{k}\sum_{i=1}^{k} x_i \qquad and \qquad \sigma^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$$

4.2 Binomial Distributions

An experiment often consists of repeated trials, each with two possible outcomes that may be labelled success or failure. The most obvious application deals with the testing of items as they come off an assembly line, where each test or trial may indicate a defective or a non-defective item. We may choose to define either outcome as a success. The process is referred to as a Bernoulli process.

4.2.1 The Bernoulli Process

Strictly speaking, the Bernoulli process must possess the following properties:

1. The experiment consists of n repeated trials.
2. Each trial results in an outcome that may be classified as a. success or a failure.
3. The probability of success, denoted by p, remains constant from trial to trial.
4. The repeated trials arc independent.

A Bernoulli trial can result in a success with probability by p and a failure with Distribution probability by $q = 1 - p$. Then the probability distribution of the binomial random variable by X, the number of successes in by n independent trials, is by

$$b(x;n,p) = \binom{n}{x} p^x q^{n-x} \quad x = 0,1,2,3, \dots., n$$

Theorem: The mean and variance of the binomial distribution $b(x; n, p)$ are

$$p = np \qquad and \qquad \sigma^2 = npq$$

Exercise

34. An employee is selected from a staff of 10 to supervise a certain project by selecting a tag at random from a box containing 10 tags numbered from 1 to 10. Find the formula for the probability distribution of X representing the number on the tag that, is drawn. What is the probability that the number drawn is less than 4?
35. In a certain city district the need for money to buy drugs is stated as the: reason for 75% of all thefts. Find the probability that among the next 5 theft cases reported in this district,
    a) Exactly 2 resulted from the need for money to buy drugs
    b) At most 3 resulted from the need for money to buy drugs.
36. According to Chemical Engineering Progress (Nov. 1990), approximately 30% of all pipework failures in chemical plants are caused by operator error.
    a) What is the probability that out of the next 20 pipework failures at least 10 are due to operator error?
    b) What is the probability that no more than 4 out of 20 such failures are due to operator error?
    c) Suppose, for a particular plant, that, out of the random sample of 20 such failures, exactly 5 are operational errors. Do you feel that the 30% figure stated above applies to this plant? Comment.
37. When a light bulb is selected at random from a box that contains a 40-watt bulb, a 60-watt bulb, a 75-watt bulb, and a 100-watt bulb, each element of the sample1 space S = {40, 60, 75, 100} occurs with probability 1/4. Therefore, we have a uniform distribution, with $f(x; 4) = \frac{1}{4} \qquad x = 40, 60, 75, 100$. Find the mean and variance of the random variable A'

4.3 Poisson Distribution and the Poisson Process

Experiments yielding numerical values of a random variable X, the number of outcomes occurring during a given time interval or in a specified region, are called ***Poisson experiments.***

4.3.1 Properties of Poisson Process

i. The number of outcomes occurring in one time interval or specified region is independent of the number that occurs in any other disjoint time interval or region of space. In this way we can say that the Poisson process has no memory.
ii. The probability that a single outcome will occur during a very short time interval or in a small region is proportional to the length of the time interval or the size of the region and does not depend on the number of outcomes occurring outside this time interval or region.
iii. The probability that more than one outcome will occur in such a short time interval or fall in such a small region is negligible.

The number X of outcomes occurring during a Poisson experiment is ***called a Poisson random variable***, and its probability distribution is called ***the Poisson distribution.***

The probability distribution of the Poisson random variable X, representing Distribution the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$P(x; Xt) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, a = 0, 1, 2, \ldots,$$

Theorem: Both the mean and variance of the Poisson distribution $P(x; Xt)\ are\ Xt$

Theorem: Let X be a binomial random variable with probability distribution $b(x; n, p)$. When $n \to oc, p \to 0\ and\ np \xrightarrow{n\to\infty} p(x; \mu)$ remains constant.

Exercise

38. Potholes on a highway can be a serious problem and are in constant need of repair. With a particular type of terrain and make of concrete, past experience suggests that there are, on the average, 2 potholes per mile after a certain amount of usage. It is assumed that the Poisson process applies to the random variable "number of potholes."
    a) What is the probability that no more than one pothole will appear in a section of one mile?
    b) What is the probability that no more than 4 pot holes will occur in a given section of 5 miles?
39. Hospital administrators in large cities anguish about problems with traffic in emergency rooms in hospitals. For a particular hospital in a large city, the staff on hand cannot accommodate the patient traffic if there are more than 10 emergency cases in a given hour. It is assumed that patient arrival follows a Poisson process and historical data suggest that, on the average, 5 emergencies arrive per hour.
    a) What is the probability that in a given hour the staff can no longer accommodate the traffic?
    b) What is the probability that more than 20 emergencies arrive during a 3-hour shift of personnel?

4.4 Geometric Distributions

If repeated independent trials can result in a success with probability p and Distribution a failure with probability $q = 1 - p$, then the probability distribution of the random variable X, the number of the trial on which the first success occurs, is

$$g(x; p) = pq^{x-1} \qquad\qquad x = 1,2,3, \ldots.$$

Theorem: The mean and variance of a random variable following the geometric distribution are

$$\mu = \frac{1}{P}, \qquad\qquad \sigma^2 = \frac{1-P}{P^2}$$

Exercise

40. Three people toss a fair coin and the odd man pays for coffee. If the coins all turn up the same, they are tossed again. Find the probability that fewer than 4 tosses are needed.
41. The probability that a student pilot passes the written test for a private pilot's license is 0.7. Find the probability that the student will pass the test
    a) On the third try
    b) Before the fourth try
42. A restaurant chef prepares a tossed salad containing, on average, 5 vegetables. Find the probability that the salad contains more than 5 vegetables
    a) On a given day
    b) On 3 of the next 4 days
    c) For the first time in April on April 5.

SOME CONTINUOUS PROBABILITY DISTRIBUTIONS

5.1 Continuous Uniform Distribution

One of the simplest continuous distributions in all of statistics is the continuous uniform distribution. This distribution is characterized by a density function that is "flat," and thus the probability is uniform in a closed interval, say [A, B].

The density function of the continuous uniform random variable X on the interval [A, B] is

$$f(x; A, B) = \begin{cases} \dfrac{1}{B-A}, & A < x < B \\ 0, & elsewhere \end{cases}$$

Theorem: The mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2}, \qquad \sigma^2 = \frac{(B-A)^2}{12}$$

5.2 Normal Distribution

The most important, continuous probability distribution in the entire field of statistics is the normal distribution. Its graph, called the normal curve, is the bell shaped curve.

The density of the normal random variable X, with mean p and variance σ^2, is

$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2}$$

Definition: The distribution of a normal random variable with mean 0 and variance 1 is called a ***standard normal distribution***.

5.2.1 Areas under the Normal Curve

The curve of any continuous probability distribution or density function is constructed so that the area under the curve bounded by the two ordinates $x = x_1 \; and \; x = x_2$ equals the probability that the random variable X assumes a value between $x = x_1 \; and \; x = x_2$.

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-(1/2)[(x-\mu)/\sigma]^2} dx$$

Fortunately, we are able to transform all the observations of any normal random variable X to a new set of observations of a normal random variable Z with mean 0 and variance 1. This can be done by means of the transformation

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

Exercise

43. A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that the battery lives are normally distributed, find the probability that a given battery will last less than 2.3 years.
44. In an industrial process the diameter of a ball bearing is an important component part. The buyer sets specifications on the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean p. = 3.0 and standard deviation a = 0.005. On the average, how many manufactured ball bearings will be scrapped?

5.3 Gamma Distributions

The continuous random variable X has a gamma distribution, with parameters $\alpha \; and \; \beta$, if its density function is given by

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0 \quad elsewhere & \end{cases}$$

Theorem: The mean and variance of the gamma distribution are

$$\mu = \alpha\beta \qquad and \qquad \mu\alpha\beta^2$$

5.4 Exponential Distributions

The continuous random variable X has an exponential distribution, with Distribution parameter β, if its density function is given by

$$f(x; \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0 \quad elsewhere & \end{cases}$$

Theorem: The mean and variance of the exponential distribution are

$$\mu = \beta \qquad and \qquad \sigma^2 = \beta^2$$

5.5 Chi-Squared Distribution

The continuous random variable X has a chi-squared distribution, with υ degrees of freedom, if its density function is given by

$$f(x;\upsilon) = \begin{cases} \dfrac{1}{2^{\upsilon/2}\Gamma(\upsilon/2)}\, x^{\upsilon/2-1}e^{-x/2}, & x > 0 \\ 0 & elsewhere \end{cases}$$

Where υ is a positive integer

Theorem: The mean and variance of the chi-squared distribution are

$$\mu = \upsilon \qquad and \qquad \sigma^2 = 2\,\upsilon$$

5.6 Weibull Distribution

The continuous random variable X has a Weibull distribution, with parameters $\alpha\ and\ \beta$ if its density function is given by

$$f(x;\alpha,\beta) = \begin{cases} \alpha\beta x^{\alpha-1}e^{-\alpha x^{\alpha}}, & x > 0 \\ 0 \quad elsewhere & \end{cases}$$

Theorem: The mean and variance of the Weibull distribution are

$$\mu = \alpha^{-1/\beta}\Gamma\left(1+\frac{1}{\beta}\right) \qquad and \qquad \sigma^2 = \alpha^{-2/\beta}\left\{\Gamma\left(1+\frac{2}{\beta}\right) - \left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2\right\}$$

The cumulative distribution function for the Weibull distribution is given by

$$F(x) = 1 - e^{-\alpha x^{\alpha}}, \qquad for\ x > 0$$

5.6.1 The Failure Rate for the Weibull Distribution

When the Weibull distribution applies, it is often helpful to determine the failure rate (sometimes called hazard rate) in order to get a sense of wear or deterioration of the component. The failure rate at time t for the Weibull distribution is given by

$$Z(t) = \alpha\beta t^{-\beta-1}, \qquad t > 0$$

Exercise

45. Suppose that the time, in hours, taken to repair a heat pump is a random variable X having a gamma distribution with parameters $\alpha = 2$ and $\beta = 1/2$. What is the probability that the next service call will require
    a) at most 1 hour to repair the heat pump?
    b) at least 2 hours to repair the heat pump?
46. In a certain city, the daily consumption of electric power, in millions of kilowatt-hours, is a random variable X having a gamma distribution with mean $\mu = 6$ and variance $\sigma^2 = 12$.

a) Find the values of α and β.
b) Find the probability that on any given day the daily power consumption will exceed 12 million kilowatt hours

47. The length of time for one individual to be served at a cafeteria is a random variable having an exponential distribution with a mean of 4 minutes. What is the probability that a person is served in less than 3 minutes on at least 4 of the next 6 days?
48. The life, in years, of a certain type of electrical switch has an exponential distribution with an average life $\beta = 2$. If 100 of these switches are installed in different systems, what is the probability that at most 30 fail during the first year?
49. Suppose that the service life, in years, of a hearing aid battery is a random variable having a Weibull distribution with $\alpha = 1/2$ and $\beta = 2$.
    a) How long can such a battery be expected to last?
    b) What is the probability that such a battery will be operating after 2 years?
50. Form the relationship between the chi-squared random variable and the gamma random variable, prove that the mean of the chi-squared random variable is υ and the variance is 2υ.

ASSIGNMENT

INSTRUCTIONS

Answer the questions stated below from the summarized lecture given. Unless otherwise stated, all symbols carry their usual meanings. **You are to submit on or before March 31st , 2020.**

Question: 7, 12, 17, 19, 36, 39, 42, 44, 46, 48, 49