Univ. of Technology, Dept. of Applied Sciences, Applied Mathematics

Subject: PROBABILITY AND STATISTICS

For Second Year Students, 3 hours per week

Text Book: Introduction to Statistics, 3rd edition

By: Ronald E. Walpole

Syllabus:

1- Probability (4 weeks):

Sample Space, Events, Operation with events, Computing sample points, Probability of an Event, Additive Rule, Conditional Probability, Multiplication Rule, Bayes Rule.

2- Distribution of Random variables (5 weeks):

Concept of a Random variable, Discrete Probability Dist., Continuous Probability Dist., Joint Probability Dist., Expected Values, Mean and Variance of a Random variable, Properties of Mean and Variance, Moment Generating Functions.

3- Some Discrete Probability Distributions (3 weeks):

Uniform Dist., Binomial & Multinomial Dist., Hypergeometric Dist., Negative Binomial & Geometric Dist., Poisson Dist.

4- Some Continuous Probability Distributions (3 weeks)

Uniform Dist., Exponential & Gamma Dist., Normal Dist., Application of the Normal Dist., Normal Approximation to the Binomial Dist.

5- Sampling Theory (3 weeks):

Sampling Distributions, Sampling Distributions of the Mean, t Distribution, Sampling Dist. of the Difference Between Two Means.

6- Estimation of Parameters (4 weeks):

Statistical Inference, Classical Method of Estimation, Estimating the Mean, Estimating the Difference Between Two Means, Estimating a Proportion, Estimating the Difference Between two Proportions, Estimating the Variance, Estimating the Ratio of Two Variances.

7- Tests of Hypotheses (4 weeks):

Statistical Hypotheses, Testing a Statistical Hypothesis, One-Tailed and Two-Tailed Tests, Tests Concerning Means, Tests Concerning Variances, Tests Concerning Proportions, Goodness-of-Fit Test, Test of Independence.

8- Regression and Correlation (4 weeks):

Linear Regression, Regression Analysis, Inferences Concerning the Regression Coefficients, Prediction, Test of Linearity of Regression, Exponential Regression, Multiple Regression, Linear Correlation, Multiple and Partial Correlation.

Probability and Statistics

① Probability:

Def: If in a statistical experiment, all the possible outcomes can be described prior to its performance (are known), but niether of them can be predicted with certainty and if this kind of experiment can be repeated under the same conditions, it is called Random Experiment.

Def: The set of all possible outcomes of a Random Experiment is called the Sample Space and is represented by the symbol S.

Examples: ① In the toss of a coin, let the outcome tails be denoted by T and the outcome heads be denoted by H.

$\therefore \quad S = \{H, T\}$

② Tossing two coins, $S = \{HH, HT, TH, TT\}$

③ Tossing three coins, $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$

④ Tossing a die, $S = \{1, 2, 3, 4, 5, 6\}$

⑤ Tossing two dice,

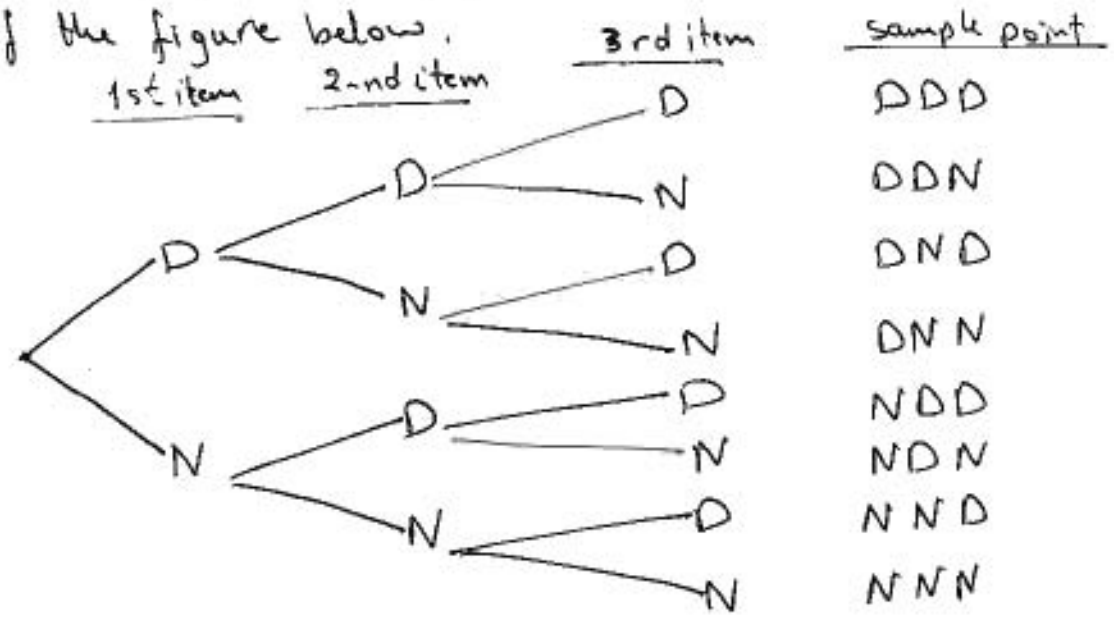
$$S = \left\{ \begin{array}{l} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6) \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{array} \right\}$$

Def:

Each outcome in a sample space is called an Element or a Member of the sample space or simply a Sample Point.

In some experiments it will be helpfull to list the elements of the sample space systematically by means of a Tree Diagram.

Ex: Suppose that 3 items are selected at random from a manufacturing process. Each item is inspected and classified, D, or nondefective, N. To List the elements of the sample space providing the most information, we construct the tree diagram of the figure below.



∴ The sample space is

$$S = \{DDD, DDN, DND, DNN, NDD, NDN, NND, NNN\}$$

Def: Any subset of a sample space is called an Event.

Ex:

$E_1 = \{HH, TT\}$, $E_2 = \{HH, HT, TH\}$

$E_3 = \{HHH, TTT\}$, $E_4 = \{HHT, THH, HTH, HTT, THT, TTH\}$

$E_5 = \{2, 4, 6\}$, $E_6 = \{1, 2, 3, 4\}$

$E_7 = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$

$E_8 = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

Def: Two events A and B are said to be Mutually Exclusive if $A \cap B = \phi$

Note: The mathematical theory of probability for finite sample spaces provides a set of real numbers called weights or Probabilities, ranging from ZERO to ONE, which allow us to evaluate the likelihood of occurrence of events.

Def: The <u>Probability of an event</u> A is the sum of the probabilities of all sample points in A. Therefore.

$$0 \leq P(A) \leq 1 \ , \quad P(\emptyset) = 0 \ , \quad P(S) = 1$$

Ex: The Probabilities of the events in the last example are:

$P(E_1) = \frac{1}{2}$, $P(E_2) = \frac{3}{4}$, $P(E_3) = \frac{1}{4}$, $P(E_4) = \frac{3}{4}$,

$P(E_5) = \frac{1}{2}$, $P(E_6) = \frac{2}{3}$, $P(E_7) = \frac{1}{6}$, $P(E_8) = \frac{1}{6}$

Theorem: 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:

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

Ex: If a card is drawn from an ordinary deck, find the probability that it is a heart.

Note An ordinary deck consists of 26 black and 26 red cards each consists of two types, from 1 to 10 and a King a Queen and a Jack. Then the sample space is: [The card which has the number 1 is called Ace.]

$$S = \left\{ \begin{array}{ccccccc} \boxed{1}, & \boxed{2}, & \ldots, & \boxed{10}, & \boxed{K}, & \boxed{Q}, & \boxed{J} \\ & & \vdots & & & & \\ & & \vdots & & & & \\ & & \vdots & & & & \end{array} \right\} \begin{array}{l} \text{Club which is black} \\ \text{Heart } \quad " \quad " \text{ red} \\ \text{Spade } \quad " \quad " \text{ black} \\ \text{Diamond } \quad " \quad " \text{ red} \end{array}$$

Sol: The number of possible outcomes is 52, of which 13 are hearts. Therefore $P(\text{heart}) = \frac{13}{52} = \frac{1}{4}$

Ex: Five cards are drawn randomly from an ordinary deck, find the probability of having 2 aces and 3 jacks

sol. The number of ways of being dealt 2 aces from 4 is: $\binom{4}{2} = \frac{4!}{2!2!} = 6$

" " " " " " " 3 jack " 4 is: $\binom{4}{3} = 4$

" " " " " " " 5 cards " 52 is $= \binom{52}{5} = 2598960$

$\therefore P(2 \text{ aces and } 3 \text{ jacks}) = \frac{6 \cdot 4}{2598960} = 0.9 \times 10^{-5}$

<u>Theorem</u>: Additive Rule. If A and B are any two events, then:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \text{(Proof ?)}$$

<u>Corollary</u>: If A and B are mutually exclusive, then

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

<u>Corollary</u>: If $A_1, A_2, A_3, \ldots, A_n$ are mutually exclusive, then:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_n).$$

<u>Note</u>: If $A_1, A_2, \ldots, A_n$ is a partition of a sample space S, then

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) &= P(A_1) + P(A_2) + P(A_3) + \cdots + P(A_n) \\ &= P(S) \\ &= 1 \end{aligned}$$

<u>Ex</u>: The probability that a student passes Mathematics is $\frac{2}{3}$, and the probability that he passes English is $\frac{4}{9}$. If the probability of passing at least one course is $\frac{4}{5}$, what is the probability that he will pass both courses?

<u>Sol</u>. Let M denotes for Mathematics
and E " " English

$$\begin{aligned} \therefore P(M \cap E) &= P(M) + P(E) - P(M \cup E) \\ &= \frac{2}{3} + \frac{4}{9} - \frac{4}{5} \\ &= \frac{14}{15} \end{aligned}$$

<u>Ex</u> What is the probability of getting a total of 7 or 11 when a pair of dice is tossed?

<u>Sol</u>. Let A denotes the event of getting a total of 7
and B " " " " " " " " 11

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \qquad \text{(why?)} \\ &= \frac{1}{6} + \frac{1}{18} \\ &= \frac{2}{9} \end{aligned}$$

<u>Theorem</u>: If A and B are complementary events, then

$$P(A) + P(B) = 1 \qquad \text{(Proof ?)}$$

<u>Ex</u> A coin is tossed 6 times in succession. What is the probability that at least 1 head occurs?

<u>Sol</u>. Let E be the event that at least 1 head occurs.
The sample space S consists of $2^6 = 64$ sample points.
Now E^c is the event that no head occurs.

$\because P(E^c) = \frac{1}{64}$ and $P(E) + P(E^c) = 1$

$\therefore P(E) = 1 - \frac{1}{64} = \frac{63}{64}$

<u>Def</u>: The <u>Conditional Probability</u> of B, given A, denoted by $P(B|A)$, is defined by the equation:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad \text{if } P(A) \neq 0$$

As an additional illustration, suppose that our sample S is the population of adults in a small town who have completed the requirements for a college degree. We shall categorize them according to sex and employment status.

	Employed	Unemployed
Male	460	40
Female	140	260

On of these individuals is to be selected at random. We shall be concerned with the following events:

M: a man is chosen and E: the chosen one is employed

Using the reduced sample space E, we find that

$$P(M|E) = \frac{460}{600} = \frac{23}{30}$$

Now if we denote $n(A)$ to the number of elements in any set A, and using this notation, we can write:

$$P(M|E) = \frac{n(E\cap M)}{n(E)} = \frac{n(E\cap M)/n(S)}{n(E)/n(S)} = \frac{P(E\cap M)}{P(E)}$$

$$= \frac{\frac{460}{900}}{\frac{600}{900}} = \frac{\frac{23}{45}}{\frac{2}{3}} = \frac{23}{30} \quad \text{(as before)}.$$

Ex: The probability that a regulary scheduled flight departs on time is $P(D)=0.83$, the probability that it arrives on time $P(A)=0.92$ and the probability that it departs and arrives on time is $P(D\cap A)=0.78$.
Find the probability that a plane (a) arrives on time given that it departed on time, and (b) departed on time given that it has arrived on time.

Sol. (a) $P(A|D) = \frac{P(A\cap D)}{P(D)} = \frac{0.78}{0.83} = 0.94$

(b) $P(D|A) = \frac{P(D\cap A)}{P(A)} = \frac{0.78}{0.92} = 0.85$

Ex: Consider the event B of getting a perfect square when a die is tossed. The die is constructed so that the even numbers are twice as likely to occur as the odd numbers.

Now we have $S=\{1,2,3,4,5,6\}$

$P(2)=P(4)=P(6)=2P(1)=2P(3)=2P(5)$

$\because P(S)=1$, hence $P(1)+2P(1)+P(1)+\cdots=1 \Rightarrow P(1)=\frac{1}{9}$

$\therefore P(1)=P(3)=P(5)=\frac{1}{9}$ and $P(2)=P(4)=P(6)=\frac{2}{9}$

$\therefore P(B)=P(1)+P(4)=\frac{1}{9}+\frac{2}{9}=\frac{1}{3}$

Now if we know that the toss of the die resulted is a number greater than 3. So to get $P(B)$ we must take into consideration that the result is the event $A=\{4,5,6\}$ a reduced sample space.

Hence we can talk only about Probability of having perfect square knowing that the result is either 4 or 5 or 6. That is we are conserned by $B|A$ which is only $\{4\}$, and hence:

$$P(B|A) = \frac{2}{5} \qquad \text{(why ?)}$$

This example illustrates that events may have different probabilities when considered relative to different sample spaces.

We can also write $P(B|A) = \frac{2}{5} = \frac{\frac{2}{9}}{\frac{5}{9}} = \frac{P(A \cap B)}{P(A)}$

<u>Ex</u>: Consider an experiment in which 2 cards are drawn in succession from an ordinary deck, with replacement. Let the event A denotes that the first card is an ace and B denotes that the second card is a spade.

Hence $P(B|A) = \frac{13}{52} = \frac{1}{4}$

and $P(B) = \frac{13}{52} = \frac{1}{4}$

$\Rightarrow P(B|A) = P(A)$ (why?)

<u>Def:</u> Two events A and B are said to be <u>Independent</u> if either $P(B|A) = P(B)$ or $P(A|B) = P(A)$.
Otherwise, A and B are <u>Dependent</u>.

<u>Theorem</u>: If in an experiment the events A and B can both occur, then: $P(A \cap B) = P(A) . P(B|A)$.
If A and B are independent then: $P(A \cap B) = P(A).P(B)$

Ex: A fuse box contains 20 fuses, of which 5 are defective. If 2 fuses are selected at random and removed from the box in succession without replacement. What is the probability that both fuses are defective?

Sol. let A denotes to the event that the first fuse is defective
and B " " " " " " Second " " "

$$\therefore P(A \cap B) = P(A) \cdot P(B|A) = \frac{1}{4} \cdot \frac{4}{19} = \frac{1}{19}$$

Note: If A and B are independent, then

$$P(A \cap B) = P(A) \cdot P(B)$$

Ex: One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?

Sol. let B_1, B_2 and W_1 represents, respectively, the drawing of a black ball from bag 1, a black ball from bag 2 and a white ball from bag 1. We are interested in the union of the mutually exclusive events $(B_1 \cap B_2)$ and $(W_1 \cap B_2)$

Now
$$P[(B_1 \cap B_2) \cup (W_1 \cap B_2)] = P(B_1 \cap B_2) + P(W_1 \cap B_2)$$
$$= P(B_1) \cdot P(B_2|B_1) + P(W_1) \cdot P(B_2|W_1)$$
$$= \frac{3}{7} \cdot \frac{6}{9} + \frac{4}{7} \cdot \frac{5}{9}$$
$$= \frac{38}{63}$$

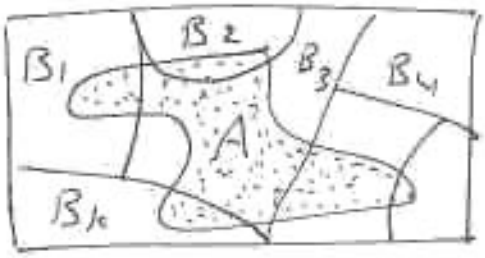
Theorem: (Total Probability)

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

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_k)P(A|B_k)$$

Proof:

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_k \cap A)$$



Theorem: (Bayes' Rule)

If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S where $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$ then for any event A in S such that $P(A) \neq 0$

$$P(B_r|A) = \frac{P(B_r) \cdot P(A|B_r)}{P(B_1)P(A|B_1) + P(B_2) \cdot P(A|B_2) + \cdots + P(B_k)P(A|B_k)}$$

for $r = 1, 2, \ldots, k$

Ex: Three members of a private club have been nominated for the office of president. The probability that A will be elected is 0.3, the probability that B will be elected is 0.5 and the probability that C is elected is 0.2. Should A be elected, the probability for an increase in membership fees is 0.8. Should B or C be elected, the corresponding probabilities for an increase in membership fees are 0.1 and 0.4. What is the probability that there will be an increase in membership fees?

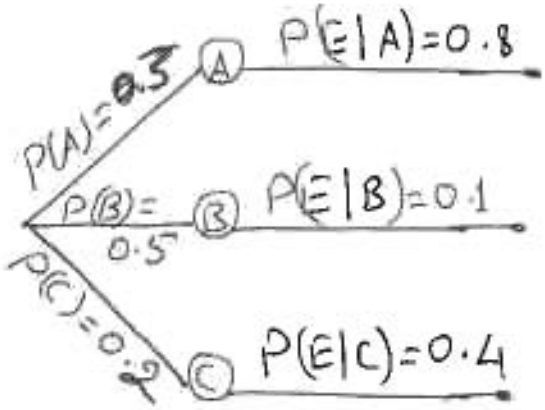
Sol. Let E denotes for an increase in membership fees.

$$P(E) = P(A).P(E|A) + P(B).P(E|B) + P(C).P(E|C)$$

$$= (0.3)(0.8) + (0.5)(0.1) + (0.2)(0.4)$$

$$= 0.24 + 0.05 + 0.08$$

$$= 0.37$$



Tree diagram

Now, what is the probability that C will be elected given that an increase in membership fees has occured.

$$P(C|E) = \frac{P(C).P(E|C)}{P(A).P(E|A) + P(B).P(E|B) + P(C).P(E|C)}$$

$$= \frac{(0.2)(0.4)}{(0.3)(0.8) + (0.5)(0.1) + (0.2)(0.4)}$$

$$= \frac{0.08}{0.37}$$

$$= \frac{8}{37}$$

Exercises:

① List the elements of each of the following sample spaces:

ⓐ The set of integers between 1 and 50 divisible by 8.

ⓑ The set $S = \{x \mid x^2 + 4x - 5 = 0\}$

ⓒ The set of outcomes when a coin is tossed until a Tail or 3 Heads appear.

ⓓ The set $S = \{x \mid x \text{ is a continent}\}$

ⓔ The set $S = \{x \mid 2x - 4 \geq 0 \text{ and } x < 1\}$

② Write the Sample Space (S) for the random experiment that consists of tossing a die and then flipping a coin once if the number on the die is even and twice if the number on the die is odd.

③ For the sample space of Exercise ②,

ⓐ List the elements corresponding to the event A that a number less than 3 occurs on the die.

ⓑ List the elements corresponding to the event B that 2 tails occur.

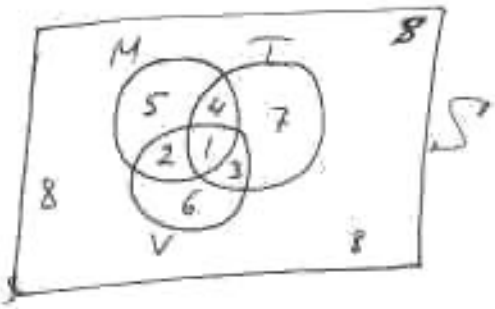
ⓒ " " " " " " " A^c.

ⓓ " " " " " " " $A^c \cap B$.

ⓔ " " " " " " " $A \cup B$.

④ Suppose you have the following Venn Diagram.

Express the figures in the diagram by the union and intersection of the events M, T, V and their complements.



⑤ If an experiment consists of throwing a die and then drawing a letter at random from the English alphabet, how many points are in the sample space?

⑥ ⓐ How many ways can 6 people be lined up to get on a bus?

ⓑ If a certain 3 persons instist on following each other, how many ways are possible?

ⓒ " " " 2 " refuse to follow each other, how many ways are possible?

(7) From a group of 4 men and 5 women, how many committees of size 3 are possible?

(a) with no restrictions. (b) with one man and two women.

(c) with two men and one women if a certain man must be on the committee.

(8) Three men are seeking public office. Candidates A and B are given about the same chance of winning, but candidate C is given twice the chance of either A or B.

(a) What is the probability that C wins?

(b) " " " " " A does not win?

(9) If A and B are mutually exclusive events and $P(A) = 0.3$, $P(B) = 0.5$. Find (a) $P(A \cup B)$, (b) $P(A^c)$, (c) $P(A^c \cap B)$

(10) If 3 books are picked at random from a shelf containing 5 novels, 3 books of poems, and a dictionary, what is the probability that (a) the dictionary is selected?

(b) 2 novals and 1 book of poems are selected?

(11) In a college of 100 students, 54 studied mathematics, 69 studied history and 35 studied both. If one of these students is selected at random, find the probability that:

(a) the student takes mathematics or history.

(b) the student does not take either of these subjects.

(c) the student takes history but not mathematics.

(12) A random sample of 200 adults are classified below according to sex and level of education attained.

A person is selected at random. Find the probability that:

	Male	Female
Elementary	38	45
Secondary	28	50
College	22	17

(a) The person is male given that the person has secondary education.

(b) The person does not have a college degree given that the person is female.

(13) Suppose that colored balls are distributed in three indistinguishable boxes as follows:

A box is selected at random from which a ball is drawn at random.

(a) Find the probability that the ball is red.

	Box		
	1	2	3
Red	2	4	3
White	3	1	4
Blue	5	3	3

(b) Given that the ball is red, what is the probability that box 3 was selected?

(14) A man owns 2 cars, 1 a compact and 1 a standard model. About $\frac{3}{4}$ of the time he uses the compact to travel to work, and about $\frac{1}{4}$ of the time the larger car is used. When he uses the compact car, he usually gets home by 5:30 p.m. about 75% of the time, if he uses the standard-sized car, he gets home by 5:30 p.m about 60% of the time. If he gets home at 5:30 p.m, what is the probability that he used the compact car?

2- Distribution of Random Variables:

Def: A function whose value is a real number determined by each element in the sample space is called a random variable.

ex. In tossing a coin three times in succession we have the sample space:

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

If one is concerned only with the number of heads that fall, then a numerical value of 0, 1, 2, or 3 will be assigned to each sample point. The numbers 0, 1, 2, and 3 are random quantities determined by the outcome of an experiment.

We shall use a capital letter, say X, to denote a random variable and its corresponding small letter, x in this case, for one of its values.

∴ We say, let X be a random variable that denotes the number of times that a head occur in the above S.

∴ $x = 0, 1, 2, 3$

Def: Discrete Sample Space. 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.

ex: $S = \{RR, RB, BR, BB\}$ drawing two balls in succession without replacement from an urn containing 4 red and 3 black balls.

ex: $S = \{F, NF, NNF, NNNF, NNNNF, \ldots\}$ throwing a die until a 5 (F) occurs.

Def: 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.

ex: time intervals, distance intervals

Def: A table or a formula listing all possible values that a discrete random variable can take on, along with the associated probabilities, is called a <u>discrete prob. dist.</u> or <u>Prob. Mass Function</u>.

ex Find the prob. dist. of the sum of the numbers when a pair of dice is tossed.

sol.

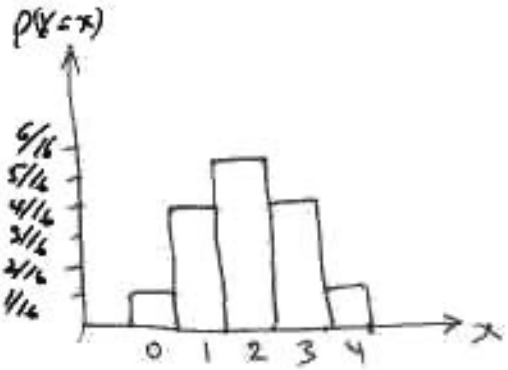
$X = x$	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Where X is a random variable denotes to the sum of the two dice.

ex Find a formula for the prob. dist. of a number of heads when a coin is tossed 4 times.

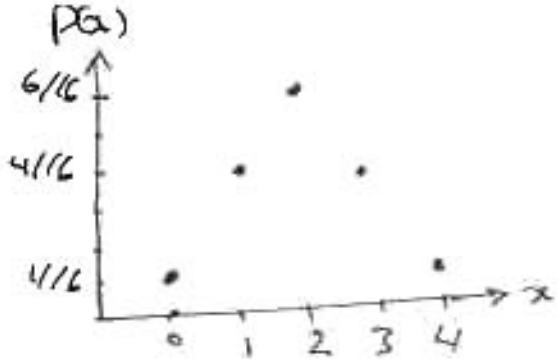
sol. Let X be a random variable denotes to the number of heads when a coin is tossed 4 times, then $x = 0, 1, 2, 3, 4$

$$P(X=x) = \frac{{}^4C_x}{16}, \quad x = 0, 1, 2, 3, 4$$



Probability Histogram.

or



Def: The function with values $f(x)$ is called a <u>probability density fun.</u> for the continuous random variable X if the total area under its curve and above the x-axis is equal to 1 and if the area under the curve between any two ordinates $x=a$ and $x=b$ gives the prob. that X lies between a and b.

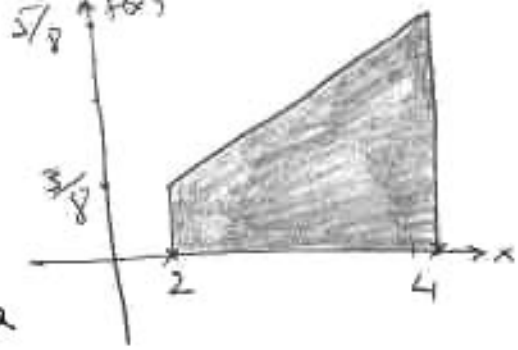
ex A continuous random variable X that can assume values between $x=2$ and $x=4$ has a density function given by

$$f(x) = \frac{x+1}{8}$$

ⓐ Show that $P(2<x<4)=1$

ⓑ Find $P(x<3.5)$

ⓒ Find $P(2.4<x<3.5)$



Sol ⓐ since the shaded region in the figure is a trapezoid

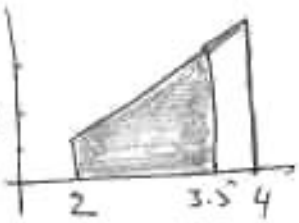
$$\therefore \text{area} = \frac{(\text{sum of the parallel sides}) * \text{base}}{2}$$

$$= \frac{[f(2)+f(4)](2)}{2} = \frac{\left(\frac{3}{8}+\frac{5}{8}\right)2}{2} = 1 \equiv P(2<x<4)$$

or $P(2<x<4) = \int_2^4 f(x)\,dx = \int_2^4 \frac{x+1}{8}\,dx = \frac{1}{8}\left[\frac{x^2}{2}+x\right]_2^4 = 1$

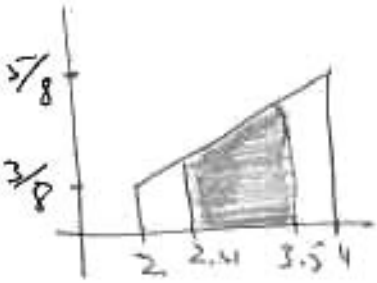
ⓑ $P(x<3.5) = \frac{\left(\frac{3}{8}+\frac{4.5}{8}\right)(1.5)}{2} = 0.70$

or $P(x<3.5) = P(2<x<3.5) = \int_2^{3.5} f(x)\,dx = 0.70$



ⓒ $P(2.4<x<3.5) = \frac{\left(\frac{3.4}{8}+\frac{4.5}{8}\right)(1.1)}{2} = 0.54$

or $P(2.4<x<3.5) = \int_{2.4}^{3.5} f(x)\,dx = \frac{1}{8}\left[\frac{x^2}{2}+x\right]_{2.4}^{3.5} = 0.54$



Exercises:

① Classify the following random variables as discrete or continuous.

ⓐ the number of car accidents each year in Baghdad.

ⓑ the length of time to play 5 points in table tennis.

ⓒ the amount of milk produced yearly by a particular cow.

ⓓ the number of eggs laid each month by 1 hen.

ⓔ the number of building permits issued each month in a certain city.

ⓕ the weight of grain in pounds produced per acre.

② Let W be a random variable giving the number of heads minus the number of tails in 3 tosses of a coin. List the elements of the sample space S for the 3 tosses of the coin and to each sample point assign a value w of W.

③ Find the prob. dist. of the random variable W in exercise 2, assuming the coin is biased so that a head is twice as likely to occur as tail.

④ Find the prob. dist. of the number of Blue balls when 4 balls are selected at random from a collection consisting of 5 Blue, 2 Red and 3 white balls. Express your results by means of a formula.

⑤ Find the prob. dist. for the number of spades, if 3 cards are drawn in succession from a deck without replacement.

⑥ If the density function of the random variable X is given by:

$$f(x) = \frac{2(1+x)}{27} \quad 2 \leq x \leq 5 \quad , \text{ zero elsewhere}$$

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

⑦ A continuous random variable X has the density function

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

ⓐ Show that $P(0<X<2)=1$, ⓑ Find $P(X<1.2)$

<u>Def</u>: A table of formula listing all possible values of x and y of a discrete random variables X and Y, together with the associated probabilities $P(x,y)$, is called a <u>joint prob. dist.</u>

When X & Y are continuous random variables the joint function will be called <u>joint density function</u>.

<u>ex</u> Two refills for a ballpoint pen are selected at random from a box that contains 3 blue, 2 red and 3 green refills. If X is the number of blue refills and Y is the number of red refills selected. Find ⓐ the joint prob. fun. $P(x,y)$ and

ⓑ $P(x,y \mid x+y \leq 1)$

<u>sol.</u> The possible pairs of values (x,y) are $(0,0), (0,1), (1,0), (1,1), (0,2)$ and $(2,0)$

So we can represent the following table for the joint prob. dist.

$P(x,y)$		x = 0	1	2	Row total
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
	2	$\frac{1}{28}$			$\frac{1}{28}$
Column totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

<u>Note that</u>: By formula we can represent $P(x,y)$ as:

$$P(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}},$$

for $x = 0,1,2$, $y = 0,1,2$ & $0 \leq x+y \leq 2$

(b) $P(x,y \mid x+y \leq 1) = P(X=0, Y=0) + P(X=1, Y=0) + P(X=0, Y=1)$

$$= \frac{3}{28} + \frac{9}{28} + \frac{3}{14} = \frac{9}{14}$$

<u>Def</u>: Given the values of the joint prob. dist. of the discrete random variables X and Y, the one-dimensional prob. dist. of X alone, with values $g(x)$, and Y alone, with values $h(y)$ are given by the column and row totals in the above table. We define two functions to be the marginal distributions of X and Y, respectively.

Hence: $g(0) = P(X=0)$

$\therefore g(0) = P(X=0, Y=0) + P(X=0, Y=1) + P(X=0, Y=2)$

$$= P(0,0) + P(0,1) + P(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$g(1) = \frac{15}{28}$, $g(2) = \frac{3}{28}$

similarly:

$h(0) = P(Y=0) = P(X=0, Y=0) + P(X=1, Y=0) + P(X=2, Y=0)$

$$= P(0,0) + P(1,0) + P(2,0)$$

$$= \frac{3}{28} + \frac{9}{28} + \frac{3}{28} = \frac{15}{28}$$

$h(1) = \frac{3}{7}$, $h(2) = \frac{1}{28}$

x	0	1	2
$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

y	0	1	2
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

<u>Def</u>: The <u>conditional dist.</u> of a discrete random variable Y, given that $X=x$ is given by:

$$P(y \mid X=x) \equiv P(y \mid x) = \frac{P(x,y)}{g(x)}, \quad g(x)>0$$

Similarly, the <u>conditional dist.</u> of a discrete random variable X, given that $Y=y$ is given by:

$$P(x \mid Y=y) \equiv P(x \mid y) = \frac{P(x,y)}{h(y)}, \quad h(y)>0$$

ex Referring to the previous example, find $P(x \mid 1)$ for all values of x and determind $P(X=0 \mid Y=1)$

<u>sol</u> since $h(1) = P(0,1) + P(1,1) + P(2,1)$
$= \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$

Then $P(x \mid 1) = \frac{P(x,1)}{h(1)} = \frac{7}{3} P(x,1)$, $x = 0,1,2$

Now
$P(0 \mid 1) = \frac{7}{3} P(0,1) = \frac{7}{3} \cdot \frac{3}{14} = \frac{1}{2}$
$P(1 \mid 1) = \frac{7}{3} P(1,1) = \frac{7}{3} \cdot \frac{3}{14} = \frac{1}{2}$
$P(2 \mid 1) = \frac{7}{3} P(2,1) = \frac{7}{3} \cdot 0 = 0$

So the conditional dist. of X given $Y=1$ is

x	0	1	2
$P(x \mid 1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Finally: $P(X=0 \mid Y=1) = P(0 \mid 1) = \frac{1}{2}$

<u>Def</u>: The random variables X and Y are said to be <u>statistically independent</u> if and only if (iff): $P(x,y) = g(x) \cdot h(y)$ for all possible values of X and Y.

Exercises:

① Determine the value of the constant c so that the following functions represent joint prob. dist. of the random variables X and Y:

ⓐ $P(x,y) = cxy$, for $x=1,2,3$; $y=1,2,3$

ⓑ $P(x,y) = c|x-y|$, for $x=-2,0,2$; $y=-2,3$

② If the joint prob. dist. of X and Y is given by:

$$P(x,y) = \frac{x+y}{30} \qquad x=0,1,2,3 \; ; \; y=0,1,2$$

. Then find:

ⓐ $P(X \leq 2, Y=1)$, ⓑ $P(X>2, Y \leq 1)$, ⓒ $P(X>Y)$, ⓓ $P(X+Y=4)$

③ From a sack of fruit containing 3 oranges, 2 apples, and 3 bananas a random sample of 4 pieces of fruit is selected. If X is the number of oranges and Y is the number of apples in the sample, find

ⓐ the joint prob. dist. of X and Y.

ⓑ $P(x,y \mid x+y \leq 2)$.

④ Let X denote the number of heads and Y denote the number of heads minus the number of tails when 3 coins are tossed. Find the joint prob. dist. of X and Y.

⑤ Refering to exercise 2, find

ⓐ the marginal dist. of X and ⓑ the marginal dist of Y.

⑥ Refering to exercise 3. Find ⓐ $P(y \mid X=2)$ for all values of y ⓑ $P(Y=0 \mid X=2)$

⑦ Suppose that X and Y have the following joint prob. function:

x \ y	1	3	5
2	0.10	0.20	0.10
4	0.15	0.30	0.15

ⓐ Find the marginal dist. of the random variable X.

ⓑ " " " " " " " " Y.

⑧ Determine whether X and Y in exercise 6 are dependent or independent.

The <u>Mean</u> of a random variable X or the <u>Expected Value</u> or the <u>Mathematical Expectation</u> is denoted by μ and defined:

$$\mu = E(X) = \begin{cases} \sum_{\text{all } x} x\, P(X=x) & \text{for discrete random variable} \\ \int_{\text{all } x} x\, f(x)\, dx & \text{" continuous " "} \end{cases}$$

and the Mean of a function of one variable $g(X)$ is

$$\mu_{g(X)} = E(g(X)) = \begin{cases} \sum_{\text{all } x} g(X)\, P(X=x) & \text{d.r.v.} \\ \int_{\text{all } x} g(x)\, f(x)\, dx & \text{c.r.v.} \end{cases}$$

and the Mean of a function of two variables $g(X,Y)$ is:

$$\mu_{g(X,Y)} = E(g(X,Y)) = \begin{cases} \sum_{i=1}^{m} \sum_{j=1}^{n} g(X_i, Y_j)\, P(x_i, y_i) \\ \int_y \int_x g(X,Y)\, f(x,y)\, dx\, dy \end{cases}$$

<u>ex</u> Find the Expected value of X, where X represents the outcome when a die is tossed.

<u>sol</u>

$X=x$	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\mu = E(X) = 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right) = 3.5$$

<u>ex</u> Find the expected value (number) of boys on a committee of 3, selected at random from 4 boys and 3 girls.

<u>sol</u> $P(X=x) = \dfrac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}$, $x = 0, 1, 2, 3$

or

$X=x$	0	1	2	3
$P(X=x)$	$\frac{1}{35}$	$\frac{12}{35}$	$\frac{18}{35}$	$\frac{4}{35}$

$$\mu = E(X) = 0\left(\tfrac{1}{35}\right) + 1\left(\tfrac{12}{35}\right) + 2\left(\tfrac{18}{35}\right) + 3\left(\tfrac{14}{35}\right) = 1.7 \approx \underline{\underline{2}}$$

ex Let X be a random variable with prob. dist. as follows:

$X = x$	0	1	2	3
$P(X=x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find. $E(X-\mu)^2$.

Sol. $\mu = E(X) = \sum_{\text{all } x} x P(x) = 0(\frac{1}{3}) + 1(\frac{1}{2}) + 2(0) + 3(\frac{1}{6}) = 1$

Now

$E(X-\mu)^2 = E(X-1)^2 = \sum_{\text{all } x} (x-1)^2 P(x) = (-1)^2(\frac{1}{3}) + (0)^2(\frac{1}{2}) + (1)^2(0) + (2)^2(\frac{1}{6}) = 1$

ex Refering to example of page 18. Find the expected value of $g(X,Y) = XY$

Sol. $E(XY) = \sum_{x=0}^{2} \sum_{y=0}^{2} xy\, P(x,y)$

$= (0)(0).P(0,0) + (0)(1)P(0,1) + (0)(2)P(0,2)$
$+ (1)(0)P(1,0) + (1)(1)P(1,1)$
$+ (2)(0)P(2,0)$

$= P(1,1) = \frac{3}{14}$

Note that if $g(X,Y) = X$ in the preceding **example**, then we have

$$\mu_X = E(X) = \sum_{x=1}^{m} \sum_{y=1}^{n} x_i P(x_i, y_j) = \sum_{i=1}^{m} x_i\, g(x_i)$$

hence $\mu_X = (0)(\frac{5}{14}) + (1)(\frac{15}{28}) + (2)(\frac{3}{28}) = 0.75$

and $\mu_Y = E(Y) = \sum_{y=1}^{n} \sum_{x=1}^{m} y_j P(x_i, y_j) = \sum_{y=1}^{n} y_j\, h(y_j)$

hence $\mu_Y = (0)(\frac{15}{28}) + (1)(\frac{3}{7}) + (2)(\frac{1}{28}) = 0.5$

Def.: The <u>Variance</u> of a random variable X is denoted by σ_x^2 and defined as:

$$V(X) \equiv \sigma_x^2 = E(X-\mu)^2 = \begin{cases} \sum_{i=1}^{n} \left((x_i-\mu)^2\right) P(x_i) & , X \text{ is d.r.v.} \\ \int_{\text{all } x} (x_i-\mu)^2 f(x)\, dx & , X \text{ is c.r.v.} \end{cases}$$

ex Calculate the variance of the random variable X in the last example on page 21.

Sol:

$$\sigma^2 = \sum_{x=0}^{3} (x-\mu)^2 P(x)$$

$$= \left(0-\frac{12}{7}\right)^2 \frac{1}{35} + \left(1-\frac{12}{7}\right)^2 \left(\frac{12}{35}\right) + \left(2-\frac{12}{7}\right)^2 \left(\frac{18}{35}\right) + \left(3-\frac{12}{7}\right)^2 \left(\frac{4}{35}\right)$$

$$= \frac{24}{49}$$

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

Proof:

$$\sigma^2 = \sum_{i=1}^{n} (x_i-\mu)^2 P(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 P(x_i) - 2\mu \sum_{i=1}^{n} x_i P(x_i) + \mu^2 \sum_{i=1}^{n} P(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 P(x_i) - 2\mu^2 + \mu^2$$

$$= E(X^2) - \mu^2 \qquad \text{(Q.E.D.)}$$

ex Consider the following prob. dist.

$X = x$	0	1	2	3
$P(X=x)$	0.51	0.38	0.10	0.01

Find $E(X)$ and $V(X)$.

Sol

$$E(X) = \mu_x = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61$$

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87$$

$$\therefore V(X) \equiv \sigma_x^2 = 0.87 - (0.61)^2 = 0.4979$$

Note that the variance of a function of X, say $g(X)$, is:

$$V(g(X)) \equiv \sigma_{g(X)}^2 = E\left(g(X) - \mu_{g(X)}\right)^2$$

Properties of the Mean and Variance:

Let X and Y be two random variables and a and b constants

① $E(aX+b) = aE(X)+b \Rightarrow E(b) = b$

② $E(X \mp Y) = E(X) \mp E(Y)$

③ If X and Y are independent then $E(XY) = E(X).E(Y)$

④ $V(X+b) = V(X) \Rightarrow V(b) = 0$

⑤ $V(aX) = a^2 V(X)$

⑥ $V(X \mp Y) = V(X) + V(Y)$

ex Suppose the number of cars, X, that pass through a car wash between 4:00 p.m and 5:00 p.m on ~~any~~ sunny Friday has the following prob. dist.

$X=x$	4	5	6	7	8	9
$P(X=x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X-1$ represent the amount of money paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

sol. $E(2X-1) = 2E(X)-1$

$= 2\left[4(\frac{1}{12}) + 5(\frac{1}{12}) + 6(\frac{1}{4}) + 7(\frac{1}{4}) + 8(\frac{1}{6}) + 9(\frac{1}{6})\right] - 1$

$= 2\ \frac{41}{6} - 1 = 12.67$

ex If X and Y are independent random variables with variances $\sigma_X^2 = 1, \sigma_Y^2 =$ Find the variance of the random variable $Z = 3X - 2Y + 5$.

sol $V(Z) = V(3X-2Y+5) = V(3X-2Y)$

$= 9V(X) + 4V(Y)$

$= 9.1 + 4.2$

$= 17$

Exercises:

① A shipment of 7 television sets contains 2 defectives. A hotel makes a random purchase of 3 of the sets. If X is the number of defective sets purchased by the hotel, find the mean of X, then find its variance.

② The Prob. dist. of the discrete random variable X is:

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

Find the mean and variance of X.

③ Find the expected number of red balls when 4 balls are selected at random from a collection consisting of 5 Red, 2 White and 3 Blue bal

④ In a gambling game a man is paid 3 units if he draws a jack or queen and 5 units if he draws a king or ace from an ordinary deck of 52 playing cards. If he draws any other card, he loses. How much should he pay to play if the game is fair?

⑤ A race-car driver wishes to insure his car for the racing season for 50.000 (units of money). The insurance company estimates a total loss may occur with prob. (0.002), a 50% loss with prob. (0.01) and a 25% loss with prob. (0.1). Ignoring all the other partial losses. What premium should the insurance company charge each season to realize an average profit of 500 units?

⑥ Let X represent the outcome when a balanced die is tossed. Find $\mu_{g(x)}$, where $g(X) = 3X^2+4$

⑦ Suppose that X and Y have the following joint prob. function:

$y \backslash x$	2	4
1	0.10	0.15
3	0.20	0.30
5	0.10	0.15

Find the expected value of XY^2.

⑧ Suppose that X and Y have the following joint prob. fun.

$x \backslash y$	2	3	4
1	0.06	0.15	0.09
2	0.14	0.35	0.21

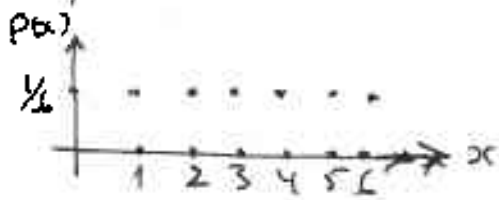
Find μ_x & μ_y.

③ Some Discrete Prob. Distributions.

Discrete Uniform Dist: If the random variable X assumes the values $x_1, x_2, \dots, x_k$, with equal probabilities, then the discrete uniform dist. is given by:

$$P(X=x) \equiv P(x;k) = \frac{1}{k} \quad \text{for } x = x_1, x_2, \dots, x_k \text{, zero elsewhere}$$

ex. ① when a die is tossed, each element of the sample space S occurs with prob. $\frac{1}{6}$, therefore, we have a uniform dist. with $P(X=x) \equiv P(x;6) = \frac{1}{6}$, for $x = 1,2,3,4,5,6$, zero elsewhere



ex. ② Suppose that sample space consists of 4 students A, B, C, and D from which 2 are to be chosen at random for a committee. Find the prob. dist. of all possible subsets of size 2.

sol. Since the no. of all possible combinations is $\binom{4}{C_2} = 6 \equiv k$, which could be listed as $S = \{AB, AC, AD, BC, BD, CD\}$ where each sample point has the same prob. of being drawn.

$\therefore P(x;6) = \frac{1}{6}$, $x = 1,2,3,4,5,6$, zero, elsewhere

so $P(x;5) = \frac{1}{6}$ which means $P(B \text{ and } D \text{ are chosen})$

ex. ③ Find the uniform dist. for the subsets of months of size 3.

sol. $\binom{12}{C_3} = 220 \equiv k$

$\therefore P(x;220) = \frac{1}{220}$, $x = 1,2,\dots,220$, zero elsewhere

<u>Binomial Dist.</u> If a binomial trial can result in a success with prob. p and a failure with prob. $q = 1-p$, then the prob. dist. of the <u>binomial random variable</u> X, the no. of successes in n independent trials is

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

<u>ex.</u> ① Find the prob. of obtaining exactly three 2's if an ordinary die is tossed 5 times.

<u>sol</u> $n=5 \quad x=3 \quad p=\frac{1}{6}, \quad q = 1-\frac{1}{6} = \frac{5}{6}$

$$P(X=3) \equiv P(3;5,\tfrac{1}{6}) = \binom{5}{3}\left(\tfrac{1}{6}\right)^3\left(\tfrac{5}{6}\right)^2$$

$$= \frac{5!}{3!\,2!}\,\frac{5^2}{6^5} = 0.032$$

<u>ex</u> ② A football team wins by prob. 75%. What is the prob. of winning two games from the next 4 games.

<u>sol</u> $n=4, \quad x=2, \quad p=\frac{3}{4}, \quad q=\frac{1}{4}$

$$P(X=2) \equiv P(2;4,\tfrac{3}{4}) = \binom{4}{2}\left(\tfrac{3}{4}\right)^2\left(\tfrac{1}{4}\right)^2 = 0.211$$

<u>Note</u>: The binomial dist. derives its name from the fact that the $(n+1)$ terms in the binomial expansion of $(q+p)^n$ correspond to the various values of $b(x;n,p)$ for $x=0,1,2,\dots,n$. That is,

$$(q+p)^n = \binom{n}{0}q^n + \binom{n}{1}pq^{n-1} + \binom{n}{2}p^2q^{n-2} + \dots + \binom{n}{n-1}p^{n-1}q + \binom{n}{n}p^n$$

$$= b(0;n,p) + b(1;n,p) + b(2;n,p) + \dots + b(n-1;n,p) + b(n;n,p)$$

since $p+q=1$, we see that $\sum_{x=0}^{n} b(x;n,p) = 1$

ex.③ The prob. that a patient recovers from a rare blood disease is (0.4), if 15 people are known to have contracted this disease, what is the prob. that, ⓐ at least 10 survive; ⓑ from 3 to 8 survive; and ⓒ exactly 5 survive?

sol Let X be a random variable that represents the number of people that survive.

ⓐ $P(X \geq 10) = P(X=10) + P(X=11) + \cdots + P(X=15)$

$= b(10; 15, 0.4) + b(11; 15, 0.4) + \cdots + b(15; 15, 0.4)$

$= \cdots = 0.0338$

or $P(X \geq 10) = 1 - P(X < 10)$

$= 1 - \sum_{x=0}^{9} b(x; 15, 0.4)$

$= 1 - 0.9662 = 0.0338$

ⓑ $P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15, 0.4)$

$= 0.8779$

Theorem: The Mean (expected value) and Variance of the binomial distribution $b(x; n, p)$ are:

$$\mu = np \quad \& \quad \sigma^2 = npq$$

Proof: Let the outcome on the j-th trial be represented by the random variable I_j, which assumes the values 0 and 1 with probabilities q and p, respectively. This is called a Bernoulli variable or an Indicator variable, since $I_j = 0$ indicates a failure and $I_j = 1$ indicates a success.

Therefore, in a binomial experiment the number of successes can be written as the sum of the n independent indicator variables.

Hence, $X = I_1 + I_2 + \cdots + I_n$

The mean of any I_j is $E(I_j) = 0 \cdot q + 1\, p = p$

$$\therefore \mu = E(X) = E[I_1 + I_2 + \cdots + I_n]$$
$$= E(I_1) + E(I_2) + \cdots + E(I_n) = \underbrace{p + p + \cdots + p}_{n \text{ times (terms)}}$$
$$= np$$

Now
$$V(X) = \sigma_X^2 = \sigma_{I_1}^2 + \sigma_{I_2}^2 + \cdots + \sigma_{I_n}^2$$

Since $\sigma_{I_j}^2 = E(I_j - p)^2 = E(I_j^2) - p^2 = 0^2 q + 1^2 p - p^2 = p - p^2$
$$= p(1-p) = pq.$$

Hence $\sigma_X^2 = \underbrace{pq + pq + \cdots + pq}_{n \text{ times (terms)}} = npq$

<u>Chebyshev's Theorem</u>: At least the fraction $(1 - \frac{1}{k^2})$ of the measurements of any set of data must lie within k standard deviations of the mean.

ex For $k=2$ the theorem states that at least $1 - \frac{1}{2^2} = \frac{3}{4}$, or 75%, of the measurements must lie within 2 standard deviations on either side of the mean. That is, $\frac{3}{4}$ or more of the observations must lie in the interval $\mu \mp 2\sigma$. Considering our set of data to be a sample, then for $k=2$ the theorem states that at least $\frac{3}{4}$ of the measurements must lie in the interval $\bar{x} \mp 2s$.

<u>ex. 4</u> Using Chebyshev's Theorem. Find and interpret the interval $\mu \mp 2\sigma$ for ex. 3.

Set $\mu = (15)(0.4) = 6$, $\sigma^2 = (15)(0.4)(0.6) = 3.6 \Rightarrow \sigma = 1.897$

$\therefore \mu \mp 2\sigma = 6 \mp (2)(1.897) \Rightarrow$ the interval is $(2.206, 9.794)$

It means that number of recoveries of 15 patients is $2.206 < \mu < 9.794$ with prob. at least 75%.

<u>Multinomial Distribution</u>: If a given trial can result in the k outcomes $E_1, E_2, \ldots, E_k$, with probabilities $p_1, p_2, p_3, \ldots, p_k$, then the prob. dist. of the random variables $X_1, X_2, \ldots, X_k$, representing the number of occurences of $E_1, E_2, \ldots, E_k$ in n independent trials, is

$$P(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k; n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

where $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} p_i = 1$

<u>ex.</u> If a pair of dice is tossed 6 times, what is the probability of obtaining a total of 7 or 11 twice, a matching pair once, and any other combination 3 times?

<u>sol.</u> Let E_1 be the event of having a total of 7 or 11.

E_2 " " " " " a matching pair.

E_3 " " " " " neither a pair nor a total of 7 or 11.

Now since $p_1 = \frac{2}{9}$, $p_2 = \frac{1}{6}$, $p_3 = \frac{11}{18}$ (why?)

and $x_1 = 2$, $x_2 = 1$, $x_3 = 3$ then:

$$P\left(x_1 = 2, x_2 = 1, x_3 = 3; p_1 = \frac{2}{9}, p_2 = \frac{1}{6}, p_3 = \frac{11}{18}; n = 6\right) =$$

$$\binom{6}{2, 1, 3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3$$

$$= \frac{6!}{2!\,1!\,3!} \frac{2^2}{9^2} \frac{1}{6} \frac{11^3}{18^3} = 0.01127$$

Exercises:

① Find a formula for the dist. of the random variable X which represents the number of a tag drawn from a box containing 12 tags numbered 1 to 12. What is the prob. that the number drawn is less than 4 ?

② Find the uniform dist. for the random samples of committees of size 4 chosen from 6 students.

③ A baseball player's batting avarage is 0.25. What is the prob. that he gets exactly 1 hit in his next 5 times at bat?

④ A multiple-choice quiz has 15 questions, each with 4 possible answers of which one is correct. What is the prob. that a student yields from 5 to 10 correct answers?

⑤ One-fourth of the female entering a College are out-of-city students. If the students are assigned to the dormitories, 3 to a room, what is the prob. that in one room at most 2 of the 3 roommates are out-of-city students?

⑥ If 64 coins are tossed a large number of times, how many heads can we expect on the average per toss? Then using Chebyshev's theorem, between what two values would you expect the number of heads to fall at least $\frac{3}{4}$ of the time?

⑦ According to the theory of genetics a certain cross of guinea pigs will result in red, black and white offspring in the ratio 8:4:4. Find the probability that among 8 such offspring 5 will be red, 2 black and 1 white.

⑧ The probabilities are 0.4, 0.2, 0.3 and 0.1, respectively, that a delegate to a certain convention arrived by air, bus, automobile or train. What is the Prob. that among 9 delegates randomly selected at this convention, 3 arrived by air, 3 by bus, 1 by automobil and 2 by train?

<u>Hypergeometric Dist.</u> If a population of size N contains k items labeled "success" and N − k items labeled "failure", then the prob. dist. of the hypergeometric random variable X, the number of success in a random sample of size n, is

$$P(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad , x = 0, 1, 2, \ldots, n$$

<u>ex①</u> A committee of size 5 is to be selected at random from 3 women and 5 men. Find the probability dist. for the number of women in the committee.

<u>Sol.</u> Let X be a random variable denotes to the no. of women in the committee.

Hence $P(X=x) = \dfrac{\binom{3}{x}\binom{5}{5-x}}{\binom{8}{5}} \quad x = 0, 1, 2, 3$

where $N = 8$, $k = 3$, $n = 5$ and we have

$X=x$	0	1	2	3
$P(X=x)$	$\frac{1}{56}$	$\frac{15}{56}$	$\frac{30}{56}$	$\frac{10}{56}$

<u>ex②</u>: If 5 cards are dealt from a standard deck of 52 playing cards, what is the prob. that 3 will be hearts?

<u>Sol.</u> $N = 52$, $k = 13$, $n = 5$, $x = 0, 1, 2, 3, 4, 5$

$$\therefore P(X=3) = \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} = 0.0815$$

<u>Theorem</u>: The Mean and Variance of the hypergeometric dist. are

$$\mu = \frac{nk}{N} \quad , \quad \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

ex③ Calculate the mean and variance for the prob. dist. of ex.① page 32, then compare with the values of mean and variance of the last theorem.

Sol. $\mu = E(X) = \sum_{\text{all } x} x\, P(x)$

$$= 0\left(\tfrac{1}{56}\right) + 1\left(\tfrac{15}{56}\right) + 2\left(\tfrac{30}{56}\right) + 3\left(\tfrac{10}{56}\right) = \frac{15}{8} = \frac{5 \cdot 3}{8} = \frac{n \cdot k}{N}$$

$$E(X^2) = \frac{225}{56} \Rightarrow \sigma^2 = V(X) = E(X^2) - \left(E(X)\right)^2$$

$$\sigma^2 = \frac{225}{56} - \left(\frac{15}{8}\right)^2 = \frac{225}{448}$$

$$\frac{N-n}{N-1}\, n \frac{k}{N}\left(1 - \frac{k}{N}\right) = \frac{8-5}{8-1} \cdot 5 \cdot \frac{3}{8} \cdot \left(1 - \frac{3}{8}\right) = \frac{3}{7} \cdot 5 \cdot \frac{3}{8} \cdot \frac{5}{8} = \frac{225}{448}$$

ex④ Using Chebyshev's theorem, find and interpret the interve $\mu \mp 2\sigma$ for ex② page 32.

Sol $\mu = \frac{(5)(13)}{52} = \frac{5}{4} = 1.25$

$$\sigma^2 = \frac{52-5}{52-1}\, 5\, \frac{13}{52}\left(1 - \frac{13}{52}\right) = 0.864 \Rightarrow \sigma = \sqrt{0.864} = 0.93$$

$\therefore$ $\mu \mp 2\sigma \Longrightarrow$ an interval $(-0.61, 3.11)$.

Chebyshev's theorem states that the number of hearts obtained when 5 cards are dealt from an ordinary deck of 52 playing cards has a prob. of at least $\frac{3}{4}$ of falling between -0.6 and 3.11; that is at least $\frac{3}{4}$ of the time the 5 cards include less than 4 hearts.

Note. If n is small relative to N, the prob. for each drawing will change only slightly. Hence we essentially have a binomial experiment and can approximate the hypergeometric distribution by using the binomial dist. with $p = \frac{k}{N}$, the mean and variance can also be approximated by the formulas,

$$\mu = np = \frac{nk}{N}, \quad \sigma^2 = npq = n\frac{k}{N}\left(1 - \frac{k}{N}\right)$$

ex. (4) The telephone company reports that among 5000 telephones installed in a new subdivision 4000 have pushbuttons. If 10 people are called at random, what is the prob. that exactly 3 will be talking on dial telephones?

sol: since the population size $N=5000$ is large relative to the sample size $n=10$, we shall approximate the desired prob. by using the binomial dist.

The prob. of calling someone with a dial telephone is (0.2

$$\therefore P(X=3;\ N=5000,\ n=10,\ k=4000) \simeq P(x=3;\ n=10,\ p=0.2)$$
$$= \sum_{x=0}^{3} \binom{3}{x} (0.2)^{x} (0.8)^{3-x} = 0.2013$$

Exercises:

(1) If 7 cards are dealt from ordinary deck, what is the prob. that:
(a) exactly 2 of them will be face cards?
(b) at least 1 " " " " a queen?

(2) A random committee of size 3 is selected from 4 men and 2 women. Write a formula for the prob. dist. of the random variable X representing the number of men on the committee, then find $P(2 \leq X \leq 3)$.

(3) A person is dealt 13 cards from an ordinary deck several times How many hearts per hand can he expect? Between what two ~~intervals~~ values would you expect the number of hearts to fall at least 75% of the time?

(4) A car rental agency at a local airport has available 5 Fords, 7 chevrolets, 4 Dodges, 3 Datsuns and 4 Toyotas. If the agency randomly selected 9 of these cars, find the probability that 2 Fords, 3 chevrolets, 1 Dodges, 1 Datsun and 2 Toyotas are used.

Negative Binomial Dist. If repeated independent trials can result in a success with prob. p and a failure with prob. $q = 1-p$, then the prob. dist. of a random X, the number of trial on which the k-th success occurs is given by:

$$P(X=x) = b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k} \quad , x = k, k+1, k+2, \ldots$$

ex. Find the prob. that a person tossing 3 coins will get either all heads or all tails for the second time on the fifth toss.

Sol. $x=5$, $k=2$, $p = P(3 \text{ heads or } 3 \text{ tails})$

$$= P(3 \text{ heads}) + P(3 \text{ tails}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

$$b^*(5;2,\tfrac{1}{4}) = \binom{5-1}{2-1}\left(\tfrac{1}{4}\right)^2\left(\tfrac{3}{4}\right)^3$$

$$= \frac{4!}{1!\,3!}\,\frac{3^3}{4^3} = \frac{27}{256}$$

Geometric Dist.: If repeated independent trials can result in a success with prob. p and a failure with prob. $q = 1-p$, then the prob. dist. of random variable X, the number of the trial on which the first success occurs, is given by:

$$P(X=x) = g(x;p) = p\,q^{x-1} \quad , x = 1, 2, 3, \ldots$$

ex. Find the prob. that a person flipping a balanced coin requires 4 tosses to get a head.

Sol. $x=4$, $p=\frac{1}{2}$

$$g(4;\tfrac{1}{2}) = \tfrac{1}{2}\left(\tfrac{1}{2}\right)^3 = \frac{1}{16}$$

Poisson Dist.

The Poisson experiment has the following properties:

1) The number of outcomes occuring in one time interval or specified region is independent of the number that occur in any other disjoint time interval or region of space.
2) The prob. 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 occuring outside this time interval or region.
3) The prob. 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 occuring in a Poisson experiment is called a Poisson Random Variable and its prob. dist. is called the Poisson Dist.

Def. The prob. dist. of the Poisson random variable X, representing the number of outcomes occuring in a given time interval or specified region, is

$$P(X=x) = p(x;\mu) = \frac{e^{-\mu}\mu^{x}}{x!}, \quad x=0,1,2,\ldots$$

where μ is the average number of outcomes occuring in the given time interval or specified region.

Ex ① The average number of days school is closed due to snow during the winter in a certain city is 4 days. What is the prob. that the schools in the city will close for 6 days during a winter?

sol. $p(6;4) = \frac{e^{-4}4^6}{6!} = 0.1042$

ex. ② The average number of field mice per acre in a S.ach wheat field is estimated to be 10. Find the prob. that a given acre contains more than 15 mice.

sol. let X be the number of mice per acre. Then using poisson dist. we get:

$$P(X>15) = 1 - P(X \leq 15)$$
$$= 1 - \sum_{x=0}^{15} p(x;10)$$
$$= 1 - \sum_{x=0}^{15} \frac{e^{-10}10^x}{x!}$$
$$= 1 - 0.9513 = 0.0487$$

ex. ③ The variance in Poisson dist. can be shown to be equal to the mean (i.e. $E(X) = V(X) = \mu$ H.W.), Now using Chebyshev's theorem find and interpret the interval $\mu \mp 2\sigma$ for ex. ①

sol $\mu = \sigma^2 = 4 \Rightarrow \sigma = 2 \Rightarrow \mu \mp 2\sigma = 4 \mp (2)(2)$

$\therefore$ The required interval is $(0, 8)$

$\therefore$ We conclude that at least $\frac{3}{4}$ of the time the schools of the given city will be closed anywhere from 0 to 8 days during the winter season.

Note when $n \to \infty$ and $p \to 0$ in a Binomial dist. then it can be approximated to Poisson dist. with $\mu = np$ and $\sigma^2 = npq$.

Ex. ④ Suppose that on the average 1 person in every 1000 is an alcoholic. Find the prob. that a random sample of 8000 people will yield fewer than 7 alcoholics.

Sol $P(X<7) = \sum_{x=0}^{6} b(x; 8000, 0.001)$

$= \sum_{x=0}^{6} p(x; 8)$ since $n = 8000$ & $p = 0.001$

$= 0.3134$ $\mu = np = (8000)(0.001) = 8$

Exercises:

① The prob. that a person living in a certain city joins to a company is (0.3). Find the prob. that the tenth person randomly interviewed in the city is the fifth one to join the company.

② Suppose 0.8 is the prob. that any given person will believe a tale about life after death. What is the prob. that:

ⓐ the sixth person to hear this tale is the fourth to believe it?

ⓑ " third " " " " " " first " " ?

③ Three people toss a coin and the odd man pays for the coffee. If the coins all turn up the same, they are tossed again. Find the prob. that fewer than 4 tosses are needed.

④ On the average a certain intersection results in 3 traffic accidents per month. What is the prob. that in any given month at this intersection.

ⓐ exactly 5 accidents will occur?

ⓑ less than 3 " " " ?

ⓒ at least 2 " " " ?

⑤ Suppose that on the average 1 person in 1000 makes a numerical error in preparing his income tax return. If 10.000 forms are selected at random and examined, find the prob. that 6, 7 or 8 of the forms will be in error.