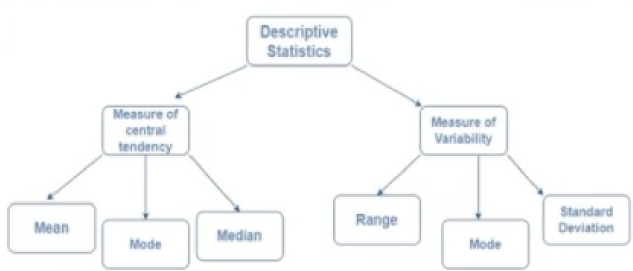
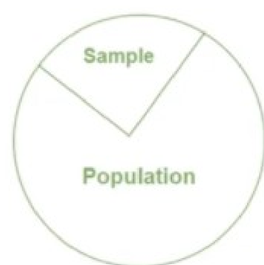
INFERENTIAL STATISTICS

DESCRIPTIVE STATISTICS Vs INFERENTIAL STATISTICS

DESCRIPTIVE	INFERENTIAL
It is the analysis of data that helps to describe, show and summarize data under study	It is the analysis of random sample of data taken from a population to describe and make inference about the population
Organize, analyze and present data in a meaningful way	Compares, test and predicts data
It is used to describe a situation	It is used to explain the chance of occurrence of an event
It explain already known data and limited to a sample or population having small size	It attempts to reach the conclusion about the population
Types: Measure of central tendency & Measure of variability	Types: Estimation of parameters & Testing of hypothesis
Results are shown with help of charts, graphs, tables etc.	Results are shown with help of probability scores





Descriptive Statistics summarize and organize characteristics of a data set. A data set is a collection of observations from a sample or entire population.

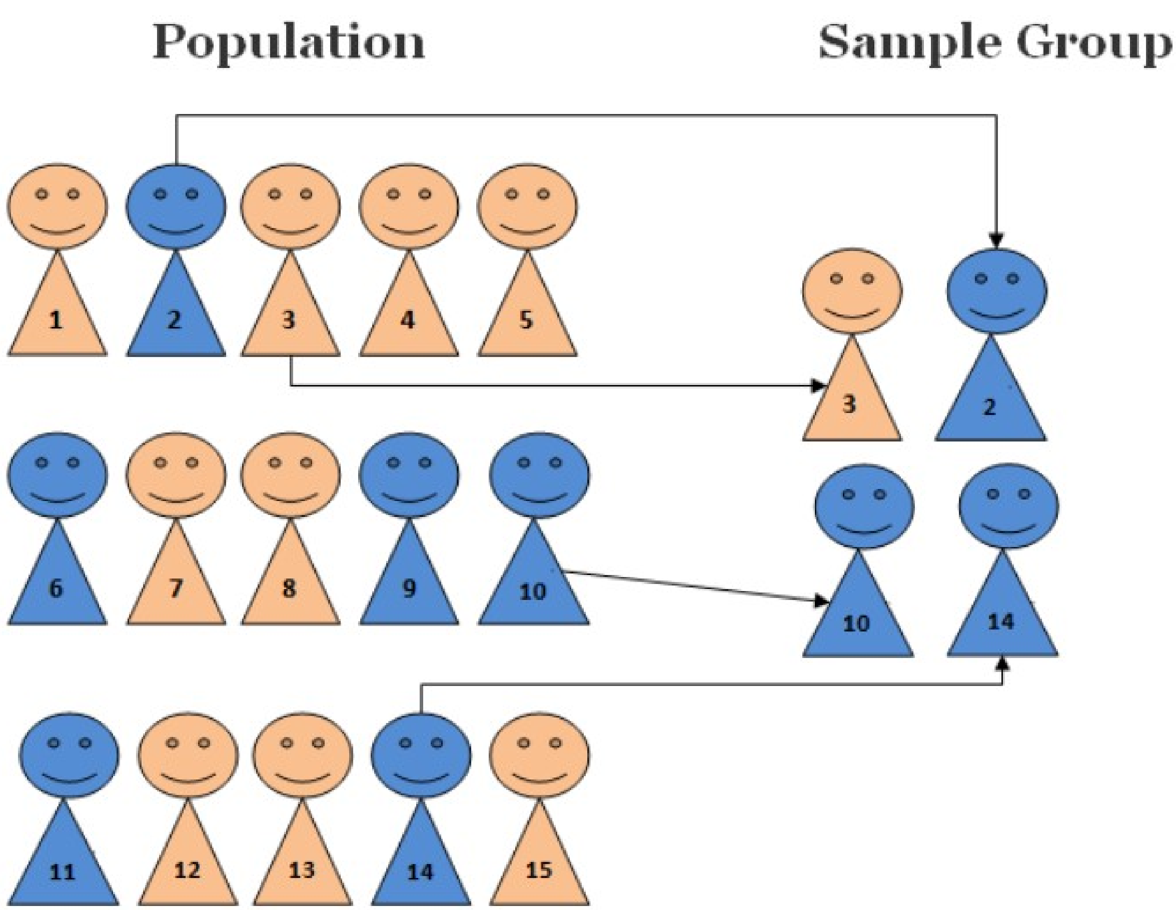
Population, sample and processes

Population: It is the entire data about which you want to draw conclusion. In fact, a large number of well-defined objects or data from which the observation / investigation is to be done is called the population.

Sample: It is the specific group that you will select from a population. The size of the sample size is always less than the total size of population. In other words, sample is a subset of population.

Example-1: In selecting 10 defective screws from a lot of 1000 screws, the population size is 1000 where as the sample size is 10.

Example-2:



Variable: A variable is any characteristic whose value may change from one object to another in the population. For example:
For a company, the variables x, y, z are

x = the brand of the object

y = quality of the object;

z = number of defective objects etc.

Data can be observed into 3 types in the sense of variable such as

A. **Univariate:** When the observations are made on a single variable, the data set is called an univariate data.

For example, the sample of pulse rates (heart beats per minute) for the patients recently admitted to an adult ICU is a numerical univariate data set: 85, 81,80, 102, 115, 95, 90, 105.

B. **Bivariate:** When the observations are made on each of two variables, the data set is called bivariate.
For example, the sample of saturation (the amount of oxygen bound to hemoglobin in the blood measured in percentage of the maximum binding capacity) and pulse rates (heart beats per minute) for the patients recently admitted to an adult ICU is a numerical bivariate data set: (85, 92), (81,87), (80,85), (95,102), (94,115), (95,120), (90.95), (98,102) .

C. **Multivariate:** When the observations are made on more than one variables, the data set is called multivariate. Univariate, bivariate are particular cases of multivariate.
For example, the sample of maximum blood pressure, saturation (the amount of oxygen bound to hemoglobin in the blood measured in percentage of the maximum binding capacity) and pulse rates (heart beats per minute) for the patients recently admitted to an adult ICU is a numerical trivariate data set: (130, 85, 92), (115, 81,87), (140, 80,85), (102, 95,102), (148, 94,115), (205, 95,120), (145, 90.95), (125, 98,102) .
In general, if $\mathrm{X}_1, \mathrm{X}_2, \mathrm{X}_3, \cdots$ are different type of categories of the object, then multivariate data set is of form

$$X = X_1 \times X_2 \times X_3 \times \cdots = \{(x_1, x_2, \cdots : x_i \in X_i, i = 1,2, \cdots)\}.$$

Pictorial and tabular methods in Descriptive statistics

Given a data set consisting of n observations on some variable x, the individual observations will be denoted by $x_1, x_2, \cdots, x_n$. The subscript bears no relation to the magnitude of a particular observation. Thus x_1 will not in general be the smallest observation in the set, not will x_n typically be the largest. In many applications, x_1 will be the first observation gathered by the experimenter, x_2 the second, and so on. The i^{th} observation in the data set will be denoted by x_i.

I. Stem-and-leaf display

Consider a numerical data set $x_1, x_2, \cdots, x_n$ for which each x_i consists at least two digits. A quick way to obtain the informative visual representation of the data set is to construct stem-and-leaf display as follows.

i. Select one or more leading digits for the stem values. The trailing digits becomes the leaves.
ii. List possible stem values in a vertical column.
iii. Record the leaf for each observation beside the corresponding stem value.
iv. Indicate the units for the stem and leaves someplace in the display.

Example: For a random sample of lengths of golf courses (yard) that have been designated by Golf Megazine as among the most challenging in the United States. Express the sample of 40 courses

6435, 6526, 7131, 6790, 6464, 6527, 6605, 6433, 6694, 6506, 6890. 6583, 6870, 6770, 6614, 6873, 6900, 6700, 6927, 7051, 6850, 7011, 6798, 7040, 7050, 7022, 6770, 6936, 6904, 7169, 7168, 6745, 7105, 7005, 7280, 6470, 6713, 7209, 7113, 7165
by stem-and-leaf display.

Solution: Among the sample of 40 courses, the shortest is 6433 yards long, and the longest is 7280. Taking Thousand-Hundred digits for stem, and Tens-Ones digits for leaf, we obtain the stem-and-leaf display with two-digit leaves shown in Figure-(a).

64	35	64	33	70				Stem: Thousands and hundreds digits
65	26	27	06	83				Leaf: Tens and ones digits
66	05	94	14					
67	90	70	00	98	70	45	13	
68	90	70	73	50				
69	00	27	36	04				
70	51	05	11	40	50	22		
71	31	69	68	05	13	65		
72	80	09						

(a)

Statistical software packages do not generally produce displays with multiple digit stems. The Minitab in Figure-(b) results from truncating each observation by deleting the one digit which is stem-and-leaf display with truncated one-digit leaves.

```
Stem-and-leaf of yardage  N = 40
Leaf Unit = 10
            4          64   3367
            8          65   0228
           11          66   019
           18          67   0147799
          (4)          68   5779
           18          69   0023
           14          70   012455
            8          71   013666
            2          72   08
```

(b)

Q. Every score in the following batch of exam scores is in the 60s, 70s, 80s, or 90s. A stem-and leaf display with only the four stems 6, 7, 8 and 9 would not give a very detailed description of the distribution of scores. Here we could repeat the stem 6 twice, using 6L for scores in the low 60s (leaves 0, 1,2,3, and 4) and 6H for scores in the high 60s (leaves 5, 6,7,8, and 9). Similarly other stems can be repeated twice to obtain a display consisting of 8 rows. Construct such a display for the given scores. What feature of the data is highlighted by this display

74	89	80	93	64	67	72	70	66	85	89	81	81
71	74	82	85	63	72	81	81	95	84	81	80	70
69	66	60	83	85	98	84	68	90	82	69	72	87
88												

Solution: For stem 6L, the leaves are 4 (in first row), 3 (in second row) and 0 (in third row) because leave L caries the values from 0 to 4. For stem 6H, the leaves are 7, 6 (in first row), and 9, 6, 8, 9 (in third row) because leave H caries the values from 5 to 9. Similarly we have following stem-and-leaf display with one-digit leaves.

6L	430
6H	769689
7L	42014202
7H	
8L	011211410342
8H	9595578
9L	30
9H	58

The gap in the data—no scores in the high 70s.

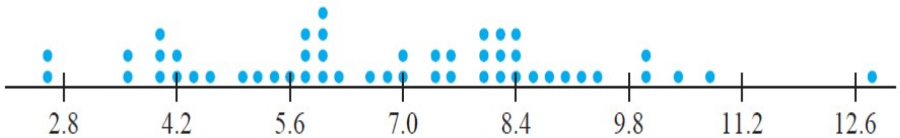
II. Dot plots

A dot plot is an attractive summary of numerical data when the data set is reasonably small or there are relatively few distinct data values. Each observation is represented by a dot above the corresponding location on a horizontal measurement scale. When a value occurs more than once, there is a dot of each occurrences, and these dots are stacked vertically. As with a stem-and-leaf display, a dot plot gives information about location, spread, extremes, and gaps.

Example: Sketch the dot plots of the data

10.8	6.9	8.0	8.8	7.3	3.6	4.1	6.0	4.4	8.3
8.1	8.0	5.9	5.9	7.6	8.9	8.5	8.1	4.2	5.7
4.0	6.7	5.8	9.9	5.6	5.8	9.3	6.2	2.5	4.5
12.8	3.5	10.0	9.1	5.0	8.1	5.3	3.9	4.0	8.0
7.4	7.5	8.4	8.3	2.6	5.1	6.0	7.0	6.5	10.3

Solution: In the data, number of numbers is $n = 50$. Smallest value is 2.5 and largest value is 12.8. Taking first value as 2.8, last value as 12.6 with space length 1.4, we have the dot plot as follows:



III. Histograms

Some numerical data is obtained by counting to determine the value of a variable (the number of traffic citations a person received during the last year, the number of customers arriving for service during a particular period), whereas other data is obtained by taking measurements (weight of an individual, reaction time to a particular stimulus). The prescription for drawing a histogram is generally different for these two cases: discrete or continuous.

Discrete and Continuous variables: A numerical variable is **discrete** if its set of possible values are either finite or else can be listed in an infinite sequence (one in which there is a first number, a second number, and so on). A numerical value is **continuous** if its possible values consist of an entire interval on the number line.

Frequency: In a discrete data set, the frequency of a variable x is the number of times that value exists.

Relative Frequency: A discrete data set (A) is obtained by some observations. The relative frequency of a value $x \in A$ is the fraction/proportion of times of the value occurs. Let number of times the value x occurs is m and number of observations in the data set is n, then

$$\text{Relative frequency of a value x} = \frac{\text{number of times the value occurs}}{\text{number of observations in the data set}} = \frac{m}{n} = f_x.$$

Theoretically sum of relative frequencies of the values is 1.

Example: A data set consists of 200 observations on $x =$ the number of courses a college student is taking this term. If 70 of these x values are 3, then frequency of the x value 3 is 70 and relative frequency of the x value 3 is 70/200=0.35, i.e.. 35% of the students for the sample are taking the three courses.

Note: In general sum of the relative frequencies differs slightly from 1 because of rounding.

Frequency Distribution: A frequency distribution is a tabulation of the frequencies and/or relative frequencies.

1.

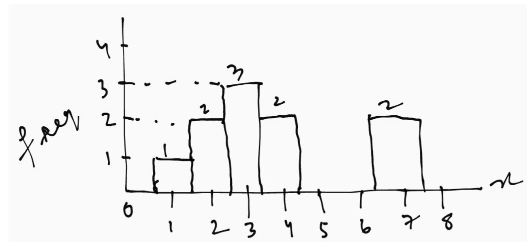
A rectangle can be plot at the point (x, f_x) with height f_x at value x such that the area of the rectangle is proportional to f_x of the value x. A **histogram** is an approximate representation of the distribution of numerical data which was introduced by Karl Pearson.

Q. Find the relative frequency of each value of the data 1, 3, 4, 8, 3, 7, 2, 7, 4, 2, 3 and draw the histogram,

Solution:

x	frequency	Relative frequency
1	1	0.1
2	2	0.2
3	3	0.3
4	2	0.2
7	2	0.2

Total frequency=10



Histogram

IV. Pareto Diagrams / Charts

A Pareto diagram is a graph that indicates the frequency of defects, as well as their cumulative impact. Pareto diagrams are useful to find the defects to prioritize in order to observe the greatest overall improvement.

A Pareto diagram is a combination of a bar graph and a line graph. Notice the presence of both bars and a line on the Pareto diagram below.

Each bar usually represents a type of defect or problem. The height of the bar represents any important unit of measure — often the frequency of occurrence or cost.

The bars are presented in descending order (from tallest to shortest). Therefore, you can see which defects are more frequent at a glance.

The line represents the cumulative percentage of defects.

Let's look at the table of data for the Pareto Chart above to understand what cumulative percentage is.

TYPE OF DEFECT	FREQUENCY OF DEFECT	% OF TOTAL	CUMULATIVE %
Button Defect	23	39.0	39.0
Pocket Defect	16	27.1	66.1
Collar Defect	10	16.9	83.0
Cuff Defect	7	11.9	94.9
Sleeve Defect	3	5.1	100
Total	59	-	-

For % of Defects, we use the formula (Frequency of Defect/Total no. of defects)*100.

For example, the cumulative % corresponding to Collar defect is the sum of all percentages previous to and including Collar Defects. In this case, this would be the sum of the percentages of Button Defects, Pocket Defects, and Collar Defects (39% + 27.1% + 16.9%).

The last cumulative percentage will always be 100%.

Cumulative percentages indicate what percentage of all defects can be removed if the most important types of defects are solved.

In the example above, solving just the two most important types of defects — Button Defects and Pocket Defects – will remove 66% of all defects.

Thus, Pareto diagram is a quality tool in analyzing & prioritizing the issues.

Probability

Trial: A trial is nothing but the statististical experminent whose outcomes are uncertain.

Statistical experiment: Describing an act for several times under given conditions is called statistical experiment.

eg: tossing of a coin once or several times

Picking a card from a deck

Random experiment: It is an experiment whose outcomes are by chance.

sample space: The set of all possible outcomes is called the sample space.

In tossing two fair coins, $S = \{HH, HT, TH, TT\}$

$|S| = 4 = 2^2$

In rolling a fair die once, $S = \{1, 2, 3, 4, 5, 6\}$, $|S| = 6$

event: All possible outcomes of a random experiment are called events.

$A = \{HH, HT, TH\}$ be the event of getting atleast one head. $A \subset S$

simple event: $\{HH\}, \{HT\}, \{TH\}, \{TT\}$ are simple events

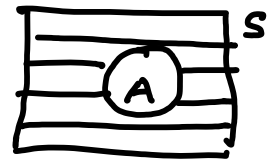
compound event: $A = \{HH, HT, TH\}$ is a compound event.

some relations from set theory

Any event is just a set.

If A is an event then the complement of A is written as

A' or A^c and is defined as the set of all possible outcomes in S which are not in A.



$S = \{1,2,3,4,5\}$, $A = \{1,2,3\}$

A' or $A^c = \{4,5\}$

S' or $S^c = \phi$. $\phi^c = S$, $(A^c)^c$ or $(A')' = A$

$A \cup B$ = the set of all outcomes that are either in A or in B or in both the events.

$B = \{4,5\}$ $\quad A \cup B = \{1,2,3\} \cup \{4,5\} = \{1,2,3,4,5\}$

A or B

$A \cap B$ = the set of all outcomes that belong to both A & B.

A and B $\quad A \cap B = \{1,2,3\} \cap \{4,5\} = \phi$

<u>Disjoint events or Mutually exclusive events</u>

Let A and B be two events. Then A and B are said to be mutually exclusive if $A \cap B = \phi$

$A = \{2,4,6\}$, $B = \{1,3,5\}$ $\quad A \cap B = \phi$

<u>Equally likely events</u>

Let A and B be the two events in a sample space 'S'. Then A and B are said to be equally likely if there are equal chance of occurance of both the events.

<u>Probability</u>: The probability of an event 'A' is denoted as $p(A)$ and is defined as

$$p(A) = \frac{|A|}{|S|} = \frac{\text{No. of favourable cases}}{\text{Total no. of possible outcomes}}$$

Axioms of Probability

i) For any event A, $0 \leq p(A) \leq 1$

ii) If S is the sample space, $p(S) = 1$

iii) For a null/impossible event $p(\phi) = 0$

iv) If $A_1, A_2, A_3, \cdots$ is an infinite collection of mutually exclusive events then

$$p(A_1 \cup A_2 \cup \cdots) = p(A_1) + p(A_2) + \cdots$$

In particular, if A and B are mutually exclusive then

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

Q. If the probability that on any work day a garage will get 10–20, 21–30, 31–40, over 40 cars to service is 0.20, 0.35, 0.25, 0.12 respectively then what is the probability that on a given work day the garage will get at least 21 cars to service.

Ans: p(getting at least 21 cars to service)

$$= 0.35 + 0.25 + 0.12$$

$$= 0.72$$

Complementation rule

If 'A' is an event in a sample space 'S' and A^c is the complement of A then

$$p(A^c) = 1 - p(A)$$

$p(S) = 1$

$P(A \cup A^c) = 1$

$\Rightarrow p(A) + p(A^c) = 1$

eg: 4 fair coins are tossed simultaneously, find the probability of an event A (at least 1 Head turns up)?

Ans: Given A: be the event of getting at least 1 head turns up.

$|S| = 2^4 = 16$

A^c: the event of getting no head

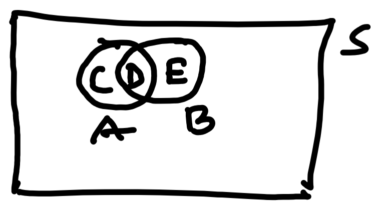
$$p(A^c) = \frac{|A^c|}{|S|} = \frac{1}{16}$$

$$p(A) = 1 - p(A^c) = \frac{15}{16}$$

Addition rule for arbitrary events

For any 2 events A and B in the sample space 'S',

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$



Pf: Let C, D & E be the mutually exclusive events such that $A = C \cup D$, $B = D \cup E$, $C \cap D = \phi$, $D \cap E = \phi$, $C \cap E = \phi$

$$p(A) = p(C \cup D) = p(C) + p(D) \quad \text{--- } ①$$

$$p(B) = p(D \cup E) = p(D) + p(E) \quad \text{--- } ②$$

Now, $p(A \cup B) = p(C \cup D \cup E) = p(C) + p(D) + p(E)$

$= p(A) + p(E)$ (By ①)

$= p(A) + p(B) - p(D)$ (By ②)

$= p(A) + p(B) - p(A \cap B)$

(As $A \cap B = D$)

Similarly, if A, B and C are any three arbitrary events in a sample space 'S' then

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

eg: In throwing a fair die, what is the probability of getting an odd no or a no. less than 4.

<u>Ans</u>: $|S| = 6$

Let A: getting an odd no., $A = \{1, 3, 5\}$, $|A| = 3$

B: getting a no. less than 4, $B = \{1, 2, 3\}$, $|B| = 3$

$A \cap B = \{1, 3\}$, $|A \cap B| = 2$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$= \frac{|A| + |B| - |A \cap B|}{|S|} = \frac{3 + 3 - 2}{6} = \frac{2}{3}.$$

<u>Sampling</u> The process of drawing one object at a time from a given set of objects is called sampling.

There are 2 types of sampling

i) sampling with replacement

ii) sampling without replacement.

eg! A box contains 10 screws out of which 3 are defective. 2 screws are drawn at random. Find the probability that none of the two screws is defective.

<u>Ans</u>:

10 screws

7ND 3D

<u>sampling with replacement</u>

$P(\text{getting a non-defective screw}) = \frac{{}^7C_1}{{}^{10}C_1} = \frac{7}{10}$

$P(\text{getting both screws non defective in 2 draws}) = \frac{{}^7C_1}{{}^{10}C_1} \cdot \frac{{}^7C_1}{{}^{10}C_1} = \frac{49}{100}$

$= 0.49$

<u>sampling without replacement</u>

$P(\text{getting both screws non-defective in 2 draws}) = \frac{{}^7C_1}{{}^{10}C_1} \cdot \frac{{}^6C_1}{{}^9C_1}$

$= \frac{7}{10} \cdot \frac{6}{9} = \frac{42}{90}$

$\simeq 0.47$

11. A mutual fund company offers its customers a variety of funds: a money-market fund, three different bond funds (short, intermediate, and long-term), two stock funds (moderate and high-risk), and a balanced fund. Among customers who own shares in just one fund, the percentages of customers in the different funds are as follows:

Money-market	20%	High-risk stock	18%
Short bond	15%	Moderate-risk stock	25%
Intermediate bond	10%	Balanced	7%
Long bond	5%		

A customer who owns shares in just one fund is randomly selected.

a. What is the probability that the selected individual owns shares in the balanced fund?

b. What is the probability that the individual owns shares in a bond fund?

c. What is the probability that the selected individual does not own shares in a stock fund?

<u>Ans:</u> $P(MM) = 0.2$, $P(SB) = 0.15$, $P(IB) = 0.1$, $P(LB) = 0.05$
$P(HRS) = 0.18$, $P(MRS) = 0.25$, $P(B) = 0.07$

a. p (the selected individual owns shares in the balanced fund)

$= 0.07$

b. p (the individual owns shares in a bond fund)

$= p(SB) + p(IB) + p(LB) = 0.15 + 0.1 + 0.05 = 0.3$

c. p (the selected individual does not own shares in a stock fund)

$= 1 - p$ (an individual owns shares in a stock fund)

$= 1 - \{p(HRS) + p(MRS)\}$

$= 1 - \{0.18 + 0.25\}$

$= 1 - 0.43 = 0.57.$

12. Consider randomly selecting a student at a certain university, and let A denote the event that the selected individual has a Visa credit card and B be the analogous event for a MasterCard. Suppose that $P(A) = .5$, $P(B) = .4$, and $P(A \cap B) = .25$.

a. Compute the probability that the selected individual has at least one of the two types of cards (i.e., the probability of the event $A \cup B$).

b. What is the probability that the selected individual has neither type of card?

c. Describe, in terms of A and B, the event that the selected student has a Visa card but not a MasterCard, and then calculate the probability of this event.

Ans: Given, A denote the event that the selected individual has a visa credit card and

B denote the event that the selected individual has a Master card

$p(A) = 0.5$, $p(B) = 0.4$ & $p(A \cap B) = 0.25$

a. $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$= 0.5 + 0.4 - 0.25$

$= 0.65$

b. p(the selected individual has neither type of card)

$= 1 - P(A \cup B)$

$= 1 - 0.65 = 0.35$

c. p(the selected student has a visa credit card but not a master card) $A \cap B' = A - B$

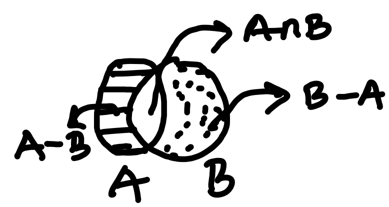
$= P(A \cap B')$

$= P(A - B)$

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

$= 0.5 - 0.25$

$= 0.25$



$A = (A - B) \cup (A \cap B)$

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

$\Rightarrow P(A-B) = P(A) - P(A \cap B)$

13. A computer consulting firm presently has bids out on three projects. Let $A_i = \{\text{awarded project } i\}$, for $i = 1, 2, 3$, and suppose that $P(A_1) = .22$, $P(A_2) = .25$, $P(A_3) = .28$, $P(A_1 \cap A_2) = .11$, $P(A_1 \cap A_3) = .05$, $P(A_2 \cap A_3) = .07$, $P(A_1 \cap A_2 \cap A_3) = .01$. Express in words each of the following events, and compute the probability of each event:

a. $A_1 \cup A_2$

b. $A_1' \cap A_2'$ [*Hint:* $(A_1 \cup A_2)' = A_1' \cap A_2'$] (De-Morgan's law)

c. $A_1 \cup A_2 \cup A_3$ **d.** $A_1' \cap A_2' \cap A_3'$

e. $A_1' \cap A_2' \cap A_3$ **f.** $(A_1' \cap A_2') \cup A_3$

Ans: a. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

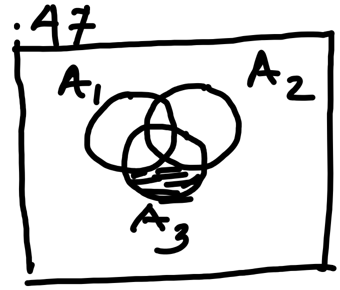
$= .22 + .25 - .11 = 0.36$

b. $P(A_1' \cap A_2') = P((A_1 \cup A_2)')$

$= 1 - P(A_1 \cup A_2) = 1 - 0.36 = 0.64$

c. $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$
$- P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_3 \cap A_1)$
$+ P(A_1 \cap A_2 \cap A_3)$
$= (.22 + .25 + .28) - (.11 + .07 + .05) + .01$
$= 0.53$

d. $P(A_1' \cap A_2' \cap A_3') = 1 - P(A_1 \cup A_2 \cup A_3)$
$= 1 - 0.53 = 0.47$



e. $P(A_1' \cap A_2' \cap A_3)$
$= P(A_3) - P(A_1 \cap A_3) - P(A_2 \cap A_3)$
$+ P(A_1 \cap A_2 \cap A_3)$
$= .28 - .05 - .07 + .01$
$= .17$

f. $P((A_1' \cap A_2') \cup A_3) = P(A_1' \cap A_2') + P(A_3)$
$- P(A_1' \cap A_2' \cap A_3)$
$= 0.64 + 0.28 - 0.17$
$= 0.75$

14. Suppose that 55% of all adults regularly consume coffee, 45% regularly consume carbonated soda, and 70% regularly consume at least one of these two products.

a. What is the probability that a randomly selected adult regularly consumes both coffee and soda?

b. What is the probability that a randomly selected adult doesn't regularly consume at least one of these two products?

Ans: Let A be the event representing the adults regularly consume coffee

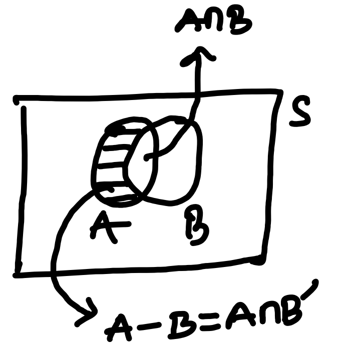
B be the event representing the adults regularly consume carbonated soda.

$(A \cup B)$ is the event representing the adults who consumes either cofee or carbonated soda.

Given $p(A) = 0.55$, $p(B) = 0.45$, $p(A \cup B) = 0.7$

a. $p(A \cap B) = p(A) + p(B) - p(A \cup B)$

$= 0.55 + 0.45 - 0.7 = 0.3$



b. $p\left((A \cap B') \cup (A' \cap B) \cup (A' \cap B')\right)$

$= p\left((A \cap B') \cup (A' \cap B) \cup (A \cup B)'\right)$

$= p(A \cap B') + p(A' \cap B) + p((A \cup B)')$

$= p(A) - p(A \cap B) + p(B) - p(A \cap B)$
$+ 1 - p(A \cup B)$

$A = (A \cap B) \cup (A \cap B')$

$\Rightarrow p(A) = p(A \cap B) + p(A \cap B')$

$= 0.55 - 0.3 + 0.45 - 0.3 + 1 - 0.7$

$= 0.7$

Conditional Probability

Let A and B be two events of a sample space 'S'. Then the probability of A under the condition that the event 'B' has already occured is denoted as $P(A|B)$ and is given by

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

similarly, $$P(B|A) = \frac{P(A\cap B)}{P(A)}, \quad P(A) > 0$$

Note: $P(A|B) \neq P(A)$, $P(B|A) \neq P(B)$

$P(A|B) = P(B|A)$ when A & B are mutually exclusive i.e; $A\cap B = \phi$.

eg: If $S = \{1,2,3,4,5,6,7,8,9\}$, $A = \{1,2,3,4,5\}$

$B = \{4,5,6,7,8,9\}$ then $A\cap B = \{4,5\}$

and $$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{2/9}{6/9} = \frac{2}{6} = \frac{1}{3}$$

$$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{2/9}{5/9} = \frac{2}{5}$$

Multiplication Rule

If A and B are any two arbitrary events in a sample space with $P(A) \neq 0$ and $P(B) \neq 0$ then

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

Note: If A & B are two independent events

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

As $P(A|B) = P(A)$ & $P(B|A) = P(B)$

45. The population of a particular country consists of three ethnic groups. Each individual belongs to one of the four major blood groups. The accompanying *joint probability table* gives the proportions of individuals in the various ethnic group–blood group combinations.

		Blood Group			
		O	**A**	**B**	**AB**
Ethnic Group	**1**	.082	.106	.008	.004
	2	.135	.141	.018	.006
	3	.215	.200	.065	.020

Suppose that an individual is randomly selected from the population, and define events by A = {type A selected}, B = {type B selected}, and C = {ethnic group 3 selected}.

a. Calculate $P(A)$, $P(C)$, and $P(A \cap C)$.

b. Calculate both $P(A|C)$ and $P(C|A)$, and explain in context what each of these probabilities represents.

c. If the selected individual does not have type B blood, what is the probability that he or she is from ethnic group 1?

Ans: Given, $A = \{\text{type A selected}\}$

$B = \{\text{type B selected}\}$

$C = \{\text{ethnic group 3 selected}\}$

a. $P(A) = .106 + .141 + .200 = .447$

$P(C) = .215 + .200 + .065 + .020 = 0.5$

$P(A \cap C) = .2$

b. $P(A|C) = \frac{P(A\cap C)}{P(C)} = \frac{.2}{.5} = \frac{2}{5} = .4$

$P(C|A) = \frac{P(A\cap C)}{P(A)} = \frac{.2}{.447} \simeq 0.447$

c. $B' = \{$selected individual does not have type B blood$\}$

$E = \{$Ethnic group 1 selected$\}$

$P(E|B') = \frac{P(E\cap B')}{P(B')} = \frac{0.192}{0.909} = 0.211$

$P(B') = 1 - P(B) = 1 - 0.091 = 0.909$

$P(E\cap B') = 0.192$

46. Suppose an individual is randomly selected from the population of all adult males living in the United States. Let A be the event that the selected individual is over 6 ft in height, and let B be the event that the selected individual is a professional basketball player. Which do you think is larger, $P(A|B)$ or $P(B|A)$? Why?

Ans: A be the event that the selected individual is over 6ft in height

B be the event that the selected individual is a professional basketball player

S is the population of all adult males in US.

clearly, $B \subseteq A$

so, $P(B) < P(A)$

$\Rightarrow \frac{1}{P(A)} < \frac{1}{P(B)}$

$\Rightarrow \frac{P(A\cap B)}{P(A)} < \frac{P(A\cap B)}{P(B)}$

$\Rightarrow \; p(B|A) < p(A|B)$

so, $p(A|B)$ is larger.

Partition of sample space

Let $A_1, A_2, \ldots, A_K$ be the events of a sample space 'S'. Then it is called the partition of the sample space 'S' under the following conditions

i) $A_i \cap A_j = \phi$ for $i, j = 1, 2, \ldots, K$ & $i \neq j$ (Mutually exclusive)

ii) $A_1 \cup A_2 \cup \cdots \cup A_K = S$ (Mutually exhaustive)

iii) $p(A_i) > 0$ for $i = 1, 2, \ldots, K$

Total Probability

Let $A_1, A_2, \ldots, A_K$ be the partitions of a sample space 'S'. For any event 'B', the probability of B is given by

$$p(B) = p(B|A_1)\cdot p(A_1) + p(B|A_2)\cdot p(A_2) + \cdots + p(B|A_K)\cdot p(A_K)$$

$$= \sum_{i=1}^{K} p(B|A_i)\cdot p(A_i)$$

$p(B|A_i)$ is called the prior probability of A_i

$p(A_i|B)$ " " " posterior " " ".

eg: A person has undertaken a construction work. Probability that there will be strike is 0.65. The Probability that construction job will be completed on time is 0.32 if there is strike and 0.8 if there is no strike. what is the probability that the construction job will completed on time.

Ans: Let B : be the event that the construction job is completed

A : be the event that there will be strike

A^c : " " " " " " " no strike.

$$P(A) = 0.65, \quad P(A^c) = 1 - P(A) = 1 - 0.65 = 0.35$$

$$P(B|A) = 0.32, \quad P(B|A^c) = 0.8$$

$$P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)$$

$$= (0.32)(0.65) + (0.8)(0.35)$$

$$= 0.488$$

Baye's Theorem:

Let $A_1, A_2, \ldots, A_K$ be both mutually exclussive and exhaustive events with prior probabilities $P(A_i)$ for $i = 1, 2, \ldots, K$. Then for any event 'B' for which $P(B) > 0$, the posterior probability of A_j given that 'B' has already occured is

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j) \cdot P(A_j)}{P(B)}$$

$$\Rightarrow P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^{K} P(B|A_i) \cdot P(A_i)} \; ; \; j = 1, 2, \ldots, K$$

59. At a certain gas station, 40% of the customers use regular gas (A_1), 35% use plus gas (A_2), and 25% use premium (A_3). Of those customers using regular gas, only 30% fill their tanks (event B). Of those customers using plus, 60% fill their tanks, whereas of those using premium, 50% fill their tanks.

a. What is the probability that the next customer will request plus gas and fill the tank ($A_2 \cap B$)?

b. What is the probability that the next customer fills the tank?

c. If the next customer fills the tank, what is the probability that regular gas is requested? Plus? Premium?

Ans: Let A_1: the customers use regular gas

A_2: the customers use plus gas

A_3: " " " premium.

B: " " who fill their tanks

$P(A_1) = 0.4$, $P(A_2) = 0.35$, $P(A_3) = 0.25$

$P(B|A_2) = 0.6$, $P(B|A_1) = 0.3$, $P(B|A_3) = 0.5$

a. $P(A_2 \cap B) = P(B|A_2) \cdot P(A_2) = (0.6)(0.35) = 0.21$

b. $P(B) = P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3)$

$= (0.3)(0.4) + (0.6)(0.35) + (0.5)(0.25)$

$= 0.455$

c. $P(A_1|B) = \dfrac{P(A_1 \cap B)}{P(B)} = \dfrac{P(B|A_1) \cdot P(A_1)}{P(B)}$

$= \dfrac{(0.3)(0.4)}{0.455} = 0.263$

$P(A_2|B) = \dfrac{P(A_2 \cap B)}{P(B)} = \dfrac{P(B|A_2) \cdot P(A_2)}{P(B)}$

$$= \frac{(0.6)(0.35)}{0.455} = 0.461$$

$$p(A_3|B) = \frac{P(A_3 \cap B)}{P(B)} = \frac{P(B|A_3) \cdot P(A_3)}{P(B)}$$

$$= \frac{(0.5)(0.25)}{0.455} = 0.275$$

71. An oil exploration company currently has two active projects, one in Asia and the other in Europe. Let A be the event that the Asian project is successful and B be the event that the European project is successful. Suppose that A and B are independent events with $P(A) = .4$ and $P(B) = .7$.

a. If the Asian project is not successful, what is the probability that the European project is also not successful? Explain your reasoning.

b. What is the probability that at least one of the two projects will be successful?

c. Given that at least one of the two projects is successful, what is the probability that only the Asian project is successful?

Ans: Given, A: be the event that the Asian project is successful.

B: be the event that the Europian project is successful

Events A and B are independent with

$P(A) = 0.4$, $P(B) = 0.7$

a. $p(B^c|A^c) = p(B^c) = 1 - p(B) = 0.3$

b. $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$= p(A) + p(B) - p(A) \cdot p(B)$

$= 0.4 + 0.7 - (0.4)(0.7)$

$= 1.1 - 0.28 = 0.82$

$P(C|D) = \frac{P(C \cap D)}{P(D)}$

c. $P\{(A \cap B') | (A \cup B)\} = \frac{P\{(A \cap B') \cap (A \cup B)\}}{P(A \cup B)}$

$= \frac{P\{(A \cap B' \cap A) \cup (A \cap B' \cap B)\}}{P(A \cup B)}$

$= \frac{P\{(A \cap B' \cap A) \cup \phi\}}{P(A \cup B)}$

$= \frac{P\{(A \cap B' \cap A)\}}{P(A \cup B)} = \frac{P(A \cap B')}{P(A \cup B)}$

$= \frac{P(A) \cdot P(B')}{P(A \cup B)} = \frac{(0.4)(0.3)}{0.82}$

$= 0.146$

<u>Random</u> <u>variable</u>

A random variable is a fn. $X : S \rightarrow \mathbb{R}$ where domain is the sample space and the range is the set of real nos.

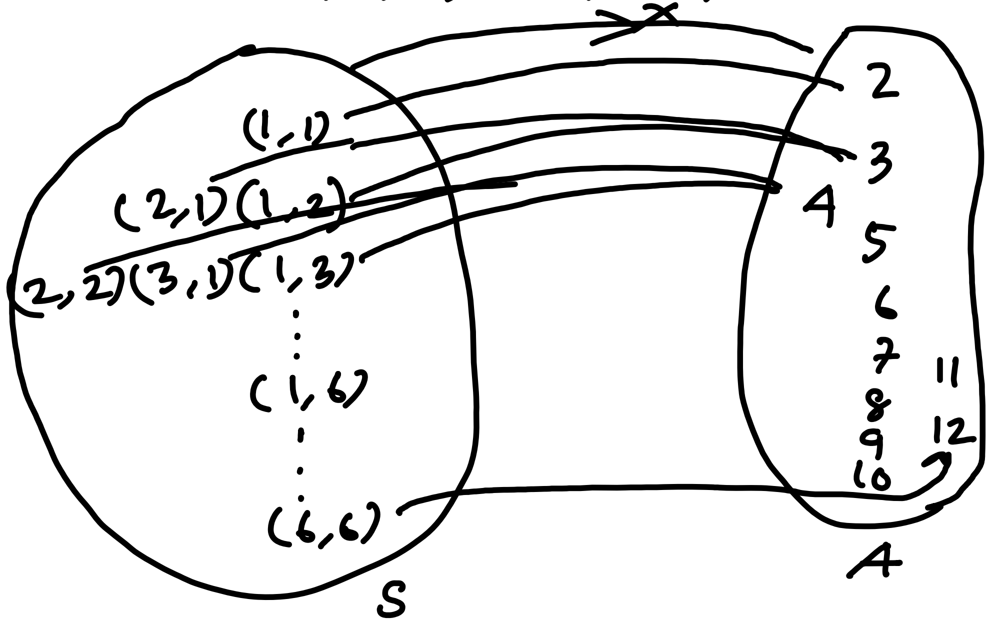
eg: In rolling 2 fair dice, the sample space is

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

$|S| = 6^2 = 36$

Let x be the sum of 2 faces turned up

$x = 2, 3, 5, 6, 7, 8, 9, 10, 11, 12$



<u>Note:</u> Any random variable which takes the values 0 and 1 only is called Bernoulli random variable.

<u>Discrete random variable (DRV)</u>

A random variable 'X' is said to be discrete if it assumes the values $x_1, x_2, \dots, x_n$ with probabilities $p_1 \{= P(X=x_1)\}$, $p_2 \{= P(X = x_2)\}, \dots;$ $p_n \{= P(X=x_n)\}$.

The probability mass function (pmf) of the discrete random variable is denoted as $f(x_j)$ or $p(x_j)$ and is defined as

$$p(x_j) = \begin{cases} p_j & \text{for } x = x_j \\ 0 & \text{otherwise} \end{cases}$$

It may noted that $\sum_{j=1}^{n} p_j = 1$

i.e; $\sum_{j=1}^{n} p(x_j) = 1$

Here $x = x_1, x_2, \ldots, x_n$

The probability distribution fn. (pdf) is denoted as $F(x)$ and is defined as

$$F(x) = p(X \leq x)$$

where $p(X \leq x_j) = p(x_1) + p(x_2) + \ldots + p(x_j)$

eg: If two coins are tossed and X is equal to the no of heads turned up then find $p(0)$, $p(1)$, $p(2)$ (probability mass fns.) and $p(X \leq 0)$, $p(X \leq 1)$, $p(X \leq 2)$ (probability distribution fns.)

Ans: $S = \{HH, HT, TH, TT\}$, $|S| = 2^2 = 4$

$X = 0, 1, 2$

$p(0) = p(X = 0) = \frac{1}{4}$

$p(1) = p(X = 1) = \frac{2}{4} = \frac{1}{2}$

$p(2) = p(X = 2) = \frac{1}{4}$

$p(X \leq 0) = p(0) = \frac{1}{4}$

$p(X \leq 1) = p(X = 0) + p(X = 1)$
$= p(0) + p(1)$
$= \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$

$p(X \leq 2) = p(0) + p(1) + p(2)$
$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$

The cummulative distribution fn. can be given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \le x < 1 \\ 3/4 & \text{if } 1 \le x < 2 \\ 1 & \text{if } 2 \le x \end{cases}$$

<u>properties of pdf</u>

1. $P(X \le x_j) = p(x_1) + p(x_2) + \cdots + p(x_j)$

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

3. $P(X \ge x) = 1 - P(X < x)$

4. $P(a < X \le b) = F(b) - F(a)$

5. $P(a \le X \le b) = P(a-1 < X \le b)$
$= F(b) - F(a-1)$

eg: For $p(x) = cx^2$ for $x = 1, 2, \ldots, 5$; find the probability mass fns. and the probability distribution fns.

<u>Ans:</u> $p(1) = c$, $p(2) = 4c$, $p(3) = 9c$, $p(4) = 16c$, $p(5) = 25c$

we know $p(1) + \cdots + p(6) = 1$

$\Rightarrow c + 4c + 9c + 16c + 25c = 1$

$\Rightarrow 55c = 1 \Rightarrow c = \frac{1}{55}$

$\therefore\ p(1) = \frac{1}{55},\ p(2) = \frac{4}{55},\ p(3) = \frac{9}{55}$

$p(4) = \frac{16}{55},\ p(5) = \frac{25}{55}.$

$F(1) = p(X \le 1) = p(1) = \frac{1}{55}$

$F(2) = p(X \le 2) = p(1) + p(2) = \frac{5}{55}$

$F(3) = p(X \le 3) = p(1) + p(2) + p(3) = \frac{14}{55}$

$F(4) = p(X \le 4) = \frac{30}{55},\ F(5) = p(X \le 5) = 1$

14. A contractor is required by a county planning department to submit one, two, three, four, or five forms (depending on the nature of the project) in applying for a building permit. Let $Y =$ the number of forms required of the next applicant. The probability that y forms are required is known to be proportional to y—that is, $p(y) = ky$ for $y = 1, \dots, 5$.

a. What is the value of k? [*Hint:* $\sum_{y=1}^{5} p(y) = 1$.]

b. What is the probability that at most three forms are required?

c. What is the probability that between two and four forms (inclusive) are required?

d. Could $p(y) = y^2/50$ for $y = 1, \dots, 5$ be the pmf of Y?

Ans: Given,

Y = the no. of forms required of the next applicant

$p(y) = ky$ for $y = 1, 2, \dots, 5$

where $p(y) = p$(that y forms are reqd.)

a) $\sum_{y=1}^{5} p(y) = 1$

$\Rightarrow p(1) + p(2) + p(3) + p(4) + p(5) = 1$

$\Rightarrow K + 2K + 3K + 4K + 5K = 1$

$\Rightarrow 15K = 1$

$\Rightarrow K = \frac{1}{15}$

$\therefore p(y) = \frac{1}{15} y$

b) p(at most 3 forms are reqd.)

$= P(Y \leq 3) = p(1) + p(2) + p(3)$

$= \frac{1}{15} + \frac{2}{15} + \frac{3}{15} = \frac{6}{15} = \frac{2}{5} = 0.4$

c) $p(2 \leq y \leq 4) = p(2) + p(3) + p(4)$

$= \frac{2}{15} + \frac{3}{15} + \frac{4}{15} = \frac{9}{15} = .6$

d) Given $p(y) = \frac{y^2}{50}$ for $y = 1, 2, --, 5$

Now,

$$\sum_{y=1}^{5} p(y) = \frac{1}{50}(1^2 + 2^2 + 3^2 + 4^2 + 5^2) = \frac{55}{50} \neq 1$$

$\therefore p(y) = \frac{y^2}{50}$; $y = 1, 2, \cdots, 5$; can not be the pmf of Y.

13. A mail-order computer business has six telephone lines. Let X denote the number of lines in use at a specified time. Suppose the pmf of X is as given in the accompanying table.

x	0	1	2	3	4	5	6
$p(x)$	.10	.15	.20	.25	.20	.06	.04

Calculate the probability of each of the following events.

a. {at most three lines are in use}
b. {fewer than three lines are in use}
c. {at least three lines are in use}
d. {between two and five lines, inclusive, are in use}
e. {between two and four lines, inclusive, are not in use}
f. {at least four lines are not in use}

Ans: Given X denote the no of lines in use at a specified time.

x:	0	1	2	3	4	5	6
$p(x)$:	.10	.15	.2	.25	.2	.06	.04

a. p (at most 3 lines are in use) $= p(x \le 3)$

$$= p(x=0) + p(x=1) + p(x=2) + p(x=3)$$
$$= p(0) + p(1) + p(2) + p(3)$$
$$= .1 + .15 + .2 + .25 = .7$$

b. p (fewer than 3 lines are in use)

$$= p(x<3) = p(0) + p(1) + p(2)$$
$$= .1 + .15 + .2 = .45$$

c. p (at least 3 lines are in use)

$$= p(x \ge 3) = 1 - p(x<3) = 1 - 0.45 = 0.55$$

d. p (between 2 and 5 lines are in use)

$$= p(2 \le x \le 5) = p(2) + p(3) + p(4) + p(5)$$
$$= .2 + .25 + .2 + .06$$
$$= 0.71$$

e. p (between 2 & 4 lines (inclusive) are not in use)

$x \to$ the no. of lines that are in use

then $(6-x)$ no of lines are not in use.

$$p(2 \le 6-x \le 4) = p(-4 \le -x \le -2)$$
$$= p(2 \le x \le 4)$$
$$= p(2) + p(3) + p(4)$$
$$= .2 + .25 + .2$$
$$= 0.65$$

f. p(at least 4 lines are not in use)

$= p(6-X \geq 4) = p(-X \geq -2)$
$= p(X \leq 2)$
$= p(0) + p(1) + p(2)$
$= .1 + .15 + .2$
$= 0.45$

<u>Expected value</u> (or Mean value) (or Expectation) (or Mean)

Let X be a drv with the set of possible values $D = \{x_1, x_2, \ldots, x_n\}$ and pmf $p(x)$. Then the expected value of 'X' is denoted as $E(X)$ (or μ_X) (or μ) and is defined as

$$E(X) = \sum_{i=1}^{n} x_i \, p(x_i) = \sum_{x \in D} x \, p(x)$$

<u>Note</u>: 1. $E\{h(X)\} = \sum_{x \in D} h(x) . p(x)$

2. $E(aX + b) = aE(X) + b$
where 'a' & 'b' are any constants

3. $E(a) = aE(1) = a$

4. $E(1) = 1$

<u>variance</u>:

Let X be a drv with pmf $p(x)$ and expected value μ. Then the variance of X is denoted as $V(X)$ (or σ_X^2 or σ^2) and is defined as

$$V(x) = \sum_{x \in D} (x-\mu)^2 \cdot p(x)$$

$$= E\{(x-\mu)^2\} = \sigma_x^2 \text{ or } \sigma^2$$

The standard deviation of x is denoted as σ or σ_x and is defined as

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

eg:

x :	1	2	3	4	5	6
$p(x)$:	.3	.25	.15	.05	.1	.15

find μ_x, σ_x^2 and σ_x

Ans: $\mu_x = \sum_{x=1}^{6} x\, p(x) = (1)(.3) + 2(.25) + 3(.15)$
$+ (4)(.05) + (5)(.1) + (6)(.15)$

$= 2.85$

$$\sigma_x^2 = \sum_{x=1}^{6} (x-\mu_x)^2\, p(x)$$

$= (1-2.85)^2(.3) + (2-2.85)^2(.25) + (3-2.85)^2(.15)$
$+ (4-2.85)^2(.05) + (5-2.85)^2(.1) + (6-2.85)^2(.15)$

$= 3.2275$

$\sigma_x = \sqrt{3.2275} \simeq 1.7965$ (4d)

Q. Prove that

$$V(X) = \sigma_x^2 = E(X^2) - \{E(X)\}^2 \text{ (shortcut formula)}$$

Pf: $V(X) = E\{(X-\mu)^2\}$

$$= E\{X^2 - 2X\mu + \mu^2\}$$

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

$$= E(X^2) - 2\mu E(X) + \mu^2 E(1)$$

$$= E(X^2) - 2\mu \cdot \mu + \mu^2 \cdot 1$$

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

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

$$= E(X^2) - \{E(X)\}^2$$

Note: 1. $V(aX+b) = a^2 V(X)$

or

$$\sigma_{aX+b}^2 = a^2 \sigma_x^2$$

2. $\sigma_{aX+b} = |a| \sigma_x$

29. The pmf of the amount of memory X (GB) in a purchased flash drive was given in Example 3.13 as

x	1	2	4	8	16
$p(x)$	.05	.10	.35	.40	.10

Compute the following:

a. $E(X)$

b. $V(X)$ directly from the definition

c. The standard deviation of X

d. $V(X)$ using the shortcut formula

Ans: Given $x \rightarrow$ amount of memory in a purchased flash drive.

a. $E(X) = \sum_{x \in D} x\, p(x) = 1(.05) + 2(.1) + 4(.35) + 8(.4) + 16(.1)$

$= .05 + .2 + 1.4 + 3.2 + 1.6$

$= 6.45 = \mu$

b. $V(X) = \sum_{x \in D} (x-\mu)^2 \cdot p(x)$

$= (1-6.45)^2(.05) + (2-6.45)^2(.1) + (4-6.45)^2(.35) + (8-6.45)^2(.4) + (16-6.45)^2(.1)$

$= 15.6475 = \sigma_x^2$

c. $\sigma_x = \sqrt{V(X)} = \sqrt{\sigma_x^2} = \sqrt{15.6475} \simeq 3.956$ (3d)

d. $V(X) = E(X^2) - \{E(X)\}^2 = 57.25 - (6.45)^2$

$= 15.6475$

Now,

$E(X^2) = \sum_{x \in D} x^2 p(x)$

$= (1)^2(.05) + (2)^2(.1) + (4)^2(.35) + (8)^2(.4) + (16)^2(.1)$

$= 57.25$

33. Let X be a Bernoulli rv with pmf as in Example 3.18. $p(0) = 1-p$ and $p(1) = p$

$D = \{0, 1\}$

a. Compute $E(X^2)$.

b. Show that $V(X) = p(1-p)$.

c. Compute $E(X^{79})$.

Ans: a. $E(x^2) = \sum_{x \in D} x^2 p(x)$

$= (0)^2 p(0) + (1)^2 p(1)$

$= 0 + p = p$

b. To show $V(x) = p(1-p)$

we know by the shortcut formula for variance,

$V(x) = E(x^2) - \{E(x)\}^2$

$= p - \{\sum_{x \in D} x\, p(x)\}^2$

$= p - \{0 \cdot p(0) + 1 \cdot p(1)\}^2$

$= p - \{0 + p\}^2$

$= p - p^2 = p(1-p)$

c. $E(x^{79})$

$= \sum_{x \in D} x^{79} \cdot p(x)$

$= (0)^{79} p(0) + (1)^{79} p(1)$

$= 0 + p = p.$

$E\{h(x)\}$

$= \sum_{x \in D} h(x) \cdot p(x)$

38. Let X = the outcome when a fair die is rolled once. If before the die is rolled you are offered either (1/3.5) dollars or $h(X) = 1/X$ dollars, would you accept the guaranteed amount or would you gamble? [*Note:* It is not generally true that $1/E(X) = E(1/X)$.]

Ans: Given, x = the outcome when a fair die is rolled once.

$E\{h(x)\} = \sum_{x \in D} h(x) \cdot p(x)$

$S = \{1,2,3,4,5,6\}$
probability of getting each face is $\frac{1}{6}$.

$$\therefore E\{h(x)\} = \sum_{x=1}^{6} \frac{1}{x} \cdot \frac{1}{6}$$
$$= \frac{1}{6}\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}\right)$$
$$\simeq 0.4083 \ (4d)$$

Again, $\frac{1}{3.5} \simeq 0.2857$

Since, $0.4083 > \frac{1}{3.5} \ (\simeq 0.2857)$

It is suggested to gamble.

39. A chemical supply company currently has in stock 100 lb of a certain chemical, which it sells to customers in 5-lb batches. Let X = the number of batches ordered by a randomly chosen customer, and suppose that X has pmf

x	1	2	3	4
$p(x)$	.2	.4	.3	.1

Compute $E(X)$ and $V(X)$. Then compute the expected number of pounds left after the next customer's order is shipped and the variance of the number of pounds left. [*Hint:* The number of pounds left is a linear function of X.]

Ans: Given x = the no. of batches ordered by a randomly chosen customer

$$E(X) = \sum_{x=1}^{4} x\,p(x) = 1(.2)+2(.4)+3(.3)+4(.1)$$
$$= 2.3$$

$$E(x^2) = \sum_{x=1}^{4} x^2 p(x) = 1^2(.2)+2^2(.4)+3^2(.3)+4^2(.1)$$
$$= .2+1.6+2.7+1.6 = 6.1$$

$V(X) = E(X^2) - \{E(X)\}^2$
$= 6.1 - (2.3)^2 = 0.81$

Since, each batch weighs 5lb, so the no of lbs left can be written as $100 - 5X$

$E(ax+b)$
$= aE(x) + b$

$E(100 - 5X)$
$= 100 - 5E(X)$
$= 100 - 5(2.3)$
$= 88.5$

$V(ax+b)$
$= a^2 V(X)$

$V(100 - 5X)$
$= (-5)^2 V(X)$
$= 25(0.81)$
$= 20.25$

Moment generating fn.

The moment generating fn. of a random variable x is defined as

$$E(e^{tX}) = \sum_x e^{tx} p(x) = G(t) \text{ (say)}$$

Note: $E(X^K) = \left\{ \frac{\partial^K}{\partial t^K} G(t) \right\}_{t=0}$

when $K=1$,

$$E(X) = \left\{ \frac{\partial}{\partial t} G(t) \right\}_{t=0}$$
$$= \left\{ \frac{\partial}{\partial t} \sum_x e^{tx} p(x) \right\}_{t=0}$$

$$= \left[\sum_x \left\{ \frac{\partial}{\partial t} e^{tx} \right\} p(x) \right]_{t=0}$$

$$= \left[\sum_x x e^{tx} p(x) \right]_{t=0}$$

$$= \sum_x x\, p(x)$$

when $k=2$

$$E(x^2) = \left[\frac{\partial^2}{\partial t^2} G(t) \right]_{t=0}$$

$$= \left[\frac{\partial}{\partial t} \left\{ \frac{\partial}{\partial t} G(t) \right\} \right]_{t=0}$$

$$= \left[\frac{\partial}{\partial t} \left\{ \sum_x x e^{tx} p(x) \right\} \right]_{t=0}$$

$$= \left[\sum_x x^2 e^{tx} p(x) \right]_{t=0}$$

$$= \sum_x x^2 p(x)$$

<u>Binomial Experiment:</u>

An experiment is called a binomial experiment if it satisfies the following conditions.

a) The experiment concists of n trials where 'n' is fixed in advance of the experiment.

b) The trials are independent

c) Each trial results in one of the possible outcomes: success(p) and failure(q)

$p+q=1 \Rightarrow q=1-p$

where $p \to$ the probability of success
and $q \to$ " " " failure

Binomial distribution

suppose 'A' be the event which occurs in 'n' trials of a random experiment.
Let 'x' be the no. of trials in which the event 'A' occurs.
The pmf corresponding to the random variable x is given by

$$p(x) = {}^nc_x \, p^x q^{n-x} = b(x; n, p)$$

p(= success) $\to$ probability of the occurance of the event A

q (= failure) $\to$ probability of the non-occurance of the event A

The moment generating fn. is

$$G(t) = E\{e^{tx}\}$$

$$= \sum_{x=0}^{n} e^{tx} \, b(x; n, p)$$

or

$$= \sum_{x=0}^{n} e^{tx} p(x)$$

$$= \sum_{x=0}^{n} e^{tx} \, {}^nc_x \, p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nc_x (pe^t)^x q^{n-x}$$

$$\Rightarrow \boxed{G(t) = (pe^t + q)^n}$$

$$\left(\sum_{x=0}^{n} {}^nc_x p^x q^{n-x} = (p+q)^n \right.$$

$$E(x) = \left\{ \frac{\partial}{\partial t} G(t) \right\}_{t=0}$$

$$= \left\{ \frac{\partial}{\partial t} (pe^t + q)^n \right\}_{t=0}$$

$$= \left\{ n(pe^t + q)^{n-1} pe^t \right\}_{t=0}$$

$$= \left\{ n(p+q)^{n-1} p \right\} \qquad (\because p+q=1)$$

$$= np = \mu$$

$$E(x^2) = \left\{ \frac{\partial^2}{\partial t^2} G(t) \right\}_{t=0}$$

$$= \left\{ \frac{\partial}{\partial t} \left(\frac{\partial}{\partial t} G(t) \right) \right\}_{t=0}$$

$$= \left[\frac{\partial}{\partial t} \left\{ npe^t (pe^t + q)^{n-1} \right\} \right]$$

$$= \left[np \left\{ e^t (pe^t + q)^{n-1} + (n-1) e^t (pe^t + q)^{n-2} pe^t \right\} \right]_{t=0}$$

$$= \left[np \left\{ (p+q)^{n-1} + (n-1)(p+q)^{n-2} p \right\} \right]$$

$= [np\{1 + (n-1)p\}]$ $\quad (\because p+q=1)$

$= np + n^2p^2 - np^2$

$\therefore V(X) = E(X^2) - \{E(X)\}^2$

$= np + n^2p^2 - np^2 - (np)^2$

$= np - np^2$

$= np(1-p) = npq$ $\quad$ (As $q = 1-p$)

$= \sigma^2$

$S.D = \sqrt{V(X)} = \sqrt{\sigma^2} = \sqrt{npq}$

46. Compute the following binomial probabilities directly from the formula for $b(x; n, p)$:

a. $b(3; 8, .35)$

b. $b(5; 8, .6)$

c. $P(3 \leq X \leq 5)$ when $n = 7$ and $p = .6$

d. $P(1 \leq X)$ when $n = 9$ and $p = .1$

Ans:

a) $b(3; 8, .35)$; $n = 8, x = 3, p = .35$

$q = 1-p = .65$

$= {}^8C_3 (.35)^3 (.65)^{8-3}$

$= {}^8C_3 (.35)^3 (.65)^5 = 0.2786$

b) $b(5; 8, .6)$ $\quad x = 5, n = 8, p = .6$

$q = .4$

$= {}^8C_5 (.6)^5 (.4)^3$

$= .2787$

c. $p(3 \leq x \leq 5)$ where $n=7$ & $p=0.6$

$\Rightarrow q=0.4$

$= p(x=3)+p(x=4)+p(x=5)$

$= b(3;7,.6)+b(4;7,.6)+b(5;7,.6)$

$= {}^7C_3(.6)^3(.4)^4 + {}^7C_4(.6)^4(.4)^3 + {}^7C_5(.6)^5(.4)^2$

$= 0.7451$

d) $p(1 \leq x)$ when $n=9$ and $p=.1$

$q=.9$

$= p(x \geq 1)$

$= p(x=1)+p(x=2)+ \cdots + p(x=9)$

$= 1-p(x=0)$

$= 1-{}^9C_0(.1)^0(.9)^9 = 0.6126$ (4d)

50. A particular telephone number is used to receive both voice calls and fax messages. Suppose that 25% of the incoming calls involve fax messages, and consider a sample of 25 incoming calls. What is the probability that

a. At most 6 of the calls involve a fax message?

b. Exactly 6 of the calls involve a fax message?

c. At least 6 of the calls involve a fax message?

d. More than 6 of the calls involve a fax message?

Ans: Let x be the no. of calls that involve fax messages

Given 25% of the incoming calls involve fax message

$p=0.25$, $q=0.75$, $n=25$

a. p (at most 6 of the calls involve a fax message)

$$= p(X \leqslant 6) = \sum_{x=0}^{6} b(x; 25, .25)$$

$$= \sum_{x=0}^{6} {}^{25}c_x (.25)^x (.75)^{25-x}$$

$$= {}^{25}c_0 (.25)^0 (.75)^{25} + {}^{25}c_1 (.25)^1 (.75)^{24} + \cdots + {}^{25}c_6 (.25)^6 (.75)^{19}$$

$$= 0.5611 \ (4d)$$

b. p (exactly 6 of the calls involve a fax message)

$$= p(X=6) = {}^{25}c_6 (.25)^6 (.75)^{19}$$

$$= 0.1828 \ (4d)$$

c. p (at least 6 of the calls involve a fax message)

$$= p(X \geqslant 6)$$

$$= 1 - p(X < 5)$$

$$= 1 - \{p(X=0) + p(X=1) + p(X=2) + \cdots + p(X=4)\}$$

$$= 1 - \{{}^{25}c_0 (.25)^0 (.75)^{25} + {}^{25}c_1 (.25)^1 (.75)^{24} + \cdots + {}^{25}c_4 (.25)^4 (.75)^{21}\}$$

$$= 1 - 0.3783 = 0.6217$$

d. p (more than 6 of the calls involve a fax message)

$= p(x>6) = 1 - p(x \leq 6)$

$= 1 - 0.5611$

$= 0.4389$

51. Refer to the previous exercise.

a. What is the expected number of calls among the 25 that involve a fax message?

b. What is the standard deviation of the number among the 25 calls that involve a fax message?

c. What is the probability that the number of calls among the 25 that involve a fax transmission exceeds the expected number by more than 2 standard deviations?

Ans: a. $E(x) = \mu = np = 25(.25) = 6.25$

b. $\sigma = \sqrt{npq} = \sqrt{25(.25)(.75)} = 2.1651$ (4d)

c. $p(x > E(x) + 2\sigma) = p(x > 10.5802)$

$= 1 - p(x \leq 10.5802)$

$= 1 - p(x \leq 10)$

$= 1 - 0.97$

$= 0.03$

poisson's distribution

A discrete rv 'x' is said to have poisson's distribution with parametre μ ($\mu > 0$) if the pmf of x is

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

The moment generating fn. is

$$G(t) = E(e^{tx})$$

$$= \sum_{x=0}^{\infty} e^{tx}\, p(x;\mu)$$

$$= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\mu} \cdot \mu^{x}}{\lfloor x}$$

$$= \sum_{x=0}^{\infty} \frac{e^{-\mu} \cdot (\mu e^{t})^{x}}{\lfloor x}$$

$$= e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu e^{t})^{x}}{\lfloor x}$$

$$= e^{-\mu} \left\{ 1 + \frac{\mu e^{t}}{\lfloor 1} + \frac{(\mu e^{t})^{2}}{\lfloor 2} + \cdots \right\}$$

$$= e^{-\mu} \cdot e^{\mu e^{t}}$$

$$\Rightarrow G(t) = e^{\mu(e^{t}-1)}$$

$$E(X) = \left\{ \frac{\partial}{\partial t} G(t) \right\}_{t=0}$$

$$= \left[\frac{\partial}{\partial t} \left\{ e^{-\mu} e^{\mu e^{t}} \right\} \right]_{t=0}$$

$$= e^{-\mu} \left\{ \frac{\partial}{\partial t} \left(e^{\mu e^{t}} \right) \right\}_{t=0}$$

$$= e^{-\mu} \cdot \left\{ \mu e^{t}\, e^{\mu e^{t}} \right\}_{t=0}$$

$$= e^{-\mu} \cdot \mu\, e^{\mu} = \mu$$

$$E(x^2) = \left\{ \frac{\partial^2}{\partial t^2} G(t) \right\}_{t=0}$$

$$= \left\{ \frac{\partial}{\partial t} \left(\frac{\partial}{\partial t} G(t) \right) \right\}_{t=0}$$

$$= \left[\frac{\partial}{\partial t} \left\{ \mu e^{t} e^{\mu e^t} \right\} \right]_{t=0}$$

$$= e^{-\mu} \left[\frac{\partial}{\partial t} \left(e^{t} e^{\mu e^t} \right) \right]_{t=0}$$

$$= \mu e^{-\mu} \left[e^{t} \cdot e^{\mu e^t} + e^{t} \cdot \mu e^{t} e^{\mu e^t} \right]_{t=0}$$

$$= \mu e^{-\mu} \left[e^{\mu} + \mu e^{\mu} \right]$$

$$= \mu + \mu^2$$

$$\therefore V(x) = E(x^2) - \{E(x)\}^2$$

$$= \mu + \mu^2 - \mu^2$$

$$= \mu$$

Note: 1. when 'n' is large i.e; when $n \to \infty$

$$b(x; n, p) \to P(x; \mu)$$

so, $\mu = np$

$\therefore$ The mean & variance of poisson's distribution becomes $\mu = V(x) = E(x)$

i.e; $\mu = np = \sigma^2$

2. The no. of events during a time interval (t) is a poisson rv with parametre (μ) then expected no. of events $\mu = \alpha t$ where 'α' specifies the rate.

79. Let X, the number of flaws on the surface of a randomly selected boiler of a certain type, have a Poisson distribution with parameter $\mu = 5$. Use Appendix Table A.2 to compute the following probabilities:

a. $P(X \leq 8)$ **b.** $P(X = 8)$ **c.** $P(9 \leq X)$
d. $P(5 \leq X \leq 8)$ **e.** $P(5 < X < 8)$

Ans: Let $X \rightarrow$ the no. of flaws on the surface of a randomly selected boiler

$\mu = 5$.

a. $$p(X \leq 8) = \sum_{x=0}^{8} p(x;\mu)$$

$$= \sum_{x=0}^{8} \frac{e^{-\mu}\mu^{x}}{\lfloor x}$$

$$= \sum_{x=0}^{8} \frac{e^{-5}5^{x}}{\lfloor x}$$

$$= 0.932$$ (see page A-5)

b. $$p(X = 8) = \frac{e^{-5}5^{8}}{\lfloor 8} = 0.0653$$

c. $$P(9 \leq X) = 1 - P(X < 9)$$
$$= 1 - P(X \leq 8)$$
$$= 1 - 0.932 = 0.068$$

d. $p(5 \le x \le 8) = \sum_{x=5}^{8} p(x;5)$

$$= \sum_{x=5}^{8} \frac{e^{-5} \cdot 5^{x}}{\lfloor x}$$

$$= 0.4914$$

e. $p(5 < x < 8) = p(6 \le x \le 7)$

$$= p(6;5) + p(7;5)$$

$$= \frac{e^{-5}\, 5^{6}}{\lfloor 6} + \frac{e^{-5}\, 5^{7}}{\lfloor 7}$$

$$= 0.2506$$

87. The number of requests for assistance received by a towing service is a Poisson process with rate $\alpha = 4$ per hour.

a. Compute the probability that exactly ten requests are received during a particular 2-hour period.

b. If the operators of the towing service take a 30-min break for lunch, what is the probability that they do not miss any calls for assistance?

c. How many calls would you expect during their break?

Ans: Let x be the no. of requests/calls received for assistance by a towing service.

Given $\alpha = 4/\text{hour}$

a. $t = 2$, $\mu = \alpha t = (4)(2) = 8$

p(Exactly 10 requests are received)

$$= p(x=10) = \frac{e^{-8}\, 8^{10}}{\lfloor 10} = 0.0993 \text{ (4d)}$$

b. $t = \frac{1}{2}$, $\mu = \alpha t = (4)\left(\frac{1}{2}\right) = 2$

$$P(X=0) = \frac{e^{-2} 2^0}{\lfloor 0} = 0.1353 (4d)$$

c. $E(X) = \mu = 2$

86. The number of people arriving for treatment at an emergency room can be modeled by a Poisson process with a rate parameter of five per hour.

a. What is the probability that exactly four arrivals occur during a particular hour?

b. What is the probability that at least four people arrive during a particular hour?

c. How many people do you expect to arrive during a 45-min period?

Ans: Let X be the no. of people arriving for treatment at an emergency room.

Given $\alpha = 5$

a. $t = 1$, $\mu = \alpha t = 5$

$$P(X=4) = \frac{e^{-5} 5^4}{\lfloor 4} = 0.1755 (4d)$$

b. $t = 1$, $\mu = \alpha t = 5$

$$\begin{aligned} P(X \geq 4) &= 1 - P(X < 4) \\ &= 1 - P(X \leq 3) \\ &= 1 - 0.265 \\ &= 0.735 \end{aligned}$$

c. $t = 45 \text{ min} = \frac{3}{4} \text{ hr.}$

$\mu = \alpha t = 5\left(\frac{3}{4}\right) = 3.75 \approx 4$ persons

so, we expect around 4 people during the 45 min. period.

continuous random variable

A random variable 'x' is said to be continuous if it is defined in the interval.
The probability distribution or probability density fn. (pdf) of x is a fn. of $f(x)$ such that for any two nos. a and b with $a \leq b$,

$$p(a \leq x \leq b) = \int_a^b f(x)\,dx = p(a < x \leq b)$$

or

$$p(a \leq x < b)$$

or

$$p(a < x < b)$$

properties

1. $f(x) \geq 0$ for all x.
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ i.e; Area under the entire curve is equal to 1.

uniform distribution

A continuous rv 'x' is said to have an uniform distribution in the interval $[A, B]$ if the probability density fn. (pdf) of x is given by

$$f(x; A, B) = \begin{cases} \dfrac{1}{B-A}, & A \leq x \leq B \\ 0, & \text{otherwise.} \end{cases}$$

1. The current in a certain circuit as measured by an ammeter is a continuous random variable X with the following density function:

$$f(x) = \begin{cases} .075x + .2 & 3 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

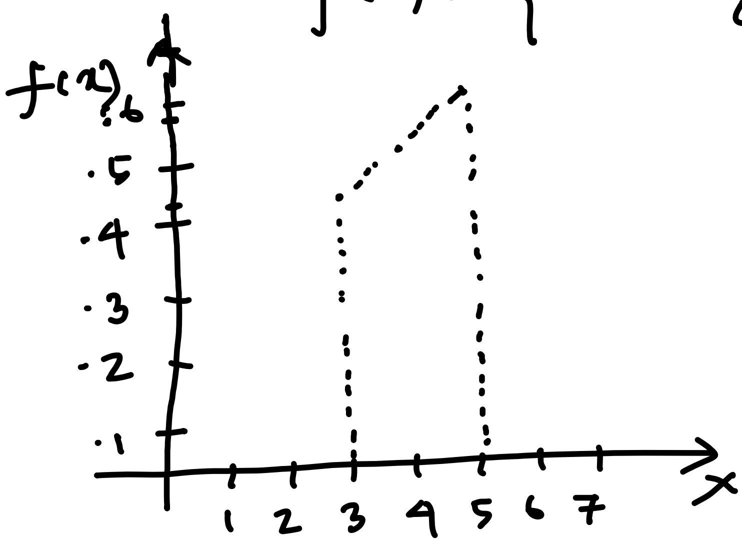
a. Graph the pdf and verify that the total area under the density curve is indeed 1.

b. Calculate $P(X \le 4)$. How does this probability compare to $P(X < 4)$?

c. Calculate $P(3.5 \le X \le 4.5)$ and also $P(4.5 < X)$.

Ans: Given

$$f(x) = \begin{cases} 0.075x + .2, & 3 \le x \le 5 \\ 0, & \text{otherwise} \end{cases}$$



$$A = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{3} f(x)\,dx + \int_{3}^{5} f(x)\,dx + \int_{5}^{\infty} f(x)\,dx$$

$$= \int_{3}^{5} (.075x + .2)\,dx$$

$$= \left[.075\frac{x^2}{2} + .2x\right]_3^5$$

$$= \frac{.075}{2}(25-9) + .2(5-3)$$

$$= 1$$

b. $$P(X \leq 4) = \int_{-\infty}^{4} f(x)\,dx$$

$$= \int_{-\infty}^{3} f(x)\,dx + \int_{3}^{4} f(x)\,dx$$

$$= \int_{3}^{4} (.075x + .2)\,dx$$

$$= \left[\frac{.075}{2}x^2 + .2x\right]_3^4$$

$$= \frac{.075}{2}(16-9) + .2(4-3)$$

$$= 0.4625$$

$$= P(X < 4)$$

c. $$P(3.5 \leq X \leq 4.5) = \int_{3.5}^{4.5} (.075x + .2)\,dx$$

$$= \frac{.075}{2}\{(4.5)^2 - (3.5)^2\} + (.2)(4.5-3.5)$$

$$= 0.5$$

$$p(4.5 < X) = p(4.5 < X < \infty)$$

$$= \int_{4.5}^{\infty} f(x)\, dx$$

$$= \int_{4.5}^{5} (0.075x + .2)\, dx + \int_{5}^{\infty} 0\, dx$$

$$= \left.\frac{0.075}{2} x^2 + .2x \right]_{4.5}^{5}$$

$$= \frac{0.075}{2} \{5^2 - (4.5)^2\} + (.2)(5 - 4.5)$$

$$= 0.278 \text{ (3d)}$$

5. A college professor never finishes his lecture before the end of the hour and always finishes his lectures within 2 min after the hour. Let X = the time that elapses between the end of the hour and the end of the lecture and suppose the pdf of X is

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a. Find the value of k and draw the corresponding density curve. [*Hint*: Total area under the graph of $f(x)$ is 1.]

b. What is the probability that the lecture ends within 1 min of the end of the hour?

c. What is the probability that the lecture continues beyond the hour for between 60 and 90 sec?

d. What is the probability that the lecture continues for at least 90 sec beyond the end of the hour?

Ans: Let X = the time that elapses between the end of the hour and end of the lecture.

The pdf is given by

$$f(x) = \begin{cases} Kx^2, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

a. we Know

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\Rightarrow \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{2} f(x)\,dx + \int_{2}^{\infty} f(x)\,dx = 1$$

$$\Rightarrow \int_{0}^{2} Kx^2\,dx = 1$$

$$\Rightarrow K\left[\frac{x^3}{3}\right]_0^2 = 1$$

$$\Rightarrow \frac{K}{3}(8) = 1 \Rightarrow \boxed{K = \frac{3}{8}}$$

b. $$P(0 \leq X \leq 1) = \int_0^1 f(x)\,dx = \int_0^1 Kx^2\,dx$$

$$= \int_0^1 \frac{3}{8}x^2\,dx = \frac{3}{8}\cdot\left\{\frac{x^3}{3}\right\}_0^1 = \frac{1}{8} = 0.125$$

c. $$P(1 \leq X \leq 1.5) = \int_1^{1.5} f(x)\,dx = \int_1^{1.5} \frac{3}{8}x^2\,dx$$

$$= \frac{3}{8}\cdot\left[\frac{x^3}{3}\right]_1^{1.5} = \frac{1}{8}\left\{(1.5)^3 - 1\right\}$$

$$= 0.297$$

d. $$P(X \geq 1.5) = P(1.5 \leq X < \infty) = \int_{1.5}^{\infty} f(x)\,dx$$

$$= \int_{1.5}^{2} \frac{3}{8}x^2\,dx + \int_{2}^{\infty} 0\,dx$$

$$= \frac{3}{8}\left[\frac{x^3}{3}\right]_{1.5}^{2}$$

$$= \frac{1}{8}\left\{8-(1.5)^3\right\} = 0.578\ (3d)$$

7. The time X (min) for a lab assistant to prepare the equipment for a certain experiment is believed to have a uniform distribution with $A = 25$ and $B = 35$.

a. Determine the pdf of X and sketch the corresponding density curve. **1**

b. What is the probability that preparation time exceeds 33 min?

c. What is the probability that preparation time is within 2 min of the mean time? [*Hint*: Identify μ from the graph of $f(x)$.]

d. For any a such that $25 < a < a + 2 < 35$, what is the probability that preparation time is between a and $a + 2$ min?

Ans: Let X (in min.) $\rightarrow$ be the time reqd. for a lab assistant to prepare the equipment for a certain experiment.

The experiment is beleived to have uniform distribution with $A=25$ and $B=35$.

a. we know the pdf for the uniform distribution is

$$f(x;A,B) = \begin{cases} \frac{1}{B-A}, & A \le x \le B \\ 0, & \text{otherwise.} \end{cases}$$

$$\Rightarrow f(x;25,35) = \begin{cases} \frac{1}{10}, & 25 \le x \le 35 \\ 0, & \text{otherwise} \end{cases}$$

b. $P(X>33) = P(33<X<\infty) = \int_{33}^{\infty} f(x;A,B)\,dx$

$$= \int_{33}^{35} f(x;25,35)\,dx + \int_{35}^{\infty} f(x;25,35)\,dx$$

$$= \int_{33}^{35} \frac{1}{10}\,dx + \int_{35}^{\infty} 0\,dx$$

$$= \frac{1}{10}(35-33) = 0.2$$

c. $\mu = \frac{25+35}{2} = 30$

$$P(28 \leq X \leq 32) = \int_{28}^{32} f(x;25,35)\,dx$$

$$= \frac{1}{10}\int_{28}^{32} dx = \frac{1}{10}(32-28) = 0.4$$

d. $P(a<X<a+2) = \int_{a}^{a+2} f(x;25,35)\,dx$

$$= \frac{1}{10}\int_{a}^{a+2} dx = \frac{1}{10}(a+2-a) = 0.2$$

<u>cumulative distribution function</u> (for continuous rv)

The cumulative distribution fn. (cdf) for a continuous rv 'X' (defined for every x) is denoted as $F(x)$ and is defined as

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)\, dy$$

<u>Note:</u>

If x is a continuous rv with pdf $f(x)$ and cdf $F(x)$ then

1. For any 'a', $P(X > a) = 1 - F(a)$
$$= 1 - P(X \leq a)$$

2. For any two nos. 'a' and 'b' with $a < b$
$$P(a \leq X \leq b) = F(b) - F(a)$$

3. $F'(x) = f(x)$ for every x at which the derivatives $F'(x)$ exists.

4. $x = a$ is called the median of the continuous rv if $F(a) = \frac{1}{2}$

eg: For uniform distribution $f(x; a, b) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$

The cdf $F(x)$ is given by

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

Expected value (or Mean value)

Expected value of a continuous rv 'x' with pdf $f(x)$ is given by

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

If $h(x)$ is any function of 'x'

$$E\{h(x)\} = \int_{-\infty}^{\infty} h(x) \cdot f(x)\, dx$$

Variance

The variance of a continuous rv 'x' with pdf $f(x)$ and mean μ (or μ_x) is given by

$$\sigma_x^2 = V(x) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

$$= E\{(x-\mu)^2\}$$

standard deviation (S.D) $= \sigma_x = \sqrt{V(x)} = \sqrt{\sigma_x^2}$

1. $V(X) = E(X^2) - \{E(X)\}^2$

2. $E(aX+b) = a\,E(X) + b$

3. $V(aX+b) = a^2\, V(X)$

4. $\sigma^2_{aX+b} = |a|\, \sigma_x$

Exercise/section 4.2

Q - **11.** Let X denote the amount of time a book on two-hour reserve is actually checked out, and suppose the cdf is

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \le x < 2 \\ 1 & 2 \le x \end{cases}$$

Use the cdf to obtain the following:

a. $P(X \le 1)$
b. $P(.5 \le X \le 1)$
c. $P(X > 1.5)$
d. The median checkout duration $\tilde{\mu}$ [solve $.5 = F(\tilde{\mu})$]
e. $F'(x)$ to obtain the density function $f(x)$
f. $E(X)$
g. $V(X)$ and σ_X
h. If the borrower is charged an amount $h(X) = X^2$ when checkout duration is X, compute the expected charge $E[h(X)]$.

Ans: Given X = the amount of time a book on two hr. reserve is actually checked out

The cdf is given by $F(x) = \begin{cases} 0, & x < 0 \\ x^2/4, & 0 \le x < 2 \\ 1, & x \ge 2 \end{cases}$

a. $P(X \le 1) = F(1) = \frac{1}{4} = 0.25$

b. $P(.5 \le X \le 1) = F(1) - F(.5)$

$= \frac{1^2}{4} - \frac{(.5)^2}{4} = \frac{1 - .25}{4}$

$= 0.1875$

c. $P(X > 1.5) = 1 - P(X \le 1.5)$

$= 1 - F(1.5)$

$= 1 - \frac{(1.5)^2}{4}$

$= \frac{4 - 2.25}{4} = 0.4375$

d. Let $0 \leq \tilde{\mu} < 2$ s.t. $F(\tilde{\mu}) = 0.5$

$$\Rightarrow \frac{(\tilde{\mu})^2}{4} = 0.5$$

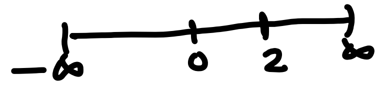
$$\Rightarrow (\tilde{\mu})^2 = 2$$

$$\Rightarrow \tilde{\mu} = \sqrt{2} = 1.414$$

e. we know $f(x) = F'(x)$

where $F(x) = \begin{cases} 0, & x < 0 \\ x^2/4, & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$

$$F'(x) = \begin{cases} 0, & x < 0 \\ x/2, & 0 \leq x < 2 \\ 0, & x \geq 2 \end{cases}$$



i.e; $F'(x) = \begin{cases} x/2, & 0 \leq x < 2 \\ 0, & \text{otherwise} \end{cases}$

so, $f(x) = \begin{cases} x/2, & 0 \leq x < 2 \\ 0, & \text{otherwise} \end{cases}$

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

$$= \int_{-\infty}^{0} x f(x)\, dx + \int_{0}^{2} x f(x)\, dx + \int_{2}^{\infty} x f(x)\, dx$$

$$= \int_{0}^{2} x \cdot \frac{x}{2}\, dx$$

$$= \left.\frac{x^3}{6}\right]_0^2 = \frac{4}{3} = 1.333 \text{ (3d)}$$

g. $V(X) = E(X^2) - \{E(X)\}^2$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^2 x^2 \cdot \frac{x}{2}\,dx$$

$$= \left.\frac{x^4}{8}\right]_0^2 = \frac{16}{8} = 2$$

$$\therefore V(X) = 2 - \{(1.333)^2\} = 0.223 \text{ (3d)}$$

h. Given $h(X) = x^2$,

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

$$= \int_0^2 x^2 \cdot \frac{x}{2}\,dx = \left.\frac{x^4}{8}\right]_0^2 = 2$$

15. Let X denote the amount of space occupied by an article placed in a 1-ft^3 packing container. The pdf of X is

$$f(x) = \begin{cases} 90x^8(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Graph the pdf. Then obtain the cdf of X and graph it.
b. What is $P(X \le .5)$ [i.e., $F(.5)$]?
c. Using the cdf from (a), what is $P(.25 < X \le .5)$? What is $P(.25 \le X \le .5)$?
d. What is the 75th percentile of the distribution?
e. Compute $E(X)$ and σ_X.
f. What is the probability that X is more than 1 standard deviation from its mean value?

Ans: Let x denote the amount of space occupied by an article in a 1-ft^3 container.

The pdf of x is

$$f(x) = \begin{cases} 90x^8(1-x), & 0<x<1 \\ 0, & \text{otherwise} \end{cases}$$

a. $F(x) = \int 90x^8(1-x)\,dx, \quad 0<x<1$

$= 90\int (x^8 - x^9)\,dx, \quad 0<x<1$

$= 90\left[\frac{x^9}{9} - \frac{x^{10}}{10}\right], \quad 0<x<1$

$= 10x^9 - 9x^{10}, \quad 0<x<1$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 10x^9 - 9x^{10}, & 0<x<1 \\ 1, & x \geq 1 \end{cases}$$

b. $P(X \leq .5) = F(.5) = 10\cdot(.5)^9 - 9\cdot(.5)^{10}$

$= 0.0107$ (4d)

c. $P(.25 \leq X \leq .5)$ (using cdf)

$= F(.5) - F(.25)$

$= \{10(.5)^9 - 9(.5)^{10}\} - \{10(.25)^9 - 9(.25)^{10}\}$

$= 0.0107 - 0.00003$

$= 0.0107$ (4d)

$P(.25 < X \leq .5) = \int_{.25}^{.5} f(x)\,dx$ (using pdf)

$$= \int_{.25}^{.5} 90x^8(1-x)\,dx$$

$$= 90\int_{.25}^{.5} (x^8 - x^9)\,dx$$

$$= 90\left[\frac{x^9}{9} - \frac{x^{10}}{10}\right]_{.25}^{.5}$$

$$= \left[10x^9 - 9x^{10}\right]_{.25}^{.5}$$

$$= \{10(.5)^9 - 9(.5)^{10}\} - \{10(.25)^9 - 9(.25)^{10}\}$$

$$= 0.0107 \text{ (4d)}$$

d. The 75^{th} percentile of the distribution

= the value of x for which $F(x) = 0.75$

for $0 < x < 1$, $F(x) = 0.75$

$$\Rightarrow 10x^9 - 9x^{10} = 0.75$$

$$\Rightarrow x = 0.09036 \text{ (using Mathematica Software)}$$

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

$$= \int_0^1 x\, 90x^8(1-x)\,dx$$

$$= 90 \int_0^1 (x^9 - x^{10})\, dx$$

$$= 90 \left[\frac{x^{10}}{10} - \frac{x^{11}}{11} \right]_0^1$$

$$= \frac{90}{110} \left[11x^{10} - 10x^{11} \right]_0^1$$

$$= 0.818 \text{ (3d)}$$

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

we know $V(X) = E(X^2) - \{E(X)\}^2$

Now, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$

$$= \int_0^1 x^2 \, 90x^8 (1-x)\, dx$$

$$= 90 \int_0^1 (x^{10} - x^{11})\, dx$$

$$= 90 \left[\frac{x^{11}}{11} - \frac{x^{12}}{12} \right]_0^1$$

$$= \frac{90}{132} \left[12x^{11} - 11x^{12} \right]$$

$$= 0.682$$

$$\therefore V(X) = 0.682 - (0.818)^2$$

$$= 0.013 \text{ (3d)}$$

Thus, $\sigma_x = \sqrt{V(X)} = 0.114 \text{ (3d)}$

f. $P(X > \mu \pm \sigma) = 1 - P(X \leq \mu \pm \sigma)$

$= 1 - P(\mu - \sigma \leq X \leq \mu + \sigma)$

$= 1 - P(.818 - .114 \leq X \leq .818 + .114)$

$= 1 - P(.704 \leq X \leq .932)$

$= 1 - \{F(.932) - F(.704)\}$

$= 1 - \left[\{10(.932)^9 - 9(.932)^{10}\} - \{10(.704)^9 - 9(.704)^{10}\}\right]$

$= 0.3003$ (4d)

21. An ecologist wishes to mark off a circular sampling region having radius 10 m. However, the radius of the resulting region is actually a random variable R with pdf

$$f(r) = \begin{cases} \frac{3}{4}[1 - (10 - r)^2] & 9 \leq r \leq 11 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected area of the resulting circular region?

Ans: Radius of the circular region 'R' is a random variable with pdf

$$f(r) = \begin{cases} \frac{3}{4}\{1 - (10 - r)^2\}, & 9 \leq r \leq 11 \\ 0, & \text{otherwise.} \end{cases}$$

Let $h(R)$ = Area of the circular region

$= \pi R^2$

$\therefore E\{h(R)\} = \int_{-\infty}^{\infty} h(r) f(r)\, dr$

$$= \int_9^{11} \pi r^2 \frac{3}{4}\{1-(10-r)^2\}\, dr$$

$$= \frac{3\pi}{4}\int_9^{11} r^2\{1-100-r^2+20r\}\, dr$$

$$= \frac{3\pi}{4}\int_9^{11} r^2(-r^2+20r-99)\, dr$$

$$= \frac{3\pi}{4}\int_9^{11} (-r^4+20r^3-99r^2)\, dr$$

$$= \frac{3\pi}{4}\left[-\frac{r^5}{5}+5r^4-33r^3\right]_9^{11}$$

$$= \frac{3\pi}{4}\left[-\frac{(11)^5}{5}+5(11)^4-33(11)^3 + \frac{(9)^5}{5}-5(9)^4+33(9)^3\right]$$

$$= 314.79 \text{ m}^2$$

23. If the temperature at which a certain compound melts is a random variable with mean value 120°C and standard deviation 2°C, what are the mean temperature and standard deviation measured in °F? [*Hint*: °F = 1.8°C + 32.]

Ans: Let $X \rightarrow$ temperature measured in °C

$Y \rightarrow$ " " " °F

so, $Y = 1.8X + 32$

$\mu_x = 120$, $\sigma_x = 2$

$\mu_y = 1.8\mu_x + 32$

$\Rightarrow \mu_y = 1.8(120) + 32$

$= 248^0 F$

and $\sigma_y = |1.8| 2$

$= 3.6^0 F$.

$\mu_y = ?$ $\sigma_y = ?$

$\left[\begin{array}{l} \text{If } y = ax + b \\ \mu_y = a\mu_x + b \\ E(ax+b) = aE(x) + b \end{array}\right.$

$\left(\begin{array}{l} \sigma_{ax+b} = |a|\sigma_x \\ \text{ie; } \sigma_y = |a|\sigma_x \end{array}\right.$

27. When a dart is thrown at a circular target, consider the location of the landing point relative to the bull's eye. Let X be the angle in degrees measured from the horizontal, and assume that X is uniformly distributed on [0, 360]. Define Y to be the transformed variable $Y = h(X) = (2\pi/360)X - \pi$, so Y is the angle measured in radians and Y is between $-\pi$ and π. Obtain $E(Y)$ and σ_Y by first obtaining $E(X)$ and σ_X, and then using the fact that $h(X)$ is a linear function of X.

Ans: Given, $x \rightarrow$ be the angle in degrees measured from the horizontal.

Also, x is uniformly distributed on $[0, 360]$.

y is the transformed variable

$$y = h(x) = \frac{2\pi}{360} x - \pi$$

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

$$\left(f(x) = \begin{cases} \frac{1}{360-0}, & 0 \le x \le 360 \\ 0, & \text{otherwise} \end{cases} \right.$$

$$= \int_0^{360} \frac{x}{360}\,dx = \frac{1}{360}\left[\frac{x^2}{2}\right]_0^{360} = \frac{360}{2} = 180^\circ$$

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

$$= \int_0^{360} (x-180)^2 \frac{1}{360}\,dx$$

$$= \frac{1}{360}\int_0^{360} (x-180)^2\,dx = \frac{1}{360} \left.\frac{(x-180)^3}{3}\right]_0^{360}$$

$$= \frac{1}{360}\left\{\frac{(180)^3}{3} - \frac{(-180)^3}{3}\right\}$$

$$= \frac{1}{360} \cdot \frac{2(180)^3}{3} = \frac{(180)^2}{3}$$

$$\therefore\ \sigma_x = \frac{180}{\sqrt{3}}$$

$$\left(\begin{array}{l} \text{If } y = ax + b \\ E(y) = a\,E(x) + b \\ \sigma_y = |a|\,\sigma_x \end{array}\right.$$

$$\therefore\ E(y) = \frac{2\pi}{360}(180) - \pi$$
$$= 0$$

$$\sigma_y = \left|\frac{2\pi}{360}\right| \sigma_x$$
$$= \frac{2\pi}{360} \cdot \frac{180}{\sqrt{3}} = \frac{\pi}{\sqrt{3}} \simeq 1.814\ (3d)$$

Normal distribution:

A continuous rv X is said to have a normal distribution with parameters μ $(-\infty<\mu<\infty)$ and σ $(\sigma>0)$ if the pdf of X is

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

The cdf of X is

$$P(X\le x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}dy = F(x)$$

The normal distribution with $\mu=0$ and $\sigma=1$ is called the standard normal distribution

A random variable having a standard normal distribution is called a standard normal random variable and is denoted as Z. The pdf of Z is

$$f(z;0,1)=\frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}, \quad -\infty<z<\infty$$

where $z=\frac{x-\mu}{\sigma}$

The cumulative distribution fn. (cdf) of Z is

$$P(Z\le z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-y^2/2}\,dy=\phi(z)$$

If X has a normal distribution with mean μ and S.D. σ then $Z\left(=\frac{X-\mu}{\sigma}\right)$ has a standard normal distribution.

Thus,

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$

or

$P(a < X \le b)$

or

$$P(a \le X < b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

or

$$P(a < X < b) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \le a) = \phi\left(\frac{a-\mu}{\sigma}\right) = P(X < a)$$

$$P(X > b) = 1 - P(X \le b)$$
$$= 1 - \phi\left(\frac{b-\mu}{\sigma}\right)$$

Q. show that $\phi(-z) = 1 - \phi(z)$

Pf: we know

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2}\, dy \quad \text{— (1)}$$

$$\therefore \phi(-z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z} e^{-y^2/2}\, dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{z} e^{-v^2/2} (-dv)$$

Let $y = -v$

$\Rightarrow dy = -dv$

$v = \infty$ when $y = -\infty$

$v = z$ when $y = -z$

$$= \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{-v^2/2}\, dv$$

$$\left(\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx\right)$$

$$= \int_{z}^{\infty} f(v)\, dv$$

where $f(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$

i.e; $\phi(-z) = \int_{z}^{\infty} f(v)\, dv$ ——— ②

Also, we know $\int_{-\infty}^{\infty} f(v)\, dv = 1$

$$\Rightarrow \int_{-\infty}^{z} f(v)\, dv + \int_{z}^{\infty} f(v)\, dv = 1$$

$$\Rightarrow \int_{z}^{\infty} f(v)\, dv = 1 - \int_{-\infty}^{z} f(v)\, dv$$

$$\Rightarrow \boxed{\phi(-z) = 1 - \phi(z)} \quad (\text{By ① \& ②})$$

Exercise/Section 4.3

28. Let Z be a standard normal random variable and calculate the following probabilities, drawing pictures wherever appropriate.

a. $P(0 \le Z \le 2.17)$
b. $P(0 \le Z \le 1)$
c. $P(-2.50 \le Z \le 0)$
d. $P(-2.50 \le Z \le 2.50)$
e. $P(Z \le 1.37)$
f. $P(-1.75 \le Z)$
g. $P(-1.50 \le Z \le 2.00)$
h. $P(1.37 \le Z \le 2.50)$
i. $P(1.50 \le Z)$
j. $P(|Z| \le 2.50)$

$P(a \le Z \le b) = \phi(b) - \phi(a)$

Ans: a. $P(0 \le Z \le 2.17)$

$= \phi(2.17) - \phi(0)$

$= 0.9850 - 0.5000$

$= 0.4850$

b. $p(0 \leq Z \leq 1) = \phi(1) - \phi(0)$

$= .8413 - .5 = .3413$

c. $p(-2.50 \leq Z \leq 0) = \phi(0) - \phi(-2.5)$

$= \phi(0) - \{1 - \phi(2.5)\}$

$= \phi(0) + \phi(2.5) - 1$

$= .5 + .9938 - 1$

$= 0.4938$

d. $p(-2.5 \leq Z \leq 2.5) = \phi(2.5) - \phi(-2.5)$

$= \phi(2.5) - \{1 - \phi(2.5)\}$

$= 2\phi(2.5) - 1$

$= 2\{.9938\} - 1$

$= 0.9876$

f. $P(-1.75 \leq Z) = P(Z \geq -1.75)$

$= 1 - P(Z < -1.75)$

$= 1 - \phi(-1.75)$

$= 1 - \{1 - \phi(1.75)\}$

$= \phi(1.75)$

$= .9599$

j. $p(|Z| < 2.5) = p(-2.5 < Z < 2.5)$

$= \phi(2.5) - \phi(-2.5)$

$= 2\phi(2.5) - 1 = 0.9876$

29. In each case, determine the value of the constant c that makes the probability statement correct.

a. $\Phi(c) = .9838$ **b.** $P(0 \le Z \le c) = .291$

c. $P(c \le Z) = .121$ **d.** $P(-c \le Z \le c) = .668$

e. $P(c \le |Z|) = .016$

Ans: a. $\phi(c) = .9838$

$\Rightarrow c = 2.14$

b. $P(0 \le Z \le c) = .291$

$\Rightarrow \phi(c) - \phi(0) = .291$

$\Rightarrow \phi(c) - .5 = .291$

$\Rightarrow \phi(c) = .791$

$\Rightarrow c = 0.81$

e. $P(c \le |Z|) = .016$

$\Rightarrow P(|Z| \geq c) = .016$

$\Rightarrow 1 - P(|Z| < c) = .016$

$\Rightarrow 1 - P(-c < Z < c) = .016$

$\Rightarrow 1 - \{\phi(c) - \phi(-c)\} = .016$

$\Rightarrow 1 - \{2\phi(c) - 1\} = .016$

$\Rightarrow 2\phi(c) = 2 - .016$

$\Rightarrow \phi(c) = 0.992$

$\Rightarrow c = 2.41$

c. $P(c \le Z) = .121$

$\Rightarrow P(Z \geq c) = .121$

$\Rightarrow 1 - P(Z < c) = .121$

$\Rightarrow 1 - \phi(c) = .121$

$\Rightarrow \phi(c) = 0.879$

$\Rightarrow c = 1.17$

41. The automatic opening device of a military cargo parachute has been designed to open when the parachute is 200 m above the ground. Suppose opening altitude actually has a normal distribution with mean value 200 m and standard deviation 30 m. Equipment damage will occur if the parachute opens at an altitude of less than 100 m. What is the probability that there is equipment damage to the payload of at least one of five independently dropped parachutes?

Ans: Let x be the opening altitude of the parachute

Given $\mu = 200$, $\sigma = 30$

Equipment damage will occur if the parachute opens at an altitude less than 100m.

Now, P(that the equipment damage will occur)

= P(that the parachute opens at an altitude less than 100m)

$= P(x < 100)$

$= P\left(\frac{X-\mu}{\sigma} < \frac{100-\mu}{\sigma}\right)$

$= P\left(Z < \frac{100-200}{30}\right)$

$= P(Z < -3.33)$

$= \phi(-3.33)$

$= 1 - \phi(3.33)$

$= 1 - .9996 = 0.0004$

so, P(that the equipment damage will not occur)

= 1 − P(that the equipment damage will occur)

$= 1 - 0.0004 = 0.9996$

p(that the equipment damage will not occur in the 5 independent drops)

$= (0.9996)^5$

Thus,

p(that there is equipment damage to the payload of at least one of the 5 independently dropped parachutes)

$= 1 - p$(that the equipment damage will not occur in the 5 independently dropped parachutes)

$= 1-(0.9996)^5$

$= 0.002$

55. Suppose only 75% of all drivers in a certain state regularly wear a seat belt. A random sample of 500 drivers is selected. What is the probability that

a. Between 360 and 400 (inclusive) of the drivers in the sample regularly wear a seat belt?

b. Fewer than 400 of those in the sample regularly wear a seat belt?

Ans: $n = 500,\ p = 0.75,\ q = 1-p = 0.25$

$\mu = np = (500)(.75) = 375,\ \sigma = \sqrt{npq} = \sqrt{(500)(.75)(.25)}$
$= 9.682$

Let $x \rightarrow$ be the no. of drivers wearing seat belt.

a. $p(360 \leq x \leq 400)$

$= P\left(\frac{360 - 0.5 - \mu}{\sigma} \leq \frac{x-\mu}{\sigma} \leq \frac{400 + 0.5 - \mu}{\sigma}\right)$

$\left[P(a \leq x \leq b) = P\left(\frac{a - .5 - \mu}{\sigma} \leq z \leq \frac{b + .5 - \mu}{\sigma}\right)\right.$

$$= P\left(\frac{360-.5-375}{9.682} \le Z \le \frac{400+0.5-375}{9.682}\right)$$

$$= P(-1.6 \le Z \le 2.63) = \phi(2.63) - \phi(-1.6)$$

$$= \phi(2.63) + \phi(1.6) - 1 \qquad (\phi(-z) = 1 - \phi(z))$$

$$= .9957 + .9452 - 1$$

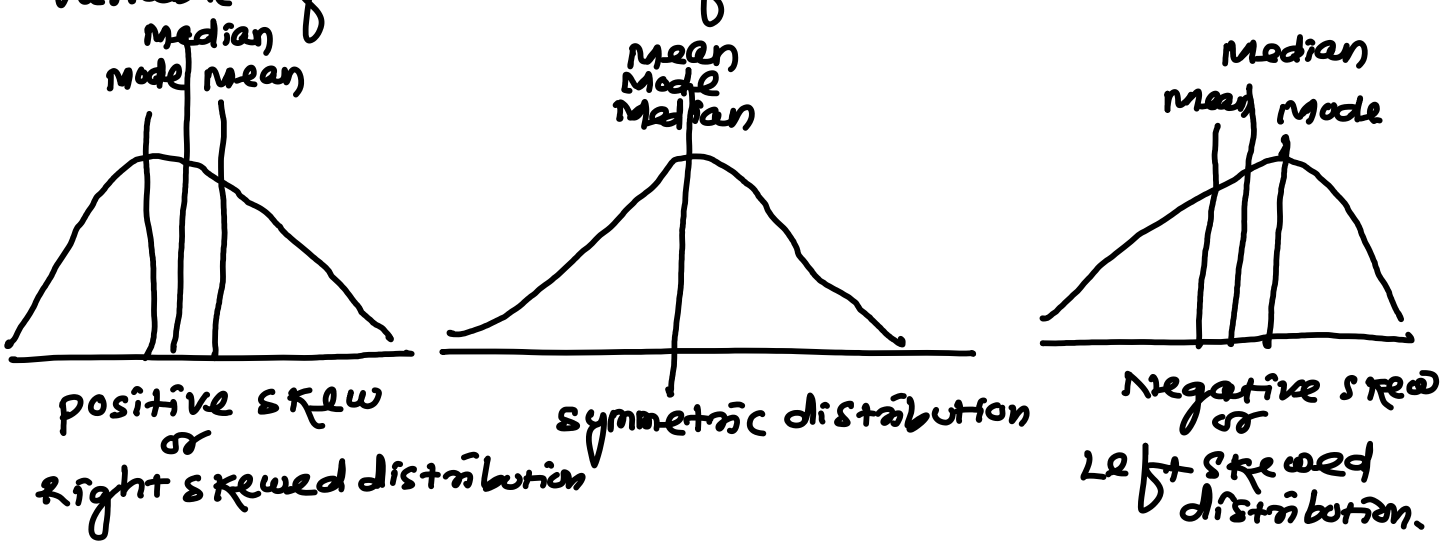
$$= 0.9409$$

b. $P(X<400) = P\left(\frac{X-\mu}{\sigma} < \frac{400-.5-\mu}{\sigma}\right)$

$$= P\left(Z < \frac{400-.5-375}{9.682}\right)$$

$$= P(Z < 2.53) = \phi(2.53) = 0.9943$$

The density curve to any normal distribution is bell shaped and therefore symmetric. sometimes the variable of interest is of skewed distribution.



Exponential distribution

A rv 'x' is said to have exponential distribution with parameter λ ($\lambda>0$) if pdf of x is

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

$$E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$V(X) = \frac{1}{\lambda^2}$$

The cdf of x is given by

$$F(x;\lambda) = \begin{cases} 0 & , x < 0 \\ 1-e^{-\lambda x} & , x \geq 0 \end{cases}$$

Gamma distribution

A continuous rv X is said to have a gamma distribution if the pdf of X is

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

where α and β are the parameters satisfying $\alpha>0$, $\beta>0$.

The standard gamma distribution has $\beta=1$. The pdf of a standard gamma random variable is given by

$$f(x;\alpha) = \begin{cases} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} & , x \geq 0 \\ 0 & , \text{otherwise} \end{cases}$$

The mean and the variance of a rv 'X' with gamma distribution $f(x;\alpha,\beta)$ are

$$E(X) = \mu = \alpha\beta \quad , V(X) = \sigma^2 = \alpha\beta^2$$

The cummulative distribution function (cdf) of the standard gamma rv 'X' is

$$F(x;\alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy \, , \, x > 0$$

is called the incomplete gamma function.

Note: 1. If X has a gamma distribution with parameters α and β then for any $x > 0$, the cdf of X is given by

$$P(X \leq x) = F(x;\alpha,\beta) = F\left(\frac{x}{\beta};\alpha\right)$$

where $F(\cdot;\alpha)$ represents the incomplete gamma function.

2. Let ν be a +ve integer. Then a rv X is said to have chi-squared distribution (or χ^2-distribution) with parameter 'ν' if the pdf of X is the gamma density with

$\alpha = \frac{\nu}{2}$ and $\beta = 2$. The pdf of χ^2 distribution is

$$f(x, \nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\,\Gamma\left(\frac{\nu}{2}\right)}\, x^{\frac{\nu}{2}-1}\, e^{-x/2}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Here ν represents the degrees of freedom of χ.

3. <u>t-distribution:</u>

Let ν be a +ve integer. Then a continuous rv X is said to have t-distribution with parameter 'ν' if the pdf of X is given by

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

4. <u>F-distribution:</u>

Let d_1 and d_2 be two +ve integers. Then a continuous rv X is said to have F-distribution if the pdf of X is given by

$$f(x; d_1, d_2) = \frac{1}{\beta(d_1, d_2)} \left(\frac{d_1}{d_2}\right)^{d_1/2} (x)^{\frac{d_1}{2}-1} \left(1 + \frac{d_1 x}{d_2}\right)^{-\frac{(d_1+d_2)}{2}}$$

for any real $x > 0$.

Here, $X = \dfrac{S_1/d_1}{S_2/d_2}$ where S_1 and S_2 are independent random variables with χ^2-distribution having

d_1 and d_2 degrees of freedom respectively.

Exercise / section 4.4

59. Let $X =$ the time between two successive arrivals at the drive-up window of a local bank. If X has an exponential distribution with $\lambda = 1$ (which is identical to a standard gamma distribution with $\alpha = 1$), compute the following:

a. The expected time between two successive arrivals

b. The standard deviation of the time between successive arrivals

c. $P(X \leq 4)$ **d.** $P(2 \leq X \leq 5)$

Ans: Let $x =$ the time between two successive arrivals at the drive up window of a local bank.

Given, x has an exponential distribution with $\lambda = 1$

The pdf of x is given by

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

and cdf of x is

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases}$$

a. The expected time between two successive arrivals $= E(x) = \frac{1}{\lambda} = \frac{1}{1} = 1$

b. The S.D. of the time between 2 successive arrivals $= \sigma_x = \frac{1}{\lambda} = \frac{1}{1} = 1$

c. $p(x \le 4) = F(4) = 1 - e^{-4} = 0.9817$ (4d)

d. $p(2 \le x \le 5) = F(5) - F(2)$

$= \{1 - e^{-5}\} - \{1 - e^{-2}\}$

$= e^{-2} - e^{-5}$

$= 0.1286$ (4d)

Q. suppose the reaction time x of a randomly selected individual to a certain stimulus has a standard gamma distribution with $\alpha = 2$. find the probability that

a. the reaction time is between 3 seconds & 5 seconds (both inclusive)

b. the reaction time is more than 4 seconds.

Ans: Let x = reaction time of a randomly selected individual to a certain stimulus.

Given that x has a standard gamma distribution. and $\alpha = 2$

a. $p(3 \le x \le 5) = F(5;2) - F(3;2)$ (Table A-4 p-717)

$= 0.960 - 0.801$

$= 0.159$

b. $p(x > 4) = 1 - p(x \le 4)$

$= 1 - F(4;2)$

$= 1 - .908 = 0.092$

66. Suppose the time spent by a randomly selected student who uses a terminal connected to a local time-sharing computer facility has a gamma distribution with mean 20 min and variance 80 min^2.

a. What are the values of α and β?

b. What is the probability that a student uses the terminal for at most 24 min?

c. What is the probability that a student spends between 20 and 40 min using the terminal?

Ans: Let x be the time spent by a randomly selected student who uses a terminal connected to a local sharing computer facility.

Given x has a gamma distribution with $\mu = 20$ min. and $\sigma^2 = 80$ min^2.

we know, $\mu = \alpha\beta = 20$, $\sigma^2 = \alpha\beta^2 = 80$

a. $\dfrac{\alpha\beta^2}{\alpha\beta} = \dfrac{80}{20} = 4 \Rightarrow \beta = 4$

$\therefore \alpha = 5$

b. $p(x \le 24) = F\left(\frac{24}{4}; 5\right) = F(6;5)$

$= 0.715$

c. $p(20 < x < 40) = F\left(\frac{40}{4}; 5\right) - F\left(\frac{20}{4}; 5\right)$

$= F(10;5) - F(5,5)$

$= 0.971 - 0.560$

$= 0.411$

Two discrete random variables

The pmf of a single discrete rv specifies how much probability mass is placed on each possible value of x. The joint pmf of 2 discrete rvs x and y describes how much probability mass is placed on each possible pair of values (x, y)

Let x and y be two discrete rvs defined on the sample space 'S' of an experiment. The joint probability mass function $p(x, y)$ is defined for each pair of nos. (x, y) by

$$p(x, y) = P(X = x, Y = y)$$

Here $p(x,y) \geq 0$ and $\sum_x \sum_y p(x,y) = 1$

For any set 'A' consisting of pairs (x, y) values

$$P\{(x,y) \in A\} = \sum_{(x,y)\in A} p(x,y)$$

The marginal probability mass function of x is denoted $p_x(x)$ and is given by

$$p_x(x) = \sum_{y : p(x,y) > 0} p(x, y) \quad \text{for each possible value of } x.$$

Similarly, $p_y(y) = \sum_{x : p(x,y) > 0} p(x,y)$ for each possible value of y.

Two continuous random variables

Let x and y be two continuous rvs. Then the joint probability density fn. $f(x,y)$ for each pair (x,y) in the two dimensional set 'A' is given by

$$P\{(x,y) \in A\} = \iint_A f(x,y)\, dx\, dy$$

Here $f(x,y) \geq 0$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$

If 'A' is a two dimensional rectangle $\{(x,y) : a \leq x \leq b,\ c \leq y \leq d\}$ then

$$P\{(x,y) \in A\} = \int_a^b \int_c^d f(x,y)\, dy\, dx$$

The marginal density fn. of x is denoted $f_x(x)$ and is given by

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\, dy \quad \text{for } -\infty < x < \infty$$

similarly $$f_y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx \quad \text{for } -\infty < y < \infty$$

Note: If x and y are independent then

$p(x,y) = p_x(x) \cdot p_y(y)$; x & y are discrete

$f(x,y) = f_x(x) \cdot f_y(y)$; x & y are continuous

Conditional distributions

If X and Y are 2 discrete rvs with joint pmf $p(x,y)$ and marginal pmf of X ($p_X(x) > 0$) then the conditional probability mass function of Y on $X(=x)$ is

$$p_{Y|X}(y|x) = \frac{p(x,y)}{p_X(x)}, \text{ for each value of } y.$$

similarly, $p_{X|Y}(x|y) = \dfrac{p(x,y)}{p_Y(y)}$, $p_Y(y) > 0$; for each value of x.

If X and Y are 2 continuous rvs with joint pdf $f(x,y)$ and marginal pdf of X ($f_X(x) > 0$) then the conditional probability density fn. of Y on $X(=x)$ is given by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \quad -\infty < y < \infty$$

similarly, $$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad -\infty < x < \infty.$$

Exercise / Section 5.1

1. A service station has both self-service and full-service islands. On each island, there is a single regular unleaded pump with two hoses. Let X denote the number of hoses being used on the self-service island at a particular time, and let Y denote the number of hoses on the full-service island in use at that time. The joint pmf of X and Y appears in the accompanying tabulation.

$p(x, y)$		$y = 0$	$y = 1$	$y = 2$
x	0	.10	.04	.02
	1	.08	.20	.06
	2	.06	.14	.30

a. What is $P(X = 1 \text{ and } Y = 1)$?
b. Compute $P(X \le 1 \text{ and } Y \le 1)$.
c. Give a word description of the event $\{X \neq 0 \text{ and } Y \neq 0\}$, and compute the probability of this event.
d. Compute the marginal pmf of X and of Y. Using $p_X(x)$, what is $P(X \le 1)$?
e. Are X and Y independent rv's? Explain.

Ans: Let x denote the no of hoses being used on the self service island at a particular time. and y denote the no of hoses on the full service island in use at that time.

a. $p(x=1 \text{ and } y=1) = 0.20$

b. $p(x \le 1 \text{ and } y \le 1) = p(0,0) + p(0,1) + p(1,0) + p(1,1)$
$= 0.10 + .04 + .08 + .20$
$= 0.42$

c. $p(x \neq 0 \text{ and } y \neq 0) = p(1,1) + p(1,2) + p(2,1) + p(2,2)$
$= .20 + .06 + .14 + .30$
$= .7$

d. $p_x(0) = \sum_{y=0}^{2} p(0,y)$ $\quad \left(p_x(x) = \sum_{y=0}^{2} p(0,y) \right)$
$= p(0,0) + p(0,1) + p(0,2)$
$= .1 + .04 + .02 = .16$

$$p_x(1) = \sum_{y=0}^{2} p(1,y) = p(1,0)+p(1,1)+p(1,2)$$
$$= .08+.2+.06 = 0.34$$

$$p_x(2) = \sum_{y=0}^{2} p(2,y) = p(2,0)+p(2,1)+p(2,2)$$
$$= .06+.14+.30 = 0.5$$

$$p_y(0) = \sum_{x=0}^{2} p(x,0) = p(0,0)+p(1,0)+p(2,0)$$
$$= .1+.08+.06 = 0.24$$

$$p_y(1) = \sum_{x=0}^{2} p(x,1) = p(0,1)+p(1,1)+p(2,1)$$
$$= .04+.20+.14 = 0.38$$

$$p_y(2) = \sum_{x=0}^{2} p(x,2) = p(0,2)+p(1,2)+p(2,2)$$
$$= .02+.06+.3 = 0.38$$

e. from the joint probability table given

$p(2,2) = 0.3$ —— (1)

from the calculation in (d)

$p_x(2) = 0.5$ and $p_y(2) = 0.38$

$p_x(2) \cdot p_y(2) = (0.5)(0.38) = 0.19$ —— (2)

from (1) and (2)

$$p(2,2) \neq p_x(2) \cdot p_y(2)$$

clearly, x and y are not independent.

2. When an automobile is stopped by a roving safety patrol, each tire is checked for tire wear, and each headlight is checked to see whether it is properly aimed. Let X denote the number of headlights that need adjustment, and let Y denote the number of defective tires.

a. If X and Y are independent with $p_X(0) = .5$, $p_X(1) = .3$, $p_X(2) = .2$, and $p_Y(0) = .6$, $p_Y(1) = .1$, $p_Y(2) = p_Y(3) = .05$, and $p_Y(4) = .2$, display the joint pmf of (X, Y) in a joint probability table.

b. Compute $P(X \leq 1 \text{ and } Y \leq 1)$ from the joint probability table, and verify that it equals the product $P(X \leq 1) \cdot P(Y \leq 1)$.

c. What is $P(X + Y = 0)$ (the probability of no violations)?

d. Compute $P(X + Y \leq 1)$.

Ans: Let x denote the no. of headlights that need adjustment and y denote the no. of defective tires

a. x and y are independent with

$p_x(0) = 0.5$, $p_x(1) = 0.3$, $p_x(2) = 0.2$,

$p_y(0) = 0.6$, $p_y(1) = 0.1$, $p_y(2) = 0.05 = p_y(3)$, $p_y(4) = .2$

As x and y are independent

$$p(x,y) = p_x(x) \cdot p_y(y)$$

$p(0,0) = p_x(0) \cdot p_y(0) = (.5)(.6) = 0.3$

$p(0,1) = p_x(0) \cdot p_y(1) = (.5)(.1) = .05$

$p(0,2) = p_x(0) \cdot p_y(2) = (.5)(.05) = .025 = p(0,3)$

$p(0,4) = p_x(0) \cdot p_y(4) = (.5)(.2) = .1$

$p(1,0) = (.3)(.6) = 0.18$, $p(1,1) = (.3)(.1) = .03$

$p(1,2) = (.3)(.05) = .015 = p(1,3)$

$p(1,4) = (.3)(.2) = .06$

$p(2,0) = (.2)(.6) = .12$, $p(2,1) = (.2)(.1) = .02$

$p(2,2) = (.2)(.05) = .01 = p(2,3)$, $p(2,4) = (.2)(.2) = .04$

$p(x,y)$		y 0	1	2	3	4
	0	.3	.05	.025	.025	.1
x	1	.18	.03	.015	.015	.06
	2	.12	.02	.01	.01	.04

b. $P(X \le 1 \text{ and } Y \le 1) = p(0,0) + p(0,1) + p(1,0) + p(1,1)$

$= .3 + .05 + .18 + .03$

$= 0.56$

$P(X \le 1) = p_X(0) + p_X(1) = 0.5 + 0.3 = 0.8$

$P(Y \le 1) = p_Y(0) + p_Y(1) = 0.6 + 0.1 = 0.7$

$P(X \le 1) \cdot P(Y \le 1) = (0.8)(0.7) = 0.56$

clearly, $P(X \le 1 \text{ and } Y \le 1) = P(X \le 1) \cdot P(Y \le 1)$

c. $P(X+Y=0) = p(0,0) = 0.3$

d. $P(X+Y \le 1) = p(0,0) + p(0,1) + p(1,0)$

$= .3 + .05 + .18 = 0.53$

9. Each front tire on a particular type of vehicle is supposed to be filled to a pressure of 26 psi. Suppose the actual air pressure in each tire is a random variable—X for the right tire and Y for the left tire, with joint pdf

$$f(x, y) = \begin{cases} K(x^2 + y^2) & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0 & \text{otherwise} \end{cases}$$

a. What is the value of K?
b. What is the probability that both tires are underfilled?
c. What is the probability that the difference in air pressure between the two tires is at most 2 psi?
d. Determine the (marginal) distribution of air pressure in the right tire alone.
e. Are X and Y independent rv's?

Ans: Let x be the actual air pressure for the right tire and y be the actual air pressure for the left tire

The joint pdf for x and y are given by

$$f(x,y) = \begin{cases} k(x^2+y^2); & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0, & \text{otherwise} \end{cases}$$

a. We know

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

$$\Rightarrow \int_{20}^{30}\int_{20}^{30} k(x^2+y^2)\,dx\,dy = 1$$

$$\Rightarrow k\int_{20}^{30}\left\{\int_{20}^{30}(x^2+y^2)\,dy\right\}dx = 1$$

$$\Rightarrow k\int_{20}^{30}\left[x^2y+\frac{y^3}{3}\right]_{20}^{30}dx = 1$$

$$\Rightarrow k\int_{20}^{30}\left[x^2\{y\}_{20}^{30}+\frac{1}{3}\{y^3\}_{20}^{30}\right]dx = 1$$

$$\Rightarrow k\int_{20}^{30}\left\{x^2(30-20)+\frac{1}{3}(30^3-20^3)\right\}dx = 1$$

$$\Rightarrow K\int_{20}^{30}\left\{10x^2+\frac{19000}{3}\right\}dx=1$$

$$\Rightarrow K\left[\frac{10x^3}{3}+\frac{19000}{3}x\right]_{20}^{30}=1$$

$$\Rightarrow K\left[\frac{10}{3}(30^3-20^3)+\frac{19000}{3}(30-20)\right]=1$$

$$\Rightarrow K\left[\frac{380000}{3}\right]=1 \Rightarrow \boxed{K=\frac{3}{380000}}$$

b. p(both tires are underfilled)

$= p(x<26 \text{ and } y<26)$

$$=\int_{20}^{26}\int_{20}^{26} f(x,y)\,dx\,dy=\int_{20}^{26}\int_{20}^{26}K(x^2+y^2)\,dx\,dy$$

$$=K\int_{20}^{26}\left\{\int_{20}^{26}(x^2+y^2)\,dy\right\}dx$$

$$=K\int_{20}^{26}\left[x^2y+\frac{y^3}{3}\right]_{20}^{26}dx$$

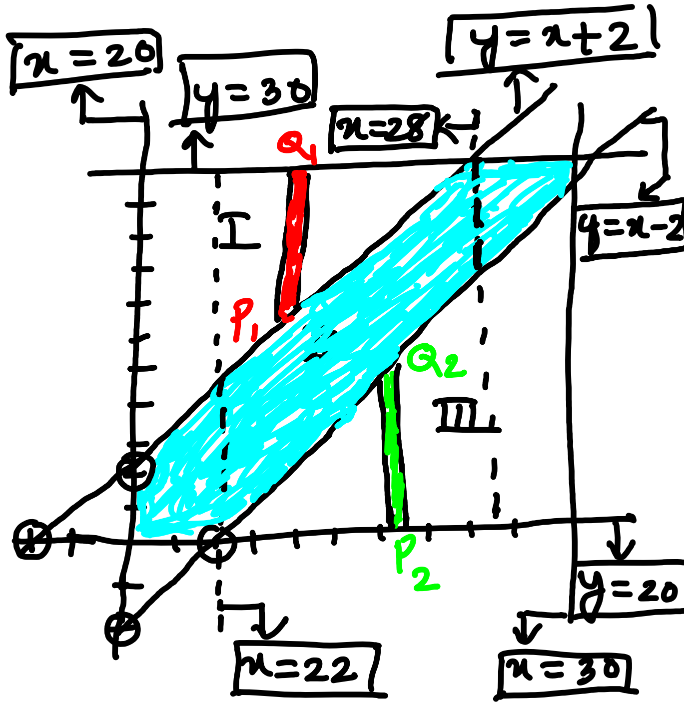
$$=K\int_{20}^{26}\left\{x^2(26-20)+\frac{1}{3}(26^3-20^3)\right\}dx$$

$$= K\int_{20}^{26}(6x^2+3192)\,dx$$

$$= K\left[2x^3+3192x\right]_{20}^{26}$$

$$= K\left[2(26^3-20^3)+3192\,(26-20)\right]$$

$$= \frac{3}{380000}\cdot 38304 = 0.3024$$



c. $P(|X-Y|\le 2)$

$$= \iint_{II} f(x,y)\,dx\,dy$$

$$= 1-\iint_{I} f(x,y)\,dx\,dy - \iint_{III} f(x,y)\,dx\,dy$$

for I

constant limits for x are

$\boxed{x=20}$ to $\boxed{x=28}$

variable limits for y.

At P_1, $y=x+2$ and at Q_1, $y=30$

so the variable limit for y are

$\boxed{y=x+2}$ to $\boxed{y=30}$

for $\underline{\text{iii}}$

constant limits for x are

$$\boxed{x = 22} \text{ to } \boxed{x = 30}$$

At P_2, $y = 20$ and at Q_2, $y = x-2$

so the variable limit for y are

$$\boxed{y = 20} \text{ to } \boxed{y = x-2}$$

$$\therefore P(|x-y| \le 2) = 1 - \int_{20}^{28} \int_{x+2}^{30} f(x,y)\,dy\,dx$$

$$- \int_{22}^{30} \int_{20}^{x-2} f(x,y)\,dy\,dx$$

$$= 1 - \int_{20}^{28} \int_{x+2}^{30} K(x^2+y^2)\,dy\,dx - \int_{22}^{30} \int_{20}^{x-2} K(x^2+y^2)\,dy\,dx$$

$$\left(K = \frac{3}{380\,000}\right)$$

$$= 1 - 0.3203 - 0.3203$$

$$= 0.3594$$

d. $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad -\infty < x < \infty$$

$$= \int_{20}^{30} K(x^2+y^2)\,dy, \quad 20 \le x \le 30$$

$$= K\left[x^2 y + \frac{y^3}{3}\right]_{20}^{30}$$

$$= k\left[x^2(30-20)+\frac{1}{3}(30^3-20^3)\right]$$

$$= 10kx^2+\frac{3}{380000}\cdot\frac{19000}{3}$$

$$= 10kx^2+.05$$

$$f_X(x)=\begin{cases}10kx^2+.05, & 20\le x\le 30\\ 0, & \text{otherwise}\end{cases}$$

e. $f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx, \quad -\infty<y<\infty$

$$= \int_{20}^{30} k(x^2+y^2)\,dx, \quad 20\le y\le 30$$

$$= k\left[\frac{x^3}{3}+xy^2\right]_{20}^{30}$$

$$= k\left[\frac{30^3-20^3}{3}+(30-20)y^2\right]$$

$$= .05+10ky^2, \quad 20\le y\le 30$$

$$f_Y(y)=\begin{cases}10ky^2+.05, & 20\le y\le 30\\ 0, & \text{otherwise.}\end{cases}$$

$f_X(x)\cdot f_Y(y)$

$$= (10kx^2+.05)\cdot(10ky^2+.05); \quad \begin{array}{l}20\le x\le 30\\ 20\le y\le 30.\end{array}$$

$\neq \quad k(x^2+y^2) = f(x,y)$

clearly, x and y are not independent.

Expected values

1. If x and y are jointly distributed rv's with pmf $p(x,y)$ then the expected value of a fn. $h(x,y)$ denoted as $E\{h(x,y)\}$ and is given by

$$E\{h(x,y)\} = \sum_x \sum_y h(x,y) \cdot p(x,y)$$

2. If x and y are jointly distributed with pdf $f(x,y)$ then the expected value of a function $h(x,y)$ is denoted as $E\{h(x,y)\}$ and is given by

$$E\{h(x,y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \cdot f(x,y)\, dx\, dy$$

covariance

If x and y are two independent rv's then the covariance of x and y, denoted as $\operatorname{cov}(x,y)$ and is given by

$$\operatorname{cov}(x,y) = E\{(x-\mu_x)(y-\mu_y)\}$$

$$= \begin{cases} \sum_x \sum_y (x-\mu_x)(y-\mu_y)\, p(x,y) \, ; & x \,\&\, y \text{ are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y)\, f(x,y)\, dx\, dy \, ; & x \,\&\, y \text{ are continuous.} \end{cases}$$

Note: 1. $cov(x, x) = E\{(x-\mu_x)^2\} = V(X) = \sigma_x^2$

2. $cov(x,y) = E(xy) - \mu_x \mu_y$

(shortcut formula to evaluate $cov(x,y)$)

Pf: $cov(x,y) = E\{(x-\mu_x)(y-\mu_y)\}$

$$= E\{xy - x\mu_y - y\mu_x + \mu_x\mu_y\}$$

$$= E(xy) - \mu_y E(x) - \mu_x E(y) + \mu_x\mu_y E(1)$$

$$= E(xy) - \mu_y\mu_x - \mu_x\mu_y + \mu_x\mu_y$$

(As $E(x) = \mu_x$, $E(y) = \mu_y$ and $E(1) = 1$)

$$= E(xy) - \mu_x\mu_y.$$

correlation

The correlation coefficient of x and y, denoted as $corr(x,y)$ (or $\rho_{x,y}$ or ρ) and is defined by

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

Note: 1. If 'a' and 'c' are positive or both negative then

$$cov(ax+b, cx+d) = cov(x,y)$$

2. For any two rv's x and y

$$-1 \le \rho_{x,y} \le 1$$

3. If x and y are independent then
$$\rho_{x,y} = 0$$

4. $\rho_{x,y} = 1$ (or $\rho_{x,y} = -1$) iff $y = ax + b$ for some nos 'a' and 'b' with $a \neq 0$.

5. $\rho_{x,y} = 0$ if x and y are uncorrelated.

22. An instructor has given a short quiz consisting of two parts. For a randomly selected student, let X = the number of points earned on the first part and Y = the number of points earned on the second part. Suppose that the joint pmf of X and Y is given in the accompanying table.

$p(x, y)$		y = 0	5	10	15
	0	.02	.06	.02	.10
x	5	.04	.15	.20	.10
	10	.01	.15	.14	.01

a. If the score recorded in the grade book is the total number of points earned on the two parts, what is the expected recorded score $E(X + Y)$?

b. If the maximum of the two scores is recorded, what is the expected recorded score?

Ans: Let $X \to$ be the no. of pts earned on the 1st part of the quiz.
$Y \to$ be the no. of pts earned on the 2nd part of the quiz.

The pmf of X & Y is give as

$p(x,y)$		y = 0	5	10	15
	0	·02	·06	·02	·10
x	5	·04	·15	·20	·10
	10	·01	·15	·14	·01

a. Let $h(x,y) = x+y$

so, $E\{h(x,y)\} = \sum_x \sum_y h(x,y) \cdot p(x,y)$

$= \sum_x \sum_y (x+y) \cdot p(x,y)$

$= (0+0)(.02) + (0+5)(.06) + (0+10)(.02) + (0+15)(.10)$
$+ (5+0)(.04) + (5+5)(.15) + (5+10)(.20) + (5+15)(.10)$
$+ (10+0)(.01) + (10+5)(.15) + (10+10)(.14) + (10+15)(.01)$

$= 14.1$

b. Let $h(x,y) = \max(x,y)$

so, $E\{h(x,y)\} = \sum_x \sum_y \max(x,y) \cdot p(x,y)$

$= 0 \cdot (.02) + 5(.06) + 10(.02) + 15(.10)$
$+ 5(.04) + 5(.15) + 10(.2) + 15(.10)$
$+ 10(.01) + 10(.15) + 10(.14) + 15(.01)$

$= 9.6$

27. Annie and Alvie have agreed to meet for lunch between noon (0:00 P.M.) and 1:00 P.M. Denote Annie's arrival time by X, Alvie's by Y, and suppose X and Y are independent with pdf's

$$f_X(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected amount of time that the one who arrives first must wait for the other person? [*Hint*: $h(X, Y) = |X - Y|$.]

Ans: Let X denote the arrival time of Annie
and Y " " " " of Alvie

Given x and y are independent with

$$f_X(x) = \begin{cases} 3x^2, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases} \qquad f_Y(y) = \begin{cases} 2y, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

we recall that if x and y are continuous and independent then

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

Thus,

$$f(x,y) = \begin{cases} 6x^2 y; & 0 \le x \le 1, \ 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Let $h(x,y) = |x-y|$

$$E\{h(x,y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)\, f(x,y)\, dx\, dy$$

$$= \int_0^1 \int_0^1 |x-y|\, 6x^2 y\, dx\, dy$$

$$|x-y| = \begin{cases} -(x-y) & \text{if } x-y \le 0 \\ x-y & \text{if } x-y > 0 \end{cases}$$

In the region I

constant limit for x are

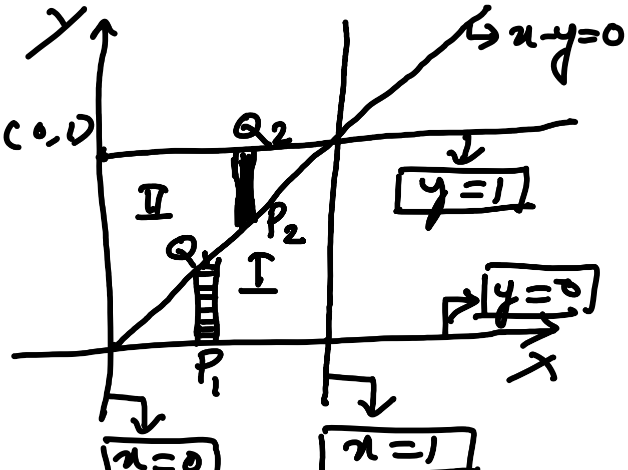
$\boxed{x=0}$ to $\boxed{x=1}$

At P_1, $y=0$; At Q_1, $y=x$

so variable limit for y are

$\boxed{y=0}$ to $\boxed{y=x}$

$|x-y| = x-y$



(Region-I represents $x-y>0$
Region-II " $x-y \le 0$

In the region – II

constant limits for x are

$\boxed{x=0}$ to $\boxed{x=1}$

At P_2, $y=x$; At Q_2, $y=1$

The variable limit for y are

$\boxed{y=x}$ to $\boxed{y=1}$

$|x-y| = y-x$

$$\therefore \int_0^1 \int_0^1 |x-y|\, 6x^2y\, dx\, dy$$

$$= \int_0^1 \int_0^x (x-y)\, 6x^2y\, dy\, dx + \int_0^1 \int_x^1 (y-x)\, 6x^2y\, dy\, dx$$

$$= \int_0^1 \left\{ \int_0^x (6x^3y - 6x^2y^2)\, dy \right\} dx + \int_0^1 \left\{ \int_x^1 (6x^2y^2 - 6x^3y)\, dy \right\} dx$$

$$= \int_0^1 \left[3x^3y^2 - 2x^2y^3 \right]_0^x dx + \int_0^1 \left[2x^2y^3 - 3x^3y^2 \right]_x^1 dx$$

$$= \int_0^1 (3x^5 - 2x^5)\, dx + \int_0^1 \{2x^2(1-x^3) - 3x^3(1-x^2)\}\, dx$$

$$= \int_0^1 x^5\, dx + \int_0^1 (2x^2 - 2x^5 - 3x^3 + 3x^5)\, dx$$

$$= \int_0^1 x^5\, dx + \int_0^1 (2x^2 - 3x^3 + x^5)\, dx$$

$$= \left.\frac{x^6}{6}\right]_0^1 + \left[\frac{2x^3}{3} - \frac{3x^4}{4} + \frac{x^6}{6}\right]_0^1$$

$$= \frac{1}{6} + \frac{2}{3} - \frac{3}{4} + \frac{1}{6} = \frac{1}{4}$$

28. Show that if X and Y are independent rv's, then $E(XY) = E(X) \cdot E(Y)$. Then apply this in Exercise 25. [*Hint*: Consider the continuous case with $f(x, y) = f_X(x) \cdot f_Y(y)$.]

Pf: If x and y are continuous & independent then

$$f(x,y) = f_x(x) \cdot f_y(y) \quad \text{——— ①}$$

Also, we know if x is a continuous rv then

$$E(x) = \mu_x = \int_{-\infty}^{\infty} x\, f_x(x)\, dx \quad \text{——— ②}$$

and $E(y) = \mu_y = \int_{-\infty}^{\infty} y\, f_y(y)\, dy$ ——— ③

if y is a cont rv.

Here, $f_x(x)$ and $f_y(y)$ represents the pdfs of x and y respectively.

Now, $E(x) \cdot E(y) = \left\{\int_{-\infty}^{\infty} x f_x(x)\, dx\right\} \cdot \left\{\int_{-\infty}^{\infty} y f_y(y) dy\right\}$

(By ② & ③)

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_x(x) f_y(y)\, dx\, dy$$

(By property of double integral)

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \, f(x,y) \, dx \, dy \quad (\text{By } ①)$$

$$= E(XY) \quad (\text{Proved})$$