HSO201A – Applied Probability & Statistics

Lecture Notes

- Prof. Partha Chakroborty

Sample Space and Events

An experiment, given certain input conditions and actions, creates an outcome.

Imagine that the outcome of this experiment cannot be predicted with certainty in advance.

Further imagine that even though the outcome of a particular experiment cannot be known in advance the set of all possible outcomes is known.

This SET of all possible outcomes is known as the SAMPLE SPACE.

A SET of possible outcomes is known as an EVENT.

So, an event is any subset of the sample space.

Sample Space and Events (Contd.)

Let S be the set that denotes the sample space.

Let E be a set that denotes an event. Note, $E \subset S$.

Let s an element of S. That is, s is a possible outcome.

If the outcome from an experiment belongs to (or is in) E, then event E is said to have occurred.

If the experiment is tossing a coin and noting the outcome of the upturned face where H means head and T means tail, then $S = \{H, T\}$.

If the experiment is tossing two coins and noting the outcomes of the upturned faces as (a,b) where a is the outcome on the first coin and b is the outcome on the second coin, then

$s = (H, H)$ or $s = (H, T)$

$$S = \{(T,T), (H,T), (T,H), (H,H)\}.$$

If an event, E is defined as upturned faces on both coins being the same, then

$$E = \{(T,T), (H,H)\}.$$

Sample Space and Events (Contd.)

If the experiment is to have a 100 m race among 4 participants, A through D, and noting down the positions in which the participants finished as (a,b,c,d) where a is the participant who finished first, b who finished second, and so on, then

$$S = \{\text{all } 4! \text{ permutations of } (A, B, C, D)\}.$$

A, B, C and D

T_1] $T_2 - T_1$
T_2] $T_3 - T_2$
T_3

$s = (B, C, A, D)$ &
$s = (D, A, B, C)$ &
$4 \times 3 \times 2 \times 1$

If the experiment is measuring the time gap (say gap) between successive vehicle arrivals at a toll booth, then the sample space contains of all positive real numbers; that is,

$$S = \{x: 0 < x < \infty\}.$$

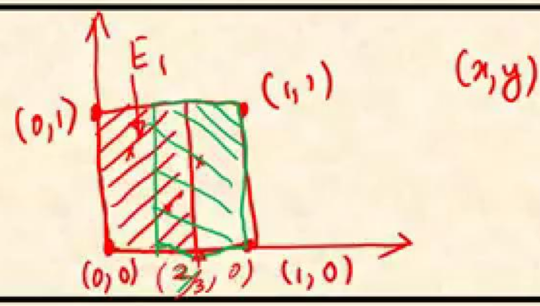
If an event, $\boldsymbol{E}$ is defined as gaps greater than 6.3 seconds, then

$$E = \{x: 6.3 < x < \infty\}.$$

Sample Space and Events (Contd.)

Imagine someone throwing darts on a unit square board with corners at $(0,0), (0,1), (1,1), (1,0)$ and noting down the coordinates (x, y) of where the dart hits the board. Then

$$S = \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

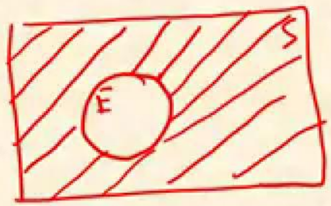
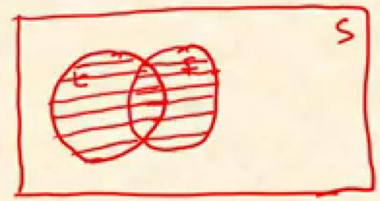
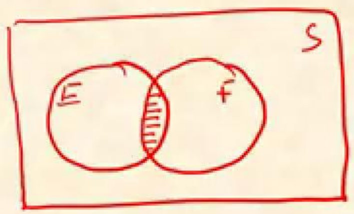
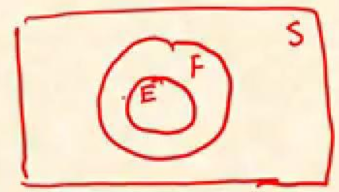


For the above situation, event E_1 is defined as the left two-thirds of the board, E_2 is defined as the right two-thirds of the board, and E_3 is defined as the middle one-third of the board. Then

$$E_1 = \left\{x: 0 \leq x \leq \frac{2}{3}, 0 \leq y \leq 1\right\} \qquad E_2 = \left\{x: \frac{1}{3} \leq x \leq 1, 0 \leq y \leq 1\right\}$$

$$E_3 = \left\{x: \frac{1}{3} \leq x \leq \frac{2}{3}, 0 \leq y \leq 1\right\}$$

Sets (Quick Recap I)

E^c Compl.		$E \cup F$ union	
$E \cap F$ Intersection		$E \subset F$	 $E \cup F = F$; $E \cap F = E$

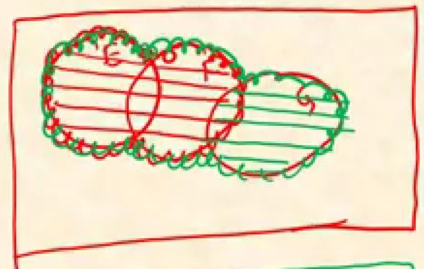
Sets (Quick Recap II)

Commutative Law $E \cup F = F \cup E$

$E \cap F = F \cap E$

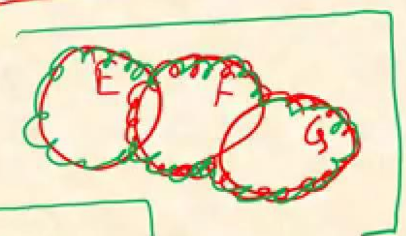
Associative Law $(E \cup F) \cup G = E \cup (F \cup G)$

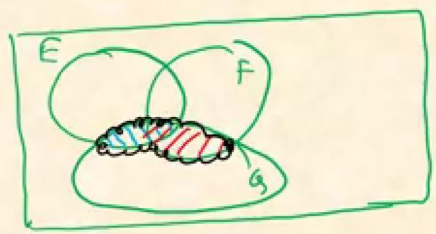
$(E \cap F) \cap G = E \cap (F \cap G)$



Distributive Law $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$

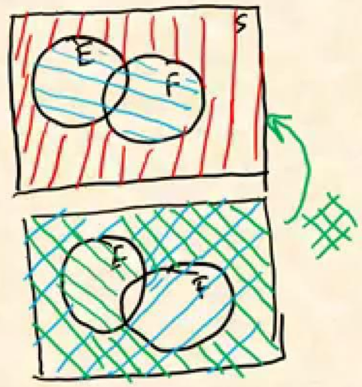
$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$





Sets (Quick Recap III)

DeMorgan's Laws $(E \cup F)^c = E^c \cap F^c$ $(E \cap F)^c = E^c \cup F^c$



DeMorgan's Laws $\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$ $\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$

$(E_1 \cup E_2 \cup E_3 \cdots \cup E_n)^c$

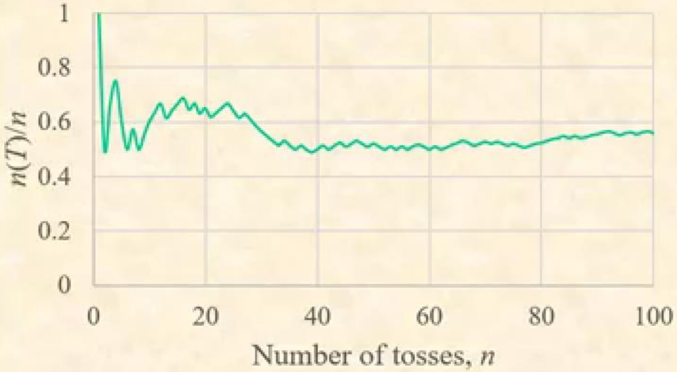
$= E_1^c \cap E_2^c \cap E_3^c \cap \cdots E_n^c$

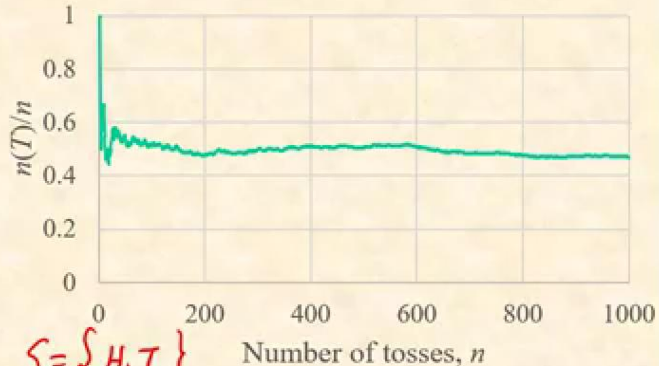
$(E_1 \cap E_2 \cap E_3 \cap \cdots \cap E_n)^c$

$= E_1^c \cup E_2^c \cup E_3^c \cup \cdots \cup E_n^c$

Probability – A definition (I)

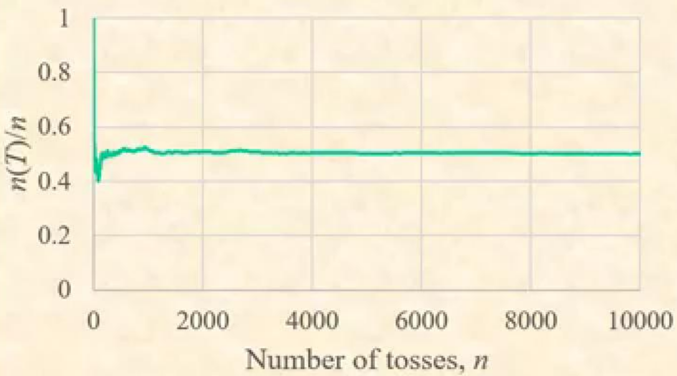
Hemachandra tosses a coin and notes down the number of times he has tossed the coin (say, n) and the number of tails he got as the outcome in the n tosses (say $n(T)$). The plot of $n(T)/n$ versus n that Hemachandra got is shown:





$S = \{H, T\}$

$E = \{T\}$



$$p = \lim_{n \to \infty} \frac{n(T)}{n} = 0.5$$

Probability – A definition (II)

Lilavati divided a circle into 6 equal pies and marked each using letters A to F. She (blindly) threw darts and noted the number of throws (*n*) and the number of times the dart hit "B" (*n*(*B*)). Plot of *n*(*B*)/n versus *n* that Lilavati got is:

$$p = \lim_{n \to \infty} \frac{n(B)}{n} = 0.167 = \frac{1}{6}$$

Probability – A Definition (III)

The number, *p* associated with an event (in earlier examples the events were "(occurrence of) Tail" or "(hitting) B") is referred to as probability that the outcome of the random experiment (tossing or throwing darts) is in the set of outcomes defining the event.

This is the relative frequency approach to defining or interpreting the probability of an event. (Note, it relies on the fact that the random experiment can be repeated, under the same condition, multiple times).

Probability can be treated also as a measure of belief. This is the interpretation that makes sense when one says, "I think the probability that Tom is guilty of lying is very high, say about 90%." Or for instance, when one says that "the probability that the events described in this ancient text are real occurrences is around 0.2."

$S = \{R, F, M\}.$

$E = \{R\}.$

Axioms

S: Sample space

E: An event

B: Collection of all events

$\emptyset$: Null set

A real valued function defined on B

Probability set function

$P(E) \xrightarrow{\text{output}}$ real value

E set

Partition

$S = \{1, 2, 3, 4, 5, 6\}.$

$E_1 = \{1, 3, 5\}. \quad E_1 \cap E_2 = \phi$

$E_2 = \{2, 4, 6\}. \quad E_1 \cup E_2 = S$

Axiom 1 $\quad P(E) \geq 0$

Axiom 2 $\quad P(S) = 1$

Axiom 3 $\quad$ **For any sequence of mutually exclusive events $E_1, E_2, \cdots$ (note, mutually exclusive events are those that do not have common outcomes; i.e., $E_i \cap E_j = \emptyset$ for all $i \neq j$)**

$F_1 = \{1, 2\}$

$F_2 = \{5, 6\}$

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

$P(F_1 \cup F_2) = P(F_1) + P(F_2)$

Some Propositions - I

P-I: $P(E^c) = 1 - P(E)$

Proof:

$S = E \cup E^c$

$E \cap E^c = \emptyset$

$\therefore P(S) = P(E) + P(E^c)$ or $1 = P(E) + P(E^c)$

$\therefore P(E^c) = 1 - P(E)$

P-II: $P(\emptyset) = 0$

Proof:

If $E = \emptyset$

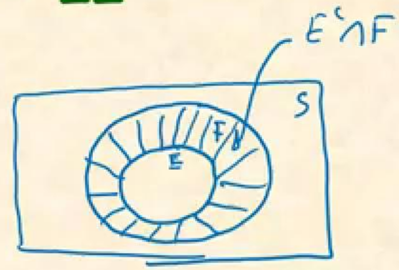
$E^c = S$

P(E)

P(E^c) $\therefore P(S) = 1 - P(\emptyset)$

$\therefore P(\emptyset) = 0$

Some Propositions - II



P-III: If $E \subset F$ then $P(E) \leq P(F)$

Proof:

$(E \cup F) = F = E \cup (E^c \cap F)$ (Since $E \subset F$)

$\therefore P(F) = P(E) + P(E^c \cap F)$ Note: $E \cap (E^c \cap F) = \emptyset$

$P(E^c \cap F) \geq 0$

$\therefore P(F) \geq P(E)$

P-IV: $P(E) \leq 1$

$[0 \leq P(E) \leq 1]$.

Proof:

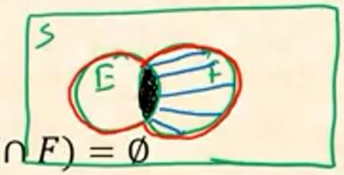
$E \subset S$

$\therefore P(E) \leq P(S)$

$\therefore P(E) \leq 1$

Some Propositions - III

P-V: $P(E \cup F) = P(E) + P(F) - P(E \cap F)$



Proof:

$E \cup F = E \cup (E^c \cap F)$ Note: $E \cap (E^c \cap F) = \emptyset$

$\therefore P(E \cup F) = P(E) + P(E^c \cap F)$

$F = (E \cap F) \cup (E^c \cap F)$

$(E \cap F) \cap (E^c \cap F) = \phi$

$\therefore P(F) = P(E \cap F) + P(E^c \cap F)$

$\therefore P(E^c \cap F) = P(F) - P(E \cap F)$

$\therefore P(E \cup F) = P(E) + P(F) - P(E \cap F)$

An aside: $P(E \cup F) \leq 1$

$\therefore P(E) + P(F) - P(E \cap F) \leq 1$

$\therefore P(E) + P(F) - 1 \leq P(E \cap F)$

Bonferroni's inequality

Some Propositions - III

P-Va: $$P(E \cup F \cup G) = \begin{cases} P(E) + P(F) + P(G) \\ -P(E \cap F) - P(E \cap G) - P(F \cap G) \\ + P(E \cap F \cap G) \end{cases}$$

sets taken one at a time

sets taken two at a time

Proof: $P(E \cup F \cup G) = P((E \cup F) \cup G)$

$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$

$$= P(E \cup F) + P(G) - P((E \cup F) \cap G)$$

$$= P(E) + P(F) - P(E \cap F) + P(G) - P((E \cap G) \cup (G \cap F))$$

$$= P(E) + P(F) - P(E \cap F) + P(G) - P(E \cap G) - P(G \cap F) + P(E \cap G \cap F)$$

$$= \underbrace{P(E) + P(F) + P(G)}_{p_1} - \underbrace{[P(E \cap F) + P(E \cap G) + P(G \cap F)]}_{p_2} + P(E \cap G \cap F)$$

$$= p_1 - p_2 + p_3$$

Inclusion-exclusion formula

By induction the following, more general, proposition can be proven:

P-Vb: $P(E_1 \cup E_2 \cup \cdots E_k) = p_1 - p_2 + p_3 - \cdots (-1)^{k+1} p_k$

Here, p_i, is the sum of probabilities of all possible intersections involving i sets.

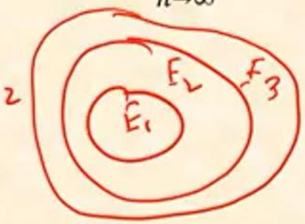
Some Propositions (A Digression)

A sequence of events $\{E_n, n \geq 1\}$ is an increasing sequence if
$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$

For such a sequence the event denoted by $\lim_{n\to\infty} E_n$ is defined as

$$\boxed{\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i}$$

$n=2$
$E_1 \cup E_2 = E_2$



$n=3$
$E_1 \cup E_2 \cup E_3 = E_3$

$$\bigcup_{i=1}^{n} E_i = E_n$$

A sequence of events $\{E_n, n \geq 1\}$ is a decreasing sequence if
$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$

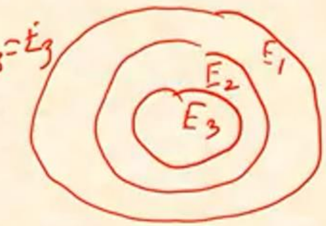
For such a sequence the event denoted by $\lim_{n\to\infty} E_n$ is defined as

$$\boxed{\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i}$$

$n=2$
$E_1 \cap E_2 = E_2$

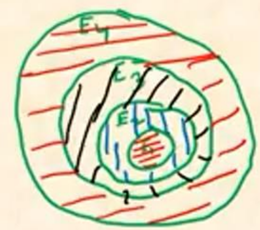
$n=3$
$E_1 \cap E_2 \cap E_3 = E_3$

$$E_n = \bigcap_{i=1}^{n} E_i$$



Some Propositions - IV

P-VI: If $\{E_n, n \geq 1\}$ is either an increasing or decreasing sequence of events, then $\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right)$



Proof:

Let $\{E_n, n \geq 1\}$ be an increasing sequence

Let $F_n, n \geq 1$ be events defined as

$$F_1 = E_1 \qquad \text{and} \qquad F_n = E_n \cap \left(\bigcup_{i=1}^{n-1} E_i\right)^c = E_n \cap E_{n-1}^c \qquad n > 1$$

From above it can be seen that F_n are mutually exclusive events that also satisfy

$$\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i \quad n \geq 1 \qquad \text{and} \qquad \bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i$$

Some Propositions - V

P-VI: (Contd.) If $\{E_n, n \geq 1\}$ is either an increasing or decreasing sequence of events, then $\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right)$

Proof (Contd.):

Recall $\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$ Hence $P\left(\lim_{n\to\infty} E_n\right) = P\left(\bigcup_{i=1}^{\infty} E_i\right)$

$$\underline{P\left(\lim_{n\to\infty} E_n\right)} = P\left(\bigcup_{i=1}^{\infty} E_i\right) = P\left(\bigcup_{i=1}^{\infty} F_i\right)$$

$$= \sum_{i=0}^{\infty} P(F_i) \quad \text{(since } F_i \text{ are mutually exclusive)}$$

$$= \lim_{n\to\infty} \sum_{i=0}^{n} P(F_i)$$

$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} F_i\right)$$

$$= \lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} E_i\right) = \lim_{n\to\infty} P(E_n) \quad \text{(since } E_1 \subset E_2 \subset \cdots \subset E_n\text{)}$$

Equally Likely Outcomes

Let O_i be an outcome. Let the sample space $S = \{O_1, O_2, \cdots O_N\}$

Let $P(O_1) = P(O_2) = \cdots = P(O_N)$ **(assumed to be equally likely outcomes)**

$F_1 = \{O_1\}, F_2 = \{O_2\} \ldots$

Think of the outcomes as events where each event has exactly one outcome. Obviously, different events have different outcomes.

$$P(O_1) + P(O_2) + \cdots + P(O_N) = P(S) = 1$$

$$P(O_i) = 1/N$$

$E = \{O_1, O_3, O_5\}$ $r = 3$

Let event *E* be one which includes *r* of these outcomes (of course *r* is less than or equal to *N*). Then,

$$P(E) = r/N$$

That is the probability of *E* occurring is the ratio of the number of ways *E* can happen (number of outcomes in *E*) to the total number of ways *S* can happen (number of outcomes in *S*).

Example 4.1

Two dice are rolled. What is the probability that the sum of the upturned faces is 7?

In order to get a sum of 7 the outcomes must be one of (1,6), (2,5), (3,4), (4,3), (5,2), and (6,1). $E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

There are 6 ways of getting 7.

In all there are 36 possible outcomes (1,1) through (6,6). These are all equally likely outcomes (assumption of a fair die). $S = \{(1,1), (1,2), \cdots, (2,1), (2,2) \cdots (6,6)\}$

$$P(\text{sum is } 7) = \frac{6}{36} = \frac{1}{6}$$

Example 4.2

Two dice are rolled. What is the probability that the sum of the upturned faces is odd? $E = \{ \underline{(a,b)} : a+b \text{ is odd} \}.$

If the outcomes are (a,b) then for a+b to be odd one of them must be odd and the other even.

That is whatever a (odd or even) is b must be the other kind.

Hence a can be any of the 6 outcomes; but whatever a is b can be only one of 3 outcomes. So total number of ways you can get the sum to be odd is $6 \times 3 = 18$

$$P(\text{sum is odd}) = \frac{18}{36} = \frac{1}{2}$$

a	b	a+b
1	1	2
1	2	3
2	1	3

It will be WRONG to do the following when the only assumption is that you have fair dice (Why?)

Possible sums are: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 (i.e., 11 possibilities)

5 of these (namely, 3, 5, 7, 9 and 11) are odd

Hence the probability that the sum is odd = 5/11

Example 4.3

Two balls are randomly drawn from a bag containing 6 white (W) and 5 black (B) balls. What is the probability that one is white and the other black? $E = \{WB, BW\}$.

Say order of draw matters. Then there are 11 ways of selecting the first ball and 10 ways of selecting the second.

In all there are $11 \times 10 = 110$ ways of getting two balls. (Assume that each of these are equally likely to be an outcome)

# of ways to get WB = $6 \times 5 = 30$; # of ways to get BW =$5 \times 6 = 30$

Total # of ways to get WB or BW = $30 + 30 = 60$. $P(\text{BW or WB}) = \dfrac{60}{110} = \dfrac{6}{11}$

Say order of draw does not matter. Then there are $\binom{11}{2} = 55$ ways of selecting two balls.

(If outcomes are equally likely when order is important then these 55 outcomes are also equally likely)

# of ways to get one W and one B = $6 \times 5 = 30$

$$P(\text{one W and one B}) = \frac{30}{55} = \frac{6}{11}$$

Example 4.4

A poker hand consists of 5 cards. A straight in poker is when all cards have consecutive values and are not from the same suit. What is the probability of being dealt a straight? (Note Ace is both a one as well as the card that comes after King).

Total # of ways you can get a hand with 5 cards (i.e, any poker hand) is $\binom{52}{5}$.

A straight can be Ace, 2, 3, 4, 5 (say type T_1), or 2, 3, 4, 5, 6 (type T_2), all the way to the one that starts with a 10, i.e., 10, Jack, Queen, King, Ace (T_{10}).

Since there are 4 cards of each value, the # of ways you can get a type $T_i = 4^5$.

Of these 4^5 ways will also include 4 ways where all the cards are from the same suit. Hence, # of ways you can get a straight = $4^5 - 4$.

# of ways you can get any of the 10 types of straights $= 10(4^5 - 4)$.

$$P(\text{straight}) = \frac{10(4^5 - 4)}{\binom{52}{5}} \approx 0.00392$$

Example 4.5

What is the probability that one player (out of 4) receives all hearts? (Note each player gets 13 cards)

Total # of ways 52 cards can be divided into 4 piles (hands) each with 13 cards is $\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{13!13!13!13!}$.

Total # of ways a particular player gets all hearts = Total # of ways that rest 39 is distributed among the other 3 players. That is in $\frac{39!}{13!13!13!}$.

2, 3, 4

H N ways

Total # of ways any player gets all hearts is $4 \times \frac{39!}{13!13!13!}$.

$$P(\text{all hearts to a player}) = \frac{4 \times 39!/(13!\,13!\,13!\,13!)}{52!/(13!\,13!\,13!\,13!)} = \frac{4 \times 39!}{52!} \approx 6.3 \times 10^{-12}$$

Example 4.6

A person (X) proposes to bet Rs. 100 (with his friend) on Y happening. Y is defined as at least two students having the same birthday in a classroom of 30 students. If Y happens then the friend will give him Rs. 100 else X will give his friend Rs. 100. Do you think X will make money in the long run (i.e., if they visit a large number of classes)? (Assume there are no leap years).

Since X wins the same amount as he loses, X will make money if

the probability of Y happening is more than 0.5

Assume there are *n* students in a classroom (obviously $n \leq 365$; why?). Each student can have a birthday in 365 ways.

Therefore, the total # of ways birthdays can be distributed in the class is $(365)^n$.

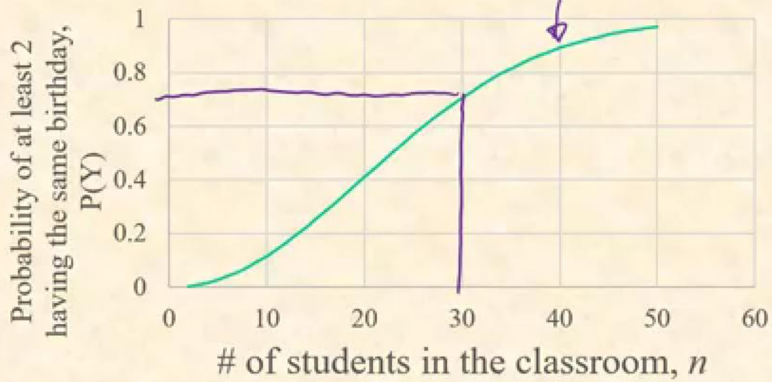
Total # of ways Y cannot happen = Total # of ways no two students have the same birthday $= 365 \times 364 \times \cdots \times (365 - (n-1))$

$$P(Y^c) = \frac{365 \times 364 \times \cdots \times (365 - (n-1))}{(365)^n}$$

$$P(Y) = 1 - P(Y^c) = 1 - \frac{365 \times 364 \times \cdots \times (365 - (n-1))}{(365)^n}$$

Example 4.6 (Contd.)

$$P(Y) = 1 - P(Y^c) = \underbrace{1 - \frac{365 \times 364 \times \cdots \times (365-(n-1))}{(365)^n}}$$



P(*Y*) at *n* = 30 is greater than 0.5. So, X will win in the long run.

Example 5.1 (Matching problem)

Each of N people at a party threw their party hats into the centre of the room. These hats were then tossed and mixed. Later each of the N went and picked up a hat from the mix. What is the probability that none selects his/her own hat?

Let E be the event that at least one person selects his/her own hat.

The event that none selects his/her own hat is E^c.

Let E_i be the event that Person i selects his/her own hat.

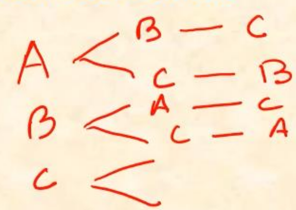
A
B
C

Therefore, $E = E_1 \cup E_2 \cup \cdots \cup E_N = \cup_{i=1}^{N} E_i$.

Recall from P-Vb: $P(E_1 \cup E_2 \cup \cdots E_N) = p_1 - p_2 + p_3 - \cdots (-1)^{N+1} p_N$

Here, p_i, is the sum of probabilities of all possible intersections involving i sets.

Since the first person can pick any of the N hats and second any of the (N-1) hats and so on this experiment can have $N!$ outcomes.



Example 5.1 (Contd.)

In order to evaluate $P(E)$ we need to evaluate p_1, p_2, p_3, etc.

$E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}$ is the event that each of the persons i_1, i_2, through i_n selects his/her own hat.

Now, after persons i_1, i_2, through i_n selects his/her own hat the rest $N - n$ persons can select their hats in $(N-n) \times \big(N-(n-1)\big) \cdots 2 \times 1 = (N-n)!$ ways.

So, the event that each of the persons i_1, i_2, through i_n selects his/her own hat can happen in $(N-n)!$ ways.

So, $P\big(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n}\big) = (N-n)!/N!$

Recall p_n is the sum of probabilities of all possible intersections involving n events (sets).

Hence, p_n has $\binom{N}{n} = \frac{N!}{(N-n)!n!}$ terms. Note, each has a probability of $(N-n)!/N!$.

$$\therefore p_n = \frac{\cancel{N!}}{\cancel{(N-n)!}\, n!} \times \frac{\cancel{(N-n)!}}{\cancel{N!}} = \frac{1}{n!}$$

Example 5.1 (Contd.)

Now, $P(E) = p_1 - p_2 + p_3 - \cdots (-1)^{N+1} p_N$

$$\therefore P(E) = 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{N+1}\frac{1}{N!}$$

Recall, required probability is $P(E^c)$.

$$\therefore P(E^c) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{N}\frac{1}{N!}$$

For large N, $P(E^c) \approx e^{-1} = \frac{1}{e}$

$$e^{-1} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \cdots$$

Example 5.2 (Matching Problem)

Each of *N* people at a party threw their party hats into the centre of the room. These hats were then tossed and mixed. Later each of the *N* went and picked up a hat from the mix. What is the probability that exactly *k* persons select their own hats?

Imagine a particular set of k persons. For only them to select their own hats mean that none from the other (N - k) persons select their own.

The total # of ways these (N - k) can select hats is (N - k)!

From the previous example, the fraction of (N - k)! ways where none of the (N - k) pick up his/her own hat can be obtained as:

$$\left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{N-k}\frac{1}{(N-k)!}\right]$$

Thus, the # of ways this particular set of k persons pick up their own hat is = the # of ways none of the other (N - k) persons pick up their own =

$$(N-k)!\left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{N-k}\frac{1}{(N-k)!}\right]$$

Example 5.2 (Contd.)

The # of ways this particular set of k persons pick up their own hat =

$$(N-k)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+(-1)^{N-k}\frac{1}{(N-k)!}\right]$$

The # of ways you can select these k persons is $\binom{N}{k}$

Therefore, the TOTAL # of ways k persons pick up their own hat =

$$\binom{N}{k}(N-k)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+(-1)^{N-k}\frac{1}{(N-k)!}\right]$$

Therefore, the probability that k persons pick up their own hat =

$$\frac{\binom{N}{k}(N-k)!\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+(-1)^{N-k}\frac{1}{(N-k)!}\right]}{N!}$$

$$\frac{\frac{N!}{k!\cancel{(N-k)!}}\times\cancel{(N-k)!}}{\cancel{N!}}$$

$$=\frac{\left[1-1+\frac{1}{2!}-\frac{1}{3!}+\cdots+(-1)^{N-k}\frac{1}{(N-k)!}\right]}{k!}$$

For large N the probability that k persons pick up their own hat $\approx \dfrac{e^{-1}}{k!}$

Example 5.3 (Runs Problem)

On a True-False question paper (QP) there are m statements that are true (T) and n statements that are false (F). What is the probability that there will be exactly r runs of Ts?

Terminology:

A run of Ts would mean a string of consecutive Ts.

For example, in a QP with 20 questions and 11 Ts and 9 Fs the following series of correct answers has 4 runs of Ts:

FTTTFFTTTTFFFFFTFTTT

The first run is of size 3, the second of size 4, the third of size 1 and the fourth of size 3.

Example 5.3 (Contd.)

Digression:

How to find the # of integer valued solutions to the equation

$$x_1 + x_2 + \cdots + x_r = n$$

This is same as asking what is the # of ways of distributing n indistinguishable balls into r urns.

When $x_i > 0 \quad i \geq 1$ (i.e., no urn can go empty).

A simple way is to think of a string of n balls which must be divided into r silos by drawing $r - 1$ walls. Note, there are $n - 1$ locations where you can draw the walls. (The number of balls in Silo i is the value of x_i.) So, say for $n = 10$ and $r = 3$ the possible solutions are:

So, # of solutions or # of ways of distributing = $\binom{n-1}{r-1}$

Example 5.3 (Contd.)

Digression:

$x_i > 0$

$\binom{n-1}{r-1}$

How to find the # of integer valued solutions to the equation

$$x_1 + x_2 + \cdots + x_r = n$$

This is same as asking what is the # of ways of distributing *n* indistinguishable balls into *r* urns.

When $x_i \geq 0 \quad i \geq 1$ (i.e., urns can go empty).

A simple way is to think of another equation in y_i where $y_i = x_i + 1$.
The corresponding equation in y_i becomes:

$$y_1 + y_2 + \cdots + y_r = n + r$$

We again have to divide the *n+r* "balls" into *r* silos.

So, # of solutions or # of ways of distributing = $\binom{n+r-1}{r-1}$

Example 5.3 (Contd.)

On a True-False question paper (QP) there are *n* statements that are true (T) and *m* statements that are false (F). What is the probability that there will be exactly *r* runs of Ts?

Imagine the runs to have size x_i, $i = 1,2,\cdots r$. That is, $x_1 + x_2 + \cdots x_r = n$.

The # of ways the above can be satisfied (note, $x_i > 0$) is $\binom{n-1}{r-1}$.

But since number of Fs between successive runs of Ts can be different; each of the $\binom{n-1}{r-1}$ run sets of Ts can occur in multiple ways.

The next task is to determine how many ways could each of these runs occur. Let's start with an example of a QP with 10 questions out of which 6 are T and 4 are F (i.e., *n* = 6, and *m* = 4).

Let's take the case of 3 runs (i.e., *r* = 3) of Ts with sizes $x_1 = 2, x_2 = 3 \text{ and } x_3 = 1$.

Examples of ways in which it can occur:

FTTFFTTTFT FFTTFTTTFT TTFFTTTFFT

Example 5.3 (Contd.)

So, in general, when in a run set of Ts there are r runs of Ts with size x_i, $i = 1,2,\cdots r$ (i.e., $x_1 + x_2 + \cdots x_r = n$) they can happen as follows:

$$\underbrace{FF\cdots F}_{y_1}\ \underbrace{TT\cdots T}_{x_1}\ \underbrace{FF\cdots F}_{y_2}\ \underbrace{T\cdots T}_{x_2}\cdots\ \underbrace{T\cdots T}_{x_r}\ \underbrace{F\cdots F}_{y_{r+1}}$$

y_r

Note, $x_i > 0,\ i = 1,,2\cdots,r;\ y_1 \geq 0;\ y_j > 0,\ \ j = 2,3,\cdots,r;\ y_{r+1} \geq 0$

So, the # of ways a particular run set of Ts can occur is the number of ways the sum y_i's can equal m

Consider the following transformation:

$y_1^* = y_1 + 1;\ y_j^* = y_j,\ \ j = 2,3,\cdots,r;\ y_{r+1}^* = y_{r+1} + 1.$

$\therefore y_1^* + y_2^* + \cdots y_{r+1}^* = m + 2$

So, the # of ways a particular run set of Ts can occur is the number of ways the sum y_i^* 's can equal m+2 $= \binom{(m+2)-1}{(r+1)-1} = \binom{m-1}{r}$

Example 5.3 (Contd.)

Recall,

The # of ways a particular run set of Ts can occur is $\binom{m-1}{r}$

The # of run sets with r runs of Ts is $\binom{n-1}{r-1}$.

Therefore, the total # of ways r runs of Ts can occur is $\binom{n-1}{r-1}\binom{m-1}{r}$.

The total # of ways n Ts and m Fs can be written in a string is $\binom{n+m}{n}$.

$$P(r \text{ runs of Ts}) = \frac{\binom{n-1}{r-1}\binom{m-1}{r}}{\binom{n+m}{n}}.$$

Example 5.4 (Runs Problem - A Possible Use)

In a 12-month period daily traffic on a road (daily demand for coffee in a cafe) is estimated from its annual average as 20000 vehicles per day (50 cups per day). Over the year it is seen that 5 times this is an overestimate (that is the error is positive) and 7 times it is an underestimate (error is negative). Do you think there is some seasonal pattern in the demand if there are 2 runs of overestimates?

5 O's ; 7 U's

If there is no seasonal pattern, then all $\binom{5+7}{5} = 792$ outcomes of stringing Os and Us are equally likely.

O O O U U O U O U U O U

U U O O U U O U O U O U

$$\frac{\binom{n-1}{r-1}\binom{m-1}{r}}{\binom{n+m}{n}}$$

$$P(2 \text{ runs of Os}) = \frac{\binom{5-1}{1}\binom{7-1}{2}}{792} = \frac{5}{66} \approx 0.076$$

2 runs is highly unlikely under the assumption that all 792 outcomes are equally likely. Yet we got 2 runs.

Conditional Probability

Two "fair" dice are rolled. Let the number on left die be a and the right be b. The outcome of a roll is written as (a,b). Do you think the answers to the following questions are same?

(i) What is the probability of getting (3,6)?

(ii) What is the probability of getting (3,6) if it is also known that the sum of the upturned faces (i.e., $a + b$) is 9?

(iii) What is the probability of getting (3,6) if it is also known that the sum of the upturned faces (i.e., $a + b$) is 12?

(iv) What is the probability of getting $a = 3$?

(v) What is the probability of getting $a = 3$ if it is also known that $b = 4$?

Conditional Probability II

Let the probability of event *E* happening when it is known (or when it is given that) event *F* has happened be represented as: $\boxed{P(E|F)}$.

Let $E = \{(3,6)\}$

Let $F = \{(3,6), (4,5), (5,4), (6,3)\}$

$P(E|F) = ?$ $= \frac{1}{4}$

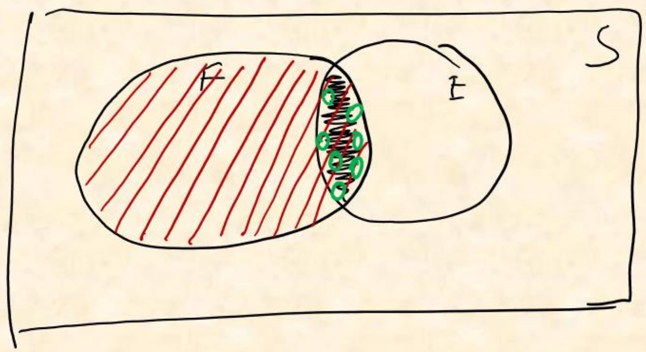
Given that *F* has happened for *E* to happen both *E* and *F* (i.e., $E \cap F$) has to happen.

The question then is what are the chances of $E \cap F$ happening given that the sample space is now *F*.

Conditional Probability (Definition)

The question then is what are the chances of $E \cap F$ happening given that the sample space is now *F*. (Note since *F* has happened implies *P*(*F*) cannot be zero; the following definition is only valid when $P(F) > 0$)

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$



From the definition of conditional probability, the multiplication rule of probabilities readily follows:

$$P(E \cap F) = P(F) \times P(E|F)$$

Example 6.1

A coin is flipped twice. What is the probability that both tosses result in heads given that the first is a head?

Assuming all four outcomes (H,H), (H,T), (T,H), and (T,T) are equally likely.

Let $E = \{(H,H)\}$ and $F = \{(H,H),(H,T)\}$

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

$$E \cap F = E$$

$$P(E \cap F) = P(E) = \frac{1}{4}$$

$$P(F) = \frac{1}{2}$$

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{1/4}{1/2} = \frac{1}{2}$$

Example 6.2

A coin was flipped twice. You are only told that at least one toss resulted in a head. What is the probability that both tosses resulted in heads?

Assuming all four outcomes (H,H), (H,T), (T,H), and (T,T) are equally likely.

Let $E = \{(H,H)\}$ and $F = \{(H,H),(H,T),(T,H)\}$

HH
HT
TH

Tail is twice as likely than head.

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

$$E \cap F = E$$

$$P(E \cap F) = P(E) = \frac{1}{4}$$

$$P(F) = \frac{3}{4}$$

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{1/4}{3/4} = \frac{1}{3}$$

Example 6.3

Rala has to take an open elective course next semester. He can either take a course on Ocean Life (O) or a course on Volcanoes (V). Based on what his seniors told him he believes the probability of him getting an A in O is 1/3 and getting an A in V is 1/2. The probability of Rala taking V next semester is 1/4. What is the probability that Rala gets an A in V next semester?

Let A be the event that Rala gets a grade A in open elective

Let V be the event that Rala takes V as open elective

Let E be the event that Rala gets grade A in V. Note $E = A \cap V$.

$$P(E) = P(A \cap V) = P(A|V)P(V)$$

$$\text{Now, } P(V) = \frac{1}{4} \text{ and } P(A|V) = \frac{1}{2}$$

$$\therefore P(E) = P(A \cap V) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$$

Example 6.4

Rala has to take one open elective course next semester. He can either take a course on Ocean Life (O) or a course on Volcanoes (V). Based on what his seniors told him he believes the probability of him getting an A in O is 1/3 and getting an A in V is 1/2. The probability of Rala taking V next semester is 1/4. What is the probability that Rala gets an A in the open elective next semester?

Let A be the event that Rala gets a grade A in open elective

$E \cap F = \phi$

Let V be the event that Rala takes V as open elective

Let O be the event that Rala takes O as open elective

Let E be the event that Rala gets grade A in V. Note $E = A \cap V$.

Let F be the event that Rala gets grade A in O. Note $F = A \cap O$.

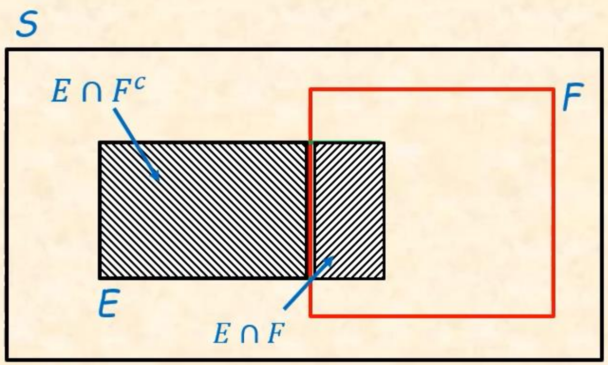
Event A can occur if Rala gets either grade A in V (Event E) or he gets grade A in O. (Note he cannot get As in both the courses since he can only take one of them.

$$\therefore P(A) = P(E) + P(F) = P(A \cap V) + P(A \cap O)$$

$$\therefore P(A) = P(A|V)P(V) + P(A|O)P(O)$$

$$\therefore P(A) = \frac{1}{2} \times \frac{1}{4} + \frac{1}{3} \times \frac{3}{4} = \frac{3}{8}$$

Law of Total Probability I



$$\overbrace{E = (E \cap F) \cup (E \cap F^c)}^{\sim\sim\sim \; + \; \sim\sim\sim}$$

$$(E \cap F) \cap (E \cap F^c) = \emptyset$$

$$P(E) = P(E \cap F) + P(E \cap F^c)$$

$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

because $P(E \cap F) = P(F) \times P(E|F)$

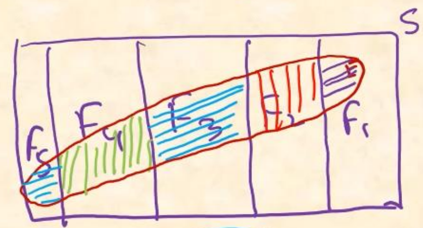
$$= P(E|F)P(F) + P(E|F^c)(1 - P(F))$$

Law of Total Probability II

The previous equation can be generalized thus:

Let $F_1, F_2 \cdots F_n$ be mutually exclusive and collectively exhaustive sets (i.e., $\bigcup_{i=1}^{n} F_i = S$).

$$P(E) = P(E \cap F_1) + P(E \cap F_2) + \cdots \; \cdots P(E \cap F_5)$$



Note:

$$E = \bigcup_{i=1}^{n} (E \cap F_i)$$

Also, $(E \cap F_i)$ for $i = 1, 2, \cdots, n$ are mutually exclusive (why?)

$$E = \bigcup_{i=1}^{n} (E \cap F_i)$$

$$\therefore P(E) = \sum_{i=1}^{n} P(E \cap F_i)$$

$$= \sum_{i=1}^{n} P(E|F_i)\, P(F_i)$$

Example 6.5

An insurance company believes that 20% of the population has a higher risk of accident (say, H is the event that a person who buys a policy from this company is in the high-risk group). Probability that people in this group will have an accident in a one-year period is 0.5; for others this probability is 0.15. What is the probability that a person who buys the policy will have an accident within one year of buying the insurance policy?

Let A be the event that a policy holder will have an accident within a year of purchasing the policy.

$$P(A) = \overbrace{P(A|H)}^{0.5}\overset{0.2}{P(H)} + \overbrace{P(A|H^c)}^{0.15}P(H^c)$$

$$\therefore P(A) = 0.5 \times 0.2 + 0.15 \times 0.8 = 0.22$$

Example 6.6

An insurance company believes that 20% of the population has a higher risk of accident (say, *H* is the event that a person who buys a policy from this company is in the high-risk group). Probability that people in this group will have an accident in a one-year period is 0.5; for others this probability is 0.15. What is the probability that a person who had an accident within one year of buying the insurance policy belongs to the higher risk group?

Let *A* be the event that a policy holder will have an accident within a year of purchasing the policy.

$P(H|A) = \frac{P(H \cap A)}{P(A)}$

The required probability here is: $P(H|A)$

$P(A \cap H) = P(A|H) . P(H)$

We know the following: $\underline{P(H) = 0.2}$; $P(A|H) \underline{= 0.5}$; $P(A|H^c) = 0.15$

$\underline{P(A) = 0.22}$ (from previous example)

$P(A) = P(A|H) . P(H) + P(A|H^c) . P(H^c)$

$$P(H|A) = \frac{P(H \cap A)}{P(A)} = \frac{P(A \cap H)}{P(A)} = \frac{P(A|H)P(H)}{P(A)}$$

$$= \frac{0.5 \times 0.2}{0.22} \approx 0.45$$

Example 6.7

Like with any examination the purpose is to identify those who know the subject. For any question (in a multiple-choice question paper) a student either guesses or knows the answer. Assume students who guess has an equal probability of choosing any one of the *n* options. Let *p* be the probability that a student knows the answer. An important question for the instructor always is: what is the probability that a student who answered a question correctly actually knew the answer?

Let C be the event that the student answers correctly

Let K be the event that the student actually knows the answer

The required probability here is: $P(K|C)$

We know the following: $P(K) = p$; $\boxed{P(C|K^c) = 1/n}$; $\boxed{P(C|K) = 1}$

$$P(K|C) = \frac{P(K \cap C)}{P(C)} = \frac{P(C \cap K)}{P(C)} = \frac{P(C|K)P(K)}{P(C)} = \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)}$$

$$= \frac{1 \times p}{(1 \times p) + (\frac{1}{n})(1-p)} = \frac{p}{p + (\frac{1}{n})(1-p)}$$

Example 6.8

Dala is a suspect in a murder investigation. Based on the circumstances and some evidence the detective is 60% certain that Dala is guilty of murder. The autopsy now reveals that the murderer is left-handed. It is known that 20% of the population is left-handed and Dala is too. The detective generally charges a suspect with a crime if she is at least 80% certain that the suspect is guilty. Should the detective charge Dala?

Let G be the event that Dala is guilty

Let L be the event that Dala is left-handed

The required probability here is: $P(G|L)$

We know the following: $P(G) = 0.6$; $P(L|G) = 1$; $P(L|G^c) = 0.2$

$$P(G|L) = \frac{P(G \cap L)}{P(L)} = \frac{P(L \cap G)}{P(L)} = \frac{P(L|G)P(G)}{P(L)} = \frac{P(L|G)P(G)}{P(L|G)P(G) + P(L|G^c)P(G^c)}$$

$$= \frac{1 \times 0.6}{(1 \times 0.6) + (0.2)(1 - 0.6)} \approx 0.88$$

Example 6.9

A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability 0.01, the test result will imply that he or she has the disease.) If 0.5 percent of the population actually has the disease, what is the probability that a person has the disease given that the test result is positive? (Assume testing is done at random and irrespective of whether you have any symptom)

Let D be the event that a person (tested) has the disease

Let E be the event that the test result is positive

$P(D) = 0.005, \qquad P(E|D) = 0.95, \qquad P(E|D^c) = 0.01$

$$P(D|E) = \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)}$$

$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} \approx 0.323$$

Bayes' Theorem

Let $F_1, F_2 \cdots F_n$ be mutually exclusive and collectively exhaustive sets/events (i.e., $\bigcup_{i=1}^{n} F_i = S$).

Let E be another event.

Recall $P(E) = \sum_{i=1}^{n} P(E|F_i)\, P(F_i)$

$$P(F_k|E) = \frac{P(F_k \cap E)}{P(E)} = \frac{P(E \cap F_k)}{P(E)} = \frac{P(E|F_k)P(F_k)}{P(E)}$$

$$P(F_k|E) = \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

Example 7.1 (Search Problem)

A plane went missing in a remote, difficult-to-access, reasonably big area. It is presumed to have crashed. This big area is divided into 4 parts; i = 1, 2, 3, 4. Given the thick forest in the area the probability of missing the debris is m_i. What is the probability that the plane is in the i^{th} region given that a search of Region 1 could not locate the plane?

Let $D_i, i = 1,2,3,4$ be the event that the plane (debris) is in Region i

Let U be the event that search of Region 1 was unsuccessful

$P(D_1|U) = \frac{P(U \cap D_1)}{P(U)}$

Assuming $P(D_i) = \frac{1}{4}$

Also note, $P(U|D_1) = m_1$, and for $i = 2,3,4$ $P(U|D_i) = 1$

$P(U \cap D_1) = P(U|D_1).P(D_1)$

For i = 1, the required probability is $P(D_1|U)$

$$P(D_1|U) = \frac{P(U|D_1)P(D_1)}{P(U|D_1)P(D_1) + P(U|D_2)P(D_2) + P(U|D_3)P(D_3) + P(U|D_4)P(D_4)}$$

$$P(D_1|U) = \frac{m_1 \times \frac{1}{4}}{m_1 \times \frac{1}{4} + 1 \times \frac{1}{4} + 1 \times \frac{1}{4} + 1 \times \frac{1}{4}} = \frac{m_1}{m_1 + 3}$$

Example 7.1 (Search Problem – Contd.)

Assuming $P(D_i) = \frac{1}{4}$

Also note, $P(U|D_1) = m_1$, and for $i = 2,3,4$ $P(U|D_i) = 1$

For k = 2, 3, 4, the required probability

$$P(D_k|U) = \frac{P(U|D_k)P(D_k)}{P(U|D_1)P(D_1) + P(U|D_2)P(D_2) + P(U|D_3)P(D_3) + P(U|D_4)P(D_4)}$$

$$P(D_k|U) = \frac{1 \times \frac{1}{4}}{m_1 \times \frac{1}{4} + 1 \times \frac{1}{4} + 1 \times \frac{1}{4} + 1 \times \frac{1}{4}} = \frac{1}{m_1 + 3}$$

$$P(D_k|U) = \begin{cases} \frac{m_1}{m_1+3} & k=1 \\ \frac{1}{m_1+3} & k=2,3,4 \end{cases}$$

Example 7.1 (Search Problem – Contd.)

Some observations:

$P(D_k|U)$

$P(D_k)$

1. $$\sum_{i=1}^{4} P(D_i|U) = \frac{m_1}{m_1+3} + 3 \times \frac{1}{m_1+3} = 1$$

2. For k = 1, the posterior probability is never greater than the prior probability; i.e., $P(D_1|U) \leq P(D_1)$ (Why?)

$$P(D_1|U) = \frac{m_1}{m_1+3} = \boxed{\frac{1}{1+\frac{3}{m_1}}}$$

3. For k = 1, the posterior probability increases as m_1 increases (should it?) and becomes equal to the prior probability when m_1 = 1 (should it?)

4. For k = 2,3,4, the posterior probability is never lesser than the prior probability; i.e., $P(D_k|U) \geq P(D_k)$ (Why?)

5. For k = 2,3,4, the posterior probability dereases as m_1 increases (should it?) and becomes equal to the prior probability when m_1 = 1 (should it?)

Example 7.2

A loan department of a bank becomes very busy during the pre-festival season. Only 60% of the potential customers (loan seekers) get to talk to bank personnel when they call. The rest 40% leave their telephone numbers. 75% of these get a call from the bank staff on the same day while the rest get a call the next day. Experience shows that (i) 80% of those who get to speak to a staff immediately, (ii) 60% of those whose calls get returned the same day, and (iii) 40% of those whose calls get returned the next day finally come to the bank seeking loan. (a) What proportion of people who call for a loan come to the bank seeking a loan? (b) What percentage of callers who visited the bank had their calls returned immediately?

Let C be the event that a caller comes to the bank seeking loan

Let T_i be the event that a caller gets to talk to a staff immediately.

Let T_s be the event that a caller gets to talk to a staff on the same day

Let T_n be the event that a caller gets to talk to a staff on the next day

$P(T_i) = 0.6,\qquad P(T_s) = 0.4 \times 0.75 = 0.3,\qquad P(T_n) = 0.4 \times 0.25 = 0.1$

$P(C|T_i) = 0.8,\qquad P(C|T_s) = 0.6,\qquad P(C|T_n) = 0.4$

Example 7.2 (Contd.)

Let C be the event that someone comes to the bank seeking loan

Let T_i be the event that a person gets to talk to a staff immediately.

Let T_s be the event that a person gets to talk to a staff on the same day

Let T_n be the event that a person gets to talk to a staff on the next day

$P(T_i) = 0.6,$ $\quad P(T_s) = 0.4 \times 0.75 = 0.3,$ $\quad P(T_n) = 0.4 \times 0.25 = 0.1$

$P(C|T_i) = 0.8,$ $\quad P(C|T_s) = 0.6,$ $\quad P(C|T_n) = 0.4$

(a)

Note T_i, T_s and T_n are mutually exclusive but collectively exhaustive sets

$$\therefore P(C) = P(C|T_i)P(T_i) + P(C|T_s)P(T_s) + P(C|T_n)P(T_n)$$

$$= 0.8 \times 0.6 + 0.6 \times 0.3 + 0.4 \times 0.1 = 0.7$$

(b)

$$\therefore P(T_i|C) = \frac{P(C|T_i)P(T_i)}{P(C)} = \frac{0.8 \times 0.6}{0.7} \approx 0.686$$

Example 7.2

A new couple, known to have two children, has just moved into town. Suppose that the mother is encountered walking with one of her children. If this child is a girl, what is the probability that both children are girls?

Let G_o , G_y and G, respectively, be the events that the oldest child is a girl, the youngest is a girl and the child seen with the mother is a girl.

Let B_o , B_y and B, respectively, be the events that the oldest child is a boy, the youngest is a boy and the child seen with the mother is a boy.

$S = \{(G_o, G_y), (G_o, B_y), (B_o, G_y), (B_o, B_y)\}$

Required probability is: $P(\overbrace{G_o, G_y}^{E = \{(G_o, G_y)\}}|G)$

$$P(G_o, G_y|G) = \frac{P(G|G_o, G_y)P(G_o, G_y)}{P(G)}$$

$$= \frac{P(G|G_o, G_y)P(G_o, G_y)}{P(G|G_o, G_y)P(G_o, G_y) + P(G|G_o, B_y)P(G_o, B_y) + P(G|B_o, G_y)P(B_o, G_y) + P(G|B_o, B_y)P(B_o, B_y)}$$

$$P(G_o, G_y) = P(G_o, B_y) = P(B_o, G_y) = P(B_o, B_y) = \frac{1}{4}$$

$P(G|G_o, G_y) = 1.0,\quad P(G|B_o, B_y) = 0.0,\quad P(G|G_o, B_y) = ?,\quad P(G|B_o, G_y) = ?$

Example 7.2 (Contd.)

$$P(G_o, G_y) = P(G_o, B_y) = P(B_o, G_y) = P(B_o, B_y) = \frac{1}{4}$$

$$P(G|G_o, G_y) = 1.0, \quad P(G|B_o, B_y) = 0.0, \quad P(G|G_o, B_y) = ?, \quad P(G|B_o, G_y) = ?$$

$$P(G_o, G_y|G) = \frac{P(G|G_o, G_y)P(G_o, G_y)}{P(G)}$$

$$= \frac{P(G|G_o, G_y)P(G_o, G_y)}{P(G|G_o, G_y)P(G_o, G_y) + P(G|G_o, B_y)P(G_o, B_y) + P(G|B_o, G_y)P(B_o, G_y) + P(G|B_o, B_y)P(B_o, B_y)}$$

$$= \frac{\frac{1}{4}}{\frac{1}{4} + \frac{P(G|G_o, B_y)}{4} + \frac{P(G|B_o, G_y)}{4} + 0} = \frac{1}{1 + P(G|G_o, B_y) + P(G|B_o, G_y)}$$

If we are further told that the probability that the mother will choose a girl to walk with (whenever there is such a choice) is q, then

$$P(G_o, G_y|G) = \frac{1}{1 + q + q} = \frac{1}{1 + 2q}$$

(What happens if $q = 1$? Have you seen a similar problem earlier?)

$P(\cdot\,|F)$ is a Probability

The task here is to show that $\underline{\underline{P(E|F)}}$ satisfy all the three axioms of probability.

$P(\cdot)$ • $E|F$

Property 1: $P(E|F) \geq 0$

Since $P(E \cap F) \geq 0$ and $P(F) > 0$, $P(E|F) = \frac{P(E \cap F)}{P(F)} \geq 0$

Additionally note $E \cap F \subset F$ and $P(E \cap F) \leq P(F)$, $P(E|F) = \frac{P(E \cap F)}{P(F)} \leq 1$

Property 2: $P(S|F) = 1$

Since $S \cap F = F$, $P(S \cap F) = P(F)$; hence, $P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$

Additionally note $F \cap F = F$, therefore, $P(F|F) = \frac{P(F)}{P(F)} = 1$

$P(\cdot\,|F)$ is a Probability (Contd.)

Property 3: If $E_1, E_2, \cdots$ events are mutually exclusive then $P(\cup_1^\infty E_i\,|F) = \sum_1^\infty P(E_i|F)$

$$P(\cup_1^\infty E_i\,|F) = \frac{P\big((\cup_1^\infty E_i) \cap F\big)}{P(F)}$$

Now $(E_1 \cup E_2 \cup \cdots) \cap F = (E_1 \cap F) \cup (E_2 \cap F) \cup \cdots = \cup_1^\infty (E_i \cap F)$

$$\text{So}, P(\cup_1^\infty E_i\,|F) = \frac{P\big((\cup_1^\infty E_i \cap F)\big)}{P(F)}$$

Since $E_1, E_2, \cdots$ are mutually exclusive so are $E_1 \cap F, E_2 \cap F, \cdots$

$$\text{Hence}, P(\cup_1^\infty E_i\,|F) = \frac{P\big((\cup_1^\infty E_i \cap F)\big)}{P(F)} = \frac{\sum_1^\infty P(E_i \cap F)}{P(F)}$$

$$\text{Note}, P(E_i \cap F) = P(E_i|F) \times P(F)$$

$$\text{Therefore}, P(\cup_1^\infty E_i\,|F) = \frac{\sum_1^\infty P(E_i \cap F)}{P(F)} = \frac{P(F)\sum_1^\infty P(E_i|F)}{P(F)}$$

$$\text{Therefore}, P(\cup_1^\infty E_i\,|F) = \sum_1^\infty P(E_i|F)$$

$P(\cdot\,|F)$ is a Probability (Contd.)

If we define $Q(E) = P(E|F)$ then $Q(E)$ is a probability (set) function. So, all properties of $P(E)$ (derived earlier or otherwise) are true for $Q(E)$. For example,

$Q(E^c) = 1 - Q(E)$; that is, $P(E^c|F) = 1 - P(E|F)$

$Q(E_1 \cup E_2) = Q(E_1) + Q(E_2) - Q(E_1 \cap E_2)$, that is
$P((E_1 \cup E_2)|F) = P(E_1|F) + P(E_2|F) - P((E_1 \cap E_2)|F)$

If we define $Q(E_1|E_2) = Q(E_1 \cap E_2)/Q(E_2)$

then $Q(E_1) = Q(E_1|E_2)Q(E_2) + Q(E_1|E_2^c)Q(E_2^c)$

or,
$P(E_1|F) = P(E_1|(E_2 \cap F))P(E_2|F) + P(E_1|(E_2^c \cap F))P(E_2^c|F)$

Note,
$$Q(E_1|E_2) = \frac{Q(E_1 \cap E_2)}{Q(E_2)}$$
$$= \frac{P((E_1 \cap E_2)|F)}{P(E_2|F)}$$
$$= \frac{P(E_1 \cap E_2 \cap F)/P(F)}{P(E_2 \cap F)/P(F)}$$
$$= \frac{P(E_1 \cap (E_2 \cap F))}{P(E_2 \cap F)}$$
$$= P(E_1|(E_2 \cap F))$$

Example 7.3

An insurance company believes that 20% of the population has a higher risk of accident (say, *H* is the event that a person who buys a policy from this company is in the high-risk group). Probability that people in this group will have an accident in a one-year period is 0.5; for others, this probability is 0.15. What is the probability that a person who buys the policy will have an accident in the second year given that he/she has had an accident in the first year after buying the insurance policy?

Let *A* be the event that a policy holder will have an accident within a year of purchasing the policy.

Let *A2* be the event that a policy holder will have an accident in the second year after purchasing the policy.

Required probability is $P(A2|A)$

Before proceeding further let's recall Examples 6.5 and 6.6 in the next slide.

Example 6.5 (Repeated for ready reference)

An insurance company believes that 20% of the population has a higher risk of accident (say, H is the event that a person who buys a policy from this company is in the high-risk group). Probability that people in this group will have an accident in a one-year period is 0.5; for others this probability is 0.15. What is the probability that a person who buys the policy will have an accident within one year of buying the insurance policy?

Let A be the event that a policy holder will have an accident within a year of purchasing the policy.

$$P(A) = P(A|H)P(H) + P(A|H^c)P(H^c) \quad \therefore P(A) = 0.5 \times 0.2 + 0.15 \times 0.8 = 0.22$$

Example 6.6 (Repeated for ready reference)

An insurance company believes that 20% of the population has a higher risk of accident (say, H is the event that a person who buys a policy from this company is in the high-risk group). Probability that people in this group will have an accident in a one-year period is 0.5; for others this probability is 0.15. What is the probability that a person who had an accident within one year of buying the insurance policy belongs to the higher risk group?

The required probability here is: $P(H|A)$

We know the following: $P(H) = 0.2;\ P(A|H) = 0.5; P(A|H^c) = 0.15$

$P(A) = 0.22$ (from previous example)

$$P(H|A) = \frac{P(H \cap A)}{P(A)} = \frac{P(A|H)P(H)}{P(A)} = \frac{0.5 \times 0.2}{0.22} \approx 0.45$$

Example 7.3 (Going back)

An insurance company believes that 20% of the population has a higher risk of accident (say, H is the event that a person who buys a policy from this company is in the high-risk group). Probability that people in this group will have an accident in a one-year period is 0.5; for others, this probability is 0.15. What is the probability that a person who buys the policy will have an accident in the second year given that he/she has had an accident in the first year after buying the insurance policy?

A is the event that a policy holder will have an accident within a year of purchasing the policy and A2 is the event that a policy holder will have an accident in the second year after purchasing the policy.

Required probability is $P(A2|A)$

$$P(E_1|F) = P(E_1|E_2 \cap F) \cdot P(E_2|F) + P(E_1|E_2^c \cap F) \cdot P(E_2^c|F)$$

We know the following: $P(H) = 0.2;\ P(A|H) = 0.5; P(A|H^c) = 0.15$

$P(A) = 0.22; P(H|A) = 0.45; P(H^c|A) = 0.55$

The probability of having an accident in a one-year period is only dependent on whether a person is accident prone. That is,

$P(A2|(H \cap A)) = P(A2|H) = 0.5;\ P(A2|(H^c \cap A)) = P(A2|H^c) = 0.15$

E_1 F (We will return to this idea while discussing conditional independence later.)

E_2

$P(A2|A) = P(A2|(H \cap A))P(H|A) + P(A2|(H^c \cap A))P(H^c|A)$

$P(A2|A) = 0.5 \times 0.45 + 0.15 \times 0.55 \approx 0.31$

Independent Events

There may be cases where the information that F has happened has no impact on the chances of E happening. That is, $P(E|F) = P(E)$.

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = P(E)$$

$$\therefore P(E \cap F) = P(E) \times P(F)$$

Two events E and F are said to be independent if the left equation holds.

If E and F are independent, then so are (i) E and F^c, *(ii)* E^c and F, and (iii) E^c and F^c .

$P(E \cap F') = P(E) \cdot P(F')$

We will show (i) here and leave the rest as exercise.

$E = (E \cap F) \cup (E \cap F^c)$

Since, $(E \cap F)$ and $(E \cap F^c)$ are mutually exclusive

$P(E) = P(E \cap F) + P(E \cap F^c)$

$P(E) = P(E)P(F) + P(E \cap F^c)$

$P(E \cap F^c) = P(E) - P(E)P(F) = P(E)[1 - P(F)] = P(E)P(F^c)$

Example 9.1

Two fair dice are rolled. Three events are defined as follows: (i) E_1 is the event that the sum of the upturned faces is 6, (ii) E_2 is the event that the sum is 7 and (iii) F is the event that the outcome on the first die is 4. Are (a) E_1 and F independent? (b) E_2 and F independent?

Sum=6 E_1: $\begin{cases} 1,5 \\ 2,4 \\ 3,3 \\ \boxed{4,2} \\ 5,1 \end{cases}$

Sum=7 E_2: $\begin{cases} 1,6 \\ 2,5 \\ 3,4 \\ \boxed{4,3} \\ 5,2 \\ 6,1 \end{cases}$

✓

$$P(E_1) = \frac{5}{36}; \qquad P(E_2) = \frac{6}{36}; \qquad P(F) = \frac{1}{6}$$

$$P(E_1 \cap F) = \frac{1}{36}; \qquad P(E_2 \cap F) = \frac{1}{36}$$

$$\overset{1/36}{P(E_1 \cap F)} \neq \overset{5/36}{P(E_1)} \times \overset{1/6}{P(F)}$$

Hence, E_1 and F are not independent.

$$\overset{1/36}{P(E_2 \cap F)} = \overset{6/36}{P(E_2)} \times \overset{1/6}{P(F)}$$

Hence, E_2 and F are independent.

Example 9.2

Two fair dice are rolled. Three events are defined as follows: (i) E is the event that the sum of the upturned faces is 7, (ii) F is the event that the outcome on the first die is 4 and (iii) G is the event that the outcome on the second die is 3. Are (a) E and F independent? (b) E and G independent, (c) E and $(F \cap G)$ independent ?

First die is 4 AND second die is 3.

From previous example, we know E and F are independent.

From previous example, we also know E and G are independent.

But note, if both F and G occur then I can say with certainty that E occurs. That is, $P(E|(F \cap G)) = 1$. Where as, $P(E) = \frac{6}{36}$.

Hence, E and $(F \cap G)$ are NOT independent.

Notice, the question whether any three events E, F, and G are independent cannot be answered by just checking whether all the $\binom{3}{2}$ pairwise combinations are independent.

Example 9.3

Independent trials are conducted. Each trial can result in a success (with probability *p*) or a failure. What is the probability that (i) at least one success occurs in the first *n* trials, and (ii) exactly *k* successes occur in the first *n* trials?

(i)

Let F_i be the event of a failure in Trial *i*.

So, if *F* is the event of no successes in *n* trials, then $F = F_1 \cap F_2 \cap \cdots \cap F_n$

$$P(F) = P(F_1 \cap F_2 \cap \cdots \cap F_n) = P(F_1)P(F_2)\cdots P(F_n) = (1-p)^n$$

So, if I is the event of at least 1 success in first *n* trials, then $P(I) = 1 - (1-p)^n$

(ii)

$p^3 \cdot (1-p)^{10-3}$ $(1-p)\cdot p^2 \cdot (1-p)^4 \cdot p\,(1-p)^2$

SSSFFFFFFF FSSFFFFSFF

Any particular sequence of first *n* outcomes that contain *k* successes will have a probability of $p^k(1-p)^{n-k}$

There will be $\binom{n}{k}$ such sequences. So, if *E* is the event of *k* successes in first *n* trials, then

$$P(E) = \binom{n}{k} p^k (1-p)^{n-k}$$

Example 9.4

A and B are playing a game divided into many identical subgames or trials (say, matches in a set). Each subgame ends with either A winning (say, we call it a success) or B winning (say, we call it a failure). Probability of a success occurring is p. The game gets interrupted at a stage when A needed n more points to win and B needed m more points to win. What is probability that A would win the game from this stage? [What is the probability of n successes occurring before m failures?]

(This solution is due to Pascal (c. 1654))

For n successes to occur before m failures it is necessary and sufficient that there be at least n successes in n+m-1 trials. (Say X is the statement in green and Y is the statement in red without the "to".)

If X occurs, then there can be at most m-1 failures in the n+m-1 trials. So, if X occurs Y occurs.

If X does not occur (i.e., there are fewer than n successes in n+m-1 trials), then there must be at least m failures in the n+m-1 trials. That is, Y does not occur. So, if X^c occurs then Y^c occurs.

So, using the result from (ii) of the previous example $\sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k (1-p)^{n+m-1-k}$

$$P(\text{A wins the game}) = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k (1-p)^{n+m-1-k}$$

Example 9.5

A pair of dice is rolled and the sum of the upturned faces is noted. Each of these trials is independent of the other. What is the probability that a 5 occurs before a 7 (let's name it *E*)?

Approach 1

In exactly *n* trials a 5 can occur before a 7 if in the first *n-1* trials neither a 5 nor a 7 occurs and on the n^{th} trial a 5 occurs. Let this be E_n.

$P(5 \text{ on any trial}) = P(\{(1,4),(2,3),(3,2),(4,1)\}) = \frac{4}{36}$; similarly, $P(7 \text{ on any trial}) = \frac{6}{36}$

4 ways　　　　　　　　6 ways.

$$P(E_n) = \underbrace{\left(1 - \frac{10}{36}\right)^{n-1}} \times \frac{4}{36}$$

Now a 5 can occur before a 7 in 1 trial or 2 trials or 3 trials and so on.

$$P(E) = \sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} \left(1 - \frac{10}{36}\right)^{n-1} \times \frac{4}{36}$$

$$= \frac{1}{9} \times \frac{1}{1 - \frac{13}{18}} = \frac{2}{5}$$

Example 9.5 (Contd.)

A pair of dice is rolled and the sum of the upturned faces is noted. Each of these trials is independent of the other. What is the probability that a 5 occurs before a 7 (let's name it E)?

Approach 2 (Conditioning on outcome of first trial)

Let F be the event that a 5 occurs in the first trial, G be the event that a 7 occurs in the first trial and H be the event that neither a 5 nor a 7 occurs in the first trial. Note, F, G, and H are mutually exclusive and exhaustive events. Hence,

$$E = (E \cap F) \cup (E \cap G) \cup (E \cap H)$$

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

$$P(E) = \overset{1}{1} \times \frac{4}{36} + \overset{0}{0} \times \frac{6}{36} + \overset{P(E)}{P(E)} \times \frac{26}{36}$$

$$P(E) \times \frac{10}{36} = \frac{4}{36} \quad \text{or}, P(E) = \frac{4}{10} = \frac{2}{5}$$

Therefore, generally speaking,

$$P(E) = P(F) + P(E)(1 - P(F) - P(G))$$

$$\cancel{P(E)} = P(F) + \cancel{P(E)} - P(E)(P(F) + P(G))$$

$$P(E) = \frac{P(F)}{P(F) + P(G)}$$

Example 9.6 (Gambler's ruin)

A and B are playing a game where if the toss of a coin is heads (which happens with probability, *p*) A gets one unit (say a unit is Rs. 100) from B. If it is tails, then B gets one unit from A. This continues until one has nothing. A starts with *i* units and B starts with N-*i* units. What is the probability that A ends up with all the money? [Assume successive trials are independent].

Let *E* be the event that A ends up with all the money when he/she starts with *i* units (of a total pool of *N* units). P(*E*) will surely depend on *i*. Let P(*E*) be referred to as P_i.

Let *H* be the event that first flip is heads.

$E = (E \cap H) \cup (E \cap H^c)$ So, $P(E) = P(E|H)P(H) + P(E|H^c)P(H^c)$

$P(E|H) = P_{i+1}$ Trials are independent; since the first is *H*, now *A* has *i*+1 units

$P(E|H^c) = P_{i-1}$ Trials are independent; since the first is tails, now *A* has *i*-1 units

So, $P(E) = P_{i+1} \times p + P_{i-1} \times q$ where $q = 1 - p$

Also note,

$P_0 = 0$ and $P_N = 1$

Example 9.6 (Contd.)

A and B are playing a game where if the toss of a coin is heads (which happens with probability, p) A gets one unit (say a unit is Rs. 100) from B. If it is tails, then B gets one unit from A. This continues until one has nothing. A starts with i units and B starts with N-i units. What is the probability that A ends up with all the money? [Assume successive trials are independent].

$P(E) = P_{i+1} \times p + P_{i-1} \times q$ where $p + q = 1, P_0 = 0,$ and $P_N = 1$

$P_i = P_{i+1} \times p + P_{i-1} \times q \qquad i = 1,2,\cdots,N-1$

$(p+q)P_i = P_{i+1} \times p + P_{i-1} \times q \qquad i = 1,2,\cdots,N-1$

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}) \qquad i = 1,2,\cdots,N-1$$

Therefore,

$$P_2 - P_1 = \frac{q}{p}P_1$$

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

$$P_4 - P_3 = \frac{q}{p}(P_3 - P_2) = \left(\frac{q}{p}\right)^3 P_1$$

$$\vdots$$

$$P_i - P_{i-1} = \frac{q}{p}(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1$$

Adding these first i-1 terms,

$$P_i - P_1 = P_1\left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right]$$

$$P_i = P_1\left[1 + \left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right]$$

Example 9.6 (Contd.)

A and B are playing a game where if the toss of a coin is heads (which happens with probability, p) A gets one unit (say a unit is Rs. 100) from B. If it is tails, then B gets one unit from A. This continues until one has nothing. A starts with i units and B starts with N-i units. What is the probability that A ends up with all the money? [Assume successive trials are independent].

$$P_i = P_{i+1} \times p + P_{i-1} \times q \qquad i = 1,2,\cdots,N-1$$

where $p + q = 1, P_0 = 0, \text{and } P_N = 1$

$$P_i = P_1\left[1 + \overbrace{\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}}^{i-1}\right]$$

$$P_i = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)} P_1 & \text{if } \dfrac{q}{p} \neq 1 \\ iP_1 & \text{if } \dfrac{q}{p} = 1 \end{cases}$$

$$P_N = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)} P_1 = 1 & \text{if } p \neq \dfrac{1}{2} \\ NP_1 = 1 & \text{if } p = \dfrac{1}{2} \end{cases}$$

Since $P_N = 1$

$$P_1 = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^N} & \text{if } p \neq \dfrac{1}{2} \\ \dfrac{1}{N} & \text{if } p = \dfrac{1}{2} \end{cases}$$

$$P_i = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & \text{if } p \neq \dfrac{1}{2} \\ \dfrac{i}{N} & \text{if } p = \dfrac{1}{2} \end{cases}$$

Random Variables

Imagine an experiment where two coins are tossed. Then, S ={(H,H),(H,T),(T,H),(T,T)}. Let, Y denote the number of heads. Then:

Outcome (c)	Value of Y
(H,H)	2
(H,T)	1
(T,H)	1
(T,T)	0

$Y((H,H)) = 2$
$Y((H,T)) = 1$
$Y((T,H)) = 1$
$Y((T,T)) = 0$

Y is a real-valued function defined on the sample space. It assigns one and only one number to each element of the sample space. Functions such as Y are called random variables.

The range of a function such as Y is the set of real numbers $D = \{y: y = Y(c), c \in S\}$

For a random variable Y, the range D becomes a sample space. Besides inducing the sample space, Y also induces a probability which is called the distribution of Y.

Example 10.1

3 fair coins are tossed. If Y denotes the numbers of heads what is the range of Y? What are the associated probabilities?

Range of Y = {0,1,2,3}

$$P(Y = 0) = P\{(T,T,T)\} = \frac{1}{8}$$

$$P(Y = 1) = P\{(T,T,H),(T,H,T),(H,T,T)\} = \frac{3}{8}$$

$$P(Y = 2) = P\{(T,H,H),(H,T,H),(H,H,T)\} = \frac{3}{8}$$

$$P(Y = 3) = P\{(H,H,H)\} = \frac{1}{8}$$

Example 10.2

In order to get a driver's license a candidate is given a maximum of *n* attempts. Since, for every attempt the department of motor vehicles incurs a cost, the department is interested in knowing the probabilities of the different number of attempts a candidate might take if the probability of passing is *p*.

Every attempt has one of two outcomes; either you pass (you are through) or you fail. Let T stand for pass and F for fail.

Let X denote the number of attempts a candidate takes (note, the attempts stop once you pass or the number reaches n).

Range of $X = \{1,2,3,\ldots,n\}$

$$P(X=1) = P\{T\} = p$$

$$P(X=2) = P\{(F,T)\} = \underbrace{(1-p)}_{q}p$$

$$P(X=3) = P\{(F,F,T)\} = (1-p)^2 p$$

$$\vdots$$

$$P(X=n-1) = P\{(F,F,\cdots F,T)\} = \underline{(1-p)^{n-2}}p$$

$$P(X=n) = P\{\underline{(F,F,\cdots F,T)},\underline{(F,F,\cdots F,F)}\} = (1-p)^{n-1}$$

$$p + qp + q^2 p + \cdots q^{n-2}\cdot p + q^{n-1}$$

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

$$= \cancel{p}\,\frac{1-q^{n-1}}{\cancel{1-q}} + q^{n-1}$$

Example 10.3

As a strategy to boost sales, a company selling kids toothpaste started including one of *N* types of figurines in the box of a toothpaste. If you are able to collect all *N* types (i.e., collect a set), then you get a special prize. The company is interested in knowing how many toothpastes one has to buy to get all *N* types of figurines.

To get all *N* types implies one must get at least one of each type.

Let *T* be a random variable that denotes the number of toothpastes that one has to buy to complete a set.

$P(T > n)$

$P(T > n-1)$

Therefore, a quantity of interest is $P(T=n)$.

In this problem (like in many others) it is easier to determine $P(T>n)$.

For a given *n* define F_j as the event that the (first) *n* collection of figurines does not contain Type *j* (j=1,2,….,N). So,

$$P(F_j) = \left(\frac{N-1}{N}\right)^n$$

Example 10.3 (contd.)

As a strategy to boost sales, a company selling kids toothpaste started including one of N types of figurines in the box of a toothpaste. If you are able to collect all N types (i.e., collect a set), then you get a special prize. The company is interested in knowing how many toothpastes one has to buy to get all N types of figurines.

Recall, F_j is the event that the (first) n collection of figurines does not contain Type j (j=1,2,....,N), and

$$P(F_j) = \left(\frac{N-1}{N}\right)^n$$

For T to be greater than n, in the first n purchases one or more of the figurine types should be missing. That is,

$$P(T > n) = P\left(\bigcup_{j=1}^{N} F_j\right)$$

Recall: (P-Vb) $P(E_1 \cup E_2 \cup \cdots E_k) = p_1 - p_2 + p_3 - \cdots (-1)^{k+1} p_k$

Here, p_i, is the sum of probabilities of all possible intersections involving i sets.

Example 10.3 (contd.)

As a strategy to boost sales, a company selling kids toothpaste started including one of N types of figurines in the box of a toothpaste. If you are able to collect all N types (i.e., collect a set), then you get a special prize. The company is interested in knowing how many toothpastes one has to buy to get all N types of figurines.

$$P(T > n) = P(F_1 \cup F_2 \cup \cdots F_N) = p_1 - p_2 + p_3 - \cdots (-1)^{N+1} p_N$$

Here, p_i, is the sum of probabilities of all possible intersections involving i sets.

$$P(F_j) = \left(\frac{N-1}{N}\right)^n \qquad P(F_{j1} \cap F_{j2}) = \left(\frac{N-2}{N}\right)^n \qquad P(F_{j1} \cap F_{j2} \cap \cdots F_{jk}) = \left(\frac{N-k}{N}\right)^n$$

$$p_1 = N\left(\frac{N-1}{N}\right)^n \qquad p_2 = \binom{N}{2}\left(\frac{N-2}{N}\right)^n \qquad p_k = \binom{N}{k}\left(\frac{N-k}{N}\right)^n$$

p_{N-1}

$$P(T > n) = N\left(\frac{N-1}{N}\right)^n - \binom{N}{2}\left(\frac{N-2}{N}\right)^n + \binom{N}{3}\left(\frac{N-3}{N}\right)^n - \cdots + (-1)^N \binom{N}{N-1}\left(\frac{1}{N}\right)^n$$

$$= \sum_{i=1}^{N-1} (-1)^{i+1} \binom{N}{i}\left(\frac{N-i}{N}\right)^n$$

$$P(T = n) = P(T > n-1) - P(T > n)$$

Example 10.3 (An interesting aside)

As a strategy to boost sales, a company selling kids toothpaste started including one of N types of figurines in the box of a toothpaste. If you are able to collect all N types (i.e., collect a set), then you get a special prize. The company is interested in knowing how many toothpastes one has to buy to get all N types of figurines.

$$P(T > n) = \sum_{i=1}^{N-1} (-1)^{i+1} \binom{N}{i} \left(\frac{N-i}{N}\right)^n$$

Now, for $1 \le n < N$, $\quad P(T > n) = 1$

Therefore, for $\underline{1 \le n < N}$

$$\underline{\sum_{i=1}^{N-1} (-1)^{i+1} \binom{N}{i} \left(\frac{N-i}{N}\right)^n} = 1 \qquad \text{Notice,} \quad \underline{(-1)^{0+1} \binom{N}{0} \left(\frac{N-0}{N}\right)^n} = -1$$

$$\text{Therefore,} \quad \sum_{i=0}^{N-1} (-1)^{i+1} \binom{N}{i} \left(\frac{N-i}{N}\right)^n = 0$$

$$\text{Can also be written as,} \quad \sum_{j=1}^{N} (-1)^{j-1} \binom{N}{j} j^n = 0$$

[Hint: Multiply by $(-1)^N N^n$ and let $j = N - i$]

Distribution Functions

If X is a random number, then its cumulative distribution function (cdf), $F_X(b)$ is defined for all real numbers, b $(-\infty < b < \infty)$ as follows:

$$F_X(b) = P(\{c \in S: X(c) \leq b\}) = P(X \leq b)$$

Note: When there is no ambiguity possible (say you are dealing with only one random variable) then often the subscript in F_X is often dropped.

Properties of a cdf

1. F_X is a non-decreasing function. That is, if $a < b$, then $F_X(a) \leq F_X(b)$.

2. $\lim_{b\to\infty} F_X(b) = 1$ (i.e., upper limit is 1)

3. $\lim_{b\to-\infty} F_X(b) = 0$ (i.e., lower limit is 1)

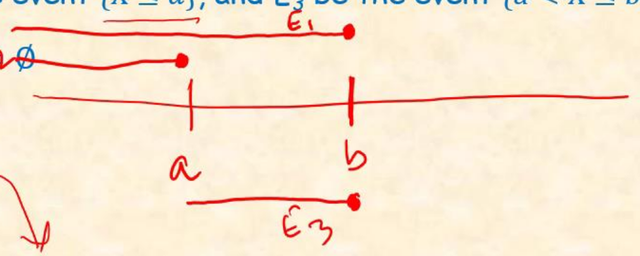
4. F_X is right continuous. For any decreasing sequence b_n that converges to b $\lim_{n\to\infty} F_X(b_n) = F(b)$

$\lim_{n\to\infty} P(X \leq b_n) = P(\lim_{n\to\infty} X \leq b_n)$

Distribution Functions (Contd.)

$$\boxed{P\{a < X \le b\} = F_X(b) - F_X(a) \ \text{ for all } a < b}$$

Let E_1 be the event $\{X \le b\}$, E_2 be the event $\{X \le a\}$, and E_3 be the event $\{a < X \le b\}$.



$$E_1 = E_2 \cup E_3 \qquad \text{and} \qquad E_2 \cap E_3 = \emptyset$$

$$P(E_1) = P(E_2) + P(E_3)$$

$$P(E_3) = P(E_1) - P(E_2)$$

$$P(E_3) = P(\{a < X \le b\}) = F_X(b) - F_X(a)$$

$$\boxed{P\{X < b\} = \lim_{n\to\infty} F_X\left(b - \frac{1}{n}\right)}$$

$$P(\{X < b\}) = P\left(\lim_{n\to\infty}\left\{X \le b - \frac{1}{n}\right\}\right)$$

$$= \lim_{n\to\infty} P\left(X \le b - \frac{1}{n}\right) \quad = \lim_{n\to\infty} F_X\left(b - \frac{1}{n}\right)$$

Example 10.4

Plot the graph of F_X when X is defined as the sum of the upturned faces of a roll of a pair of fair dice.

Range of X is $\{2,3,4,\ldots,12\}$.

$P(x \le 2.89)$
$P(x=2) + P(\text{between } 2 \text{ and } 2.89)$

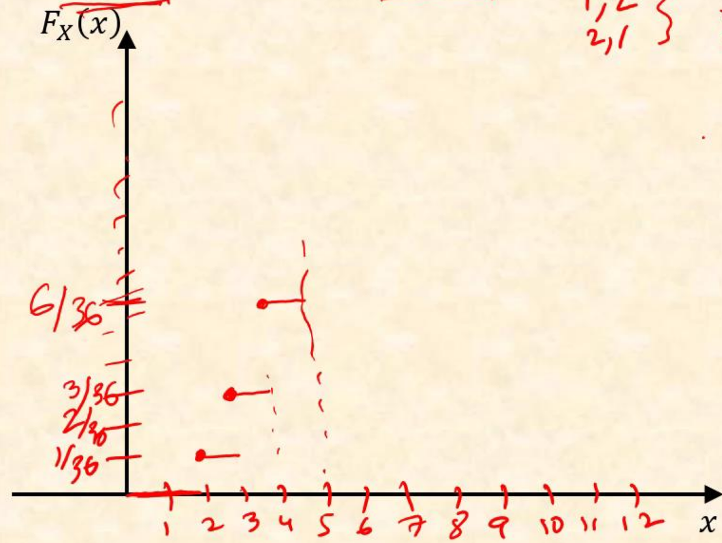
$P(x \le 1.5)$

Note, $F_X(x) = 0, \quad x < 2$; $\qquad F_X(x) = \frac{1}{36}, \qquad 2 \le x < 3$; and so on

For $3 \le x < 4 \rightarrow P(x \le 3.7) = P(x=2) + 2\text{-}3 + P(x=3) + 3\text{-}3.7$

1,2
2,1 $\frac{2}{36}$

$$F_X(x) = \begin{cases} 0 & x < 2 \\ 1/36 & 2 \le x < 3 \\ 3/36 & 3 \le x < 4 \\ 6/36 & 4 \le x < 5 \\ 10/36 & 5 \le x < 6 \\ 15/36 & 6 \le x < 7 \\ 21/36 & 7 \le x < 8 \\ 26/36 & 8 \le x < 9 \\ 30/36 & 9 \le x < 10 \\ 33/36 & 10 \le x < 11 \\ 35/36 & 11 \le x < 12 \\ 1 & 12 \le x \end{cases}$$



Example 10.5

F_X is as shown. What are (i) P(X<2), (ii) P(X=1), (iii) P(X>3/4), and (iv) P(2<X<5)?

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{3} & 0 \le x < 1 \\ \frac{1}{2} & 1 \le x < 2 \\ \frac{4}{5} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

(i) P(X<2)

$$P(\{X < 2\}) = \lim_{n\to\infty} F_X\left(2 - \frac{1}{n}\right) = \frac{1}{2}$$

(ii) P(X=1)

$$P(\{X = 1\}) = P(\{X \le 1\}) - P(\{X < 1\})$$

$$= F_X(1) - \lim_{n\to\infty} F_X\left(1 - \frac{1}{n}\right) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$$

Example 10.5 (Contd.)

F_X is as shown. What are (i) P(X<2), (ii) P(X=1), (iii) P(X>3/4), and (iv) P(2<X<5)?

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{3} & 0 \le x < 1 \\ \dfrac{1}{2} & 1 \le x < 2 \\ \dfrac{4}{5} & 2 \le x < 3 \\ 1 & 3 \le x \end{cases}$$

$$\frac{\cancel{3}/4}{\cancel{3}} = \frac{1}{4}$$

(iii) P(X>3/4)

$$P\left(\left\{X > \frac{3}{4}\right\}\right) = 1 - P\left(\left\{X \le \frac{3}{4}\right\}\right)$$

$$= 1 - F_X\left(\frac{3}{4}\right) = 1 - \frac{1}{4} = \frac{3}{4}$$

(ii) P(2<X<5)

$$P(\{2 < X < 5\}) = P(\{X < 5\}) - P(\{X \le 2\})$$

$$= \lim_{n \to \infty} F_X\left(5 - \frac{1}{n}\right) - F_X(2) \quad = 1 - \frac{4}{5} = \frac{1}{5}$$

Bernoulli Random Variable (Discrete)

Consider a trial that can be either a success or a failure. If the outcome of the trial is a success let the random variable X take the value of 1 and if it is a failure let $X = 0$.

$S = \{\text{Success}, \text{Failure}\}$
$x = \quad 1 \qquad 0$

Let the probability of success be p and that of failure be 1-p. Note, $p \in (0,1)$.

The probability mass function is:

$$p_X(x) = \begin{cases} p^x(1-p)^{1-x} & x = 0,1 \\ 0 & \text{elsewhere} \end{cases}$$

A random variable that has the above pmf is called a Bernoulli Random Variable (after James Bernoulli) with parameter p.

Binomial Random Variable

Let X denote the number of successes in n independent Bernoulli trials. Then the probability mass function of X is

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,2,\cdots,n$$

It is easy to see, from the binomial theorem that:

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-\cancel{p})]^n = 1$$

A random variable that has a pmf described by $p_X(x)$ is called a Binomial Random Variable with parameters n and p.

Example 11.3

Colour of coat in dogs is determined by a pair of genes (each of these can be of either type A or type B). The coat is brown if the dog has at least one gene A in the pair. It is white if both the genes are B. Two brown parent dogs (who are both of Type AB) have four puppies. (A puppy gets one gene in the pair from each of the parents). What is the probability mass function for a random variable, *X* denoting the number of brown puppies?

If a puppy is AA or AB or BA the coat will be brown. It will be white if the puppy is BB.

Assuming, each puppy is equally likely to get either of the two genes (A or B) from each parent, the probability a puppy will have a brown coat is $\frac{3}{4}$.

This is the same case where probability of success (brown coat) in a trial (puppy) is $\frac{3}{4}$. And the total number of trials is 4 (4 puppies).

So, X is binomially distributed with parameters 4 and $\frac{3}{4}$.

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

Example 11.4

at least

12 safety experts study a road accident site in order to determine whether the accident was caused by driver error; i.e., whether the driver was guilty (*G*). The practice is if 8 out of the 12, vote guilty then the driver is declared guilty (i.e., the accident is classified as one due to human/driver error). Each expert votes independently and has a probability *r* of making the right decision. What is the probability that a right decision will be made when (i) the driver is in fact innocent and (ii) the driver is in fact guilty.

Case (i) [Driver is innocent]

Probability that an expert declares the driver guilty = (1-*r*)

p success $(1-(1-r))$

Probability the driver is declared guilty = $\sum_{k=8}^{12}\binom{12}{k}(1-r)^k r^{12-k}$

Probability of right decision = Probability of not declaring the driver guilty =

$$1-\sum_{k=8}^{12}\binom{12}{k}(1-r)^k r^{12-k}$$

Example 11.4 (Contd.)

12 safety experts study a road accident site in order to determine whether the accident was caused by driver error; i.e., whether the driver was guilty (G). The practice is if at least 8 out of the 12, vote guilty then the driver is declared guilty (i.e., the accident is classified as one due to human/driver error). Each expert votes independently and has a probability r of making the right decision. What is the probability that a right decision will be made when (i) the driver is in fact innocent and (ii) the driver is in fact guilty.

Case (i) [Driver is innocent] Second approach

Probability that an expert declares the driver innocent = r

Probability driver is declared innocent = Probability fewer than 8 declare the driver guilty = Probability 5 or 6 or 7 .... or 12 declare the driver innocent

$$= \underbrace{\sum_{k=5}^{12} \binom{12}{k} r^k (1-r)^{12-k}}$$

← Prob. k declare driver not guilty

Of course,

$$\sum_{k=5}^{12} \binom{12}{k} r^k (1-r)^{12-k} = 1 - \sum_{k=8}^{12} \binom{12}{k} (1-r)^k r^{12-k}$$

Poisson Random Variable

A random variable that has the pmf described by the following $p_X(x)$ is called a Poisson Random Variable with parameter λ (where, $\lambda > 0$).

$$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda} \qquad x = 0,1,2,\cdots$$

It is easy to see that

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$= e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \cdots\right)$$

$$= e^{-\lambda}\, e^{\lambda} = 1$$

The common occurrence of Poisson Random Variable (or a Poisson Distribution) in nature is because occurrence of such a variable can be explained by the Law of Rare Events.

Poisson R.V. and Law of Rare Events

Imagine *n* independent Bernoulli trials each with a probability *p* of a success. Let us now imagine that we have a large number of trials but the probability of success in a given trial is low (success is a rare event).

Even though $n \to \infty$ and $p \to 0$, assume that np takes a positive constant value; i.e., $np = \lambda$ and $\lambda > 0$.

Under these conditions, the random variable *X* that counts the number of successes in *n* trials, becomes a Poisson distribution.

$np = \lambda \Rightarrow \frac{\lambda}{n} = p$

Now, $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,2,\cdots,n$

$$= \frac{n!}{(n-x)!\,x!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \qquad x = 0,1,2,\cdots,n$$

$$= \frac{n(n-1)\cdots(\overbrace{n-x+1}^{n-(x-1)})}{n^x}\left(\frac{\lambda^x}{x!}\right)\left(1-\frac{\lambda}{n}\right)^n \frac{1}{\left(1-\frac{\lambda}{n}\right)^x} \qquad x = 0,1,2,\cdots,n$$

Poisson R.V. & Law of Rare Events (Contd.)

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,2,\cdots,n$$

$$= \underbrace{\frac{\overbrace{n(n-1)\cdots(n-x+1)}^{x \text{ terms.}}}{n^x}}_{1} \left(\frac{\lambda^x}{x!}\right) \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{e^{-\lambda}} \underbrace{\frac{1}{\left(1-\frac{\lambda}{n}\right)^x}}_{1} \qquad x = 0,1,2,\cdots,n$$

Now

$$\frac{n(n-1)\cdots(n-x+1)}{n^x} = 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)$$

As $n \to \infty$

$$\frac{n(n-1)\cdots(n-x+1)}{n^x} \to 1$$

$$\left(1-\frac{\lambda}{n}\right)^x \to 1$$

$$\left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

Hence for the conditions mentioned earlier

$$p_X(x) \approx \left(\frac{\lambda^x}{x!}\right) e^{-\lambda} \qquad x = 0,1,2,\cdots$$

Poisson R.V. & Law of Rare Events (Contd.)

From the same law of rare events one can determine that in situations where the number of successes (events) in a given duration, t, is counted and λ is the per unit duration rate of occurrence of the event, then

$$p_{N(t)}(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad k = 0,1,2,\cdots$$

Where, $N(t)$ is the number of events occurring in duration t.

Example 11.4

Over 3 hours 100 vehicles cross a point on the road (assume vehicles to be dimensionless). What is the probability that no on~~g~~e crosses in three successive minutes?

Assuming rate of crossing remains constant, $\lambda = \frac{100}{180} = 0.55$ vehicles per minute.

Duration over which vehicles are being counted, $t = 3$ minutes.

Hence,

$$p_{N(3)}(0) = \frac{((0.55 \times 3)^0}{0!} e^{-0.55\times 3} = 0.192$$

Geometric Random Variable

Let N denote the number of trials required for the first success to occur. If p is the probability of success on any trial, then the probability mass function of N is

$$p_N(n) = p(1-p)^{n-1} \qquad n = 1,2,\cdots$$

$$\frac{1}{1-(1-p)}$$

It is easy to see, that:

$$\sum_{n=1}^{\infty} p(1-p)^{n-1} = p\sum_{n=1}^{\infty}(1-p)^{n-1} = p\frac{1}{1-(1-p)} = 1$$

A random variable that has a pmf described by $p_N(n)$ is called a Geometric Random Variable with parameter p.

Example 13.1

Vehicles are moving in a single file (that is, one after the other; often referred to as single lane). A pedestrian is waiting to cross this single lane. The gaps between vehicles are of varying lengths and only when a gap is sufficiently long (or is acceptable to the pedestrian) does he/she cross. If the probability of a gap being acceptable is *p* what is the probability that the pedestrian (i) crosses on the k^{th} gap (ii) has to wait at least *k* gaps before crossing. (Assume gaps or their sizes are independent of one another).

Let N be the random variable representing the number of the gap which the pedestrian accepts

(i) $$p_N(k) = p(1-p)^{k-1}$$

(ii) $$P(N \geq k) = \sum_{n=k}^{\infty} p_N(n)$$

$P(N \geq k) = P(N = k) + P(N = k+1) + P(N = k+2) + \cdots$

Hence, $$P(N \geq k) = p\sum_{n=k}^{\infty}(1-p)^{n-1} = p(1-p)^{k-1}[1 + (1-p) + (1-p)^2 \cdots]$$

$$= p(1-p)^{k-1}\frac{1}{p} = (1-p)^{k-1}$$

Note, if one has to wait at least *k* gaps that means each of the first *k*-1 gaps were unacceptable and the k^{th} gap is anything (acceptable or unacceptable). Hence, $P(N \geq k) = (1-p)^{k-1}$.

Negative Binomial Random Variable

Let N_r denote the number of trials required until r successes occur. If p is the probability of success on any trial, then the probability mass function of N_r is as shown below.

Before writing the probability mass function, note that for r successes to occur on the n^{th} trial, (r-1) successes must have occurred in the previous (n-1) trials and the n^{th} trial must be a success.

The probability that (r-1) successes occur in (n-1) trials is $\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$

Hence the mass function is:

failures = n-1 - (r-1)
= n-r.
w = n-r
n = w+r

$$p_{N_r}(n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \qquad n = r, r+1, \cdots$$

A random variable that has a pmf described by $p_{N_r}(n)$ is called a Negative Binomial Random Variable with parameters r and p.

W_r, a variable that denotes the number of failures before the r^{th} success is also a negative binomial random variable. It is written a bit differently. The pmf is written as:

$$p_{W_r}(w) = \binom{w+r-1}{r-1} p^r (1-p)^{w} \qquad w = 0, 1, \cdots$$

Example 13.2

A neighbourhood shop owner keeps notebooks in two piles. Whenever he gets a stock, he ensures that he makes two piles of say *N* each. He then sells notebooks from either pile. He chooses a pile with probability $\frac{1}{2}$. As soon as he discovers one pile is empty (this happens only when someone asks for a notebook, and he goes to a pile and finds none there) he orders a fresh stock. It takes *D* days for the new stock to arrive (i.e., if he orders on day 1 (the day he discovers a pile is empty), he can sell from the new stock from day *D*+1). What is the probability that he will not be able to sell a notebook to a prospective buyer if he invariably sells exactly one notebook everyday.

Let E_k be the event when the stop owner sees that a pile is empty while the other has k notebooks. Note k = 0,1,2,….,N (why 0?).

If D=2, then he will be able sell everyday if event E_2 or E_3 or E_4 ... or E_N occcurs.

If D=3, then he will be able sell everyday if event E_3 or E_4 or E_5 ... or E_N occcurs.

E_D . . . . .

and so on ……. Also note that if D is greater than N then the required probability is unity.

If going to the pile, (P1) which becomes empty is termed a success and going to the other pile (P2) as failure, then event E_k^{P1} (Pile P1 is empty while P2 has k) can happen if the shop owner goes to P2 N-k times before going to P1 N+1 times.

Example 13.2 (Contd.)

A neighbourhood shop owner keeps notebooks in two piles. Whenever he gets a stock, he ensures that he makes two piles of say *N* each. He then sells notebooks from either pile. He chooses a pile with probability ½ . As soon as he discovers one pile is empty (this happens only when someone asks for a notebook, and he goes to a pile and finds none there) he orders a fresh stock. It takes *D* days for the new stock to arrive (i.e., if he orders on day 1 (the day he discovers a pile is empty), he can sell from the new stock from day *D*+1). What is the probability that he will not be able to sell a notebook to a prospective buyer if he invariably sells exactly one notebook everyday.

Let E_k be the event when the stop owner sees that a pile is empty while the other has k notebooks. Note $k = 0,1,2,\dots,N$.

Event E_k^{P1} can happen if the shop owner goes to P2 N-k times before going to P1 N+1 times. That is, N-k failures have to happen before N+1 successes. So,

$$P\left(E_k^{P1}\right) = p_{W_{N+1}}(N-k) = \binom{(N-k)+(N+1)-1}{(N+1)-1} p^{N+1}(1-p)^{N-k} \qquad k = 0,1,\cdots,N$$

$$= \binom{2N-k}{N}\left(\frac{1}{2}\right)^{N+1}\left(\frac{1}{2}\right)^{N-k} = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k+1} \qquad k = 0,1,\cdots,N$$

$$P(E_k) = P\left(E_k^{P1}\right) + P\left(E_k^{P2}\right) = 2P\left(E_k^{P1}\right) = \binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k} \qquad k = 0,1,\cdots,N$$

Required probability (assuming $D \leq N$) is $= 1 - \sum_{k=D}^{N}\binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k}$

An Example from Ecology

Example 13.3

In order to estimate the population size of, say, a particular type of animal in a given region, ecologists captured r of these animals, marked them and released them back. After sufficient time (to allow the marked animals to mingle with the entire population) n animals are caught, and the number of marked animals in this set is noted. (i) What is the probability that i marked animals are found (assume that the population size of this type of animal is N)? (ii) How can N be estimated from this exercise?

(i)

Let M be the number of marked animals in the second capture.

$$P(M = i) = \frac{\binom{r}{i}\binom{N-r}{n-i}}{\binom{N}{n}} \qquad i = 0,1,\cdots,\min(n,r)$$

A random variable, (like M) that has the probability mass function described above (with parameters, N, r, and n) is referred to as a Hypergeometric random variable.

(ii)

$P(M = i)$ is a function of N, r, and n. However, in this problem n and r are fixed.

Continuous Random Variables

Recall for a discrete random variable the set of possible values is countable. There are random variables whose set of possible values are uncountable. Let X be such a random variable. X is said to be a Continuous Random Variable if there exists a non-negative function f_X defined for all real $x \in (-\infty, \infty)$ and having the property:

$$P(X \in A) = \int_A f_X(x)dx$$

Where, A is a set of real numbers.

The function, $f_X(x)$ is referred to as the probability density function (pdf) of X.

Note:

$$P\big(X \in (-\infty, \infty)\big) = \int_{-\infty}^{\infty} f_X(x)dx = 1$$

$$P(X \in [a, b]) = P(a \le X \le b) = \int_a^b f_X(x)dx$$

$$P(X = a) = \int_a^a f_X(x)dx = 0$$

$$P(X < x) = P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

$$\frac{d}{dx}F_X(x) = f_X(x)$$

Example 13.1

Random variable X has the following density function:

$$f_X(x) = \begin{cases} ax - ax^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

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

What is the probability that $X > \frac{1}{4}$?

$$\int_{-\infty}^{\infty} (ax - ax^2)dx = \int_0^1 (ax - ax^2)dx = 1 \qquad a\left[\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1\right] = a \times \frac{1}{6} = 1; \quad a = 6$$

$$P\left(X > \frac{1}{4}\right) = 6\int_{\frac{1}{4}}^{1} (x - x^2)dx = 6\left[\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_{\frac{1}{4}}^{1}\right] = 0.844$$

Example 13.2

Points inside a circle of unit radius (and centered at the origin) are chosen at random. Let X be the length of the line joining the origin to the chosen point. Assume that the probability of choosing a point within a sub-region A of the unit circle is equal to the fraction of the area of the unit circle occupied by the sub-region A. Determine the pdf of X. Also determine the probability that the chosen point lies in a ring with inner radius as $\frac{1}{2}$ and outer radius as $\frac{3}{4}$.

$0 \le x < 1$

$$P(A) = \frac{\text{area of A}}{\pi}$$

$F_X(x)$

$X \le x$ implies that the point is within a circle of area πx^2; hence, $\underline{P(X \le x) = x^2}$

So $$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$ and the pdf is $$f_X(x) = \begin{cases} 2x & \underline{0 \le x < 1} \\ 0 & \text{elsewhere} \end{cases}$$

The $P\left(\frac{1}{2} \le X \le \frac{3}{4}\right)$ can be obtained as $$\int_{\frac{1}{2}}^{\frac{3}{4}} 2x\,dx = x^2\Big|_{\frac{1}{2}}^{\frac{3}{4}} = \frac{5}{16}$$

Uniform Random Variable

X is a Uniform Random Variable on the interval (a,b) if its pdf is as follows:

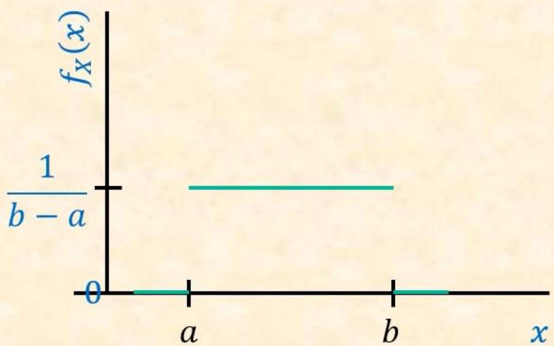
$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

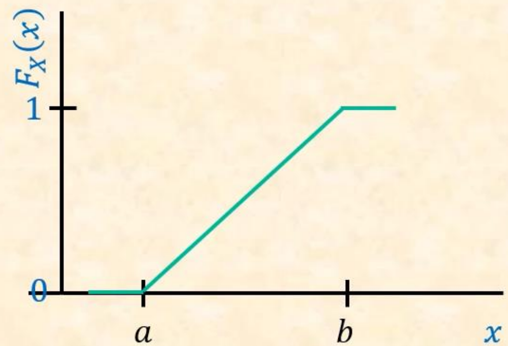
$$F_X(x) = \int_a^x \frac{1}{(b-a)}\,dt = \frac{1}{b-a}(x-a)$$

Note, $\int_a^b \frac{1}{b-a}\,dx = 1.$

The cdf of X is:

$$F_X(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$





Uniform Random Variable

X is a Uniform Random Variable on the interval (a,b) if its pdf is as follows:

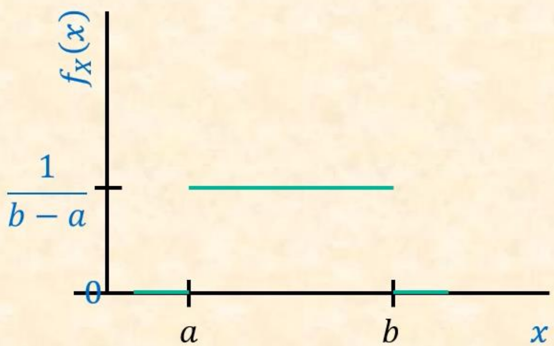
$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

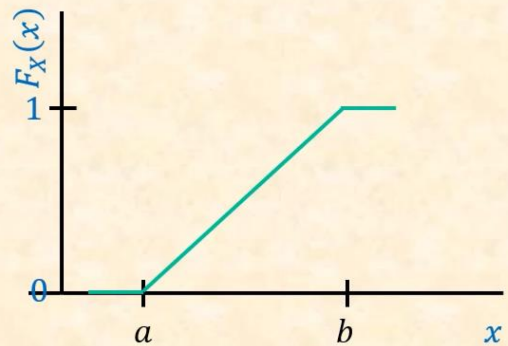
$$F_X(x) = \int_a^x \frac{1}{(b-a)}\, dt = \frac{1}{b-a}(x-a)$$

Note, $\int_a^b \frac{1}{b-a} dx = 1.$

The cdf of X is:

$$F_X(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \ge b \end{cases}$$



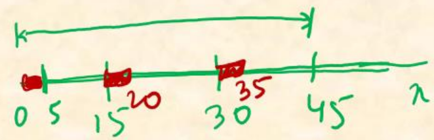


Example 13.3

Buses arrive at a stop every 15 minutes starting from 8 am. The arrival time of a commuter (regular passenger) at the stop is uniformly distributed between 8 am and 8:45 am. What is the probability that the passenger has to wait for more than 10 minutes for a bus?

Let X be the arrival time (expressed in minutes from 8 am) of the passenger.

$$f_X(x) = \begin{cases} \dfrac{1}{45-0} & 0 < x < 45 \\ 0 & \text{elsewhere} \end{cases}$$



Waiting time greater than 10 minutes is equivalent to saying $0 < X < 5$ or $15 < X < 20$ or $30 < X < 35$. So, the probability of waiting time can be evaluated as:

$$= \int_{0}^{5} \frac{1}{45}\,dx + \int_{15}^{20} \frac{1}{45}\,dx + \int_{30}^{35} \frac{1}{45}\,dx = \frac{1}{3}$$

Exponential Random Variable

A random variable that has the pdf described by the following $f_X(x)$ is called an Exponential Random Variable with parameter λ (where, $\lambda > 0$).

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

$$\lambda \left[-\frac{e^{-\lambda x}}{\lambda} \Big|_0^\infty \right].$$

$$= \frac{\cancel{\lambda}}{\cancel{\lambda}} [1 - 0] = 1$$

Note, $\int_0^\infty \lambda e^{-\lambda x} dx = 1$.

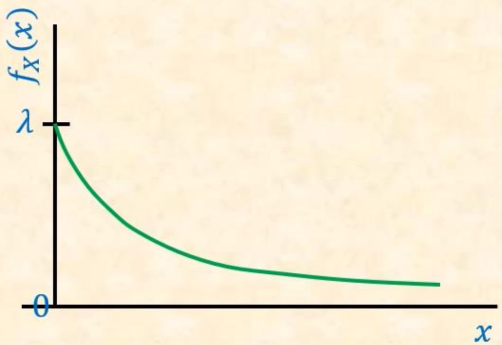
The cdf of X is:

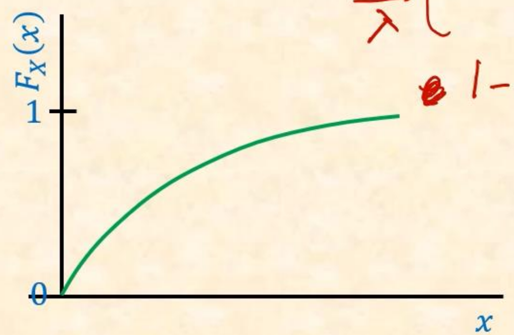
$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

$$\int_0^x \lambda e^{-\lambda t} dt$$

$$\frac{\lambda}{\lambda} \left[e^{-\lambda t} \Big|_x^0 \right]$$

$$1 - e^{-\lambda x}$$





Exponential Random Variable: Relation to Poisson

Let G be the time one has to wait for the first occurrence of an event in a Poisson process with occurrence rate is λ (where, $\lambda > 0$)

$$F_G(g) = P(G \leq g) = 1 - P(G > g)$$

$P(G > g)$ is the probability that nothing occurs in g. That is,

$$P(G > g) = p_{N(g)}(0) = \frac{(\lambda g)^0}{0!} e^{-\lambda g}$$

$$F_G(g) = 1 - e^{-\lambda g}$$

Note, G can be thought of as the gap between two successive occurrences.

Exponential R. V.: Memoryless Property

A random variable X that is exponentially distributed is memoryless (in the class of continuous distributions with positive support, exponential distribution is the only distribution that has this property). That is, $P(X > a + \ell | X > \ell) = P(X > a)$

This implies, $\dfrac{P(X > a + \ell, X > \ell)}{P(X > \ell)} = P(X > a)$

$$\frac{P(X > a + \ell)}{P(X > \ell)} = P(X > a) \quad \text{Or,} \quad P(X > a + \ell) = P(X > a)P(X > \ell)$$

Note, if X has an exponential distribution (with parameter λ) then:

$P(X > a + \ell) = e^{-\lambda(a+\ell)}$ $\qquad$ $P(X > a) = e^{-\lambda a}$ $\qquad$ $P(X > \ell) = e^{-\lambda \ell}$

Therefore: $P(X > a + \ell) = e^{-\lambda(a+\ell)} = e^{-\lambda a}e^{-\lambda \ell} = P(X > a)P(X > \ell)$

Imagine, if X is the gap between events, then the probability that at least a more units of the gap is left given you have already seen ℓ units of the gap is the same as probability that a gap is more than a units. Or,

Imagine, if X is the life of a component, then the probability that remaining life is more than a units given that it has already lived ℓ units is the same as the probability that the life is more than a units.

Example 13.4

Vehicles arrive at a location on a road with a rate of 1800 vehicles per hour (vph). The arrival follows a Poisson process. A person waiting to cross the road does so when the gap between successive arrivals is greater than 5 seconds. What is the probability that a person will accept a gap to cross the road?

Probability that a person accepts a gap = Probability gap is greater than 5 seconds.

If X is the gap between successive arrivals in seconds, then X has an exponential distribution with parameter $\lambda = \frac{1800}{3600} = \frac{1}{2}$ (vehicles per second).

$$P(X > 5) = \int_{5}^{\infty} \frac{1}{2} e^{-\frac{1}{2}x} dx = e^{-\frac{5}{2}} = 0.082$$

Example 13.5

A small rural bank is served by two tellers. When Manmohan (M) enters he sees that both the tellers are occupied; one is serving his friend Charan (C) and the other is serving Vishwanath (V). The time required to serve a customer is exponential distributed with parameter λ. What is the probability that between M, V, and C, M is not the last to leave? Assume M's service begins as soon as V or C leaves.

Since, the probability distribution is exponential, the probability distribution of the time left (say, Y) in serving whoever (between V and C) is still being served is exponential with parameter λ.

The probability distribution of the service time of M (say X) is exponential with parameter λ.

The required probability that M is not the last to leave is the same as the probability that $X < Y$.

Hazard (Failure) Rate

The memoryless property of exponential distribution implies that an old item (one which has served for ℓ units of time) is as good as new since both the old and new item has the "same amount of life left" in a probabilistic sense.

That is, failure rate of items whose duration of life is exponential, is constant (and does not change with time).

If X is a positive continuous random variable with cdf of F_X and pdf of f_X then the hazard (or failure) rate function $\omega(x)$ is defined as:

$$\omega(x) = \frac{f_X(x)}{1 - F_X(x)}$$

Hazard (Failure) Rate (Contd.)

Let's ask the question what is the probability that some item that has survived for x units will not survive another additional dx units, where X is the lifespan of the item. That is,

$$P(X \in (x, x+dx) | X > x) = \frac{P(\overbrace{X \in (x, x+dx)}^{A}, \overbrace{X > x}^{B})}{P(X > x)}$$

$$= \frac{P(X \in (x, x+dx))}{P(X > x)} \approx \frac{f_X(x)dx}{1 - F_X(x)}$$

$$= \omega(x)dx$$

If X has an exponential distribution with parameter λ, then the hazard (or failure) rate function $\omega(x)$ is defined as:

$$\omega(x) = \frac{f_X(x)}{1 - F_X(x)}$$

$$\omega(x) = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda$$

Hazard (Failure) Rate (Contd.)

The hazard rate function can be used to uniquely determine the cdf F_X.

$$\omega(t) = \frac{f_X(t)}{1 - F_X(t)} = \frac{\frac{d}{dt}F_X(t)}{1 - F_X(t)}$$

$$\omega(t) = \frac{d}{dt}\left(-\ln(1 - F_X(t))\right)$$

Integrating both sides from 0 to x

$$\int_0^x \omega(t)dt = \underline{\ln(1 - F_X(0))} - \underline{\ln(1 - F_X(x))}$$

$$e^{-\int_0^x \omega(t)dt} = 1 - F_X(x)$$

Since $F_X(0) = 0$, $\quad \int_0^x \omega(t)dt = -\ln(1 - F_X(x))$

Therefore, $\quad F_X(x) = 1 - e^{-\int_0^x \omega(t)dt}$

Example 13.6

A positive continuous random variable X has a hazard (failure) rate function given by $\omega(x) = a + bx \quad (x > 0)$. Determine the pdf of X.

$$F_X(x) = 1 - e^{-\int_0^x \omega(t)dt}$$

$$F_X(x) = 1 - e^{-ax-\frac{bx^2}{2}} \qquad x > 0$$

$$f_X(x) = (a + bx)e^{-ax-\frac{bx^2}{2}} \qquad x > 0$$

Gamma Random Variable

A random variable that has the pdf described by the following $f_X(x)$ is called a Gamma Random Variable with parameters λ and t (where, $\lambda > 0$ and $t > 0$).

$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

If $t = 1$ $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 \end{cases}$

Recall

The gamma function $\Gamma(t)$ is defined as $\int_0^\infty y^{t-1} e^{-y} dy$ $\quad (t > 0)$

$\Gamma(1) = \int_0^\infty e^{-y} dy = 1$

$\Gamma(t) = (t-1)\Gamma(t-1)$

When t takes integer values, say n, $\Gamma(n) = (n-1)!$

A random variable that has a gamma distribution with integer values of t (say, n) is referred to as an Erlang (or n-Erlang) random variable with parameters λ and n.

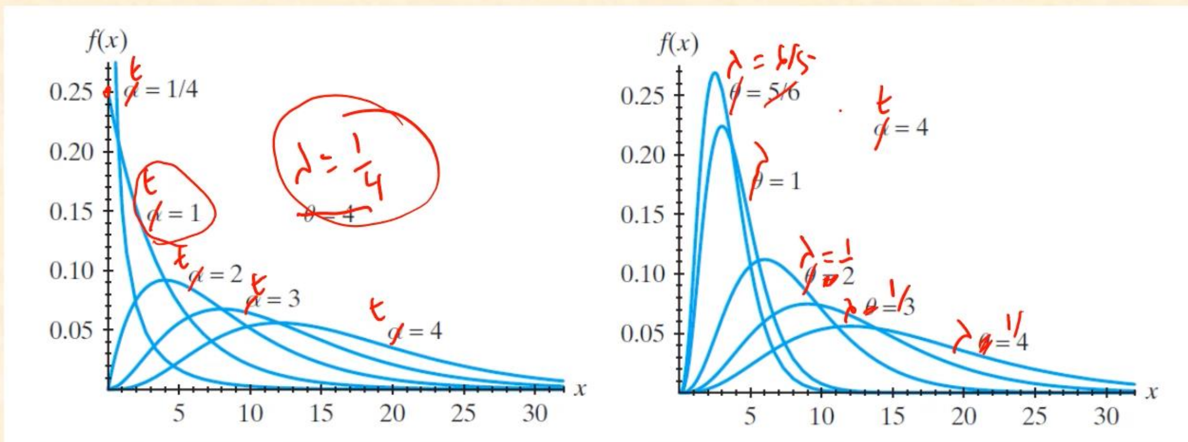
$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{n-1}}{(\mathrm{n}-1)!} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Gamma Random Variable (Contd.)

Recall a Gamma random variable has the pdf $f_X(x)$ (where, $\lambda > 0$ and $t > 0$).

$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

In the following diagram α is the same as t and $\lambda = \frac{1}{\theta}$.



Source: Hogg, Tannis and Zimmermann

Erlang Random Variable: Relation to Poisson

Let D_n be the time one has to wait for the n^{th} occurrence of an event in a Poisson process with occurrence rate is λ (where, $\lambda > 0$)

Here t is time

Not the parameter of Gamma Distr.

If number of occurrences of the event in time t, say, $N(t)$ is greater than or equal to n then surely, D_n must be less than or equal to t.

$$F_{D_n}(t) = P(D_n \le t) = P(N(t) \ge n) = \sum_{k=n}^{\infty} P(N(t) = k) = \sum_{k=n}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

$f_{D_n}(t)$ can be obtained by differentiating the above. Hence,

$$f_{D_n}(t) = \sum_{k=n}^{\infty} \frac{-\lambda(\lambda t)^k e^{-\lambda t}}{k!} + \sum_{k=n}^{\infty} \frac{k\lambda(\lambda t)^{k-1} e^{-\lambda t}}{k!} = \boxed{\sum_{k=n}^{\infty} \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!}} - \sum_{k=n}^{\infty} \frac{\lambda(\lambda t)^k e^{-\lambda t}}{k!}$$

$$= \frac{\lambda(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} + \cancel{\frac{\lambda(\lambda t)^{n} e^{-\lambda t}}{(n)!}} + \cancel{\frac{\lambda(\lambda t)^{n+1} e^{-\lambda t}}{(n+1)!}} \cdots - \cancel{\frac{\lambda(\lambda t)^{n} e^{-\lambda t}}{(n)!}} - \cancel{\frac{\lambda(\lambda t)^{n+1} e^{-\lambda t}}{(n+1)!}} \cdots$$

$$= \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$$

& Parameters are $\lambda > 0$ and $n \geq 1$

Gamma Distribution: Relation to χ_n^2 Distribution

Recall a Gamma random variable has the pdf $f_X(x)$ (where, $\lambda > 0$ and $t > 0$).

$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

A Gamma random variable with $\lambda = \frac{1}{2}$ and $t = \frac{n}{2}$ (where, *n* is a positive integer) is said to have a chi-squared distribution with *n* degrees of freedom or χ_n^2.

χ_n^2 distribution is an important distribution in statistical applications. We will see why, later.

Example 13.1

A roadside vendor selling puffed rice and fritters for people returning home from office always comes with 10 packets, sits till all the 10 are sold and then leaves. Given that customers arrive according to the Poisson distribution with arrival rate of 30 per hour and each customer buys exactly one packet, what is the probability that the vendor has to sit for (i) more than 10 minutes to sell the first two packets, and (ii) more than 1 hour before he/she can leave?

Erlang: (λ, n) $\rightarrow$ $\dfrac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}$

Here "t" is time.

(a) $\lambda = \frac{1}{2}$ customers per minute; $n = 2$; hence,

$$f_{D_2}(t) = \frac{\frac{1}{2} e^{-\frac{1}{2}t}\left(\frac{1}{2}t\right)^{2-1}}{(2-1)!} = \frac{te^{-\frac{t}{2}}}{4}$$

So $P(D_2 > 10)$ can be obtained as $\frac{1}{4}\int_{10}^{\infty} te^{-\frac{t}{2}}dt = 0.29$

(b) $\lambda = \frac{1}{2}$ customers per minute; $n = 10$; hence,

$$f_{D_{10}}(t) = \frac{\frac{1}{2} e^{-\frac{1}{2}t}\left(\frac{1}{2}t\right)^{10-1}}{(10-1)!} = \frac{t^9 e^{-\frac{t}{2}}}{2^{10}(9!)}$$

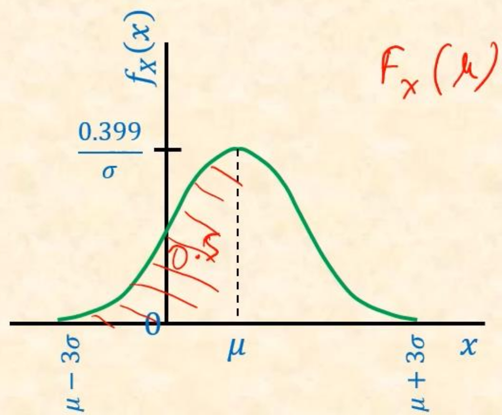
So $P(D_{10} > 60)$ can be obtained as $\frac{1}{2^{10}(9!)}\int_{60}^{\infty} t^9 e^{-\frac{t}{2}}dt$

Normal Random Variable

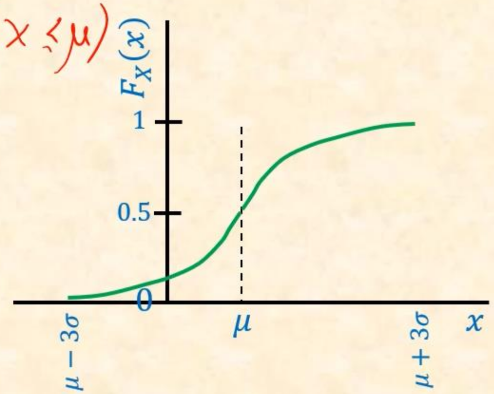
A random variable that has the pdf described by the following $f_X(x)$ is called a Normal Random Variable with parameters μ and σ^2.

$$f_X(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

This distribution is seen to arise frequently in real life possibly because it can be shown that when many independent and identically distributed random variables are added then the sum tends to have a normal distribution (this is loose statement of the Central Limit Theorem).



$F_X(\mu) = P(X \leq \mu)$



Normal Random Variable (Contd.)

$F_X(x)$ has to be evaluated numerically. If working out problems by hand one uses tables to determine the area under normal probability density function to the left of x. We will come to that soon.

But before that, a few properties:

The normal distribution is symmetric about μ.

If X is normal random variable with parameters μ and σ^2 then random variable $Y = \alpha X + \beta$ is distributed normally with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$.

$$F_Y(a) = P(Y \le a) = P(\alpha X + \beta \le a) = P\left(X \le \frac{a-\beta}{\alpha}\right) = F_X\left(\frac{a-\beta}{\alpha}\right)$$

$$= \int_{-\infty}^{\frac{a-\beta}{\alpha}} \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let $y = \alpha x + \beta$ then the upper limit of the integral becomes $y = a$;
$x = (y-\beta)/\alpha$; and the above integral can be rewritten as:

$$F_Y(a) = \int_{-\infty}^{a} \frac{1}{(2\pi)^{\frac{1}{2}}\alpha\sigma} e^{-\frac{(y-(\beta+\alpha\mu))^2}{2\alpha^2\sigma^2}} dy \quad \text{or,} \quad f_Y(y) = \frac{1}{(2\pi)^{\frac{1}{2}}\alpha\sigma} e^{-\frac{(y-(\beta+\alpha\mu))^2}{2\alpha^2\sigma^2}}$$

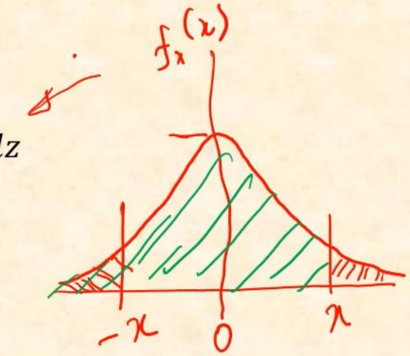
Normal Random Variable (Contd.)

If X is normal random variable with parameters μ and σ^2 then random variable $Y = \alpha X + \beta$ is distributed normally with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$.

It follows that, if $\alpha = \frac{1}{\sigma}$ and $\beta = -\frac{\mu}{\sigma}$ or, alternatively, $Z = \frac{X-\mu}{\sigma}$ then Z is distributed normally with parameters 0 and 1. Such a random variable is said to have a standard normal distribution.

It is customary to denote $F_Z(x)$ as $\Phi(x)$. That is,

$$F_Z(x) = \Phi(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$



Note,

$$\boxed{\Phi(-x) = 1 - \Phi(x)}$$

$$\boxed{F_X(a) = P(X \leq a) = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = P\left(Z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

Normal and Binomial Distributions

If X is the number of successes in n trials (with probability of success = p) then it can be shown that as $n \to \infty$

$$P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) \to \Phi(b) - \Phi(\mathrm{a})$$

Note, in this case X is a binomial random variable with parameters n and p.

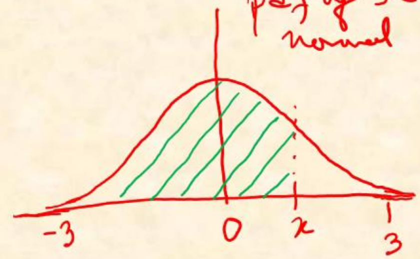
For all practical purposes this is a good approximation when $np(1-p) \geq 10$. (Recall, earlier we had seen the Poisson approximation of the binomial probabilities that worked well when n is large and np is moderate.)

Standard Normal Table

Area $\Phi(x)$ Under the Standard Normal Curve to the Left of x

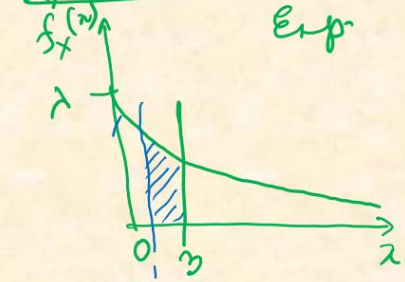
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

pdf of std. normal



Digression

Exp.



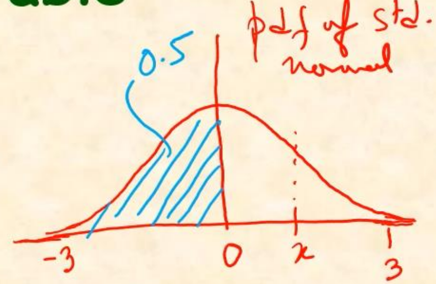
$$P(1 < X \le 3) = \int_1^3 \lambda e^{-\lambda x}\,dx$$

$$P(X \le 3) = \int_0^3 \lambda e^{-\lambda x}\,dx$$

Standard Normal Table

Area $\Phi(x)$ Under the Standard Normal Curve to the Left of x

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817



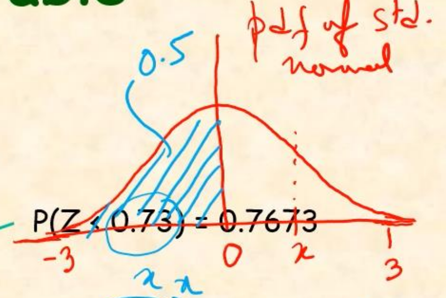
$\phi(-x) = 1 - \phi(x)$

If you have a table that lists only area between 0 and x then $\phi(x) = 0.5 + \text{Area}$

Standard Normal Table

Area $\Phi(x)$ Under the Standard Normal Curve to the Left of x

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817



$P(Z < 0.73) = 0.7673$

$P(Z < 1.77) = 0.9616$

$P(Z \le 1.453)$

$= 0.9265 + \frac{3}{10}(0.9279 - 0.9265)$

$\Phi(1.77) - \Phi(0.73)$

$P(0.73 < Z < 1.77) = P(Z < 1.77) - P(Z < 0.73) = 0.9616 - 0.7673 = 0.1943$

Example 16.1

X is a normal random variable with parameters $\mu = 5$ and $\sigma^2 = 9$. Determine the probability that (a) $X > 2$ and (b) $|X - 5| > 3$.

$P(Z > -1)$

(a)

$$P(X > 2) = P\left(\frac{X-5}{3} > \frac{2-5}{3}\right) = 1 - P(Z \le -1)$$

$$= 1 - \Phi(-1)$$

$$\Phi(-1) = 1 - \Phi(1)$$

$$1 - \Phi(-1) = \Phi(1) = 0.8413$$

(b)

$$P(|X-5| > 3) = P(X > 8) + P(X < 2)$$

$1 - P(Z \le 1)$

$$P(X > 8) = P\left(\frac{X-5}{3} > \frac{8-5}{3}\right) = P(Z > 1) = 1 - \Phi(1)$$

$$P(X < 2) = P\left(\frac{X-5}{3} < \frac{2-5}{3}\right) = P(Z < -1) = \Phi(-1) = 1 - \Phi(1)$$

$$\therefore P(|X-5| > 3) = 2 - 2\Phi(1) = 2 - 2 \times 0.8413 = 0.3174$$

Area $\Phi(x)$ Under the Standard Normal Curve to the Left of x

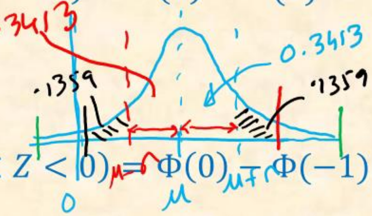
x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

Example 16.2

X is a normal random variable with parameters μ and σ^2. Determine the probability that X is between (a) μ and $\mu+\sigma$, (b) $\mu-\sigma$ and μ, (c) $\mu-2\sigma$ and $\mu-\sigma$, (d) $\mu+\sigma$ and $\mu+2\sigma$, (e) $\mu-3\sigma$ and $\mu+3\sigma$.

(a) $P(\mu < X < \mu+\sigma) = P\left(\frac{\mu-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{\cancel{\mu}+\sigma-\cancel{\mu}}{\sigma}\right) = P(0 < Z < 1) = \Phi(1) - \Phi(0)$

$$= 0.8413 - 0.5000 = 0.3413$$



(b) $P(\mu - \sigma < X < \mu) = P\left(\frac{\cancel{\mu}-\sigma-\cancel{\mu}}{\sigma} < \frac{X-\mu}{\sigma} < \frac{\cancel{\mu}-\cancel{\mu}}{\sigma}\right) = P(-1 < Z < 0) = \Phi(0) - \Phi(-1)$

$$= \Phi(0) - \big(1 - \Phi(1)\big) = 0.5000 + 0.8413 - 1 = 0.3413$$

(c) $P(\mu - 2\sigma < X < \mu - \sigma) = P(-2 < Z < -1) = \Phi(-1) - \Phi(-2) = 1 - \Phi(1) - \big(1 - \Phi(2)\big)$

$$= \Phi(2) - \Phi(1) = 0.9772 - 0.8413 = 0.1359$$

$= P(\mu+\sigma < X < \mu + 2\sigma)$

(d) $P(\mu + \sigma < X < \mu + 2\sigma) = P(1 < Z < 2) = \Phi(2) - \Phi(1) = 0.1359$

(e) $P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) = \Phi(3) - \Phi(-3) = 2\Phi(3) - 1$

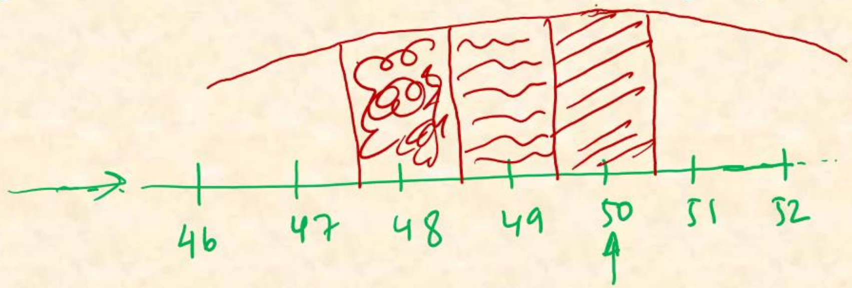
$$= 2 \times 0.9987 - 1 = 0.9974$$

Example 16.3

X is a random variable that counts the number of heads in 100 tosses. Approximate the probability of X=50 using the normal approximation. Also calculate using the exact probability.

np, $np(1-p)$

X is a discrete random variable the takes values 0, 1, 2, … 100. When approximating the probability that X=50 using a continuous random variable the following may be the best way forward:



Example 16.3

X is a random variable that counts the number of heads in 100 tosses. Approximate the probability of X=50 using the normal approximation. Also calculate using the exact probability.

np, $np(1-p)$

X is a discrete random variable the takes values 0, 1, 2, … 100. When approximating the probability that X=50 using a continuous random variable the following may be the best way forward:

$$P(X = 50) = P(49.5 < X < 50.5) \qquad np = 50; np(1-p) = 25$$

$$P(49.5 < X < 50.5) = P\left(\frac{49.5 - 50}{5} < \frac{X - 50}{5} < \frac{50.5 - 50}{5}\right)$$

$$= \Phi(0.1) - \Phi(-0.1) = 0.0796$$

Also,

$$P(X = 50) = \binom{100}{50}\left(\frac{1}{2}\right)^{50}\left(\frac{1}{2}\right)^{50} = 0.07959$$

Distribution of a Function of a Random Variable

Say you know the pdf of speeds of vehicles on a road. Also, you are told that every vehicles decelerates at the same constant rate when coming to a stop. Can you find the pdf of the stopping distance?

Say a sheet metal is cut into square pieces. If you know that the length of the sides vary randomly and you know the pdf, can you find the pdf of the area?

Not pdf.

These are all questions that require you to find the pdf of $Y = f(X)$ given that you know the pdf of X.

Example 16.4

Discs of steel are cut by an apprentice machine operator. It is seen that the radius, R is uniformly distributed over (R_1, R_2). What is the probability density function of the area of the discs?

$$f_R(r) = \begin{cases} \frac{1}{R_2 - R_1} & R_1 < r < R_2 \\ 0 & \text{elsewhere} \end{cases}$$

Let A be the area of the discs; $A = \pi R^2$

$$F_R(b) = \frac{b - R_1}{R_2 - R_1} \quad R_1 < b < R_2$$

$$
\begin{aligned}
F_A(a) &= P(A \le a) \\
&= P(\pi R^2 \le a) \qquad \pi R_1^2 < a < \pi R_2^2 \\
&= P(R^2 \le a/\pi) \\
&= P(R \le \sqrt{a/\pi}) \\
&= F_R\left(\sqrt{\frac{a}{\pi}}\right) = \frac{\sqrt{a/\pi} - R_1}{R_2 - R_1} \qquad \pi R_1^2 < a < \pi R_2^2
\end{aligned}
$$

$$\frac{1}{R_2 - R_1}\left[\frac{1}{\sqrt{\pi}} \cdot \frac{1}{2} \cdot \frac{1}{\sqrt{a}}\right]$$

Also, $\frac{d}{da} F_A(a) = f_A(a)$

Therefore,

$$f_A(a) = \begin{cases} \frac{1}{2\sqrt{a\pi}} \frac{1}{R_2 - R_1} & \pi R_1^2 < a < \pi R_2^2 \\ 0 & \text{elsewhere} \end{cases}$$

or, $$f_A(a) = \begin{cases} \frac{a^{-\frac{1}{2}}}{2\sqrt{\pi}(R_2 - R_1)} & \pi R_1^2 < a < \pi R_2^2 \\ 0 & \text{elsewhere} \end{cases}$$

Example 16.5

X is uniformly distributed over (-2,3). Determine the pdf of Y = X².

Note, here Y will have the support (space) $0 \leq y < 9$

$$F_Y(y) = P(Y \leq y) = P\left(X^2 \leq y\right)$$

For $y \leq 4$

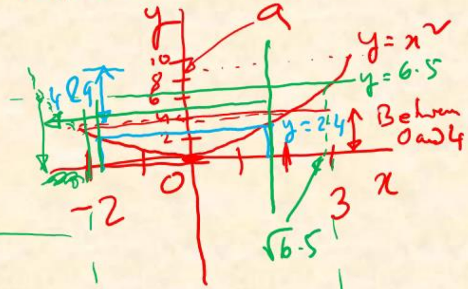
$$F_Y(y) = P\left(X^2 \leq y\right) = P(-\sqrt{y} \leq X \leq \sqrt{y})$$

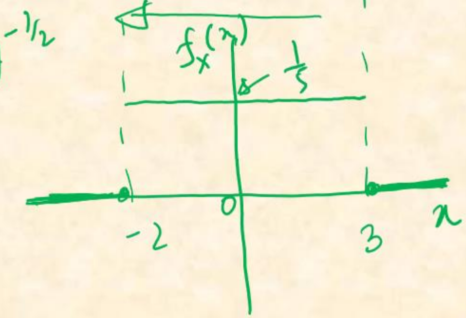
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$= \frac{\sqrt{y} - (-2)}{3-(-2)} - \frac{-\sqrt{y} - (-2)}{3-(-2)} = \frac{2\sqrt{y}}{5}$$

$$f_Y(y) = \frac{1}{5\sqrt{y}}$$

$2 \cdot \frac{1}{2} y^{-1/2}$





Example 16.5

X is uniformly distributed over (-2,3). Determine the pdf of Y = X².

Note, here Y will have the support (space) $0 \le y < 9$

$$F_Y(y) = P(Y \le y) = P\left(X^2 \le y\right)$$

For $y \le 4$

$$F_Y(y) = P\left(X^2 \le y\right) = P(-\sqrt{y} \le X \le \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$= \frac{\sqrt{y}-(-2)}{3-(-2)} - \frac{-\sqrt{y}-(-2)}{3-(-2)} = \frac{2\sqrt{y}}{5}$$

$$f_Y(y) = \frac{1}{5\sqrt{y}}$$

For $4 < y < 9$

$$F_Y(y) = P\left(X^2 \le y\right) = P(X \le \sqrt{y})$$

$$= F_X(\sqrt{y})$$

$$= \frac{\sqrt{y}-(-2)}{3-(-2)} = \frac{\sqrt{y}+2}{5}$$

$$f_Y(y) = \frac{1}{10\sqrt{y}}$$

$4 < y < 9$

$F_Y(y)$

Therefore,

$$f_Y(y) = \begin{cases} \dfrac{1}{5\sqrt{y}} & 0 \le y \le 4 \\ \dfrac{1}{10\sqrt{y}} & 4 < y < 9 \\ 0 & \text{elsewhere} \end{cases}$$

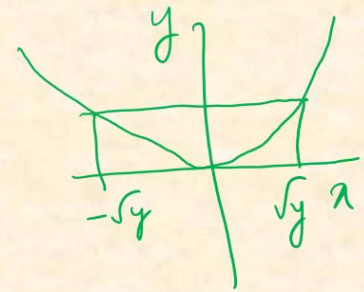
$$\int_0^9 f_Y(y)\,dy = 1$$

$$= \frac{1}{5}\int_0^4 y^{-1/2}dy + \frac{1}{10}\int_4^9 y^{-1/2}dy$$

$$= \frac{2}{5}\sqrt{y}\Big|_0^4 + \frac{2}{10}\sqrt{y}\Big|_4^9 = \frac{2}{5}\cdot 2 + \frac{1}{5}(3-2) = 1$$

Example 16.6

X has a pdf of $f_X(x)$; $-\infty < x < \infty$. Determine the pdf of $Y = X^2$.



Note, here Y will have the support (space) $y \geq 0$.

$$F_Y(y) = P(Y \leq y) = P\left(X^2 \leq y\right)$$

$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$f_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Recall

$$\frac{d}{dy} F_Y(y) = f_Y(y)$$

Now, $$\frac{d}{dy} F_Y(y) = f_X(\sqrt{y}) \times \frac{d}{dy}(\sqrt{y}) - f_X(-\sqrt{y}) \times \frac{d}{dy}(-\sqrt{y})$$

$$\frac{d}{dy} F_Y(y) = f_X(\sqrt{y}) \times \frac{1}{2\sqrt{y}} - f_X(-\sqrt{y}) \times -\frac{1}{2\sqrt{y}}$$

Therefore, $$f_Y(y) = f_X(\sqrt{y}) \times \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \times \frac{1}{2\sqrt{y}}$$

Change of Variable Technique (Contd.)

Let X be a continuous random variable with pdf of f_X with support $c_1 < x < c_2$ and let $Y = g(X)$.

Let $Y = g(X)$ be a continuous decreasing function with inverse function $X = h(Y)$.

Note, Y will have the support (space) $d_2 < y < d_1$ where, $g(c_1) = d_1$ and $g(c_2) = d_2$

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

$$= P(X \ge h(y)) \qquad d_2 < y < d_1 \qquad \text{(Because } h(y) \text{ is also decreasing)}$$

$$F_Y(y) = \int_{h(y)}^{c_2} f_X(x)dx \qquad d_2 < y < d_1$$ (Note, for $y \le d_2, F_Y(y) = 0$ and for $y \ge d_1, F_Y(y) = 1$.)

$$= F_X(c_2) - F_X(h(y))$$

$$\frac{d}{dy} F_Y(y) = f_Y(y)$$

Hence, $f_Y(y) = -f_X\big(h(y)\big)h'(y) \qquad d_2 < y < d_1$

Note, since $h(y)$ is a decreasing function, $-h'(y) = |h'(y)|$

Therefore, $f_Y(y) = f_X\big(h(y)\big)|h'(y)| \qquad d_2 < y < d_1$

Change of Variable Technique (Contd.)

So, if $Y = g(X)$ is either a continuous increasing function or a continuous decreasing with inverse function $X = h(Y)$. Then over the appropriate support of Y, $f_Y(y)$ can be written as:

$$f_Y(y) = f_X\big(h(y)\big)|h'(y)|$$

Example 16.7

X has a pdf as shown. $Y = (1 - X)^3$. Determine the pdf of Y.

$$f_X(x) = \begin{cases} 3(1-x)^2 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

$Y = g(X)$ is a decreasing function with inverse function $X = h(Y) = 1 - Y^{1/3}$.

Also, $0 < y < 1$ and $h'(y) = -\frac{1}{3}y^{-\frac{2}{3}} = -\frac{1}{3y^{\frac{2}{3}}}$ $\quad |h'(y)| = \frac{1}{3y^{2/3}} \quad h(y) = 1 - y^{1/3}$

Recall, $f_Y(y) = f_X(h(y))|h'(y)|$

Now, $f_X(h(y)) = 3\left(1 - \left(1 - y^{\frac{1}{3}}\right)\right)^2 = 3y^{\frac{2}{3}}$

Therefore, $f_Y(y) = f_X(h(y))|h'(y)| = 3y^{\frac{2}{3}}\frac{1}{3y^{\frac{2}{3}}} \qquad 0 < y < 1$

Or, $f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$

Example 16.7

X has a pdf as shown. $Y = (1 - X)^3$. Determine the pdf of Y.

$$f_X(x) = \begin{cases} 3(1-x)^2 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

$Y = g(X)$ is a decreasing function with inverse function $X = h(Y) = 1 - Y^{1/3}$.

$h(y) = 1 - y^{1/3}$

Also, $0 < y < 1$ and $h'(y) = -\frac{1}{3}y^{-\frac{2}{3}} = -\frac{1}{3y^{\frac{2}{3}}}$ $|h'(y)| = \frac{1}{3y^{2/3}}$

Recall, $f_Y(y) = f_X(h(y))|h'(y)|$

Now, $f_X(h(y)) = 3\left(1 - \left(1 - y^{\frac{1}{3}}\right)\right)^2 = 3y^{\frac{2}{3}}$

Therefore, $f_Y(y) = f_X(h(y))|h'(y)| = 3y^{\frac{2}{3}}\frac{1}{3y^{\frac{2}{3}}}$ $0 < y < 1$

Or, $f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$

That is, Y is distributed uniformly over (0,1)

Change of Variable Technique when X is discrete

So far, we dealt with continuous random variables. In such variables the probabilities are determined by integrating the pdf over suitable limits.

In discrete random variables, the probabilities are specified through the probability mass function. So, if X is a discrete random variable, with space or support C_X , then

$$p_X(x) = P\{X = x\} \qquad x \in C_X$$

Recall, C_X has a countable number of points, say, $c_1, c_2, \cdots$.

Now, let $Y=g(X)$ be a one-to-one transformation and its inverse $X=h(Y)$.

$y \quad x$

Then, $y=g(x)$ maps C_X onto C_Y where the elements of C_Y are $d_1 = g(c_1), d_2 = g(c_2), \cdots$

Therefore, $p_Y(y) = P\{Y = y\} = P\{g(X) = y\} = P\{X = h(y)\} \qquad y \in C_Y$

$$p_Y(y) = P\{X = h(y)\} = p_X\big(h(y)\big) \qquad y \in C_Y$$

Example 16.7

X is binomially distributed with $n = 4$ and $p = \frac{1}{4}$. Determine the probability mass function of $Y=X^2$.

$C_x = \{0, 1, 2, 3, 4\}$

$C_Y = \{0, 1, 4, 9, 16\}$

$Y = g(X) = X^2$ with inverse function $X = h(Y) = Y^{1/2}$.

Also, $x = 0, 1, 2, 3, 4$ and $y = 0, 1, 4, 9, 16$.

$P(Y=0) = P(X=0)$

$(Y=1) = P(X=1)$

$\boxed{(Y=4) = P(X=2)}$

Recall, $p_Y(y) = p_X(h(y))$

Therefore, $p_Y(y) = \binom{4}{\sqrt{y}}\left(\frac{1}{4}\right)^{\sqrt{y}}\left(\frac{3}{4}\right)^{4-\sqrt{y}}$ $\qquad y = 0,1,4,9,16$

$P(X=0) = \binom{4}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^4$

$P(X=2) = \binom{4}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2$

$P_Y(Y=y) =$

$P(Y=0) = \binom{4}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^4$

$P(Y=4) = \binom{4}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^{4-2}$

Example 16.7

X is a discrete random variable that takes values -3, -2, -1, 0, 1, 2. The probability mass function of X is as shown. Determine the probability mass function of $Y=X^2$.

$C_Y = \{0, 1, 4, 9\}$

x	$p_X(x)$	y
-3	0.1	9
-2	0.2	4
-1	0.1	1
0	0.3	0
1	0.2	1
2	0.1	4

y	$p_Y(y)$
0	0.3
1	0.1+0.2=0.3
4	0.2+0.1=0.3
9	0.1

A Useful Theorem and its Converse

Let $F(x)$ have the properties of a cdf of a continuous type with $F(a)=0$ and $F(b)=1$; also suppose that it is *strictly* increasing in the support $a < x < b$. Let Y be a random variable that is uniformly distributed over $(0,1)$. Then a random variable defined as $X = F^{-1}(Y)$ is a continuous type random variable with cdf $F(x)$.

To obtain the cdf of X determine $P(X \le x) = P(F^{-1}(Y) \le x) \qquad a < x < b$

$$P(X \le x) = P(F^{-1}(Y) \le x) \qquad a < x < b$$

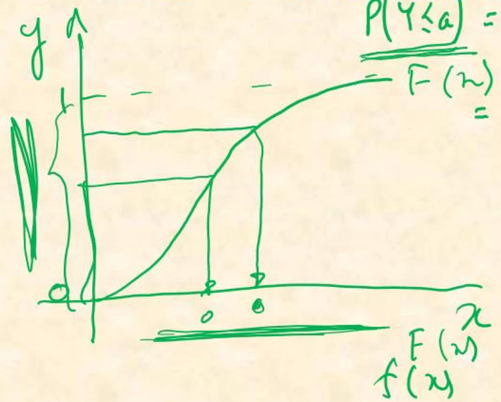
$$P(X \le x) = P(Y \le F(x)) \qquad a < x < b \qquad \text{(F is strictly increasing)}$$

$$P(X \le x) = F(x) \qquad a < x < b \qquad \text{(Since Y is uniform over $(0,1)$)}$$

$$P(Y \le a) = \int_0^a 1 \cdot dy = y \Big|_0^a = a.$$

That is the cdf of X is $F(x)$.

This result can be used to randomly generate observations from a given distribution once you have uniformly generated random numbers.



The Converse

Let X have a continuous type cdf, $F(x)$ which is *strictly* increasing on the support $a < x < b$. Then $Y = F(X)$ has a uniform distributed over (0,1).

To obtain the cdf of Y determine $P(Y \leq y) = P(F(X) \leq y)$ $\quad F(a) = 0 < y < F(b) = 1$

$$P(Y \leq y) = P(F\,(X) \leq y) \qquad 0 < y < 1$$

$$P(Y \leq y) = P(X \leq F^{-1}(y)) \qquad 0 < y < 1 \qquad \text{(F is strictly increasing)}$$

$$P(Y \leq y) = F\left(F^{-1}(y)\right) = y \qquad 0 < y < 1$$

That is the cdf of a random variable uniformly distributed over (0,1).

Jointly Distributed Random Variables

So far we talked about one random variable and its distribution.

Often, however, we may need to study the distribution of two or more random variables.

Such distributions are called joint (cumulative) distributions.

For discrete random variables we have joint probability mass functions and for continuous variables we have joint probability density functions.

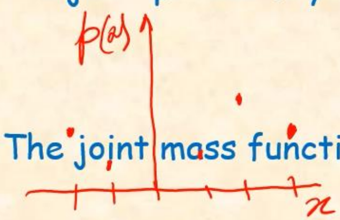
Here, we will restrict ourselves to discussing joint distributions of only two random variables.

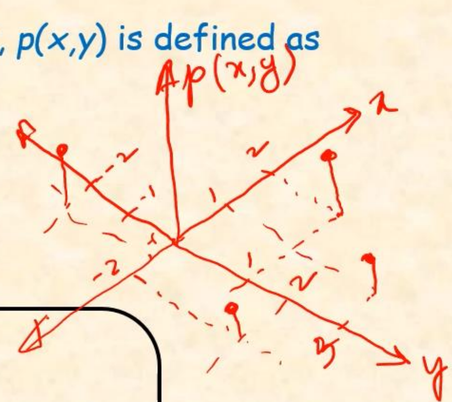
Bivariate Distributions of Discrete Type

Let X and Y be discrete random variables.

The joint probability mass function $p_{X,Y}(x,y)$, or simply, $p(x,y)$ is defined as

$$p(x,y) = P(X = x, Y = y)$$





The joint mass function has the following properties:

$$0 \leq p(x,y) \leq 1$$

$$\sum_{(x,y)\in S} p(x,y) = 1 \qquad S \text{ is the space of } (x,y)$$

$$P[(X,Y) \in A] = \sum_{(x,y)\in A} p(x,y) \qquad A \subset S$$

Discrete Bivariate Distributions (Contd.)

Let the joint probability mass function be $p(x, y) = P(X = x, Y = y)$.

$p_X(x) = P\{X = x\} = P\{X = x, Y \text{ can be anything}\}.$

From this one can determine the pmf of X alone and pmf of Y alone (often referred to as the marginal distribution) as follows:

$$p_X(x) = P(X = x) = \sum_{y:p(x,y)>0} p(x, y) \qquad x \in S_X$$

$$p_Y(y) = P(Y = y) = \sum_{x:p(x,y)>0} p(x, y) \qquad y \in S_Y$$

Example 18.1

Total = 12

3 balls are randomly selected from a box containing 3 red, 4 white and 5 blue balls. R and W, respectively denote the number of red balls and the number of white balls chosen. Determine, (a) the joint pmf, $p(i,j) = P(R = i, W = j)$, and (b) $p_R(i)$ and $p_W(j)$.

$p_{R,W}(i,j)$

Note, $S = \{(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2), (2,0), (2,1), (3,0)\}$

$$p(0,0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}; \quad p(0,1) = \frac{\binom{5}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{40}{220}; \quad p(0,2) = \frac{\binom{5}{1}\binom{4}{2}}{\binom{12}{3}} = \frac{30}{220};$$

$$p(0,3) = \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220}; \quad p(1,0) = \frac{\binom{3}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{30}{220}; \quad p(1,1) = \frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220};$$

$$p(1,2) = \frac{\binom{3}{1}\binom{4}{2}}{\binom{12}{3}} = \frac{18}{220};$$

$$p(2,0) = \frac{\binom{3}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{15}{220};$$

$$p(2,1) = \frac{\binom{3}{2}\binom{4}{1}}{\binom{12}{3}} = \frac{12}{220};$$

$$p(3,0) = \frac{\binom{3}{3}}{\binom{12}{3}} = \frac{1}{220};$$

i	*j*				Row sum =P(R=*i*)
	0	1	2	3	
0	10/220	40/220	30/220	4/220	84/220
1	30/220	60/220	18/220		108/220
2	15/220	12/220			27/220
3	1/220				1/220
Col. sum =P(*W*=*j*)	56/220	112/220	48/220	4/220	

Example 18.2

The joint pmf of X and Y is $p(x,y) = \frac{xy^2}{k}$ $\quad x = 1,2,3 \quad y = 1,2$. Find (a) k, (b) $p_X(x)$ and (c) $p_Y(y)$.

$S = \{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2)\}$

(a) $$\sum_{y=1}^{2}\sum_{x=1}^{3}\frac{xy^2}{k} = 1; \quad \frac{1}{k}[(1+2+3)+(4+8+12)] = \frac{30}{k} = 1; \quad k = 30$$

(b) $$p_X(x) = \sum_{y=1}^{2}\frac{xy^2}{30} = \frac{x}{30} + \frac{4x}{30} = \frac{x}{6} \qquad x = 1,2,3.$$

$p_X(1) = \frac{1}{6}$, $p_X(2) = \frac{2}{6} = \frac{1}{3}$

$p_X(3) = \frac{3}{6} = \frac{1}{2}$

(c) $$p_Y(y) = \sum_{x=1}^{3}\frac{xy^2}{30} = \frac{y^2}{30} + \frac{2y^2}{30} + \frac{3y^2}{30} = \frac{y^2}{5} \qquad y = 1,2.$$

$$p_Y(y) = \sum_{x=1}^{3}\frac{xy^2}{30} = \frac{y^2}{30} + \frac{2y^2}{30} + \frac{3y^2}{30} = \frac{6y^2}{30} = \frac{y^2}{5}$$

An aside:

Here, $p_X(x)p_Y(y) = \frac{x}{6} \times \frac{y^2}{5} = \frac{xy^2}{30} = p(x,y)$

Example 18.3

The joint pmf of X and Y is $p(x,y) = \frac{x+y}{21}$ $\quad x = 1,2,3 \quad y = 1,2$. Find (a) $p_X(x)$ and (b) $p_Y(y)$.

(a) $p_X(x) = \sum_{y=1}^{2} \frac{x+y}{21} = \frac{x+1}{21} + \frac{x+2}{21} = \frac{2x+3}{21}$ $\qquad x = 1,2,3.$

$6+3y = 3(2+y)$

(b) $p_Y(y) = \sum_{x=1}^{3} \frac{x+y}{21} = \frac{1+y}{21} + \frac{2+y}{21} + \frac{3+y}{21} = \frac{2+y}{7}$ $\qquad y = 1,2.$

An aside:

Here, $p_X(x)p_Y(y) = \frac{2x+3}{21} \times \frac{2+y}{7} \neq p(x,y)$

Bivariate Distributions of Continuous Type

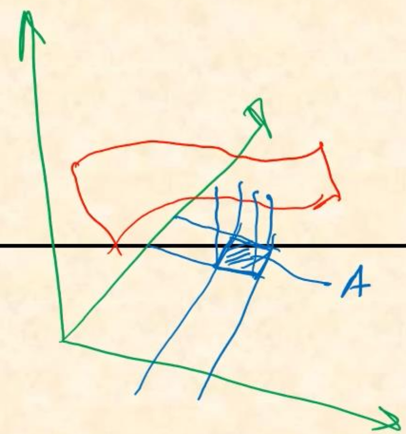
Let X and Y be continuous random variables.

The joint probability density function $f_{X,Y}(x,y)$, or simply, $f(x,y)$ is an integrable function with the following properties:

$f(x,y) \geq 0$ where, $f(x,y) = 0$ when (x,y) is not in the support or space, S

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

$$P[(X,Y) \in A] = \iint_A f(x,y)\,dx\,dy$$



Continuous Bivariate Distributions (Contd.)

Let the joint probability density function be $f(x, y)$.

From this one can determine the pdf of X alone and pdf of Y alone (often referred to as the marginal distribution) as follows:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \qquad x \in S_X$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx \qquad y \in S_Y$$

The joint distribution function can be obtained as:

$$F(a, b) = P(X \leq a, Y \leq b) = P(X \in (-\infty, a], Y \in (-\infty, b]) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x, y)dxdy$$

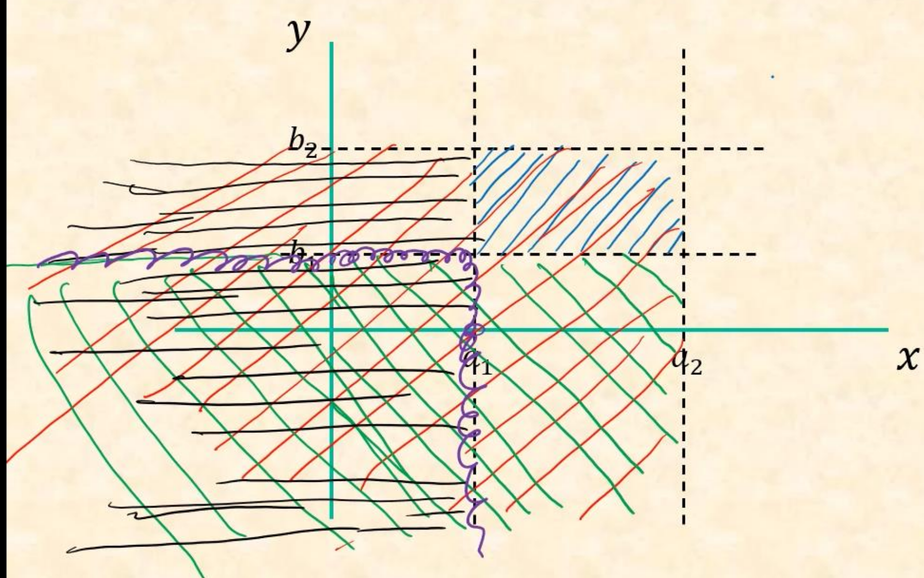
Of course, $\dfrac{\partial^2}{\partial a \partial b} F(a, b) = f(a, b)$

Continuous Bivariate Distributions (Contd.)

Note:

$$F_X(a) = F(a, \infty) \quad \text{and} \quad F_Y(b) = F(\infty, b)$$

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F(a_2, b_2) - F(a_2, b_1) - F(a_1, b_2) + F(a_1, b_1)$$



Example 18.4

The joint pdf of X and Y is $f(x,y) = 2e^{-x}e^{-2y}$ $\quad 0 < x < \infty \quad 0 < y < \infty$.
Find (a) $P(X>1, Y<1)$, (b) $f_X(x)$, (c) $f_Y(y)$, and (d) $P(X<a)$.

(a) $P(X > 1, Y < 1) = \int_1^\infty \int_0^1 2e^{-x}e^{-2y}\,dy\,dx = \int_1^\infty \cancel{2}e^{-x}\left(\frac{e^{-2y}}{\cancel{2}}\bigg|_1^0\right)dx$

$= \int_1^\infty e^{-x}(1 - e^{-2})\,dx = (1 - e^{-2})e^{-x}\Big|_\infty^1 = (1 - e^{-2})e^{-1}$

(b) $f_X(x) = \int_0^\infty 2e^{-x}e^{-2y}\,dy = e^{-x}$ $\quad 0 < x < \infty$

(d) $P(X < a) = \int_0^a e^{-x}\,dx = 1 - e^{-a}$

(c) $f_Y(y) = \int_0^\infty 2e^{-x}e^{-2y}\,dx = 2e^{-2y}$ $\quad 0 < y < \infty$

An aside:

Here, $f_X(x)f_Y(y) = e^{-x} \times 2e^{-2y} = f(x,y)$