FUNDAMENTALS OF MATHEMATICS

(COUNTING PRINCIPLES, PROBABILITY AND STATISTICS)

UNIT–I

ELEMENTARY COUNTING PRINCIPLE : Product rule, Binomial and multinomial theorem, Stirling's formula, Principle of inclusion and exclusion, Permutations and combinations, Dearrangements, Marriage problem.

UNIT-II

RECURRENCES : Recurrences and generating functions, Solution of recurrences using generating functions

UNIT-III

DISCRETE PROBABILITY: Discrete probability, Applications of counting principles to calculate discrete probability.

UNIT–IV

PROBABILITY DISTRIBUTIONS: Definition of a random variable, Probability distribution and density function, Mathematical Expectation. mean, median, mode. Skewness and Kurtosis, Higher moments, Various probability distributions, Normal, Binomial, Poisson, and Cauchy distributions, and their properties.

UNIT–V

CORRELATION AND STATISTICAL INDEPENDENCE : Correlation and statistical independence, Conditional probability, Numerical generation of random variables with a given distribution, Statement of the central limit theorem, and numerical test of the central limit theorem.

UNIT-VI

Basics of sampling theory, Sample mean and variance, Sampling biases, with special reference to Internet sampling, Stratified sampling.

UNIT-VII

Introduction to Monte Carlo methods

Binomial Theorem

An algebraic expression consisting of two terms with a positive or negative sign between them is called a binomial expression.

Example: $(a+b)$, $(P / x^2) - (Q / x^4)$ etc.

Binomial Theorem:

When a binomial expression is raised to a power 'n' we would like to be able to expand it. The binomial theorem assists us in doing this. It converts such an expression into a series.

Binomial Theorem for positive integral index:

$\mathbf{(x+y)^n = x^n + {}^nC_1x^{n-1}y+{}^nC_2x^{n-2}y^2+\text{-----}+{}^nC_{rx}{}^{n-r}y^r+ \text{-------}+\text{---------}+{}^nC_{n-1}xy^{n-1} + {}^nc_ny^n.}$

It can be represented **as:**

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r} y^r$$

Particular – Cases :

(i) Replacing 'y' by '-y', we have :

$(x-y)^n = {}^nC_oxy^{o-n}C_1x^{n-1}y+{}^nC_2x^{n-2}y^2\text{-------}+(-1)^r\,{}^nC_rx^{n-r}y^r+\text{------}+(-1)^n\,{}^nC_nx^oy^n.$

It can be represented as :

$$(x+y)^n = \sum_{r=0}^{n} (-1)^r \binom{n}{r} x^{n-r} y^r$$

(ii) Replacing 'x' by '1' and 'y' by 'x', we have :

$(1+x)^n = {}^nC_ox^o+{}^nC_1x+{}^nC_2x^2+\text{---------}+{}^nC_rx^r+\text{------}+{}^nC_{n-1}x^{n-1}+{}^nC_nx^n.$

or

$$(1+x)^n = \sum_{r=0}^{n} \binom{n}{r} x^r$$

(ii) Replacing 'x' by '-x', we have :

$(1+x)^n = {}^nC_ox^o-{}^nC_1x^1+{}^nC_2x^2 - \text{---------}+(-1)^r\,{}^nC_rx^r+\text{------}+{}^nC_{n-1}(-1)^{n-1} \quad +(-1)^n \quad {}^nC_nx^n.$

or

$$(1+x)^n = \sum_{r=0}^{n} (-1)^r \binom{n}{r} x^r$$

Properties of Binomial – Expansion $(x+y)^n$:

(i) There are (n+1) terms in the expansion.

(ii) In each term, sum of the indices of 'x' and 'y' is equal to 'n'.

(iii) In any term, the lower suffix of 'c' is equal to the index of 'y', and the index of x = n-(lower suffix of c).

(iv) Because $\mathbf{{}^{n}C_{r} = {}^{n}C_{n-r}}$,

so we have :

${}^{n}C_{o} = {}^{n}C_{n}$

${}^{n}C_{1}={}^{n}C_{n-1}$

${}^{n}C_{2}={}^{n}C_{n-2}$ etc.

It follows that the coefficients of terms equidistant from the beginning and the ends **are equal.**

Example:-

(1) Simplify $(x+v(x^2-1)) + (x- v(x^2-1))^6$

Solution: let $vx^2-1 = a$, so we have:

$(x=a)^6 + (x-a)^6$

$= [x^6+{}^6C_1x^5.a+{}^6C_2.x^4.a^2 + {}^6C_3x^3a^3 + {}^6C_4x^2a^4 + {}^6C_5xa^5 +{}^6C_6a^6]$

$+ [x^6-{}^6C_1x^5a+{}^6C_2.x^4.a^2 – {}^6C_3x^3a^3 + {}^6C_4x^2a^4 – {}^6C_5xa^5 +{}^6C_6a^6]$

$= 2[x^6+{}^6C_2x^4a^2+{}^6C_4x^2a^4+{}^6C_6a^6]$

$= 2[x^6+15x^4(x^2-1)+15x^2(x^2-1)^2+(x^2-1)^3]$

$= 2[x^6+15x^6-15x^4+15x^6+15x^2-30x^4+x^6-1-3x^4+3x^3]$

$= 2[32x^6-48x^4+18x^2-1]$

General Terms : (r +1) th term from beginning in

$(x+y)^n$ is called general – term, and

it is denoted by

$$\mathbf{T_{r+1} = {}^{n}C_{r}x^{n-r}y^{r}}$$

Explanation: We know

$(x-y)^n = {}^nC_ox^ny^o+{}^nC_1x^{n-1}y^1+{}^nC_2x^{n-2}y^2+----+{}^nC_nx^oy^o$

Here:

First term $T = {}^{n}C\,x^{n}y^{o}$

$T_2 = {}^nC_1 x^{n-1} y^1$

$T_3 = {}^nC_2 x^{n-2} y^2$

$T_r = {}^nC_{r-1}\ x^{n-(r-1)}\ y^{r-1}$

Putting r = r+1 in this expression, we get:

General Term: $T_{r+1} = {}^nC_r\, x^{n-r}\, y^r$

Note : ‘T_r’ can be used as general terms also.

Problem based on General Terms

Type I:

Example: Find the 7th term in the expansion of

$[4x - (1 / 2vx)]^{13}$

Sol : $T_7 \quad = T_{6+1} = \ {}^{13}C_6(4x)^{13-6} \quad - (1/2vx)^{6}$

$= \ {}^{13}C_6.4^7x^7. \quad 1/(2^6.x^3)$

$= \ {}^{13}C_6.\ 2^8.x^4$

$= \ 13!/(6!x7!) \quad .\ 2^8.\ x^4$

$= \ T_7 = 439296x^4$

Type II : Find the coefficient of x-7 in the expansion of $(ax - \ 1/bx^2)^{\ 11}$

Sol.: General Term , $T_{r+1} = {}^{11}C_r(ax)^{11-r} \quad - \ (1/ bx^2\ )^{\,r}$

$T_{r+1} \ = (-1)^r\, {}^{11}C_r.\ (a^{11-r} / b^r)\ \ x^{11-3r}$ -------------> (i)

Putting $11 - 3r = -7$

Or $3r = 18$

⇨ $r = 6$

From (i) to $T_7 = (-1)^6\, {}^{11}C_6.\ (\ a^5 / b^6)\ \ x^{-7}$ --------------> (i)

Hence, the coefficient of x^{-7} in ax- $(1 / b\, x^2)^{11}$ is ${}^{11}C_6 a^5 b^{-6}$

Type III : Find the term independent of 'x' in $[(3\,x^2 / 2) - (1/ 3x)]^9$

Sol.: General Term, T_{r+1} $= {}^9C_r\ (3\,x^2 / 2)^{9-r} - (1/3x)^{r}$

$= (-1)^r\, {}^9C_r\, (3/2)^{9-r} x^{18-2r} \quad (1 / 3^r. x^r)$

$T_{r+1} = (-1)^{r9}\, C_r\, (3^{9-2r} / 2^{9-r}).\ x^{18-3r}$ -------> (i)

Putting 18- 3 r = o

⇨r = 6

So, from (i), 7th term is independent of 'x', and its value is:

$T_7 = (-1)^6 .\ {}^9C_6.\ (3^{-3} / 2^3)\, x^o$

= 9 ! /(6! X 3!) . $1/ (3^3 \times 2^3)$

= $T_7 =$ (7/18)

P^{th} term from end:

'P'th term from end in the expansion of $(x+y)^n$ is (n-P+2)th term from beginning.

Example: .: Find the 4th term from the end in the expansion of $[\,(x^3/2) - (2/x^2)\,]^7$

Sol.: 4th term from end = (7-4+2)th or 5th term from beginning.

$T_5 = T_{4+1} = {}^7C_4\ (x^3/2)^{7-4} . (-2/x^4)^4$

$= {}^7C_4\ (x^3 /2)^3 \quad (-2/ x^2)^4$

= 7! / (4! X 3!) . $(x^9/8) . (16/ x^8)$

= (7.6.5 / 3.2.1) .2x

$T_5 = 70x$

Hence '4' term, from the end = 70x.

Middle Terms: It depends upon the value of 'n'.

Case -1 : When 'n' is even, then total number of terms in $(x+y)^n$ is odd. So there is only one middle term i.e. [(n/2) + 1] th them is the middle term.

So we find $(T_{n+1}/2)$. th term in this case, if 'n' is even.

Case II : When 'n' is odd, then total number of terms in $(x+y)^n$ is even. So there are two middle terms i.e. (n+1) /2 th and (n+3) /2 th are true middle terms.

so we find $T_{(n+1)/2}$ th and $T_{(n+3)/2}$ th in this case if 'n' is odd.

Example: Find the middle – term in the expansion of $[\ 3x - (x^3 / 6)]^9$

Solution.: Here total no. of terms are 10 (even). So there are true middle-terms

i.e (9+1) / 2 th and (9+3) / 2 th. So we have to find – out 'T_5' and 'T_6'.

$T_5 \quad = \ T_{4+1} = {}^9C_4(3x)^{9-4} \ (-x^3 / 6)^{\ 4}$

$= \quad 9! / (4! \ X \ 5!) \qquad .3^5 x^5 \ (x^{12} / 6^4)$

$= (9.8.7.6 / 4.3.2.1) \quad 3^5 / (2^4 \ x \ 3^4) \ x^{17}$

$T_5 \quad = \ (189 / 8) \quad x^{17}$

$T_6 \quad = \ T_{5+1} = {}^9C_5(3x)^{9-5} \ (-x^3 / 6)^{\ 5}$

$= \quad 9! / (5! \ X \ 4!) \qquad .3^4 x^4 \quad (x^{15} / 6^5)$

$= -(9.8.7.6 / 4.3.2.1) \quad 3^4(2^5 \ x \ 3^5) \ x^{19}$

$T_6 \quad = - (\ 21 / 16)\ x^{19}$

Greatest – term in $(1+x)^n$: If 'T_r' and 'T_{r+1}' be the 'r' th and (r+1)th terms in the

Expansion of $(1+x)^n$, then :

$T_{r+1} \qquad = {}^nC_r(1)^{n-r}\ x^r = {}^nC_r\ x^r$

And $T_r \qquad = {}^nC_{r-1}.\ x^{r-1}$

So: $T_{r+1} / T_r \ = ({}^nC_r \ x^r / {}^nC_{r-1}\ x^{r-1}) \quad = (n-r+1)/r \ |x|$

If 'T_{r+1} be the greatest term, then $T_{r+1}\ T_r$

Or $T_{r+1} / T_r \quad 1$

since $(n-r+1) / r.\ |x| \ >= 1$, where 'r' is a '+' ve integer.

This inequality, changes either to the form r<=m+f pr r <= m, where 'm' is a '+' ve integer and 'f' is a fraction. So we get:

r <= m + f ---------------> (i)

or r <= m ------------------> (ii)

In case (i), 'T'$_{m+1}$ is the greatest term, and in case (i) 'T'$_m$ and 'T'$_{m+1}$ are the greatest terms, and both are equal.

Short-cut: First calculate $m \ = \ |\ x\ (n+1) / (x + 1)\ |$

Case (1) If ‘m’ is an integer, then ‘T’$_{m}$ and ‘T’$_{m+1}$ are the greatest terms and both are equal.

Case (2) If ‘m’ is not an integer, then $T_{[m]+1}$ will be the greatest term, where [.] denotes greatest integer function.

Example: Find numerically the greatest term in the expansion of (2+3x),

when x = (3 / 2)

solution 1 Method : $(2+3x)^{9} = 2^{9}\,[1+ 3x / 2]^{9}$

In the expansion of $[(1 + 3x) / 2]^{9}$, we have :

$T_{r+1} / T_{r} = \quad (9-r+1)/ r \quad |3x / 2|$

$= ((10 - r)/r) \quad |\,(3/2)\, x(3/2)\,|^{3}$

$= (10 - r) / r \;\; x \;\; 9 / 4$

$T_{r+1} / T_{r} = (90- 9r) / 4r$

Putting $T_{r+1} / T_{r} >= 1$

⇨$(90-9r) / 4r >= 1$

or $90 >= 13\, r$

or $r \quad 90 / 13$

or $r <= 6 + 12 / 13$

⇨T_{6+1} or ‘T_{7}’ is the greatest term.

‘T_{7}’ in $[1 + (3x / 2)]^{9}$

$T_{7} = T_{6+1} = {}^{9}C_{6}\, (3x / 6)^{6}$

$= 9! / (3!\ X\ 6!)\,.[\,(3 / 2) \;\; x \;\; (3 / 2)]^{6}$

$= (9\,.8.7 / 3.2.1) \;\; x \;\; (9^{6} / 4^{6})$

$= (3 \;\; x \;\; 7\ x\ 9^{6}) / 4^{5} \quad = (3\ x\ 7\ x\ 3^{12}) / 2^{10}$

$= 7.\ (3^{13} / 2^{10})$

So greatest term in $(2+ 3x)^{9}$ is :

$= 2^{9}.\ 7.\ (3^{13} / 2^{10})$

$= (7\ x\ 3^{13}) / 2$

II- Method : $(2+3x)^9 = 2^9 [(1 + 3x) / 2]^9$

$= 2^9 [1 + 9 / 4]^9$

since $x = 3 / 2$

Here $m = | x (n + 1) / (x + 1)| = | 9/4 (9+1) / 9/4 + 1|$

$= 90 / 13$

So greatest term in the expansion is $T_{[m]+1} = T_{3+1} = T_7$

Now the method is same as in method (1)

Greatest Coefficient : In any binomial expansion middle-term has the greatest.

Coefficient. So

(i) If 'n' is even, then greatest – coefficient $= {}^{n}C_{n/2}$

(ii) If 'n' is odd, then greatest – coefficients are ${}^{n}C_{(n+1)/2}$ **and** ${}^{n}C_{(n-1)/2}$

Properties of Binomial coefficients :

(1) The sum of binomial coefficient in $(1 + x)^n$ is 2^n.

Proof $(1 + x)^n = C_o+C_1x+C_2x^2 + \text{-----} + C_nx^n$-----------> (i)

Putting $x = 1$:

$\mathbf{2^n = C_o + C_1 + C_2 + \text{-----------} + C_n}$ -----------> (ii)

Ex.: Prove that the sum of the coefficients in the expression $(1+x – 3x^2)^{2163}$

is '-1'.

Sol.: Putting $x = 1$ in $(1 + x – 3x^2)^{2163}$

Some of the coefficients $= (1 + 1 – 3)^{2163}$

$= (-1)^{2163} = -1$

(2) The sum of the coefficients of the odd-terms in $(1+x)^n$ is equal to the sum of coefficients of the even terms and each is equal to 2^{n-1}.

Proof: Putting $x = -1$, in eg(1) :

$O = C_o – C_1 + C_2 – C_3 + \text{------} + (-1)^nC_n$

and from (ii): $2^n = C_o + C_1 + C_2 +$ --------- $+ C_n$

Adding these eg^n:

$2^n = 2\ (C_o + C_2 + C_4 +$ ---------------$)$

or **$C_o + C_2 + C_4 +$ ------- $= 2^{n-1}$** ------------> (ii)

Subtracting these eg^n:

$2^n = 2\ (C1 + C3 + C5 +$ --------------$)$

or **$C_1 + C_3 + C_5 +$ ------- $= 2^{n-1}$** ------------> (iv)

From (iii) and (iv) :

$C_0 + C_2 + C_4 +$ ------- $= C_1 + C_3 + C_5 +$ ------- $= 2^{n-1}$

Example: Evaluate the sum of the $^8C_1 + {}^8C_3 + {}^8C_5 + {}^8C_7$

Solution.: since $^nC_1 + {}^nC_3 + {}^nC_5 + {}^nC_7 +$ -------- $= 2^{n-1}$

Here n = 8

$\Rightarrow {}^8C_1 + {}^8C_3 + {}^8C_5 + {}^8C_7 = (2^{8-1})$

$= 2^7$

$= 128$

($^8C_9$, $^8C_{11}$ etc. are not possible)

Some important results:

(i) In the expansion of $(1+x)^n$, coefficient of $x^r = {}^nC_r$

(ii) In the expansion of $(1-x)^n$, coefficient of $x^r = (-1)^r.\ {}^nC_r$

(iii) If ‘n’ is a negative integer or fraction, then

$(1+x)^n = 1 + (n / 1!)\ x + [\ n\ (n-1)/ 2!]\ x^2 + [n(n-1)(n-2) / 3!]\ x^3 +$ -------------

$+ [n(n-1)(n-2)$ -----------$(n-r+1) / r!]x^r +$ -------------- ∞

Here $|\ x\ | <1$, i.e. $-1<x<1$ is necessary for its validity.

(iv) In $(1+x)^n$, general – term $T_{r+1} = [n(n-1)(n-2)$ -------------$(n-r+1) / r!]$. x2

(v) $^nC_r + {}^nC_{r-1} = {}^{n+1}C_r$

(vi) $^nC_x = {}^nC_y \Rightarrow x = y$ or $x + y = n$

(vii) $^nC \quad / \quad {}^{n-1}C$

Multinomial theorem : .

(For a ‘+’ve integral index):

If nN, and x_1, x_2, x_3, --------x_m C, then

$(x_1 + x_2 + x_3 +$ ---------$+x_m)^n$ = ? n! / (n_1! n_2! ---n_m!) x_1^{n1}, x_2^{n2} ….x_m^{nm}

Where n_1, n_2, n_3 --------, n_m are non-negative integers, satisfying the condition

$n_1 + n_2 +$ -----------$+n_m = n$

Note: The coefficient of x_1^{n1}. x_2^{n2}. ---------x_m^{nm} in the expansion of

$(x_1 + x_2 + x_3 +$ ------------------- $+ x_m)^n$ is :

= n! / (n_1! x n_2! ---n_m!)

So, **general-term in $(a+b+c+d)^n$ = n! / (p! x q! x r! x s!). $a^p.b^q.c^r.d^s$.**

Where p+q+r+s = n, and p, q, r, s W.

(2) Number of terms in $(x_1 + x_2 + x_3 +$ --------- $+ {}^x m)^n$: $^{n+m-1}C_{m-1}$.

Example: Find the number of terms in the expansion of $(2x – 3y + 4z)^{100}$

Solution.: Number of terms = $^{100+3-1}C_{3-1} = {}^{102}C_2$

= 102 ! / (2! X 100!)

= (102 x 101) / (2 x 1) = 5151

General term of a multinomial – theorem :

T_{r+1} = n! / (n_1! x n_2! ---^{n}m!) x_1^{n1}. x_2^{n2} -----------x_m^{nm}

Example: Find the coefficient of $x^3\ y^4\ z^2$ in the expansion of $(2x – 3y + 4z)^9$.

Solution. General Term in $(2x – 3y + 4z)^9$

= 9! / (n_1! X n_2! X n_3!). $(2x)^{n1}$. $(-3y)^{n2}$. $(4z)^{n3}$

= 9! / (n_1! X n_2! X n_3!). $2^{n1}\ (-3)^{n2}.\ (4)^{n3}.\ x^{n1}.\ y^{n2}.\ z^{n3}$

Putting n1 = 3, n2 = 4, n3 = 2 :

= 9! / (3! x 4! x 2!). $2^3\ (-3)^4.\ (4)^2.\ x^3\ y^4\ z^2$

= [9. 8. 7. 6. 5. 4!/ (3.2.1. 4!.2)] x 8 x 81 x 16 $x^3\ y^4\ z^2$

➡Coefficient of $x^3\ y^4\ z^2$ = 9 x 8 x 7 x 5 x 8 x 81 x 8

= 13063600

Greatest coefficient in the expansion of $(x1 + x2 + \text{--------} + xm)^n$ is

= $n! / (q!)^{m-r} (q+1!)$

Where ‘q’ is the quotient and ‘r’ is the remainder, when ‘n’ is divided by ‘m’.

Example: Find the greatest coefficient in the expansion of $(a + b + c + d)^{15}$.

Solution: Here n = 15, m = 4

15/4 is quotient 3 and remainder 3.

since q = 3 and r = 3

Hence greatest – coefficient = $15! / [(3!)^{4-3} \times (3+1!)^3]$

= $15! / [(3!) \times (4!)^3]$

= $15! / (3! \times 4! \times 4! \times 4!)$

Example: Find the coefficient of x^7 in the expansion of $(1+3x-2x^3)^{10}$.

Solution: General term in $(1+3x-2x^3)^{10}$

= $10! / (n_1! \times n_2! \times n_3!). (1)^{n1} (3x)^{n2} (-2x^3)^{n3}$

= $10! / (n_1! \times n_2! \times n_3!). 3^{n2} (-2)^{n3} x^{n2+3n3}$

Where $n_1 + n_2 + n_3 = 10$ --------------> (i)

For coefficient of x^7 : $n_2 + 3n_3 = 7$ ------------> (ii)

From (ii), possible non-negative integral values of ‘n_2’ and ‘n_3’ are :

$n_2 = 7, n_3 = o$ since from (i) : $n_1 = 3$

$n_2 = 1, n_3 = 2$ since from(i) : $n_1 = 7$

or $n_2 = 4, n_3 = 1$ since from (i): $n_1 = 5$

So required coefficient of x^7 :

$10! / (3! \times 7! \times 0!) . (3)^7 (-2)^0 + 10! / (7! \times 1! \times 2!). (3)^1(-2)^2 + 10! /(5! \times 4! \times 1!). 3^4 (-2)^1$

$(10. 9. 8\ 7!) / (7!.3.2.1).3^7 + (10. 9. 8. 7!) / (7! . 2) \times 3 \times 4 - [(10. 9. 8. 7. 6. 5!) / (5!. 4.3.2.1.)] .3. 2.$

= $10 \times 9 \times 4 \times 3^6 + 10 \times 9 \times 4 \times 3 \times 4 – 10 \times 9 \times 7 \times 6 \times 3^3 \times 2$

= $10 \times 9 \times 4 (3^6 + 12 – 7 \times 3^4)$

= $360 \times (729 + 12 – 567)$

= 62640

Some tips on the solution of binomial – coefficients:

(1) If the difference of the lower suffixes of binomial coefficients in each term is same.

For Ex.: $C_1 C_3 + C_2 C_4 + C_3 C_5 +$ ------ etc.

Then :

Case -1 : If each term is positive, then

$(1+x)^n = C_0 + C_1x + C_2x^2 +$ ------------ $C_n x^n$ -----------------> (i)

Interchanging '1' and 'x':

$(x+1)^n = C_0 x^n + C_1 x^{n-1} + C_2 x^{n-2} +$ --------- $+ C_n$ ------------->(ii)

Then multiplying (i) and (ii), and equate the coefficient to suitable power of 'x' on both sides.

Case –II : If terms are alternately positive and negative

Then:

$(1-x)^n = C_0 – C_1 x + C_2 x^2 -$ -------------- $+ (-1)^n C_n x^n$ ---------------> (1)

and $(x+1)^n = C_0 x^n + C_1 x^{n-1} + C_2 x^{n-2} +$ ---------- $+ C_n$ ------------> (2)

The multiplying (1) and (2), and equate the coefficient of suitable power of 'x' on both sides.

Note : [(Odd – number) / 2] = 8

(2) If the sum of the lower suffixes of binomial – coefficients in each term is same.

For Ex.: $C_0 C_n + C_1 C_{n-1} + C_2. C_{n-2} +$ ------- $+ C_n C_0$

Then:

Case – 1 : If each term is positive, then

$(1+x)^n = C_0 + C_1 x + C_2 x^2 -$ -------------- $+ C_n x^n$ ---------------> (1)

and $(1+x)^n = C_0 x^n + C_1 x + C_2 x^2 +$ ---------- $+ C_n x^n$-------------> (2)

Then multiplying (i) and (ii), and equate the coefficient of suitable power of 'x' on both sides.

Case –II : If terms are alternately positive and negative,

The $(1+x)^n = C_0 + C_1 x + C_2 x^2 -$ -------------- $+ C_n x^n$ ---------------> (1)

and $(1-x)^n = C_0 - C_1 x + C_2 x^2 +$ ---------- $+ (-1)^n C_n x^n$-------------> (2)

Then multiplying (i) and (ii) and equating the coefficient of suitable power of 'x' on both side.

PROBLEMS

(1) Show that the middle – term in the expansion of $(1+x)^{2n}$ is

$1.\ 3.\ 5$ ------- $(2n-1) / (n!)\ .\ 2^n x^n$, 'n' being a positive integer.

Solution: The no. of terms in $(1+x)^{2n}$ = $2n+1$ (odd).

It's ,middle-term = $(2n + 1) / 2$ = $(n+1)$th term.

⇨T_{n+1} = ${}^{2n}C_n\, x^n$

= $2n! / (n! \times n!).\ x^n$

= $2n\,(2n-1)$ ------ $4.3.2.1 / (n! \times n!).\ x^n$

= $[\{(2n-1)(2n-3)$ ----- $3.1.\}\ \{2n\,(2n-2)$ ------ $4.2.\}] / (n! \times n!).\ x^n$

= $[\{1.\ 3.\ 5.$ ---- $(2n-1)\}\ 2^n\ \{1.2$ ---- $n\}] / (n! \times n!)\ .\ x^n$

= $[\{1.3.5$----$(2n-1)\}.\ 2^n] / (n! \times n!).\ x^n$

T_{n+1} = $1.\ 3.\ 5 - (2n-1) / (n!).\ \ 2^n x^n$

(2) Find the term independent of 'x' in the expansion of

(i) $(1+x+2x^3)\ \ [(3x^2 / 2) - (1/3x)]^9$

Solution: (i) $(1+x+2x^3)\ [(3/2)x^2 - (1/3x)]^9$

= $(1+x+2x^3)\ \{ [(3/2)x^2]^9 - {}^9C_1 [(3/2)x^2]^8\ 1/3x\ +$ ---------- $+$

$+\ {}^9C_6 [(3/2)x^2]^3 (1/3x)^6 - {}^9C_7 [(3/2)x^2]^2 (1/3x)^7$ ---$\}$

= $(1+x+2x^3)\ \{ [(3/2)x^2]^9 - {}^9C_1 (3^7 / 2^8)x^{15} +$ ---- $+\ {}^9C_6 (1 \times 1 / 2^3 \times 3^3) - {}^9C_7\, 1/ (2^2 \times 3^5)\ 1/ x^3\ +$ ---- $\}$

Term independent of 'x' :

${}^9C_6 \times\ 1/ (2^3 \times 3^3) - {}^9C_7\ \ 2 / (2^2 \times 3^5)$

= $9! / (6! \times 3!)\ .\ 1/ (8 \times 27) - 9!/(7! \times 2!)\ .\ \ 1/ (2 \times 243)$

= $(9.\ 8.\ 7.\ 6!) / (6!.\ 3.\ 2.\ 1.) \times 1/(8.\ 27) - (9.\ 8.\ 7!) / (7!\ .\ 2).\ 1/ (2.243)$

= $7 / 18 - 2 / 27 = 17 / 54$

(3) Find the coefficient of 'x' in the expansion of $(1-2x^3 + x^5)\ [1 + (1/x)]^8$

Solution: $(1-2x^3 + 3x^5)\ [1 + (1/x)]^8$

$= (1-2x^3 + 3x^5)\ [1 + {}^8C_1 (1/x) + {}^8C_2\ (1/ x^2) + {}^8C_3 (1/ x^3) + {}^8C_4 (1/ x^4) + {}^8C_5 (1/ x^5) +$ --- $+\ {}^8C_8 (1/ x^8)$

coefficient of x = $-2.\ {}^8C_2 + 3\ {}^8C_4$

$= -2.\ \ 8! / (2! \times 6!) +\ 3.\ \ 8! / (4! \times 4!)$

$= -2. \ (8.\ 7) / 2 \quad + \ 3 \quad (8.\ 7.\ 6.\ 5.) / (4.3.2.1)$

$= -56 + 210$

$= 154$

(4) Prove that the ration of the coefficient of x^{10} in $(1-x^2)^{10}$ and the term independent of ‘x’ in $[x – (2/x)]^{10}$ is 1 : 32.

Solution: In $(1-x)^2$: $T_{r+1} = {}^{10}C_r (-1)^r (x^2)^r$

Putting $r = 5$

$T_6 = -{}^{10}C_5 \, x^{10}$

⇨Coefficient of $x^{10} = -{}^{10}C_5$

In $[x – (2/x)]$: $T_{r+1} = {}^{10}C_r (-1)^r (x)^{10-r} (2/x)^r$

$= (-1)^r \, {}^{10}C_r .\ 2^r .\ x^{10-2r}$

Putting $10 – 2r = 0$

⇨ $r = 5$

So term independent of x : $T_6 \quad = (-1)^5 \, {}^{10}C_5 .\ 2^5$

Hence their ratio $\quad = (-{}^{10}C_5) : (-32.\ {}^{10}C_5)$

$= 1 : 32$

or $2z^2 + 3z – 5 = 0$

(7) If in the expansion of $(1+x)^m \ (1-x)^n$, the coefficients of ‘x’ and ‘x^2’ are ‘3’ and ‘-6’ res. Find the value of ‘m’.

Solution: $(1+x)^m \ (1-x)^n = [{}^mC_0 + {}^mC_1x + {}^mC_2x^2 + ---- + \ {}^mC_m \, x^m]$

$[{}^nC_o – {}^nC_1x + {}^nC_2x^2 + ------- + \ (-1)^n \, {}^nC_nx^n]$

Coefficient of x $= {}^mC_1 \text{ x } {}^nC_o – {}^mC_0 .\ {}^nC_1$

$= \quad m! / (1! \text{ x } m-1!) \ \text{ x } 1 – 1 \text{ x } n! / (1! \text{ x } n-1!)$

$= m – n = 3$ ---------------> (i)

Coefficient of $x^2 \quad = -{}^mC_1 \text{ x } {}^nC_1 + {}^nC_0 \text{ x } {}^mC_2 + {}^mC_0 \text{ x } {}^nC_2$

$= \ - \quad m!/ (1! \text{ x } m-1!) \text{ x } n! / (1! \text{ x } n-1!) + 1 \text{ x } m! / (2! \text{ x } m-2!) \quad + 1 \text{ x } n! / (2! \text{ x } n-2!)$

$= -mn + m (m-1) / 2 \ + \ n(n-1) / 2 \ = \ -6$

or $– 2mn + m(m-1) + n(n-1) \ = \ -12$

or $2mn + m^2 \ \ m + n^2 \ \ n = 12$

or $(m-n)^2 - (m+n) = -12$

From (i), putting the value of (m-n) :

$- 9 + (m + n) = 12$

or $m + n = 21$ -----------> (ii)

eg^n (i) + eg^n(ii) $= 2m = 24$

$m = 12$

Q11. Find the coefficient of x^r in the expansion of $[x + (1/x)]^n$, if it occurs.

Solution: General term : $T_{p+1} = {}^nC_p (x)^{n-p} (1/x)^p$

$T_{p+1} = {}^nC_p x^{n-2p}$ ---------------> (i)

Putting $n-2p = r$

⇨$p = (n - r) / 2$

From: (i) $T_{(n-r)/2 +1} = {}^nC_{(n-r)/2} x^r$

Coefficient of $x^r = {}^nC_{(n-r)/2}$

Q. Let ‘n’ be a positive integer. If the coefficients of second, third and fourth terms in $(1+x)^2$ are in arithmetic progression, then find the value of ‘n’.

Solution: General Term : $T_{r+1} = {}^nC_r \ x^r$

⇨2nd Term : $T_2 = {}^nC_1 x$

Coefficient $= {}^nC_1$

⇨3rd Term : $T_3 = {}^nC_2 x^2$

Coefficient $= {}^nC_2$

Similarly coefficient of 4th term = nC3

These are in A. P., so.

$2\ {}^{n}C_2 = {}^{n}C_1 + {}^{n}C_3$

⇨ $2 [n! / \{2! \text{ x } (n-2)!\}] = n! / \{1! \text{ x } (n-1!)\} + n! / \{3! \text{ x } (n-3!)\}$

or $n! / (n-2!) = n! [1 / (n-1!) + 1 / \{6 (n-3!\}]$

or $1/ [(n-3) \text{ x } (n-3!) = 1 / [(n-1)\text{x } (n-2) \text{ x}(n - 3!)] +1/ [6! (n-3!)])$

or $1 / (n - 2) - 1 / [(n-1) (n-2)] = 1 / 6$

or $(n - 1 - 1) / [(n-1) (n-2)] = 1 / 6$

or $(n- 2) / [(n-1) (n-2)] = 1 / 6$

or $n - 1 = 6$

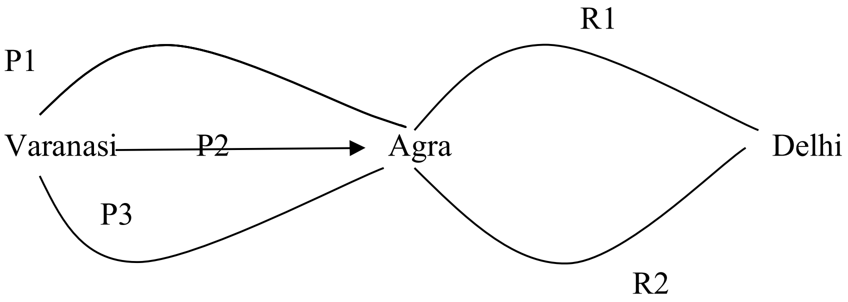
⇨$n = 7$

Principle of counting

There are two principle of counting:-

1) **Multiplication or AND principle :-**

   If the two works w1 and w2 individually can perform in m1 and m2 wave respectively then both work can be perform simultaneously in **m1.m2** ways.
   This principle is also known as **Product Rule.**
   For e.g. Consider the following diagram.



As shown in the diagram there are three routes from Varanasi to Aga and also there are two routes from Agra to Delhi. Thus, a person can perform a journey from Varanasi to Agra in 3 ways y selecting either of the routes P1,P2 and P3.Similarly, he can perform the journey Agra to Delhi in 2 ways as R1 and R2 at a time.

Therefore, By Product Rule the no of ways in which the person can perform the journey from Varanasi to Agra and from Agra to Delhi is 3×2.

2) **Addition or OR principle :-**

   According to this principle the two works either of them at a time can be performing in m1+m2 ways .This principle is also known as **Sum Rule**.
   For e.g. in the above example the now ways are performing the journey 3+2= 5 ways.

Combinations, Permutations, and Counting Events

The solution to many statistical experiments involves being able to count the number of points in a sample space. Counting points can be hard, tedious, or both.

Fortunately, there are ways to make the counting task easier. This lesson focuses on three rules of counting that can save both time and effort - event multiples, permutations, and combinations.

Event Multiples

The first rule of counting deals with event multiples. An **event multiple** occurs when two or more *independent* events are grouped together. The first rule of counting helps us determine how many ways an event multiple can occur.

Rule 1. Suppose we have k independent events. Event 1 can be performed in n_1 ways; Event 2, in n_2 ways; and so on up to Event k (which can be performed in n_k ways). The number of ways that these events can be performed together is equal to $n_1 n_2 \ldots n_k$ ways.

Example 1
How many sample points are in the sample space when a coin is flipped 4 times?

Solution: Each coin flip can have one of two outcomes - heads or tails. Therefore, the four coin flips can land in (2)(2)(2)(2) = 16 ways.

Example 2
A business man has 4 dress shirts and 7 ties. How many different shirt/tie outfits can he create?

Solution: For each outfit, he can choose one of four shirts and one of seven ties. Therefore, the business man can create (4)(7) = 28 different shirt/tie outfits.

Permutations

Often, we want to count all of the possible ways that a single set of objects can be arranged. For example, consider the letters X, Y, and Z. These letters can be arranged a number of different ways (XYZ, XZY, YXZ, etc.) Each of these arrangements is a permutation.

- In general, *n* objects can be arranged in $n(n - 1)(n - 2) \ldots (3)(2)(1)$ ways. This product is represented by the symbol *n*!, which is called **n factorial**. (By convention, 0! = 1.)
- A **permutation** is an arrangement of all or part of a set of objects, *with* regard to the order of the arrangement. This means that XYZ is considered a different permutation than ZYX.
- **The number of permutations of *n* objects taken *r* at a time is denoted by $_nP_r$.**

Rule 2. The number of permutations of *n* objects taken *r* at a time is

$$_nP_r = n(n - 1)(n - 2) \ldots (n - r + 1) = n! / (n - r)!$$

Example 1
How many different ways can you arrange the letters X, Y, and Z? (Hint: In this problem, order is important; i.e., XYZ is considered a different arrangement than YZX.)

Solution: One way to solve this problem is to list all of the possible permutations of X, Y, and Z. They are: XYZ, XZY, YXZ, YZX, ZXY, and ZYX. Thus, there are 6 possible permutations.

Another approach is to use Rule 2. Rule 2 tells us that the number of permutations is n! / (n - r)!. We have 3 distinct objects so n = 3. And we want to arrange them in groups of 3, so r = 3. Thus, the number of permutations is 3! / (3 - 3)! or 3! / 0!. This is equal to (3)(2)(1)/1 = 6.

Example 2
In horse racing, a trifecta is a type of bet. To win a trifecta bet, you need to specify the horses that finish in the top three spots in the exact order in which they finish. If eight horses enter the race, how many different ways can they finish in the top three spots?

Solution: Rule 2 tells us that the number of permutations is n! / (n - r)!. We have 8 horses in the race. so n = 8. And we want to arrange them in groups of 3, so r = 3. Thus, the number of permutations is 8! / (8 - 3)! or 8! / 5!. This is equal to (8)(7)(6) = 336 distinct trifecta outcomes. With 336 possible permutations, the trifecta is a difficult bet to win.

Combinations

Sometimes, we want to count all of the possible ways that a single set of objects can be selected - without regard to the order in which they are selected.

- A **combination** is a selection of all or part of a set of objects, *without* regard to the order in which they were selected. This means that XYZ is considered the same combination as ZYX.
- The number of combinations of *n* objects taken *r* at a time is denoted by ${}_nC_r$.

Rule 3. The number of Combinations of *n* objects taken *r* at a time is

$$\mathbf{{}_nC_r = n(n - 1)(n - 2) \ldots (n - r + 1)/r! = n! / r!(n - r)! = {}_nP_r / r!}$$

Example 1
How many different ways can you select 2 letters from the set of letters: X, Y, and Z? (Hint: In this problem, order is NOT important; i.e., XY is considered the same selection as YX.)

Solution: One way to solve this problem is to list all of the possible selections of 2 letters from the set of X, Y, and Z. They are: XY, XZ, and YZ. Thus, there are 3 possible combinations.

Another approach is to use Rule 3. Rule 3 tells us that the number of combinations is n! / r!(n - r)!. We have 3 distinct objects so n = 3. And we want to arrange them in groups of 2, so r = 2. Thus, the number of combinations is 3! / 2!(3 - 2)! or 3! /2!1!. This is equal to (3)(2)(1)/(2)(1)(1) = 3.

Example 2
Five-card stud is a poker game, in which a player is dealt 5 cards from an ordinary deck of 52 playing cards. How many distinct poker hands could be dealt? (Hint: In this problem, the order in which cards are dealt is NOT important; For example, if you are dealt the ace, king, queen, jack, ten of spades, that is the same as being dealt the ten, jack, queen, king, ace of spades.)

Solution: For this problem, it would be impractical to list all of the possible poker hands. However, the number of possible poker hands can be easily calculated using Rule 3.

Rule 3 tells us that the number of combinations is n! / r!(n - r)!. We have 52 cards in the deck so n = 52. And we want to arrange them in groups of 5, so r = 5. Thus, the number of permutations is 52! / 5!(52 - 5)! or 52! / 5!47!. This is equal to 2,598,960 distinct poker hands.

Hint: Although this problem is not hard to solve, it does involve a lot of computations. You can use Stat Trek's Event Counter to do the actual computations. The tool is free, and it reduces the drudgery.

Inclusion-Exclusion

The principle of Inclusion-Exclusion generalized the sum rule the the case of non-empty intersection:

INCLUSION-EXCLUSION: If A and B are sets, then

$$|A \cup B| = |A|+|B| - |A \cap B|$$

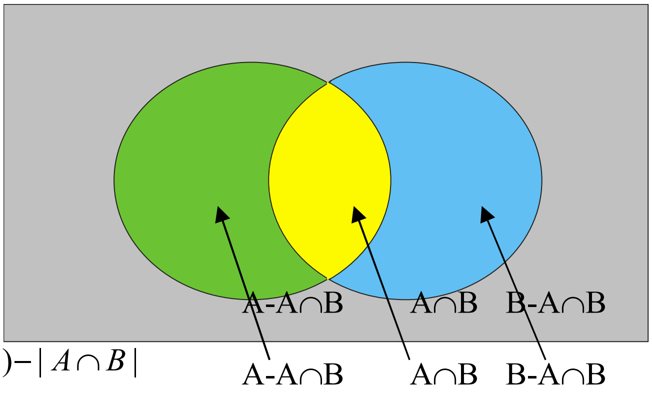
This says that when counting all the elements in A or B, if we just add the the sets, we have double-counted the intersection, and must therefore subtract it out.

Inclusion-Exclusion

Visualize Diagram

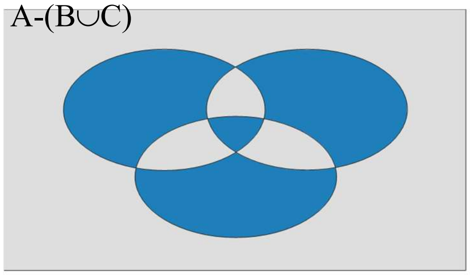
gives proof

Inclusion-Exclusion principle:



$$\begin{aligned}|A \cup B| &= |A - A\cap B| + |A\cap B| + |B - A\cap B| \\ &= (|A - A\cap B| + |A\cap B|) + (|B - A\cap B| + |A\cap B|) - |A\cap B| \\ &= |A| + |B| - |A\cap B|\end{aligned}$$

Inclusion-Exclusion-Inclusion” principle:



$$|A\cup B\cup C| = |A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|+|A\cap B\cap C|$$

Proof of Inclusion-Exclusion-Inclusion

$$\begin{aligned}&|A\cup B\cup C| \\ &= 1+2+3+4+5+6+7 \\ &= (1+4+5+7)+(2+4+6+7)-4-7+(3+5+6+7)-4-6-7 \\ &= |A|+|B|+|C|-(4+5+6+7+7) \\ &= |A|+|B|+|C|-([4+7]+[5+7]+[6+7]-7) \\ &= |A|+|B|+|C|-(|A\cap B|+|A\cap C|+|B\cap C|-|A\cap B\cap C|) \\ &= |A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|+|A\cap B\cap C|\end{aligned}$$

Counting Derangements

Definition: A derangement of {1,2,3,…,n} is a permutation f on the set such that for no element i does f (i) = i.

So the answer to the witch problem is:

|{ derangements of {1,2,3,4} }| / |{permutations}|

Define: Ai = {permutations which bring i to i }

Inclusion-Exclusion and symmetry imply:

|{witch←→poison derangements}|

= |{4-perm’s}|-C (4,1)|A1|+C (4,2)|A1∩A2|

-C (4,3)|A1∩A2∩A3|+C (4,4)|A1∩A2∩A3∩A4|

$= 4! - C(4,1)3! + C(4,2)2! - C(4,3)1! + C(4,4)0!$

$= 4! - 4\cdot 3!/1! + 4\cdot 3\cdot 2!/2! - 4\cdot 3\cdot 2\cdot 1!/3! + 4\cdot 3\cdot 2\cdot 1/4!$

$$= 4!\left(1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}\right)$$

Finally, divide the number of derangements by the number of permutations to get the probability that all die:

$$= 4!\left(1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}\right)/4!$$

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

$$= \frac{1}{2}-\frac{1}{6}+\frac{1}{24} \approx 0.375$$

Counting Derangements General Formula

THM: The number of derangements of a set with n elements is given by:

$$D_n = n!\left(1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}+\cdots+(-1)^n\frac{1}{n!}\right)$$

The proof is just a generalization of the argument in the witch party problem.

Marriage Problem

Latin squares could be used by dating services to organize meetings between a number n of girls and the same number n of boys. Having met all the boys, each girl come up with a list of boys she would not mind marrying. The dating service is faced now with the task of arranging marriages so as to satisfy each girl preferences. Call the set of boys listed by the i-th girl A_i. The problem is then to pick boys, one from each list, without selecting the same boy more than once. An abstract formulation of this problem and its constructive solution by W. McWorter, Jr. is given below.

Let $A_1, ..., A_n$ be subsets of an n-set M. The family of sets $A_1, ..., A_n$ has a system of distinct representatives (SDR) iff there exist distinct elements $x_1, ..., x_n$, such that x_i is in A_i, for each i = 1, ..., n.

Let $A_1, ..., A_n$ be subsets of an n-set M. The family of sets $A_1, ..., A_n$ satisfies the marriage condition iff the union of any k of the sets contains at least k elements, for all k = 1, ..., n.

MARRIAGE THEOREM

A family $A_1, ..., A_n$ of sets has a system of distinct representatives iff the family satisfies the marriage condition.

Remark

Strictly speaking, the proof below does not require the sets of boys and girls to be equipotent. However, this is a sensible assumption. For, if there are fewer boys the marriage condition fails. On the other hand, if there are fewer girls, the difference can be made up by "desperate" girls willing to marry any boy without disturbing the truth or falsity of the marriage condition.

Proof

I will prove this theorem by describing a procedure for constructing a system of distinct representatives which succeeds iff the marriage condition is satisfied. Another elegant, though less procedural, proof illuminates various phases of the algorithm. It is convenient for me to describe everything in terms of zero-one matrices. Let $A = (a_{ij})$ be an n×n matrix whose columns are labelled by the sets $A_1, ..., A_n$ and whose rows are labelled by the elements of M (which is the union of the A_i's). For each i, j = 1, ..., n, set $a_{ij} = 1$ if the i^{th} element of M is in A_j and set $a_{ij} = 0$ otherwise.

Then the family has a SDR iff the matrix A has a transversal of 1's. Here is the procedure together with a proof that it constructs a transversal of 1's iff the family satisfies the marriage condition. Assume matrix A is in the form below, where, at the beginning, the upper left block with 1's on the diagonal may or may not be empty. Successive steps of the algorithm are designed to increase the size of the block.

Marriage Theorem, Solutions to Problems

Problem # 1

Consider a matrix with 1s in squares next to the diagonal and 0s everywhere else, $a_{ij} = 1$ iff $|i - j| = 1$, and $a_{ij} = 0$, otherwise. Prove that if n, the size of the matrix, is even, the complete match possible. The complete match is impossible if n is odd.

Solution #1 (by W.McWorter)

For n even, the matrix contains a diagonal of blocks of 2 by 2 permutation matrices. The 1's in these matrices make a complete matching. For n odd, the (n+1)/2 columns with odd indices together only contain 1's from the (n-1)/2 rows with even indices. Hence the marriage condition is violated because (n-1)/2 < (n+1)/2. The n even case shows that a maximal matching with only one girl going unmarried exists for n odd.

Solution #2 (by W.McWorter)

The problem can also be solved by row and column changes. For n > 0 even, interchange rows 2i-1 and 2i, for i = 1, ..., n/2. The main diagonal is now all 1's. For n>1 odd, do the same thing. Interchange rows 2i-1 and 2i, for i = 1, ..., (n-1)/2. You get the picture

```
1 0 1               0
0 1 0    . . .
   .  .    .
     .  .
       . . 0
        1 0 1
0 . . 1 0 1 0
0 . .   0 1 0
```

None of the 1's below the main diagonal correspond to a 1 in the last column, so no way for the algorithm to get a 1 to percolate (from under the G submatrix) into the lower right corner. Hence there is no complete matching.

Solution #3

We may apply induction. Depending on the parity of n induction starts either with n = 1 or 2. For the inductive step, match girls {1, 2} with boys {1, 2}. Remove the columns 1, 2 and rows 1, 2. What remains is a matrix of exactly the same kind (bidiagonal) with the order reduced by 2.

Solution #4

Apply Konig's Theorem: The minimal number of lines that could be drawn through all the 1s equals the maximal number of marriages that can be arranged. For a 3×3 matrix one only needs 2 lines - through 2nd column and row. By induction, as before, for a (2n+1)×(2n+1) matrix, it will take only 2n lines to cover all 1s.

Example 1: Consider $S = \{A_1, A_2, A_3\}$ with

$A_1 = \{1, 2, 3\}$
$A_2 = \{1, 4, 5\}$

$A_3 = \{3, 5\}$.

A valid SDR would be {1, 4, 5}. (Note this is not unique: {2, 1, 3} works equally well, for example.)

Example 2: Consider $S = \{A_1, A_2, A_3, A_4\}$ with

$A_1 = \{2, 3, 4, 5\}$
$A_2 = \{4, 5\}$
$A_3 = \{5\}$
$A_4 = \{4\}$.

No valid SDR exists; the marriage condition is violated as is shown by the subcollection $\{A_2, A_3, A_4\}$.

Example 3: Consider $S= \{A_1, A_2, A_3, A_4\}$ with

$A_1 = \{a, b, c\}$
$A_2 = \{b, d\}$
$A_3 = \{a, b, d\}$
$A_4 = \{b, d\}$.

The only valid SDR's are (c, b, a, d) and (c, d, a, b).

Standard example with men and women

The standard example of an application of the marriage theorem is to imagine two groups; one of n men, and one of n women. For each woman, there is a subset of the men any one of which she would happily marry; and any man would be happy to marry a woman who wants to marry him. Consider whether it is possible to pair up (in marriage) the men and women so that every person is happy.

If we let A_i be the set of men that the i-th woman would be happy to marry, then the marriage theorem states that each woman can happily marry a man if and only if the collection of sets $\{A_i\}$ meets the marriage condition.

Note that the marriage condition is that, for any subset I of the women, the number of men whom at least one of the women would be happy to marry, $|\bigcup_{i\in I} A_i|$, be at least as big as the number of women in that subset, $|I|$. It is obvious that this condition is necessary, as if it does not hold, there are not enough men to share among the I women. What is interesting is that it is also a sufficient condition.

Probability

The **probability** of a sample point is a measure of the likelihood that the sample point will occur.

Probability of a Sample Point

By convention, statisticians have agreed on the following rules.

- The probability of any sample point can range from 0 to 1.
- The sum of probabilities of all sample points in a sample space is equal to 1.

Example 1
Suppose we conduct a simple statistical experiment. We flip a coin one time. The coin flip can have one of two outcomes - heads or tails. Together, these outcomes represent the sample space of our experiment. Individually, each outcome

Solution: The sum of probabilities of all the sample points must equal 1. And the probability of getting a head is equal to the probability of getting a tail. Therefore, the probability of each sample point (heads or tails) must be equal to 1/2.

Example 2
Let's repeat the experiment of Example 1, with a die instead of a coin. If we toss a fair die, what is the probability of each sample point?

Solution: For this experiment, the sample space consists of six sample points: {1, 2, 3, 4, 5, 6}. Each sample point has equal probability. And the sum of probabilities of all the sample points must equal 1. Therefore, the probability of each sample point must be equal to 1/6.

Probability of an Event

The probability of an event is a measure of the likelihood that the event will occur. By convention, statisticians have agreed on the following rules.

- The probability of any event can range from 0 to 1.
- The probability of event A is the sum of the probabilities of all the sample points in event A.
- The probability of event A is denoted by P(A).

Thus, if event A were very unlikely to occur, then P(A) would be close to 0. And if event A were very likely to occur, then P(A) would be close to 1.

Example 1
Suppose we draw a card from a deck of playing cards. What is the probability that we draw a spade?

Solution*:* The sample space of this experiment consists of 52 cards, and the probability of each sample point is 1/52. Since there are 13 spades in the deck, the probability of drawing a spade is

$$P(\text{Spade}) = (13)(1/52) = 1/4$$

Example 2
Suppose a coin is flipped 3 times. What is the probability of getting two tails and one head?

Solution: For this experiment, the sample space consists of 8 sample points.

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

Each sample point is equally likely to occur, so the probability of getting any particular sample point is 1/8. The event "getting two tails and one head" consists of the following subset of the sample space.

$$A = \{TTH, THT, HTT\}$$

The probability of Event A is the sum of the probabilities of the sample points in A. Therefore,

$$P(A) = 1/8 + 1/8 + 1/8 = 3/8$$

What is Probability?

The **probability** of an event refers to the likelihood that the event will occur.

How to Interpret Probability

Mathematically, the probability that an event will occur is expressed as a number between 0 and 1. Notationally, the probability of event A is represented by P(A).

- If P(A) equals zero, event A will almost definitely not occur.
- If P(A) is close to zero, there is only a small chance that event A will occur.
- If P(A) equals 0.5, there is a 50-50 chance that event A will occur.
- If P(A) is close to one, there is a strong chance that event A will occur.
- If P(A) equals one, event A will almost definitely occur.

In a statistical experiment, the sum of probabilities for all possible outcomes is equal to one. This means, for example, that if an experiment can have three possible outcomes (A, B, and C), then P(A) + P(B) + P(C) = 1.

How to Compute Probability: Equally Likely Outcomes

Sometimes, a statistical experiment can have *n* possible outcomes, each of which is equally likely. Suppose a subset of *r* outcomes are classified as "successful" outcomes.

The probability that the experiment results in a successful outcome (S) is:

P(S) = (Number of successful outcomes) / (Total number of equally likely outcomes)

= r / n

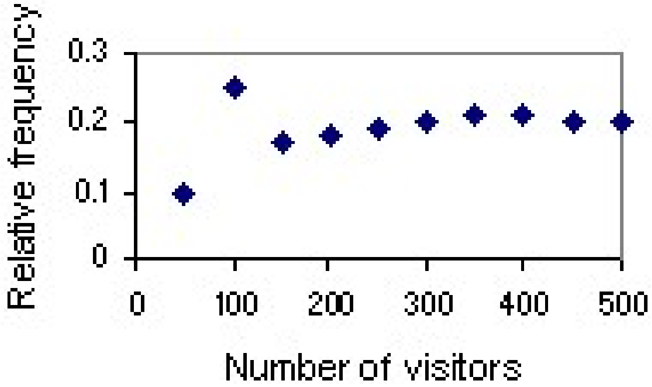
Consider the following experiment. An urn has 10 marbles. Two marbles are red, three are green, and five are blue. If an experimenter randomly selects 1 marble from the urn, what is the probability that it will be green?

In this experiment, there are 10 equally likely outcomes, three of which are green marbles. Therefore, the probability of choosing a green marble is 3/10 or 0.30.

How to Compute Probability: Law of Large Numbers

One can also think about the probability of an event in terms of its *long-run* relative frequency. The relative frequency of an event is the number of times an event occurs, divided by the total number of trials.

P(A) = (Frequency of Event A) / (Number of Trials)



For example, a merchant notices one day that 5 out of 50 visitors to her store make a purchase. The next day, 20 out of 50 visitors make a purchase. The two relative frequencies (5/50 or 0.10 and 20/50 or 0.40) differ. However, summing results over many visitors, she might find that the probability that a visitor makes a purchase gets closer and closer 0.20.

The scatterplot (above right) shows the relative frequency as the number of trials (in this case, the number of visitors) increases. Over many trials, the relative frequency converges toward a stable value (0.20), which can be interpreted as the probability that a visitor to the store will make a purchase.

The idea that the relative frequency of an event will converge on the probability of the event, as the number of trials increases, is called the **law of large numbers**.

Test Your Understanding of This Lesson

Problem

A coin is tossed three times. What is the probability that it lands on heads *exactly* one time?

(A) 0.125
(B) 0.250
(C) 0.333
(D) 0.375
(E) 0.500

Solution

The correct answer is (D). If you toss a coin three times, there are a total of eight possible outcomes. They are: HHH, HHT, HTH, THH, HTT, THT, TTH, and TTT. Of the eight possible outcomes, three have exactly one head. They are: HTT, THT, and TTH. Therefore, the probability that three flips of a coin will produce *exactly* one head is 3/8 or 0.375.

Rules of Probability

Often, we want to compute the probability of an event from the known probabilities of other events. This lesson covers some important rules that simplify those computations.

Definitions and Notation

Before discussing the rules of probability, we state the following definitions:

- Two events are **mutually exclusive** or **disjoint** if they cannot occur at the same time.
- The probability that Event A occurs, given that Event B has occurred, is called a **conditional probability**. The conditional probability of Event A, given Event B, is denoted by the symbol P(A|B).
- The **complement** of an event is the event not occuring. The probability that Event A will not occur is denoted by P(A').
- The probability that Events A and B *both* occur is the probability of the **intersection** of A and B. The probability of the intersection of Events A and B is denoted by P(A ∩ B). If Events A and B are mutually exclusive, P(A ∩ B) = 0.
- The probability that Events A or B occur is the probability of the **union** of A and B. The probability of the union of Events A and B is denoted by P(A ∪ B) .
- If the occurence of Event A changes the probability of Event B, then Events A and B are **dependent**. On the other hand, if the occurence of Event A does not change the probability of Event B, then Events A and B are **independent**.

Rule of Subtraction

In a previous lesson, we learned two important properties of probability:

- The probability of an event ranges from 0 to 1

- The sum of probabilities of all possible events equals 1.

The rule of subtraction follows directly from these properties.

Rule of Subtraction The probability that event A will occur is equal to 1 minus the probability that event A will not occur.

$$\mathbf{P(A) = 1 - P(A')}$$

Suppose, for example, the probability that Bill will graduate from college is 0.80. What is the probability that Bill will not graduate from college? Based on the rule of subtraction, the probability that Bill will not graduate is 1.00 - 0.80 or 0.20.

Rule of Multiplication

The rule of multiplication applies to the situation when we want to know the probability of the intersection of two events; that is, we want to know the probability that two events (Event A and Event B) both occur.

Rule of Multiplication The probability that Events A and B both occur is equal to the probability that Event A occurs times the probability that Event B occurs, given that A has occurred.

$$\mathbf{P(A \cap B) = P(A)\ P(B|A)}$$

Example
An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *without replacement* from the urn. What is the probability that both of the marbles are black?

Solution: Let A = the event that the first marble is black; and let B = the event that the second marble is black. We know the following:

- In the beginning, there are 10 marbles in the urn, 4 of which are black. Therefore, P(A) = 4/10.
- After the first selection, there are 9 marbles in the urn, 3 of which are black. Therefore, P(B|A) = 3/9.

Therefore, based on the rule of multiplication:

$$P(A \cap B) = P(A)\ P(B|A)$$
$$P(A \cap B) = (4/10) * (3/9) = 12/90 = 2/15$$

Rule of Addition

The rule of addition applies to the following situation. We have two events, and we want to know the probability that either event occurs.

Rule of Addition The probability that Event A or Event B occurs is equal to the probability that Event A occurs plus the probability that Event B occurs minus the probability that both Events A and B occur.

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

Note: Invoking the fact that $P(A \cap B) = P(\ A\ )P(\ B\ |\ A\ )$, the Addition Rule can also be expressed as

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

Example
A student goes to the library. The probability that she checks out (a) a work of fiction is 0.40, (b) a work of non-fiction is 0.30, and (c) both fiction and non-fiction is 0.20. What is the probability that the student checks out a work of fiction, non-fiction, or both?

Solution: Let F = the event that the student checks out fiction; and let N = the event that the student checks out non-fiction. Then, based on the rule of addition:

$$P(F \cup N) = P(F) + P(N) - P(F \cap N)$$
$$P(F \cup N) = 0.40 + 0.30 - 0.20 = 0.50$$

Test Your Understanding of This Lesson

Problem 1

An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *with replacement* from the urn. What is the probability that both of the marbles are black?

(A) 0.16
(B) 0.32
(C) 0.36
(D) 0.40
(E) 0.60

Solution

The correct answer is A. Let A = the event that the first marble is black; and let B = the event that the second marble is black. We know the following:

- In the beginning, there are 10 marbles in the urn, 4 of which are black. Therefore, $P(A) = 4/10$.
- After the first selection, we replace the selected marble; so there are still 10 marbles in the urn, 4 of which are black. Therefore, $P(B|A) = 4/10$.

Therefore, based on the rule of multiplication:

$$P(A \cap B) = P(A)\ P(B|A)$$
$$P(A \cap B) = (4/10)*(4/10) = 16/100 = 0.16$$

Problem 2

A card is drawn randomly from a deck of ordinary playing cards. You win $10 if the card is a spade or an ace. What is the probability that you will win the game?

(A) 1/13
(B) 13/52
(C) 4/13
(D) 17/52
(E) None of the above.

Solution

The correct answer is C. Let S = the event that the card is a spade; and let A = the event that the card is an ace. We know the following:

- There are 52 cards in the deck.
- There are 13 spades, so $P(S) = 13/52$.
- There are 4 aces, so $P(A) = 4/52$.
- There is 1 ace that is also a spade, so $P(S \cap A) = 1/52$.

Therefore, based on the rule of addition:

$$P(S \cup A) = P(S) + P(A) - P(S \cap A)$$

$$P(S \cup A) = 13/52 + 4/52 - 1/52 = 16/52 = 4/13$$

Bayes Theorem (aka, Bayes Rule)

Bayes' theorem (also known as Bayes' rule) is a useful tool for calculating conditional probabilities. Bayes' theorem can be stated as follows:

Bayes' theorem. Let $A_1, A_2, \ldots, A_n$ be a set of mutually exclusive events that together form the sample space S. Let B be any event from the same sample space, such that $P(B) > 0$. Then,

$$P(A_k \mid B) = \frac{P(A_k \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B)}$$

Note: Invoking the fact that $P(A_k \cap B) = P(A_k)P(B \mid A_k)$, Baye's theorem can also be expressed as

$$P(A_k \mid B) = \frac{P(A_k)\,P(B \mid A_k)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2) + \ldots + P(A_n)\,P(B \mid A_n)}$$

Unless you are a world-class statiscian, Bayes' theorem (as expressed above) can be intimidating. However, it really is easy to use. The remainder of this lesson covers material that can help you understand when and how to apply Bayes' theorem effectively.

When to Apply Bayes' Theorem

Part of the challenge in applying Bayes' theorem involves recognizing the types of problems that warrant its use. You should consider Bayes' theorem when the following conditions exist.

- The sample space is partitioned into a set of mutually exclusive events $\{ A_1, A_2, \ldots, A_n \}$.
- Within the sample space, there exists an event B, for which $P(B) > 0$.
- The analytical goal is to compute a conditional probability of the form: $P(A_k \mid B)$.
- You know at least one of the two sets of probabilities described below.
    - $P(A_k \cap B)$ for each A_k
    - $P(A_k)$ and $P(B \mid A_k)$ for each A_k

Example 1

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

Solution: The sample space is defined by two mutually-exclusive events - it rains or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. Notation for these events appears below.

- Event A_1. It rains on Marie's wedding.
- Event A_2. It does not rain on Marie's wedding.
- Event B. The weatherman predicts rain.

In terms of probabilities, we know the following:

$P(A_1) = 5/365 = 0.0136985$ [It rains 5 days out of the year.]

- $P(A_2) = 360/365 = 0.9863014$ [It does not rain 360 days out of the year.]
- $P(B | A_1) = 0.9$ [When it rains, the weatherman predicts rain 90% of the time.]
- $P(B | A_2) = 0.1$ [When it does not rain, the weatherman predicts rain 10% of the time.]

We want to know $P(A_1 | B)$, the probability it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman. The answer can be determined from Bayes' theorem, as shown below.

$$P(A_1 | B) = \frac{P(A_1) P(B | A_1)}{P(A_1) P(B | A_1) + P(A_2) P(B | A_2)}$$

$$P(A_1 | B) = (0.014)(0.9) / [(0.014)(0.9) + (0.986)(0.1)]$$

$$P(A_1 | B) = 0.111$$

Note the somewhat unintuitive result. Even when the weatherman predicts rain, it only rains only about 11% of the time. Despite the weatherman's gloomy prediction, there is a good chance that Marie will not get rained on at her wedding.

Probability Distribution

- **What is a Random Variable?**

When the numerical value of a variable is determined by a chance event, that variable is called a **random variable**.

Discrete vs. Continuous Random Variables

Random variables can be discrete or continuous.

- **Discrete**. Within a range of numbers, discrete random variables can take on only certain values. Suppose, for example, that we flip a coin and count the number of heads. The number of heads results from a random process - flipping a coin. And the number of heads is represented by an *integer* value - a number between 0 and plus infinity. Therefore, the number of heads is a discrete random variable.
- **Continuous**. Continuous random variables, in contrast, can take on any value within a range of values. For example, suppose we flip a coin many times and compute the *average* number of heads per flip. The average number of heads per flip results from a random process - flipping a coin. And the average number of heads per flip can take on any value between 0 and 1, even a non-integer value. Therefore, the average number of heads per flip is a continuous random variable.

Discrete Variables: Finite vs. Infinite

Some references state that continuous variables can take on an infinite number of values, but discrete variables cannot. This is incorrect.

- In some cases, discrete variables can take on only a finite number of values. For example, the number of aces dealt in a poker hand can take on only five values: 0, 1, 2, 3, or 4.
- In other cases, however, discrete variables can take on an infinite number of values. For example, the number of coin flips that result in heads could be infinitely large.

When comparing discrete and continuous variables, it is more correct to say that continuous variables can always take on an infinite number of values; whereas some discrete variables can take on an infinite number of values, but others cannot.

Probability Distribution

A **probability distribution** is a table or an equation that links each possible value that a random variable can assume with its probability of occurence.

Discrete Probability Distributions

The probability distribution of a discrete random variable can always be represented by a table. For example, suppose you flip a coin two times. This simple exercise can have four possible outcomes: HH, HT, TH, and TT. Now, let the variable X represent the number of heads that result from the coin flips. The variable X can take on the values 0, 1, or 2; and X is a discrete random variable.

The table below shows the probabilities associated with each possible value of X. The probability of getting 0 heads is 0.25; 1 head, 0.50; and 2 heads, 0.25. Thus, the table is an example of a probability distribution for a discrete random variable.

Number of heads, x	**Probability, P(x)**
0	0.25
1	0.50
2	0.25

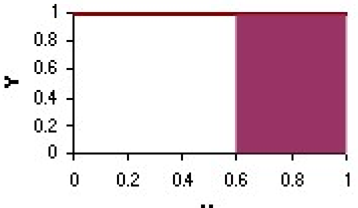
Note: Given a probability distribution, you can find cumulative probabilities. For example, the probability of getting 1 or fewer heads [$P(X < 1)$] is $P(X = 0) + P(X = 1)$, which is equal to $0.25 + 0.50$ or 0.75.

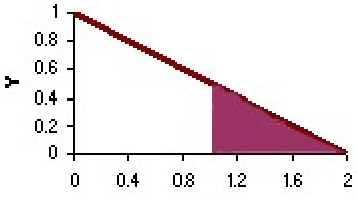
Continuous Probability Distributions

The probability distribution of a continuous random variable is represented by an equation, called the **probability density function** (pdf). All probability density functions satisfy the following conditions:

- The random variable Y is a function of X; that is, $y = f(x)$.
- The value of y is greater than or equal to zero for all values of x.
- The total area under the curve of the function is equal to one.

The charts below show two continuous probability distributions. The chart on the left shows a probability density function described by the equation $y = 1$ over the range of 0 to 1 and $y = 0$ elsewhere. The chart on the right shows a probability density function described by the equation $y = 1 - 0.5x$ over the range of 0 to 2 and $y = 0$ elsewhere. The area under the curve is equal to 1 for both charts.





y = 1 y = 1 - 0.5x

The probability that a continuous random variable falls in the interval between *a* and *b* is equal to the area under the pdf curve between *a* and *b*. For example, in the first chart above, the shaded area shows the probability that the random variable X will fall between 0.6 and 1.0. That probability is 0.40. And in the second chart, the shaded area shows the probability of falling between 1.0 and 2.0. That probability is 0.25.

Note: With a continuous distribution, there are an infinite number of values between any two data points. As a result, the probability that a continuous random variable will assume a particular value is always zero. For example, in both of the above charts, the probability that variable X will equal *exactly* 0.4 is zero.

Test Your Understanding of This Lesson

Problem 1

The number of adults living in homes on a randomly selected city block is described by the following probability distribution.

Number of adults, x	1	2	3	4 or more
Probability, P(x)	0.25	0.50	0.15	???

What is the probability that 4 or more adults reside at a randomly selected home?

(A) 0.10
(B) 0.15
(C) 0.25
(D) 0.50
(E) 0.90

Solution

The correct answer is A. The sum of all the probabilities is equal to 1. Therefore, the probability that four or more adults reside in a home is equal to 1 - (0.25 + 0.50 + 0.15) or 0.10.

Attributes of Random Variables

Just like variables from a data set, random variables are described by measures of central tendency (i.e., mean and median) and measures of variability (i.e., standard deviation and variance). This lesson shows how to compute these measures for discrete random variables.

Mean of a Discrete Random Variable

The mean of the discrete random variable X is also called the **expected value** of X. Notationally, the expected value of X is denoted by E(X). Use the following formula to compute the mean of a discrete random variable.

$$\mathbf{E(X) = \mu_x = \Sigma [x_i * P(x_i)]}$$

where x_i is the value of the random variable for outcome i, μ_x is the mean of random variable X, and $P(x_i)$ is the probability that the random variable will be outcome i.

Example 1 In a recent little league softball game, each player went to bat 4 times. The number of hits made by each player is described by the following probability distribution.

Number of hits, x	0	1	2	3	4
Probability, P(x)	0.10	0.20	0.30	0.25	0.15

What is the mean of the probability distribution?

(A) 1.00
(B) 1.75
(C) 2.00
(D) 2.25
(E) None of the above.

Solution The correct answer is E. The mean of the probability distribution is 2.15, as defined by the following equation.

$$E(X) = \Sigma [x_i * P(x_i)]$$
$$E(X) = 0*0.10 + 1*0.20 + 2*0.30 + 3*0.25 + 4*0.15 = 2.15$$

Median of a Discrete Random Variable

The median of a discrete random variable is the "middle" value. It is the value of X for which $P(X < x)$ is greater than or equal to 0.5 *and* $P(X > x)$ is greater than or equal to 0.5.

Consider the problem presented above in Example 1. In Example 1, the median is 2; because $P(X < 2)$ is equal to 0.60, and $P(X > 2)$ is equal to 0.70. The computations are shown below.

$$P(X < 2) = P(x=0) + P(x=1) + P(x=2) = 0.10 + 0.20 + 0.30 = 0.60$$

$$P(X > 2) = P(x=2) + P(x=3) + P(x=4) = 0.30 + 0.25 + 0.15 = 0.70$$

Variability of a Discrete Random Variable

The standard deviation of a discrete random variable (σ) is equal to the square root of the variance of a discrete random variable (σ^2). The equation for computing the variance of a discrete random variable is shown below.

$$\boldsymbol{\sigma^2 = \Sigma \{ [x_i - E(x)]^2 * P(x_i) \}}$$

where x_i is the value of the random variable for outcome i, $P(x_i)$ is the probability that the random variable will be outcome i, E(x) is the expected value of the discrete random variable x.

Example 2 The number of adults living in homes on a randomly selected city block is described by the following probability distribution.

Number of adults, x	1	2	3	4
Probability, P(x)	0.25	0.50	0.15	0.10

What is the standard deviation of the probability distribution?

(A) 0.50
(B) 0.62
(C) 0.79

(D) 0.89
(E) 2.10

Solution The correct answer is D. The solution has three parts. First, find the expected value; then, find the variance; then, find the standard deviation. Computations are shown below, beginning with the expected value.

$$E(X) = \Sigma [x_i * P(x_i)]$$
$$E(X) = 1*0.25 + 2*0.50 + 3*0.15 + 4*0.10 = 2.10$$

Now that we know the expected value, we find the variance.

$$\sigma^2 = \Sigma \{ [x_i - E(x)]^2 * P(x_i) \}$$
$$\sigma^2 = (1 - 2.1)^2 * 0.25 + (2 - 2.1)^2 * 0.50 + (3 - 2.1)^2 * 0.15 + (4 - 2.1)^2 * 0.10$$
$$\sigma^2 = (1.21 * 0.25) + (0.01 * 0.50) + (0.81) * 0.15) + (3.61 * 0.10) = 0.3025 + 0.0050 + 0.1215 + 0.3610 = 0.79$$

And finally, the standard deviation is equal to the square root of the variance; so the standard deviation is sqrt(0.79) or 0.889.

Rules of Probability

Often, we want to compute the probability of an event from the known probabilities of other events. This lesson covers some important rules that simplify those computations.

Definitions and Notation

Before discussing the rules of probability, we state the following definitions:

- Two events are **mutually exclusive** or **disjoint** if they cannot occur at the same time.
- The probability that Event A occurs, given that Event B has occurred, is called a **conditional probability**. The conditional probability of Event A, given Event B, is denoted by the symbol P(A|B).
- The **complement** of an event is the event not occuring. The probability that Event A will not occur is denoted by P(A').
- The probability that Events A and B *both* occur is the probability of the **intersection** of A and B. The probability of the intersection of Events A and B is denoted by P(A ∩ B). If Events A and B are mutually exclusive, P(A ∩ B) = 0.
- The probability that Events A or B occur is the probability of the **union** of A and B. The probability of the union of Events A and B is denoted by P(A ∪ B) .
- If the occurence of Event A changes the probability of Event B, then Events A and B are **dependent**. On the other hand, if the occurence of Event A does not change the probability of Event B, then Events A and B are **independent**.

Rule of Subtraction

In a previous lesson, we learned two important properties of probabilit

- The probability of an event ranges from 0 to 1.
- The sum of probabilities of all possible events equals 1.

The rule of subtraction follows directly from these properties.

Rule of Subtraction The probability that event A will occur is equal to 1 minus the probability that event A will not occur.

$$\mathbf{P(A) = 1 - P(A')}$$

Suppose, for example, the probability that Bill will graduate from college is 0.80. What is the probability that Bill will not graduate from college? Based on the rule of subtraction, the probability that Bill will not graduate is 1.00 - 0.80 or 0.20.

Rule of Multiplication

The rule of multiplication applies to the situation when we want to know the probability of the intersection of two events; that is, we want to know the probability that two events (Event A and Event B) both occur.

Rule of Multiplication The probability that Events A and B both occur is equal to the probability that Event A occurs times the probability that Event B occurs, given that A has occurred.

$$\mathbf{P(A \cap B) = P(A)\ P(B|A)}$$

Example An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *without replacement* from the urn. What is the probability that both of the marbles are black?

Solution: Let A = the event that the first marble is black; and let B = the event that the second marble is black. We know the following:

- In the beginning, there are 10 marbles in the urn, 4 of which are black. Therefore, $P(A) = 4/10$.
- After the first selection, there are 9 marbles in the urn, 3 of which are black. Therefore, $P(B|A) = 3/9$.

Therefore, based on the rule of multiplication:

$$P(A \cap B) = P(A)\ P(B|A)$$
$$P(A \cap B) = (4/10) * (3/9) = 12/90 = 2/15$$

Rule of Addition

The rule of addition applies to the following situation. We have two events, and we want to know the probability that either event occurs.

Rule of Addition The probability that Event A or Event B occurs is equal to the probability that Event A occurs plus the probability that Event B occurs minus the probability that both Events A and B occur.

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

Note: Invoking the fact that $P(A \cap B) = P(A)P(B \mid A)$, the Addition Rule can also be expressed as

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

Example A student goes to the library. The probability that she checks out (a) a work of fiction is 0.40, (b) a work of non-fiction is 0.30, and (c) both fiction and non-fiction is 0.20. What is the probability that the student checks out a work of fiction, non-fiction, or both?

Solution*:* Let F = the event that the student checks out fiction; and let N = the event that the student checks out non-fiction. Then, based on the rule of addition:

$$P(F \cup N) = P(F) + P(N) - P(F \cap N)$$
$$P(F \cup N) = 0.40 + 0.30 - 0.20 = 0.50$$

Test Your Understanding of This Lesson

Problem 1 An urn contains 6 red marbles and 4 black marbles. Two marbles are drawn *with replacement* from the urn. What is the probability that both of the marbles are black?

(A) 0.16
(B) 0.32
(C) 0.36
(D) 0.40
(E) 0.60

Solution The correct answer is A. Let A = the event that the first marble is black; and let B = the event that the second marble is black. We know the following:

- In the beginning, there are 10 marbles in the urn, 4 of which are black. Therefore, P(A) = 4/10.
- After the first selection, we replace the selected marble; so there are still 10 marbles in the urn, 4 of which are black. Therefore, P(B|A) = 4/10.

Therefore, based on the rule of multiplication:

$$P(A \cap B) = P(A)\ P(B|A)$$
$$P(A \cap B) = (4/10)*(4/10) = 16/100 = 0.16$$

Problem 2 A card is drawn randomly from a deck of ordinary playing cards. You win $10 if the card is a spade or an ace. What is the probability that you will win the game?

(A) 1/13
(B) 13/52
(C) 4/13
(D) 17/52
(E) None of the above.

Solution The correct answer is C. Let S = the event that the card is a spade; and let A = the event that the card is an ace. We know the following:

- There are 52 cards in the deck.
- There are 13 spades, so P(S) = 13/52.
- There are 4 aces, so P(A) = 4/52.
- There is 1 ace that is also a spade, so $P(S \cap A) = 1/52$.

Therefore, based on the rule of addition:

$$P(S \cup A) = P(S) + P(A) - P(S \cap A)$$
$$P(S \cup A) = 13/52 + 4/52 - 1/52 = 16/52 = 4/13$$

Combinations of Random Variables

Sometimes, it is necessary to add or subtract random variables. When this occurs, it is useful to know the mean and variance of the result.

Recommendation: Read the sample problems at the end of the lesson. This lesson introduces some important equations, and the sample problems show how to apply those equations.

Sums and Differences of Random Variables: Effect on the Mean

Suppose you have two variables: X with a mean of μ_x and Y with a mean of μ_y. Then, the mean of the sum of these variables μ_{x+y} and the mean of the difference between these variables μ_{x-y} are given by the following equations.

$$\mu_{x+y} = \mu_x + \mu_y \qquad \text{and} \qquad \mu_{x-y} = \mu_x - \mu_y$$

The above equations for general variables also apply to random variables. If X and Y are random variables, then

$$E(X + Y) = E(X) + E(Y) \qquad \text{and} \qquad E(X - Y) = E(X) - E(Y)$$

where E(X) is the expected value (mean) of X, E(Y) is the expected value of Y, E(X + Y) is the expected value of X plus Y, and E(X - Y) is the expected value of X minus Y.

Independence of Random Variables

If two random variables, X and Y, are **independent**, they satisfy the following conditions.

- $P(x|y) = P(x)$, for all values of X and Y.
- $P(x \cap y) = P(x) * P(y)$, for all values of X and Y.

The above conditions are equivalent. If either one is met, the other condition also met; and X and Y are independent. If either condition is not met, X and Y are **dependent**.

Note: If X and Y are independent, then the correlation between X and Y is equal to zero.

Sums and Differences of Independent Random Variables: Effect on Variance

Suppose X and Y are *independent* random variables. Then, the variance of (X + Y) and the variance of (X - Y) are described by the following equations

$$\text{Var}(X + Y) = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$$

where Var(X + Y) is the variance of the sum of X and Y, Var(X - Y) is the variance of the difference between X and Y, Var(X) is the variance of X, and Var(Y) is the variance of Y.

Note: The standard deviation (SD) is always equal to the square root of the variance (Var). Thus,

$$SD(X + Y) = \text{sqrt}[\ \text{Var}(X + Y)\ ] \qquad \text{and} \qquad SD(X - Y) = \text{sqrt}[\ \text{Var}(X - Y)\ ]$$

Test Your Understanding of This Lesson

Problem 1

		X		
		0	1	2
Y	3	0.1	0.2	0.2
	4	0.1	0.2	0.2

The table on the right shows the joint probability distribution between two random variables - X and Y. (In a joint probability distribution table, numbers in the cells of the table represent the probability that particular values of X and Y occur together.)

What is the mean of the sum of X and Y?

(A) 1.2
(B) 3.5
(C) 4.5
(D) 4.7
(E) None of the above.

Solution The correct answer is D. The solution requires three computations: (1) find the mean (expected value) of X, (2) find the mean (expected value) of Y, and (3) find the sum of the means. Those computations are shown below, beginning with the mean of X.

$$E(X) = \Sigma [x_i * P(x_i)]$$
$$E(X) = 0 * (0.1 + 0.1) + 1 * (0.2 + 0.2) + 2 * (0.2 + 0.2) = 0 + 0.4 + 0.8 = 1.2$$

Next, we find the mean of Y.

$$E(Y) = \Sigma [y_i * P(y_i)]$$
$$E(Y) = 3 * (0.1 + 0.2 + 0.2) + 4 * (0.1 + 0.2 + 0.2) = (3 * 0.5) + (4 * 0.5) = 1.5 + 2 = 3.5$$

And finally, the mean of the sum of X and Y is equal to the sum of the means. Therefore,

$$E(X + Y) = E(X) + E(Y) = 1.2 + 3.5 = 4.7$$

Note: A similar approach is used to find differences between means. The difference between X and Y is E(X - Y) = E(X) - E(Y) = 1.2 - 3.5 = -2.3; and the difference between Y and X is E(Y - X) = E(Y) - E(X) = 3.5 - 1.2 = 2.3

Problem 2 The table on the left shows the joint probability distribution between two random variables - X and Y; and the table on the right shows the joint probability distribution between two random variables - A and B.

		X		
		0	1	2
Y	3	0.1	0.2	0.2
	4	0.1	0.2	0.2

		A		
		0	1	2
B	3	0.1	0.2	0.2
	4	0.2	0.2	0.1

Which of the following statements are true?

I. X and Y are independent random variables.
II. A and B are independent random variables.

(A) I only
(B) II only
(C) I and II
(D) Neither statement is true.
(E) It is not possible to answer this question, based on the information given.

Solution The correct answer is A. The solution requires several computations to test the independence of random variables. Those computations are shown below.

X and Y are independent if P(x|y) = P(x), for all values of X and Y. From the probability distribution table, we know the following:

P(x=0) = 0.2; P(x=0 | y=3) = 0.2; P(x=0 | y = 4) = 0.2
P(x=1) = 0.4; P(x=1 | y=3) = 0.4; P(x=1 | y = 4) = 0.4
P(x=2) = 0.4; P(x=2 | y=3) = 0.4; P(x=2 | y = 4) = 0.4

Thus, P(x|y) = P(x), for all values of X and Y, which means that X and Y are independent. We repeat the same analysis to test the independence of A and B.

P(a=0) = 0.3; P(a=0 | b=3) = 0.2; P(a=0 | b = 4) = 0.4
P(a=1) = 0.4; P(a=1 | b=3) = 0.4; P(a=1 | b = 4) = 0.4
P(a=2) = 0.3; P(a=2 | b=3) = 0.4; P(a=2 | b = 4) = 0.2

Thus, P(a|b) is not equal to P(a), for all values of A and B. For example, P(a=0) = 0.3; but P(a=0 | b=3) = 0.2. This means that A and B are *not* independent.

Problem 3 Suppose X and Y are independent random variables. The variance of X is equal to 16; and the variance of Y is equal to 9. Let Z = X - Y.

What is the standard deviation of Z?

(A) 2.65
(B) 5.00
(C) 7.00
(D) 25.0
(E) It is not possible to answer this question, based on the information given.

Solution The correct answer is B. The solution requires us to recognize that Variable Z is a combination of two *independent* random variables. As such, the variance of Z is equal to the variance of X plus the variance of Y.

$$\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 16 + 9 = 25$$

The standard deviation of Z is equal to the square root of the variance. Therefore, the standard deviation is equal to the square root of 25, which is 5.

Linear Transformations of Random Variables

Sometimes, it is necessary to apply a linear transformation to a random variable. This lesson explains how to make a linear transformation and how to compute the mean and variance of the result.

What is a Linear Transformation?

A **linear transformation** is a change to a variable characterized by one or more of the following operations: adding a constant to the variable, subtracting a constant from the variable, multiplying the variable by a constant, and/or dividing the variable by a constant.

When a linear transformation is applied to a random variable, a new random variable is created. To illustrate, let X be a random variable, and let *m* and *b* be constants. Each of the following examples show how a linear transformation of X defines a new random variable Y.

- Adding a constant: $Y = X + b$
- Subtracting a constant: $Y = X - b$
- Multiplying by a constant: $Y = mX$
- Dividing by a constant: $Y = X/m$
- Multiplying by a constant and adding a constant: $Y = mX + b$
- Dividing by a constant and subtracting a constant: $Y = X/m - b$

Note: Suppose X and Z are variables, and the correlation between X and Z is equal to *r*. If a new variable Y is created by applying a linear transformation to X, then the correlation between Y and Z will also equal *r*.

How Linear Transformations Affect the Mean and Variance

Suppose a linear transformation is applied to the random variable X to create a new random variable Y. Then, the mean and variance of the new random variable Y are defined by the following equations.

$$Y = mX + b \quad \text{and} \quad \text{Var}(Y) = m^2 * \text{Var}(X)$$

where *m* and *b* are constants, Y is the mean of Y, X is the mean of X, Var(Y) is the variance of Y, and Var(X) is the variance of X.

Note: The standard deviation (SD) of the transformed variable is equal to the square root of the variance. That is, SD(Y) = sqrt[Var(Y)].

Test Your Understanding of This Lesson

Problem 1

The average salary for an employee at Acme Corporation is $30,000 per year. This year, management awards the following bonuses to every employee.

- A Christmas bonus of $500.
- An incentive bonus equal to 10 percent of the employee's salary.

What is the mean bonus received by employees?

(A) $500
(B) $3,000
(C) $3,500
(D) None of the above.
(E) There is not enough information to answer this question.

Solution The correct answer is C. To compute the bonus, management applies the following linear transformation to the each employee's salary.

$$Y = mX + b$$
$$Y = 0.10 * X + 500$$

where Y is the transformed variable (the bonus), X is the original variable (the salary), *m* is the multiplicative constant 0.10, and *b* is the additive constant 500.

Since we know that the mean salary is $30,000, we can compute the mean bonus from the following equation.

$$Y = mX + b$$
$$Y = 0.10 * \$30,000 + \$500 = \$3,500$$

Problem 2 The average salary for an employee at Acme Corporation is $30,000 per year, with a variance of 4,000,000. This year, management awards the following bonuses to every employee.

- A Christmas bonus of $500.
- An incentive bonus equal to 10 percent of the employee's salary.

What is the standard deviation of employee bonuses?

(A) $200
(B) $3,000
(C) $40,000
(D) None of the above.
(E) There is not enough information to answer this question.

Solution The correct answer is A. To compute the bonus, management applies the following linear transformation to the each employee's salary.

$$Y = mX + b$$
$$Y = 0.10 * X + 500$$

where Y is the transformed variable (the bonus), X is the original variable (the salary), *m* is the multiplicative constant 0.10, and *b* is the additive constant 500.

Since we know the variance of employee salaries, we can compute the variance of employee bonuses from the following equation.

$$\mathrm{Var}(Y) = m^2 * \mathrm{Var}(X) = (0.1)^2 * 4{,}000{,}000 = 40{,}000$$

where Var(Y) is the variance of employee bonuses, and Var(X) is the variance of employee salaries.

And finally, since the standard deviation is equal to the square root of the variance, the standard deviation of employee bonuses is equal to the square root of 40,000 or $200.

Frequency distribution:

Graphics Representation of Frequency distribution:

There are 2 types of frequency distribution –

1. Histogram
2. Frequency Polygon

1. **Histogram:** In drawing the histogram of a given continuous frequency distribution we first mark off along the x-axis all the class intervals on a suitable scale. On each class interval erect rectangles with heights proportional to the frequency of the class. If, however, the classes are of unequal width then the height of the rectangle will be proportional to the ratio of the frequencies to the width of the classes. The diagram of continuous rectangles so obtained is called histogram.

For Eg. Draw the histogram—

Marks	No. of students
10-20	9

20-30	11
30-40	10
40-50	4
50-60	3
60-70	6

2. Frequency Polygon : For an ungrouped distribution ,the frequency polygon is obtained by plotting points with abscissa as the variate values and ordinate as the corresponding frequencies and joining the plotted points by means of straight lines. For a grouped frequency distribution, the abscissa of points are mid values of the class intervals .For equal class intervals the frequency polygon can be obtained by joining the middle points of the upper sides of the adjacent rectangles of the histogram by means of straight lines. If the class intervals are of small width, the polygon can be obtained by drawing a smooth freehand curve through the vertices of the frequency polygon.

The frequency polygon so obtained should be extended to the base line (x-axis) at both the ends so that it meets the x-axis at the mid points of 2 hypothetical classes, viz, the class before the first class and the class after the last class, each assumed to have zero frequency.

For Eg. Draw the frequency polygon—

Marks	**No. of students**
10-20	9
20-30	11
30-40	10
40-50	4
50-60	3
60-70	6

Measure of Central Tendency:

According to Professor Bowely, averages are “statistical constant which enables us to comprehend in a single effort the significance of the whole.” They give us idea about the concentration of the values in the central part of the distribution.

According to Painly speaking, an average of statistical series is the value of the variable which is representative of the entire distribution.

Types of Measure of Central Tendency:

1. Arithmetic Mean or Simple Mean
2. Geometric Mean
3. Harmonic Mean
4. Median
5. Mode

Characteristics of Measure of Central Tendency:

- **a.** It should be rigidly defined
- **b.** It should readily comprehensible and easy to calculate
- **c.** It should be based on all the observation
- **d.** It should not be affected much by extreme values
- **e.** It should be suitable for mathematical treatment

1. ARITHMETIC MEAN:

A.M mean of a set of observation is their sum divided by the number of observations. If the arithmetic mean x of the n observation
x1, x2,x3,------,xn is given by:

$$\overline{X}=(x1+x2+x3+x4+\text{-----}+xn)/n$$

$$\overline{\mathbf{X}} = \sum_{i=0}^{n} \mathbf{xi/n}$$

In case of frequency distribution xi/fi, i-1,2,------,n , where fi is the frequency of the variable xi.

$$\overline{X} = (f1x1+f2x2+\text{-------}+fnxn)/f1+f2+\text{---------}+fn$$

$$\overline{\mathbf{X}} = \frac{\mathbf{1}}{\mathbf{N}}\sum_{i=0}^{n} \mathbf{fixi}\text{ , where N} = \sum_{i=0}^{n} \text{fi}$$

a) In case of discrete data without frequency distribution-

Example. Find the A.M. of the following observation-

1 , 2 , 3 , 4 , 5

Solution:

$\overline{X}$= (sum of the observation) / Total no. of the observation

$\overline{X}$= (1+2+3+4+5)/5 = 3

3 is the arithmetic mean of given observation

b) In case of discrete data with frequency distribution-

Example.. Find the A.M. of the following frequency distribution:

X: 1 2 3 4 5

F: 2 3 5 1 8

Solution:

X	F	xf
1	2	2
2	3	6
3	5	15
4	1	4
5	8	40

Total	**19**	**67**

$\overline{X} = \frac{1}{N}\sum_{i=0}^{n}$ fixi , where N = $\sum_{i=0}^{n}$ fi

$\overline{X}$ = 67/19

A.M = 3.52 → Answer

c) In case of continuous data without frequency distribution-

Example. Find the A.M. of the following frequency distribution:

Marks:	0-10	10-20	20-30	30-40	40-50	50-60
Frequency:	12	18	27	20	17	6

Solution:

Marks	**f**	**x**	**Xf**
0-10	12	5	60
10-20	18	15	270
20-30	27	25	675
30-40	20	35	700
40-50	17	45	765
50-60	6	55	330
Total	100		2800

$\overline{X} = \frac{1}{N}\sum_{i=0}^{n}$ fixi , where N = $\sum_{i=0}^{n}$ fi

= 2800/100
= 28 →Answer

. GEOMETRIC MEAN:

G.M of a set of n observation is the nth root of their product.
Thus the G.M of n observation xi; i=1,2,----,n is given by—
G =(x1,x2,x3,-----xn)1/n
Taking log of both sides-
Log G = log(x1.x2.-----.xn)/n
Log G = (log x1 + log x2 + ----- + log xn)/n
Log G = $\frac{1}{n}\sum_{i=1}^{n} logxi$
G = Antilog $[\frac{1}{n}\sum_{i=1}^{n} logxi]$

In case of frequency distribution

xi/fi (i = 1 ,2 ,------,n)
G = $(x_1^{f1}.x_2^{f2}.x_3^{f3}\ldots\ldots\ldots x_n^{fn})$
Taking log of both sides-
Log G = (f1.log x1 + f2.log x2 + ----- + fn.log xn)/n

$$\text{Log } G = \frac{1}{n}\sum_{i=1}^{n} fi.logxi$$

$$\mathbf{G = Antilog}\ [\frac{1}{n}\sum_{i=1}^{n} fi.logxi]$$

3. HARMONIC MEAN:

H.M of a number of observations, none of which is zero , is the reciprocal of the arithmetic mean of the reciprocal of the given values. Thus, H.M(H) of n observations xi , i = 1,2,3,-----,n is given by-

H = 1/A.M

4. MEDIAN:

Median of a distribution is the value of the variable which divides into 2 equal parts, which separates the ascending or descending order.

In other words , median is the middle value that separates the higher half from lower half of the data set.

For e.g. - Calculate the median of the given data-

17 , 40 , 38 , 21 , 41

Solution:

Arrange the given data in ascending order –

17 , 21 , 38 , 40 , 41

Median = 38

Note

A) If number of the given data is even then how will we calculate the median---

Median = [(N/2) term + {(N/2) term + 1} term]/2

B) If number of the given data is odd then how will we calculate the median---

Median = [(N/2) term + 1] term

i) For Discrete Frequency Distribution :

In case of Discrete Frequency Distribution median is obtained by considering the cumulative frequencies. The steps for calculating median are given below :

1. Find N/2 where N = f1 + f2 + f3 + ----- + fn
2. See the (less than) cumulative frequency (c.f.) just greater than N/2.
3. The corresponding value of x is median

For e.g. - Calculate the median for the following frequency distribution:

X:	1	2	3	4	5	6	7	8	9
F:	8	10	11	16	20	25	15	9	6

Solution:

X	F	c.f.
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90

7	15	105
8	9	114
9	6	120
Total	**N= 120**	

Here N/2 = 60

The cumulative frequency just greater than N/2 is 65 and the value of x corresponding to 65 is 5.

Therefore, median = 5

ii) For Continuous Frequency Distribution:

In case of frequency distribution, the class corresponding to the c.f. just greater than N/2 is called the median class and the value of median is obtained by the following formula:

Median = l + [(N/2 – c)h]/f

Where, l = is the lower limit of the median class ; f = is the frequency of the median class ; h = is the interval of the median class (upper limit – lower limit) ; c = is the c.f. of the class preceding the median class

And N = f1 + f2 +f3 + ------ + fn

For e.g. - Calculate the median for the following frequency distribution :

Wages	**No. of employees**
2000 – 3000	3
3000 – 4000	5
4000 – 5000	20
5000 – 6000	10
6000 - 7000	5

Solution:

wages	**No. of employees**	**c.f.**
2000 – 3000	3	3
3000 – 4000	5	8
4000 – 5000	20	28
5000 – 6000	10	38
6000 - 7000	5	43
Total	**43**	

Here N/2 = 21.5

c.f. just greater than 21.5 is 28 and the corresponding class is 4000 – 5000.

Thus the median class is 4000 – 5000

Here l = 4000 , h = 1000 , c = 8 , f = 20

Median = 4000 + [(21.5 – 8)1000]/20

Median = 4675

5. MODE:

Mode is the most frequent value of the given data.

For e.g. – 1, 2 , 2 , 6 ,4

Solution: Mode = 2

A) For Discrete Frequency Distribution :

Example.

X :	1	2	3	4	5	6	7	8
F :	4	9	16	25	22	15	7	3

Solution:

Mode = 25, the corresponding value is 4

Hence, Mode = 4 → answer

B) For Continuous Frequency Distribution :

The value of mode is obtained by the following formula:

Mode = l + h (f1 – f0)/(2f1 – f0 – f2)

Where, l = is lower limit of the modal class h = is the interval of the modal class ; f1 = is the frequency of the modal class ;

f0 and f2 = are frequency of the classes preceding and succeeding the modal class.

For e.g. –

Class - Interval	Frequency
0 – 10	5
10 – 20	8
20 – 30	7
30 – 40	12
40 – 50	28
50 – 60	20
60 – 70	10
70 – 80	10

Solution:-

Here, maximum frequency = 28

Thus the mode class is 40 – 50

Here l = 40, h =10, f1 = 28, f0 = 12, f2 = 20

By formula –

Mode = 40 + 10(28 – 12)/ (2*28 – 12 – 20)

Mode = 46.67 → answer

Dispersion:

Averages (or the measures of central tendency) give us idea of the concentration of the observations about the central part of the distribution. If we know the average alone, we cannot form a complete idea about the distribution. If we know the average alone, we cannot form a complete idea about the distribution as will be clear from the following example.

Consider the series (i) 7,8,9,10,11 (ii) 3,6,9,12,15 (iii) 1,5,9,13,17. In all these cases we see that n, the number of observations, is 5 and the mean is 9. If we are given that the mean of 5 observations is 9.we can'nt form an idea as to whether it is the average of first series or second series or third series or of any other series of 5 observations whose sum is 45. Thus we see that the measures of central tendency are inadequate to give us a complete idea of the distribution. They must be supported and supplemented by some other measures. One such measures is **Dispersion.**

Literal meaning of dispersion is 'scatterdness'. We study dispersion to have an idea about the homogeneity or heterogeneity of the distribution. In the above case we say series (i) is more homogeneous(less dispersed) than the series (ii) or (iii) or we say that series (iii) is more heterogeneous (more scattered) than the series (i) or (ii).

Some important definition of dispersion is given below:

(i) "Dispersion is the measures of extent to which individual items vary."

(ii) "Measures of variation tells us how widely the data scatter about their mean"

Characteristics for Measures of central Tendency:

- **a)** It should be rigidly defined
- **b)** It should readily comprehensible and easy to calculate
- **c)** It should be based on all the observation
- **d)** It should not be affected much by extreme values
- **e)** It should be suitable for mathematical treatment

MEASURES OF DISPERSION:

There are 4 types of measures of dispersion –

1. Range
2. Quartile Deviation
3. Mean Deviation
4. Standard Deviation

1. RANGE :

The range is the difference between 2 extreme observations of the distribution. If A and B the greatest and smallest observations respectively in a distribution, then its range is given by:

Range = A – B

2. Quartile Deviation :

Q.D or semi- interquartile range Q is given by-

Q = (Q3 – Q1)/2

Where Q1 and Q3 are the first and third quartiles of the distribution respectively.

Q.D is definitely a better measure than the range as it makes use of 50% of the data. But since it ignores the other 50% of the data, it can't be regarded as a reliable measure.

3. Mean Deviation :

If xi/ fi,i=1,2,…,n is the frequency distribution, then mean deviation from the average A(usually mean median or mode)is given by:

Mean deviation from average A=$\frac{1}{N}\sum_{i=0}^{n}$ fi|xi − A|

Where, **N** = $\sum_{i=0}^{n}$ **fi** and | xi-A | represents modulus or the absolute value of the deviations (xi-A), where the negative sign is ignored.

Since mean deviation is based on all the observation, it is a better measure of dispersion than rang or quartile deviation .But the step of ignoring the sighs of the deviation (xi-A) creates artificial and renders it useless for further mathematical treatment.

Standard Deviation and root mean square Deviation:

Standard deviation, usually denoted by the Greek letter small sigma(σ),is the positive square root of the arithmetic mean of the squares of the deviation of the given value from their arithmetic mean. For the frequency distribution xi|fi; i=1,2,…,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=0}^{n} fi(xi\ - \ddot{x})}^2$$

Where x is the arithmetic mean of the distribution and N = $\sum_{i=0}^{n}$ fi

The square of standard deviation is called the **variance** and is given by:

$$\sigma^2 = \frac{1}{N}\sum_{i=0}^{n} fi(xi - \dddot{x})^2$$

Root mean square deviation, denoted by 's',is given by: s= $\sqrt{\frac{1}{N}\sum_{i=0}^{n} fi(xi\ - A)}^2$

Where A is any arbitrary number .s^2 is called mean square deviation.

Example : Calculate: (i) Quartile Deviation (Q.D)

(iii) Mean Deviation (M.D) (iii) Standard Deviation (S.D)from mean , for the following data:

Marks	No. Of Student
0 – 10	6
10 – 20	5
20 – 30	8
30 – 40	15
40 – 50	7
50 – 60	6
60 – 70	3

Solution:

(ii) Here N = 50 , $\sum_{i=0}^{n}$ fi|xi − A| = 1424.8 where

Marks	Mid-value (x)	No. Of students (f)	$|x_i - \overline{x}|$ $\overline{x} = 4.9$	~~fi~~ $|x_i - \overline{x}|$	$f_i (x_i - \overline{x})^2$	c.f
0 – 10	5	6	0.1	0.6	0.36	6
10 – 20	15	5	10.1	50.5	2550.25	11
20 - 30	25	8	20.1	160.8	25856.64	19
30 – 40	35	15	30.1	451.5	203852.25	34
40 – 50	45	7	40.1	280.7	78792.49	41
50 – 60	55	6	50.1	300.6	90360.36	47
60 – 70	65	3	60.1	180.1	32436.01	50
Total		**50**		**1424.8**	**433848.36**	

A = x

Mean Deviation = $\frac{1}{N}\sum_{i=0}^{n} fi|xi - A|$

= 1424.8/50 = 28.496 → Answer

(i) Here N = 50 , N/4 = 12.75 ; 3N/4 = 37.25

The c.f just greater than 12.75 is 19. Hence, the corresponding class 20 – 30 contains Q_1.

$Q_1 = 20 + \frac{10}{8}(12.75 - 11)$

= 22.19

The c.f just greater than 37.25 is 41. Hence, the corresponding class 40 – 50 contains Q_3

$Q_3 = 40 + .\ \frac{10}{7}(37.25 - 34) = 44.64$

Hence , Q.D = $(Q_3 - Q_1)/2$ = (44.64 – 22.19)/2 = 11.23

Q.D = 11.23 → answer

(ii) We calculate standard deviation -

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=0}^{n} fi(xi - \ddot{x})^2}$$

$\sigma = \sqrt{(433848.36)/50} = 93.15$

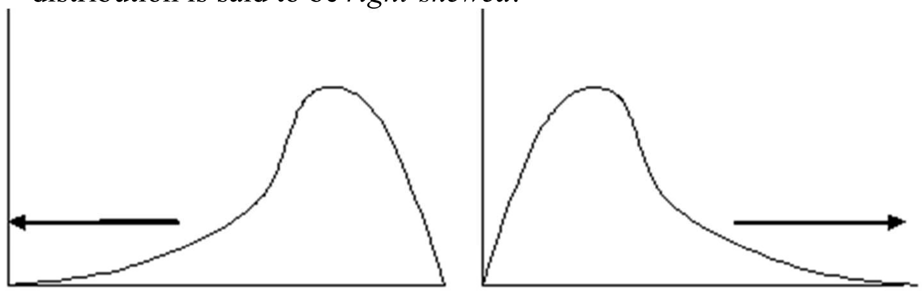
$\sigma = 93.15$ → answer

Skewness

The skewness of a distribution is defined as the lack of ***symmetry.*** In a symmetrical distribution, the Mean, Median and Mode are equal to each other and the ordinate at mean divides the distribution into two equal parts such that one part is mirror image of the other. If some observations, of very high (low) magnitude, are added to such a distribution, its right (left) tail gets elongated.

Consider the distribution in the figure. The bars on the right side of the distribution taper differently than the bars on the left side. These tapering sides are called *tails* (or snakes), and they provide a visual means for determining which of the two kinds of skewness a distribution has:

1. **Negative skew**: The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be *left-skewed*.
2. **Positive skew**: The right tail is longer; the *mass* of the distribution is concentrated on the left of the figure. The distribution is said to be *right-skewed*.



Negative Skew
Elongated tail at the **left**
More data in the left tail than would be expected in a normal distribution

Positive Skew
Elongated tail at the **right**
More data in the right tail than would be expected in a normal distribution

These observations are also known as extreme observations. The presence of extreme observations on the right hand side of a distribution makes it positively skewed and the three averages, **viz**., mean, median and mode, will no longer be equal

We shall in fact have Mean > Median > Mode when a distribution is positively skewed. On the other hand, the presence of extreme observations to the left hand side of a distribution make it negatively skewed and the relationship between mean, median and mode is: Mean < Median < Mode. we
depict the shapes of positively skewed and negatively skewed distributions.The direction and extent of skewness can be measured in various ways.

Karl Pearson's measure of Skewness

In that the mean, median and mode are not equal in a skewed distribution. The Karl Pearson's measure of skewness is based upon the ***divergence of mean from mob*** in a skewed distribution.Since Mean = Mode in a symmetrical distribution, (Mean - Mode) can be taken as an ***absolute measure of skewness.*** The absolute measure of skewness for a distribution depends upon the unit of measurement. For example, if the mean = 2.45 qetre and mode = 2.14 metre, then absolute measure of skewness will be 2.45 aetre - 2.14 metre = 0.31 metre. For the same distribution, if we change the d t of measurement to centimetres, the absolute measure of skewness is 245 centimetre - 2 14 centimetre = 3 1 centimetre. In order to avoid such a problem Karl Pearson takes a relative measure of skewness.
A relative measure, independent of the units of measurement, is defined as the Karl Pearson's Coeficient of Skewness Sk, given by

$$\mathbf{Sk} = \frac{Mean - Mode}{s.d.}$$

The sign of Sk gives the direction and its magnitude gives the extent of skewness.
If Sk > 0, the distribution is positively skewed, and if S, < 0 it is negatively skewed.
So far we have seen that Sk is strategically dependent upon mode. If mode is not defined for a distribution we cannot find Sk . But empirical relation between mean, median and mode states that, for a moderately symmetrical distribution, we have

Mean - Mode = 3 (Mean - Median)

Hence Karl Pearson's coefficient of skewness is defined in t m s of median as

$$\mathbf{Sk} = \frac{3(Mean - Median)}{s.d.}$$

Example: Compute the Karl Pearson's coefficient of skewness fiom the following data.

Height (in inches)	**Number of Persons**
58	10
59	18
60	30
61	42
62	3 5
63	28
64	16
65	8

Table for the computation of mean and s.d.

Solution:-

Height *(X)*	***u* = X - 61**	**No. of persons *(f)***	***fu***	fu^2

58	- 3	10	- 30	90
59	- 2	18	-36	72
60	- 1	30	- 30	30
61	0	42	0	0
62	1	35	35	35
63	2	28	56	112
64	3	16	48	144
65	4	8	32	128
Total		**187**	**75**	**611**

$$\text{Mean} = 61 + \frac{75}{187} = 61.4$$

$$\text{s.d.} = \sqrt{\frac{611}{187} - \left(\frac{75}{187}\right)^2} = 1.76$$

To find mode, we note that height is a continuous variable. It is assumed that the height has been measured under the approximation that a measurement on height that is, e.g., greater than 58 but less than 58.5 is taken as 58 inches while a measurement greater than or equal to 58.5 but less than 59 is taken as 59 inches. Thus the given data can be written as

Height (in inches)	**Number of Persons**
57.5 - 58.5	10
58.5 - 59.5	18
59.5 - 60.5	30
60.5 - 61.5	42
61.5 - 62.5	35
62.5 - 63.5	28
63.5 - 64.5	16
64.5 - 65.5	8

By inspection, the modal class is 60.5 - 61.5. Thus, we have
I_m= 60.5,Δ_1 = 42 – 30 = 12, Δ_2 = 42 – 35 = 7 and h = 1.

$$\text{Mode} = 60.5 + \frac{12}{12+7} X\ 1 = 61.13$$

Hence, the Karl Pearson's coefficient of skewness, ***Sk*** = $\frac{61.4 - 61.13}{1.76}$ = 0.153.
Thus the distribution is positively skewed.

Bowley's Measure of Skewness

This measure is based on quartiles. For a symmetrical distribution, it is seen that ***Q1*** and ***Q3 are*** equidistant ftom median. Thus ***(Q3 - Md)*** - *(Md – **Q1)*** can be taken as an absolute measure of skewness.

A relative measure of skewness, known as Bowley's coefficient ***(SQ),***is given by

$$S_Q = \frac{(Q_3 - Md) - (Md - Q_1)}{(Q_3 - Md) + (Md - Q_1)}$$

$$S_Q = \frac{Q_3 - 2Md + Q_1}{(Q_3 - Q_1)}$$

Moments

$$\boldsymbol{\mu_r} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \mathbf{f_i}(\mathbf{x_i} - \mathbf{x})^{\mathbf{r}} \quad \text{where r = 0, 1, 2, 3, 4,....}$$

Thus, **rth** moment about mean is the mean of the rth power of deviations of Observations from their arithmetic mean. In particular,

if r = 0, we have $\boldsymbol{\mu_0} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \mathbf{f_i}(\mathbf{x_i} -\bar{}\mathbf{x})^{\mathbf{0}}$
if r = 1, we have $\boldsymbol{\mu_1} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \mathbf{f_i}(\mathbf{x_i} - \mathbf{x})^{\mathbf{1}}$
if ***r*** = 2, we have $\boldsymbol{\mu_2} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \mathbf{f_i}(\mathbf{x_i} - \mathbf{x})^{\mathbf{2}}$
if ***r*** = 3, we have $\boldsymbol{\mu_3} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \mathbf{f_i}(\mathbf{x_i} - \mathbf{x})^{\mathbf{3}}$ and so on.

These moments are also known as central moments.

In addition to the above, we can define **raw moments** as moments about any arbitrary mean.
Let ***A*** denote an arbitrary mean, then uth moment about ***A*** is defined as

$$\boldsymbol{\mu_r'} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \mathbf{f_i}(\mathbf{x_i} - \mathbf{A})^{\mathbf{r}} \quad \text{where r = 0, 1, 2, 3, 4, ...}$$

When ***A*** = 0, we get various moments about origin.

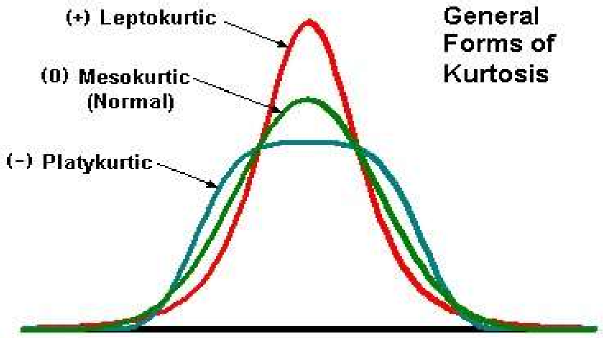
Moment Measure of Skewness

The moment measure of skewness is based on the property that, for a symmetrical distribution, all odd ordered central moments are equal to zero.We note that $\boldsymbol{\mu_0}$ = 0, for every distribution, therefore, the lowest order moment
that can provide an absolute measure of skewness 'is $\boldsymbol{\mu_3}$ **.**

Kurtosis

Kurtosis is another measure of the shape of a distribution. Whereas skewness measures the lack of symmetry of the frequency curve of a distribution, kurtosis is a measure of the relative peakedness of its fi-equency curve. Various frequency curves can be divided into three categories depending upon the shape of their peak. The three shapes are termed as Leptokurtic, Mesokurtic and Platykurhc

Kurtosis is the degree of peakedness of a distribution. A normal distribution is a mesokurtic distribution. A pure leptokurtic distribution has a higher peak than the normal distribution and has heavier tails. A pure platykurtic distribution has a lower peak than a normal distribution and lighter tails.



A measure of kurtosis is given by $\boldsymbol{\beta_2} = \frac{\mu_4}{\mu_2^2}$ a coefficient given by Karl Pearson.

The value of $\beta_2 = 3$ for a Mesokurtic curve. When $\beta_2 > 3$, the curvt: is more peaked than the Mesokurtic curve and is tenned as Leptokurtic. Similarly, when $\beta_2 < 3$**,** the curve is less peaked than the mesokurtic curve and is called as
Platykurtic curve.

Example : The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75. Examine the skewness and kurtosis of the distribution.

Soluion: To examine skewness , we computeβ_1.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^{32}} = \frac{0.7^2}{2.5^2} 0.031$$

Since $\mu_3 > 0$ and β_1**,** is small, the distribution is moderately positivelyType equation here. skewed.
Kurtosis is given by the coefficient $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.75}{2.5^2} = 3.0$
Hence the curve is mesokurtic.

CORRELATION AND REGRESSION

Introduction

So far we have confined our discussion to the distributions involving only one variable. Sometimes, in practical applications, we might come across certain set of data, where each item of the set may comprise of the values of two or more variables.

Suppose we have a set of 30 students in a class and we want to measure the heights and weights of all the students. We observe that each individual (unit) of the set assumes two values – one relating to the height and the other to the weight. Such a distribution in which each individual or unit of the set is made up of two values is called a bivariate distribution. The following examples will illustrate clearly the meaning of bivariate distribution.

(i) In a class of 60 students the series of marks obtained in two subjects by all of them.
(ii) The series of sales revenue and advertising expenditure of two companies in a particular year.
(iii) The series of ages of husbands and wives in a sample of selected married couples.

Thus in a bivariate distribution, we are given a set of pairs of observations, wherein each pair represents the values of two variables.

In a bivariate distribution, we are interested in finding a relationship (if it exists) between the two variables under study.

The concept of ‘correlation’ is a statistical tool which studies the relationship between two variables and Correlation Analysis involves various methods and techniques used for studying and measuring the extent of the relationship between the two variables.

“Two variables are said to be in correlation if the change in one of the variables results in a change in the other variable”.

5.2: Types of Correlation

There are two important types of correlation. They are (1) Positive and Negative correlation and (2) Linear and Non – Linear correlation.

5.2.1: Positive and Negative Correlation

If the values of the two variables deviate in the same direction i.e. if an increase (or decrease) in the values of one variable results, on an average, in a corresponding increase (or decrease) in the values of the other variable the correlation is said to be positive.

Some examples of series of positive correlation are:

(i) Heights and weights;
(ii) Household income and expenditure;
(iii) Price and supply of commodities;
(iv) Amount of rainfall and yield of crops.

Correlation between two variables is said to be negative or inverse if the variables deviate in opposite direction. That is, if the increase in the variables deviate in opposite direction. That is, if increase (or decrease) in the values of one variable results on an average, in corresponding decrease (or increase) in the values of other variable.

Some examples of series of negative correlation are:

(i) Volume and pressure of perfect gas;
(ii) Current and resistance [keeping the voltage constant] (R = V/L)
(iii) Price and demand of goods.

Graphs of Positive and Negative correlation:

Suppose we are given sets of data relating to heights and weights ofstudents in a class. They can be plotted on the coordinate plane using x –axis to represent heights and y – axis to represent weights. The different graphs shown below illustrate the different types of correlations.

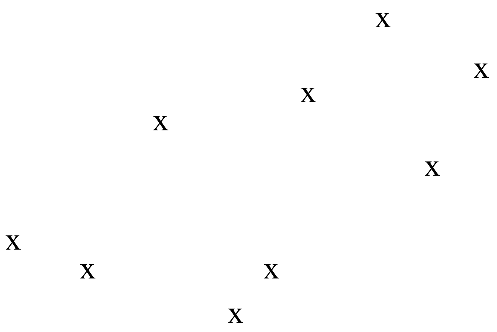


Figure for positive correlation

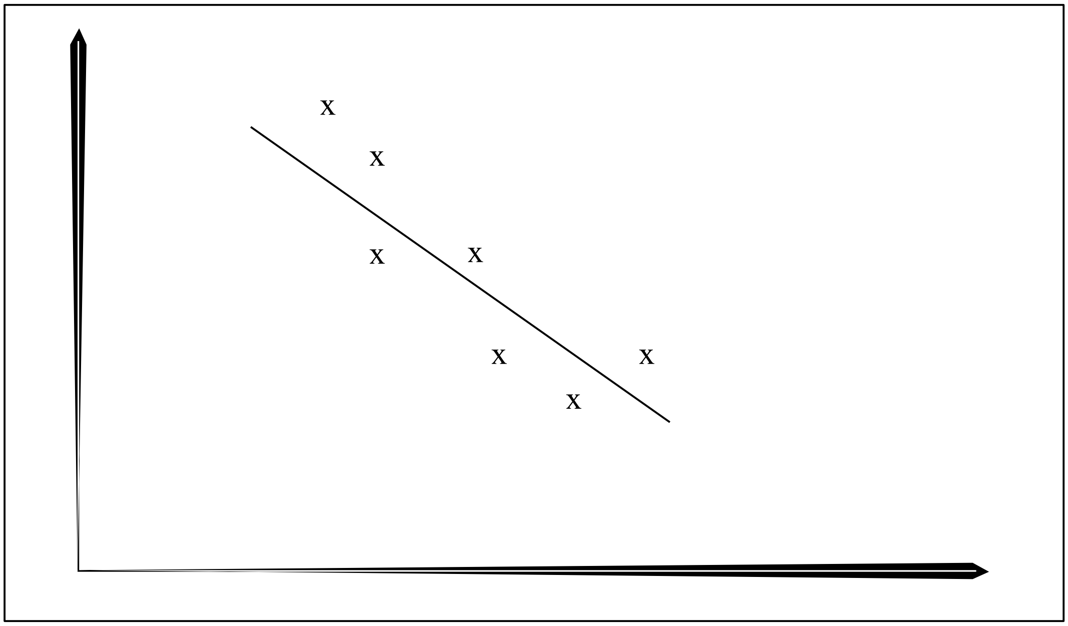


Figure for negative correlation

Note:

(i) If the points are very close to each other, a fairly good amount of correlation can be expected between the two variables. On the other hand if they are widely scattered a poor correlation can be expected between them.

(ii) If the points are scattered and they reveal no upward or downward trend as in the case of (d) then we say the variables are uncorrelated.

(iii) If there is an upward trend rising from the lower left hand corner and going upward to the upper right hand corner, the correlation obtained from the graph is said to be positive. Also, if there is a downward trend from the upper left hand corner the correlation obtained is said to be negative.

(iv) The graphs shown above are generally termed as **scatter diagrams**.

Example:1: The following are the heights and weights of 15 students of a class. Draw a graph to indicate whether the correlation is negative or positive.

Heights (cms)	Weights (kgs)
170	65
172	66
181	69
157	55
150	51
168	63
166	61
175	75
177	72
165	64
163	61
152	52
161	60
173	70
175	72

Since the points are dense (close to each other) we can expect a high degree of correlation between the series of heights and weights. Further, since the points reveal an upward trend, the correlation is positive. Arrange the data in increasing order of height and check that , as height increases, the weight also increases, except for some (stray) cases..

EXERCISES

(1) A Company has just brought out an annual report in which the capital investment and profits were given for the past few years. Find the type of correlation (if it exists).

Capital Investment (crores)	10	16	18	24	36	48	57
Profits (lakhs)	12	14	13	18	26	38	62

(2) Try to construct more examples on the positive and negative correlations.

(3) Construct the scattered diagram of the data given below and indicate the type of correlation.

(Average Value in Lakhs of Rs.)

Years	1965	1970	1975	1980	1985	1990
Raw cotton import	42	60	112	98	118	132
Cotton manufacture exports	68	79	88	86	106	114

5.3: Linear and Non – Linear Correlation

The correlation between two variables is said to be **linear** if the change of one unit in one variable result in the corresponding change in the other variable over the entire range of values.

For example consider the following data.

X	2	4	6	8	10
Y	7	13	19	25	31

Thus, for a unit change in the value of x, there is a constant change in the corresponding values of y and the above data can be expressed by the relation

$$y = 3x + 1$$

In general two variables x and y are said to be **linearly related**, if there exists a relationship of the form

$$y = a + bx$$

where ‘a’ and ‘b’ are real numbers. This is nothing but a straight line when plotted on a graph sheet with different values of x and y and for constant values of a and b. Such relations generally occur in physical sciences but are rarely encountered in economic and social sciences.

The relationship between 2 is said to be non – linear if corresponding to a unit change in one variable , the other variable doesnot change at a fluctuating rate. In such case , if the data is plotted on a graph sheet we will not get a straight line curve. For E.g. –

$$Y = a + bx + cx^2$$

Or more general polynomial.

5.4: The Coefficient of Correlation

One of the most widely used statistics is the **coefficient of correlation** 'r' which measures the degree of association between the two values of related variables given in the data set. It takes values from + 1 to – 1. If two sets or data have r = +1, they are said to be perfectly **correlated positively** if r = -1 they are said to be perfectly **correlated negatively**; and if r = 0 they are **uncorrelated.**

The coefficient of correlation 'r' is given by the formula

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}$$

The following example illustrates this idea.

Example:2: A study was conducted to find whether there is any relationship between the weight and blood pressure of an individual. The following set of data was arrived at from a clinical study. Let us determine the coefficient of correlation for this set of data. The first column represents the serial number and the second and third columns represent the weight and blood pressure of each patient.

S. No.	Weight	Blood Pressure
1.	78	140
2.	86	160
3.	72	134
4.	82	144
5.	80	180
6.	86	176
7.	84	174
8.	89	178
9.	68	128
10.	71	132

Solution:

x	y	x^2	y^2	xy
78	140	6084	19600	10920
86	160	7396	25600	13760
72	134	5184	17956	9648
82	144	6724	20736	11808
80	180	6400	32400	14400
86	176	7396	30976	15136
84	174	7056	30276	14616
89	178	7921	31684	15842
68	128	4624	16384	8704
71	132	5041	17424	9372
796	1546	63,776	243036	1242069

Then

Then

$$r = \frac{10(124206) - (796)(1546)}{\sqrt{\left[(10)63776 - (796)^2\right]\left[(10)(243036) - (1546)^2\right]}}$$

$$= \frac{11444}{\sqrt{(1144)(40244)}}$$

$$= 0.5966$$

5.4: Rank Correlation

Data which are arranged in numerical order, usually from largest to smallest and numbered 1,2,3 ---- are said to be in **ranks** or **ranked data.**. These ranks prove useful at certain times when two or more values of one variable are the same. The coefficient of correlation for such type of data is given by **Spearman rank difference correlation coefficient** and is denoted by R.

In order to calculate R, we arrange data in ranks computing the difference in rank 'd' for each pair. The following example will explain the usefulness of R. R is given by the formula

$$R = 1 - 6\frac{(\sum d^2)}{n(n^2 - 1)}$$

Example:3: The data given below are obtained from student records. Calculate the rank correlation coefficient 'R' for the data.

Subject	Grade Point Average (x)	Graduate Record exam score (y)
1.	8.3	2300
2.	8.6	2250
3.	9.2	2380
4.	9.8	2400
5.	8.0	2000
6.	7.8	2100
7.	9.4	2360
8.	9.0	2350
9.	7.2	2000
10.	8.6	2260

Note that in the G. P. A. column we have two students having a grade point average of 8.6 also in G. R. E. score there is a tie for 2000.

Now we first arrange the data in descending order and then rank 1,2,3,---- 10 accordingly. In case of a tie, the rank of each tied value is the mean of all positions they occupy. In x, for instance, 8.6 occupy ranks 5 and 6. So each has a rank $\frac{5+6}{2}=5.5$;

Similarly in 'y' 2000 occupies ranks 9 and 10, so each has rank $\frac{9+10}{2}=9.5$.

Now we come back to our formula $R=1-\frac{6\sum d^2}{n(n^2-1)}$

We compute 'd' , square it and substitute its value in the formula.

Subject	x	y	Rank of x	Rank of y	d	d^2
1.	8.3	2300	7	5	2	4
2.	8.6	2250	5.5	7	-1.5	2.25
3.	9.2	2380	3	2	1	1
4.	9.8	2400	1	1	0	0
5.	8.0	2000	8	9.5	-1.5	2.25
6.	7.8	2100	9	8	1	1
7.	9.4	2360	2	3	-1	1
8.	9.0	2350	4	4	0	0
9.	7.2	2000	10	9.5	0.5	0.25
10.	8.6	2260	5.5	6	-0.5	0.25

So here, n = 10, sum of d^2 = 12. So

$$R = 1 - \frac{6(12)}{10\,(100-1)}$$

$$= 1 - 0.0727 \quad = 0.9273$$

Note: If we are provided with only ranks without giving the values of x and y we can still find Spearman rank difference correlation R by taking the difference of the ranks and proceeding in the above shown manner.

5.5: Regression

If two variables are significantly correlated, and if there is some theoretical basis for doing so, it is possible to predict values of one variable from the other. This observation leads to a very important concept known as 'Regression Analysis'.

Regression analysis, in general sense, means the estimation or prediction of the unknown value of one variable from the known value of the other variable. It is one of the most important statistical tools which is extensively used in almost all sciences – Natural, Social and Physical. It is specially used in business and economics to study the relationship between two or more variables that are related causally and for the estimation of demand and supply graphs, cost functions, production and consumption functions and so on.

Prediction or estimation is one of the major problems in almost all the spheres of human activity. The estimation or prediction of future production, consumption, prices, investments, sales, profits, income etc. are of very great importance to business professionals. Similarly, population estimates and population projections, GNP, Revenue and Expenditure etc. are indispensable for economists and efficient planning of an economy.

Regression analysis was explained by M. M. Blair as follows: "Regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data."

5.5.1: Regression Equation

Suppose we have a sample of size 'n' and it has two sets of measures, denoted by x and y. We can predict the values of 'y' given the values of 'x' by using the equation, called the REGRESSION EQUATION.

$$y^* = a + bx$$

where the coefficients a and b are given by

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{\sum y - b \sum x}{n}$$

The symbol y* refers to the predicted value of y from a given value of x from the regression equation.

Example: 4 : Scores made by students in a statistics class in the mid - term and final examination are given here. Develop a regression equation which may be used to predict final examination scores from the mid – term score.

STUDENT	MID – TERM	FINAL
1.	98	90
2.	66	74
3.	100	98
4.	96	88
5.	88	80
6.	45	62
7.	76	78
8.	60	74
9.	74	86
10.	82	80

Solution:

We want to predict the final exam scores from the mid term scores. So let us designate ‘y’ for the final exam scores and ‘x’ for the mid – term exam scores. We open the following table for the calculations.

Stud	x	y	X^2	xy
1	98	90	9604	8820
2	66	74	4356	4884
3	100	98	10,000	9800
4	96	88	9216	8448
5	88	80	7744	7040
6	45	62	2025	2790
7	76	78	5776	5928
8	60	74	3600	4440
9	74	86	5476	6364
10	82	80	6724	6560
Total	785	810	64,521	65,071

Numerator of b = 10 * 65,071 – 785 * 810 = 6,50,710 – 6,35,850 = 14,860
Denominator of b = 10 * 64, 521 – $(785)^2$ = 6,45,210 – 6,16,225 = 28,985

Therefore, b = 14,860 / 28,985 = 0.5127

Numerator of a = 810 – 785 * 0.5127 = 810 – 402.4695 = 407.5305
Denominator of a = 10

Therefore a = 40.7531

Thus , the regression equation is given by

$$y^* = 40.7531 + (0.5127)\,x$$

We can use this to find the projected or estimated final scores of the students.
For example, for the midterm score of 50 the projected final score is
$y^* = 40.7531 + (0.5127)\,50 = 40.7531 + 25.635 = 66.3881$
which is a quite a good estimation.
To give another example, consider the midterm score of 70. Then the projected final score is
$y^* = 40.7531 + (0.5127)\,70 = 40.7531 + 35.889 = 76.6421$,
which is again a very good estimation.

Normal Distribution

The **normal distribution** refers to a family of continuous probability distributions described by the normal equation.

The Normal Equation

The normal distribution is defined by the following equation:

Normal equation. The value of the random variable Y is:

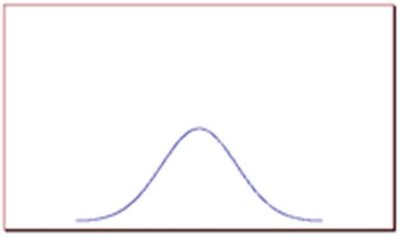
$$Y = \{ 1/[\sigma * \mathrm{sqrt}(2\pi)] \} * e^{-(x - \mu)2/2\sigma 2}$$

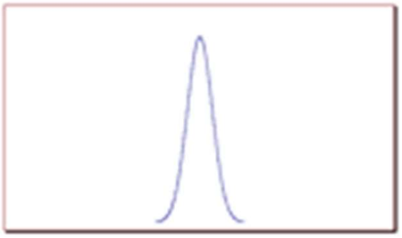
where X is a normal random variable, μ is the mean, σ is the standard deviation, π is approximately 3.14159, and e is approximately 2.71828.

The random variable X in the normal equation is called the **normal random variable**. The normal equation is the probability density function for the normal distribution.

The Normal Curve

The graph of the normal distribution depends on two factors - the mean and the standard deviation. The mean of the distribution determines the location of the center of the graph, and the standard deviation determines the height and width of the graph. When the standard deviation is large, the curve is short and wide; when the standard deviation is small, the curve is tall and narrow. All normal distributions look like a symmetric, bell-shaped curve, as shown below.





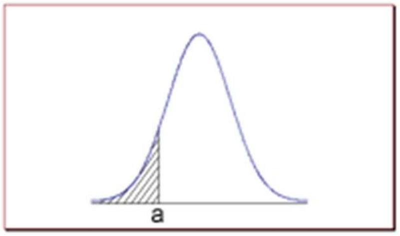
The curve on the left is shorter and wider than the curve on the right, because the curve on the left has a bigger standard deviation.

Probability and the Normal Curve

The normal distribution is a continuous probability distribution. This has several implications for probability.

- The total area under the normal curve is equal to 1.
- The probability that a normal random variable X equals any particular value is 0.
- The probability that X is greater than a equals the area under the normal curve bounded by a and plus infinity (as indicated by the *non-shaded* area in the figure below).

- The probability that *X* is less than *a* equals the area under the normal curve bounded by *a* and minus infinity (as indicated by the *shaded* area in the figure below).



Additionally, every normal curve (regardless of its mean or standard deviation) conforms to the following "rule".

- About 68% of the area under the curve falls within 1 standard deviation of the mean.
- About 95% of the area under the curve falls within 2 standard deviations of the mean.
- About 99.7% of the area under the curve falls within 3 standard deviations of the mean.

Collectively, these points are known as the **empirical rule** or the **68-95-99.7 rule**. Clearly, given a normal distribution, most outcomes will be within 3 standard deviations of the mean.

To find the probability associated with a normal random variable, use a graphing calculator, an online normal distribution calculator, or a normal distribution table. In the examples below, we illustrate the use of Stat Trek's Normal Distribution Calculator, a free tool available on this site. In the next lesson, we demonstrate the use of normal distribution tables.

Example 1 An average light bulb manufactured by the Acme Corporation lasts 300 days with a standard deviation of 50 days. Assuming that bulb life is normally distributed, what is the probability that an Acme light bulb will last at most 365 days?

Solution*:* Given a mean score of 300 days and a standard deviation of 50 days, we want to find the cumulative probability that bulb life is less than or equal to 365 days. Thus, we know the following:

- The value of the normal random variable is 365 days.
- The mean is equal to 300 days.
- The standard deviation is equal to 50 days.

We enter these values into the Normal Distribution Calculator and compute the cumulative probability. The answer is: $P(X < 365) = 0.90$. Hence, there is a 90% chance that a light bulb will burn out within 365 days.

Example 2 Suppose scores on an IQ test are normally distributed. If the test has a mean of 100 and a standard deviation of 10, what is the probability that a person who takes the test will score between 90 and 110?

Solution: Here, we want to know the probability that the test score falls between 90 and 110. The "trick" to solving this problem is to realize the following:

$$P(90 < X < 110) = P(X < 110) - P(X < 90)$$

We use the Normal Distribution Calculator to compute both probabilities on the right side of the above equation.

- To compute $P(X < 110)$, we enter the following inputs into the calculator: The value of the normal random variable is 110, the mean is 100, and the standard deviation is 10. We find that $P(X < 110)$ is 0.84.
- To compute $P(X < 90)$, we enter the following inputs into the calculator: The value of the normal random variable is 90, the mean is 100, and the standard deviation is 10. We find that $P(X < 90)$ is 0.16.

We use these findings to compute our final answer as follows:

$$P(90 < X < 110) = P(X < 110) - P(X < 90)$$
$$P(90 < X < 110) = 0.84 - 0.16$$
$$P(90 < X < 110) = 0.68$$

Thus, about 68% of the test scores will fall between 90 and 110.

Standard Normal Distribution

The **standard normal distribution** is a special case of the normal distribution. It is the distribution that occurs when a normal random variable has a mean of zero and a standard deviation of one.

Standard Score (aka, z Score)

The normal random variable of a standard normal distribution is called a **standard score** or a **z-score**. Every normal random variable X can be transformed into a z score via the following equation:

$$z = (X - \mu) / \sigma$$

Where X is a normal random variable, μ is the mean mean of X, and σ is the standard deviation of X.

BINOMIAL DISTRIBUTION:

Binomial distribution was discovered by James Bernoulli (1654-1705) in the year 1700 and was first published posthumously in the 1713, eight years after his death.

Let a random experiment be performed repeatedly, each repetition being called a trial and let the occurrence of an event in a trial be called a success and it's non – occurrence a failure. Consider a set of n independent Bernoulli an trials (n being finite) in which the probability 'p' of success in any trial is constant for each trial, then q = 1-p, is the probability of failure in any trial.

The probability of x successes and consequently (n – x) failures in n independent trials, in a specified order (say) SSFSFFFS….FSF (where S represent success and F represents failure) is given by compound probability theorem by the expression :

P(SSFSFFFS….FSF) = P(S)P(S)P(F)P(S)…..P(F)P(S)P(F)

= p.p.q.p……q.p.q

= p.p.p….p . q.q.q….q

{ x factor } . {(n –x) factor}

$= p^x .q^{n-x}$

But x successes in n trials can occur in ${}^nc_x$ ways and the probability of x successes in n trials in any order is given by the addition theorem of probability by the expression ${}^nc_x\, p^x .q^{n-x}$

He probability distribution of the number of successes, so obtained is called the Binomial probability distribution, for the obvious reason that the probabilities of 0,1,2…..,n successes , viz, q^n, ${}^nc_1q^{n-1}p$, ${}^nc_2q^{n-2}p^2$,…..,p^n, are the successive terms of the binomial expansion $\mathbf{(q+p)^n}$.

Definition:

A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by :

P(X = x) = p(x) = { $\mathbf{{}^nc_x\, p^x .q^{n-x}}$ **; x = 0, 1,2,…..,n; q = 1 – p**

{ 0 , otherwise

The 2 independent constants n and p in the distribution are known as the parameters of the distribution . 'n' is also sometimes , known as the degree of the binomial distribution.

Binomial distribution is a discrete distribution as X can take only the integral values , viz 0,1,2,….,n. Any random variable which follows binomial distribution is known as binomial variate.

We shall use the notation X ~ B(n,p) to denote that the random variable X follows binomial distribution with parameters n and p.

Remark **- 1** $= (x + a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}$

Poisson Distribution:

P.D was discovered by the French mathematician and physicist Simeon Denis Poisson (1781-1840) who published it in 1837. P.D is a limiting case of the binomial distribution under the following conditions:

(i) n, the number of trials is indefinitely large, i.e. n→∞
(ii) p, the constant probability of success for each trial is indefinitely small, i.e. , p→0.
(iii)np = λ

Definition: A random variable X is said to follow a Poisson distribution if it assumes only non negative values and its probability mass function is given by-

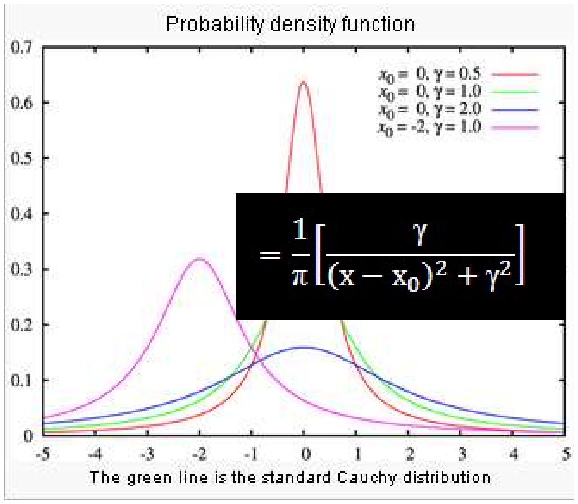
$$\mathbf{P(x,\lambda) = P(X=x)} = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} \; ; \; \mathbf{x = 0,1,2,\ldots; \; \lambda>0} \\ \mathbf{0 \;, \; otherwise} \end{cases}$$

Remark - it should be noted that

$$\sum_{x=0}^{\infty} \mathbf{P(X=x)} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{x}}{x!} = e^{-\lambda}e^{\lambda} = \mathbf{1}$$

Cauchy Distribution:

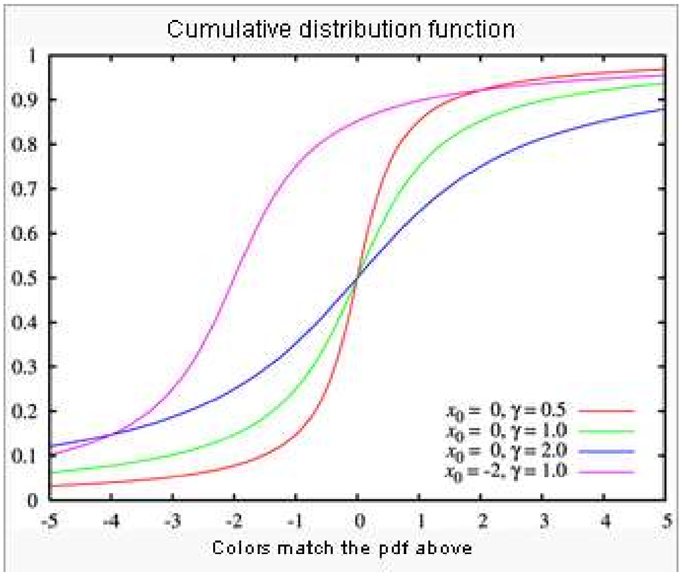
The Cauchy distribution is often used in statistics as the canonical example of a "pathological" distribution. Its mean does not exist and its variance is infinite. The Cauchy distribution does not have finite moments of order greater than or equal to one; only fractional absolute moments exist.The Cauchy distribution has no moment generating function.



$$f(x; x_0\gamma) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

$$= \frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right]$$

- $x0$: the location of the peak of the distribution
- γ : the half-width at half-maximum (HWHM).
- special case : $x0 = 0$, $\gamma = 1$ -> standard Cauchy distribution



Colors match the pdf above

- The cumulative distribution function is

$$F(x; x_0, \gamma) = \frac{1}{\pi}\arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}$$

- the inverse cumulative distribution function of the Cauchy distribution

$$F^{-1}(p; x_0, \gamma) = x_0 + \gamma \tan\left[\pi\left(p - \frac{1}{2}\right)\right]$$

Properties

- mean, variance or higher moments : not defined
- mode and median : x0
- *U, V* ~N(0, 1) -> the ratio *U*/*V* has the standard Cauchy distribution.
- *X*1, …, *Xn* : i.i.d. random variables, each with a standard Cauchy distribution -> sample mean (*X*1 + … + *Xn*)/*n* : the same standard Cauchy distribution. (the sample median, which is not affected by extreme values, can be used as a measure of central tendency).
- The standard Cauchy distribution ~
- Characteristic Function :

$$\phi_x(t; x_0, \gamma) = E\left(e^{iXt}\right) = \exp(ix_0 t - \gamma|t|)$$

- Why the mean is undefined?

$$\int_{-\infty}^{\infty} xf(x)$$
$$\int_0^{\infty} xf(x) - \int_{-\infty}^{0} |x| f(x)$$

1)

2)

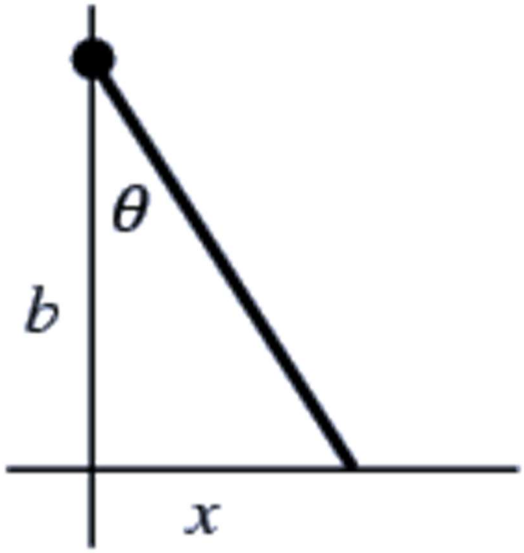
In the case of the Cauchy distribution, both the positive and negative terms of (2) are infinite. This means (2) is undefined.

- Why the 2nd moment is infinite?
  - impossible to consider the variance or standard deviation
  - But, the second moment about zero can be considered. It turns out to be infinite :

$$E(X^2) \propto \int_{-\infty}^{\infty} \frac{x^2}{1+x^2}\,dx = \int_{-\infty}^{\infty} dx - \int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = \infty - \pi = \infty$$

Example:-

- The distribution of horizontal distances at which a line segment tilted at a random angle cuts the *x*-axis.



$$\tan\theta = \frac{x}{b} \qquad \theta = \tan^{-1}\left(\frac{x}{b}\right)$$

$$d\theta = -\frac{1}{1+\frac{x^2}{b^2}}\frac{dx}{b} = -\frac{b\,dx}{b^2+x^2}$$

$$\frac{d\theta}{\pi} = -\frac{1}{\pi}\frac{b\,dx}{b^2+x^2}$$

$$\int_{-\pi/2}^{\pi/2} \frac{d\theta}{\pi} = 1$$

$$-\int_{-\infty}^{\infty} \frac{1}{\pi}\frac{b\,dx}{b^2+x^2} = \frac{1}{\pi}\left[\tan^{-1}\left(\frac{x}{b}\right)\right]_{-\infty}^{\infty}$$

$$= \frac{1}{\pi}\left[\frac{1}{2}\pi - \frac{1}{2}\pi\right]$$

Recurrence Relations

Definition:

A ***recurrence relation*** is the recursive part of a *recursive definition* of either a number sequence or integer function.

For E.g. Recall the Fibonacci sequence:

$$\{f_n\} = 0,1,1,2,3,5,8,13,21,34,55,\ldots$$

Recursive definition for $\{f_n\}$:

INITIALIZE: $f_0 = 0, f_1 = 1$

RECURSE: $f_n = f_{n-1}+f_{n-2}$ for $n > 1$.

The recurrence relation is the recursive part $f_n = f_{n-1}+f_{n-2}$.Thus a recurrence relation for a sequence consists of an equation that expresses each term in terms of lower terms.

Q: Is there another solution to the Fibonacci recurrence relation?

A: Yes, for example could give a different set of ***initial conditions*** such as $f_0=1, f_1= -1$ in which case would get the sequence

$$\{f_n\} = 1,-1,0,-1,-2,-3,-5,-8,-13,-21,\ldots$$

Q: How many solutions are there to the Fibonacci recursion relation?

A: Infinitely many solutions as each pair of integer initial conditions (a,b) generates a unique solution.

Recurrence Relations for Counting

Often it is very hard to come up with a closed formula for counting a particular set, but coming up with recurrence relation easier.

e.g. Geometric example of counting the number of points of intersection of n lines.

Q: What are the initial conditions?

A: Need to give enough initial conditions to avoid ensure well-definedness. The smallest n for which length is well defined is n=0. Thus the smallest n for which an = 2an-1 + 2n-3 - an-3 makes sense is n=3. Thus need to give a0, a1 and a2 explicitly.

a0 = a1 = 0 (strings to short to contain 00)

a2 = 1 (must be 00).

Note: example 6 on p. 313 gives the simpler recursion relation bn = bn-1 + bn-2 for strings which do not contain two consecutive 0's.

Solving Recurrence Relations

We will learn how to give closed solutions to certain kinds of recurrence relations. Unfortunately, most recurrence relations cannot be solved analytically.

EG: If you can find a closed formula for partition function tells me!

However, recurrence relations can all be solved quickly by using *dynamic programming*.

Numerical Solutions Dynamic Programming

Recursion + Lookup Table = Dynamic Programming

Consider a recurrence relation of the form:

$a_n = f(a_0, a_1, \ldots, a_{n-2}, a_{n-1})$

Then can always solve the recurrence relation for first n values by using following pseudocode:

Linear Recurrences

The only case for which telescoping works with a high probability is when the recurrence gives the next value in terms of a single previous value. But…

There is a class of recurrence relations which *can* be solved analytically in general. These are called *linear recurrences* and include the Fibonacci recurrence.

Begin by showing how to solve Fibonacci:

Solving Fibonacci

Recipe solution has 3 basic steps:

1) Assume solution of the form $a_n = r^n$
2) Find all possible r's that seem to make this work. Call these[1] r_1 and r_2. Modify assumed solution to ***general solution*** $a_n = Ar_1^n + Br_2^n$ where A,B are constants.
3) Use initial conditions to find A,B and obtain specific solution.

   A) Assume exponential solution of the form $a_n = r^n$:

Plug this into $a_n = a_{n-1} + a_{n-2}$:

$r^n = r^{n-1} + r^{n-2}$

Notice that all three terms have a common r^{n-2} factor, so divide this out:

$r^n / r^{n-2} = (r^{n-1} + r^{n-2}) / r^{n-2} \Rightarrow r^2 = r + 1$

This equation is called the ***characteristic equation*** of the recurrence relation.

B) Find all possible r's that solve characteristic

$r^2 = r + 1$

Call these r_1 and r_2.[1] General solution is $a_n = Ar_1^n + Br_2^n$ where A,B are constants.

Quadratic formula[2] gives:

$r = (1 \pm \sqrt{5})/2$

So $r_1 = (1+\sqrt{5})/2$, $r_2 = (1-\sqrt{5})/2$

General solution:

$a_n = A\,[(1+\sqrt{5})/2]^n + B\,[(1-\sqrt{5})/2]^n$

Linear Recurrences with Constant Coefficients

Previous method generalizes to solving "*linear recurrence relations with constant coefficients*":

DEF: A recurrence relation is said to be *linear* if a_n is a linear combination of the previous terms plus a function of n. I.e. no squares, cubes or other complicated function of the previous a_i can occur. If in addition all the coefficients are constants then the recurrence relation is said to have ***constant coefficients***.

Homogeneous Linear Recurrences

To solve such recurrences we must first know how to solve an easier type of recurrence relation:

DEF: A linear recurrence relation is said to be ***homogeneous*** if it is a linear combination of the previous terms of the recurrence *without* an additional function of n.

Q: Which of the following are homogeneous?

1. $a_n = 2a_{n-1}$
2. $a_n = 2a_{n-1} + 2^{n-3} - a_{n-3}$

A:

1. $a_n = 2a_{n-1}$: YES
2. $a_n = 2a_{n-1} + 2^{n-3} - a_{n-3}$: No. There's an extra term $f(n) = 2^{n-3}$

Homogeneous Complications

1) *Repeating roots* in characteristic equation. Repeating roots imply that don't learn anything new from second root, so may not have enough information to solve formula with given initial conditions. We'll see how to deal with this on next slide.
2) *Non-real number roots* in characteristic equation. If the sequence has periodic behavior, may get complex roots (for example $a_n = -a_{n-2}$)[1]. We won't worry about this case (in principle, same method works as before, except use complex arithmetic).

Note: [1] The characteristic equation of $a_n = -a_{n-2}$ is $r^2+1 = 0$ which gives $r = i, -i$ where i is the square root of -1. Nominally, this means that the general solutions has the form:

$$a_n = Ai^n + B(-i)^n$$

which is seemingly a non-real number

Complication: Repeating Roots

EG: Solve $a_n = 2a_{n-1} - a_{n-2}$, $a_0 = 1$, $a_1 = 2$

Find characteristic equation by plugging in $a_n = r^n$:

$$r^2 - 2r + 1 = 0$$

Since $r^2 - 2r + 1 = (r-1)^2$ the root $r = 1$ repeats.

If we tried to solve by using general solution

$$a_n = Ar_1^n + Br_2^n = A1^n + B1^n = A+B$$

which forces a_n to be a constant function (→←).

Solution: Multiply second solution by n so general solution looks like:

$$a_n = Ar_1^n + Bnr_1^n$$

Note: If had more repeating roots, would multiply each new solution by another n. So if r_1 is repeated four times general solution would look like:

$$a_n = Ar_1^n + Bnr_1^n + Cn^2r_1^n + Dn^3r_1^n$$

Q. Solve $a_n = 2a_{n-1} - a_{n-2}$, $a_0 = 1$, $a_1 = 2$

General solution: $a_n = A1^n + Bn1^n = A + Bn$

Plug into initial conditions

$$1 = a_0 = A + B \cdot 0 \cdot 1^0 = A$$

$$2 = a_0 = A \cdot 1^1 + B \cdot 1 \cdot 1^1 = A + B$$

Plugging first equation $A = 1$ into second:

$$2 = 1 + B \text{ implies } B = 1.$$

Final answer: $a_n = 1 + n$

The Nonhomogeneous Case

Consider the Tower of Hanoi recurrence (see Rosen p. 311-313) $a_n = 2a_{n-1} + 1$.

Could solve using telescoping. Instead let's solve it methodically. Rewrite:

$a_n - 2a_{n-1} = 1$

1) Solve with the RHS set to 0, i.e. solve the homogeneous case.
2) Add a particular solution to get general solution. I.e. use rule:

General nonhomogenous = General Homogenous + Particular nonhomogenous

The Nonhomogeneous Case

$a_n - 2a_{n-1} = 1$

1) Solve with the RHS set to 0, i.e. solve

$a_n - 2a_{n-1} = 0$

Characteristic equation: $r - 2 = 0$

so unique root is $r = 2$. General solution to homogeneous equation is

$a_n = A \cdot 2^n$

2) Add a particular solution to get general solution for $a_n - 2a_{n-1} = 1$. Use rule:

General nonhomogenous = General Homogenous + Particular nonhomogenous

There are little tricks for guessing particular nonhomogeneous solutions. For example, when the RHS is constant, the guess should also be a constant.[1]

So guess a particular solution of the form $b_n=C$.

Plug into the original recursion:

$1 = b_n - 2b_{n-1} = C - 2C = -C$. Therefore $C = -1$.

General solution: $a_n = A\cdot 2^n - 1$.

Notes: [1] If the non-homogeneous part of the equation is a polynomial $P(n)$ of degree k, then you should guess a particular solution of the form:

$A_k n^k + A_{k-1} n^{k-1} + A_{k-2} n^{k-2} + \ldots + A_1 n + A_0$

Where the A_i are constants that are determined by plugging into the recurrence relation. If instead, the inhomogeneous part is $P(n)\cdot s^n$ [polynomial·exponential] and s is not a root of the characteristic equation, the guess should be modified to:

$[A_k n^k + A_{k-1} n^{k-1} + A_{k-2} n^{k-2} + \ldots + A_1 n + A_0] \cdot s^n$

Finally, if the inhomogeneous part is $P(n)\cdot s^n$ s a root of the characteristic equation which is repeated m times [m=1 means the root occurs but is not repeating], the guess should be modified to:

$[A_k n^k + A_{k-1} n^{k-1} + A_{k-2} n^{k-2} + \ldots + A_1 n + A_0] \cdot s^n n^m$

Finally, use initial conditions to get closed solution. In the case of the Towers of Hanoi recursion, initial condition is:

$a_1 = 1$

Using general solution $a_n = A\cdot 2^n - 1$ we get:

$1 = a_1 = A\cdot 2^1 - 1 = 2A - 1$.

Therefore, $2 = 2A$, so $A = 1$.

Final answer: $a_n = 2^n - 1$

More Complicated

EG: Find the general solution to recurrence from the bit strings example: $a_n = 2a_{n-1} + 2^{n-3} - a_{n-3}$

1) Rewrite as $a_n - 2a_{n-1} + a_{n-3} = 2^{n-3}$ and solve homogeneous part:

Characteristic equation: $r^3 - 2r + 1 = 0$.

Guess root $r = \pm 1$ as integer roots   divide.

$r = 1$ works, so divide out by $(r - 1)$ to get

$r^3 - 2r + 1 = (r - 1)(r^2 + r - 1)$.

$r^3 - 2r + 1 = (r - 1)(r^2 + r - 1)$.

Quadratic formula on $r^2 + r - 1$:

$r = (-1 \pm \sqrt{5})/2$

So $r_1 = 1$, $r_2 = (-1+\sqrt{5})/2$, $r_3 = (-1-\sqrt{5})/2$

General homogeneous solution:

$a_n = A + B\,[(-1+\sqrt{5})/2]^n + C\,[(-1-\sqrt{5})/2]^n$

2) Nonhomogeneous particular solution to $a_n - 2a_{n-1} + a_{n-3} = 2^{n-3}$

Guess the form $b_n = k\,2^n$. Plug guess in:

$k\,2^n - 2k\,2^{n-1} + k\,2^{n-3} = 2^{n-3}$

Simplifies to: $k = 1$.

So particular solution is $b_n = 2^n$

General nonhomogenous = General Homogenous + Particular nonhomogenous

Final answer:

$a_n = A + B\,[(-1+\sqrt{5})/2]^n + C\,[(-1-\sqrt{5})/2]^n + 2^n$

Ordinary generating function

The *ordinary generating function* of a sequence a_n is

$$G(Z) = G(a_n; x) = \sum_{n=0}^{\infty} a_n x^n.$$

When the term *generating function* is used without qualification, it is usually taken to mean an ordinary generating function.

If a_n is the probability mass function of a discrete random variable, then its ordinary generating function is called a probability-generating function.

Polynomials are a special case of ordinary generating functions, corresponding to finite sequences, or equivalently sequences that vanish after a certain point. These are important in that many finite sequences can usefully be interpreted as generating functions, such as the Poincaré polynomial, and others.

A key generating function is the constant sequence 1, 1, 1, 1, 1, 1, 1, 1, 1, ..., whose ordinary generating function is

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}.$$

Alternatively, the right-hand side expression can be justified by multiplying the power series on the left by $1 - x$, and checking that the result is the constant power series 1, in other words that all coefficients except the one of x^0 vanish. Moreover there can be no other power series with this property. The left-hand side therefore designates the multiplicative inverse of $1 - x$ in the ring of power series.

Expressions for the ordinary generating function of other sequences are easily derived from this one. For instance, the substitution $x \to ax$ gives the generating function for the geometric sequence $1, a, a^2, a^3, \ldots$ for any constant a:

$$\sum_{n=0}^{\infty} (ax)^n = \frac{1}{1-ax}\,.$$

(The equality also follows directly from the fact that the left-hand side is the Maclaurin series expansion of the right-hand side.) In particular,

$$\sum_{n=0}^{\infty} (-1)^n x^n = \frac{1}{1+x}\,.$$

One can also introduce regular "gaps" in the sequence by replacing x by some power of x, so for instance for the sequence 1, 0, 1, 0, 1, 0, 1, 0, ... one gets the generating function

$$\sum_{n=0}^{\infty} x^{2n} = \frac{1}{1-x^2}.$$

Using Generating functions to solve recurrence relations

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

Basics of Sampling Theory

P = { $x_1, x_2, \ldots\ldots, x_N$}

where P = population

$x_1, x_2, \ldots\ldots, x_N$ are real numbers

Assuming x is a random variable;

Mean/Average of x,

$$\overline{\mathbf{X}} = \sum_{i=0}^{n} \mathbf{xi}$$

Standard Deviation,

$$\boldsymbol{\sigma} = \sqrt{\frac{1}{N}\sum_{i=0}^{n} \mathbf{fi}(\mathbf{xi} - \ddot{\mathbf{x}})}^2$$

Variance,

$$\boldsymbol{\sigma}^2 = \frac{1}{N}\sum_{i=0}^{n} \mathbf{fi}(\mathbf{xi} - \ddot{\mathbf{x}})^2$$

Basics of Sampling Theory---

Theorem About Mean

picking random numbers x, mean = x

picking random numbers y, mean = y

x = y

Picking another number z,

mean z = x = y

$z = c_1x + c_2y$; c_1, c_2 are constants

$\overline{z} = \overline{x} + \overline{y}$

Independence

two events are independent if the occurrence of one of the events gives no information about whether or not the other event will occur; that is, the events have no influence on each other

for example a, b and c are independent if:

- a and b are independent; a and c are independent; and b and c are independent

Theorem **About Variances/Sampling Theorem**

$z = (x + y)/2; \quad \sigma_z^2 = ? \qquad \sigma_z^2 < \sigma_x^2$

Taking, $z = (x + y)/2$

$\sigma_z^2 = (\sigma_x^{2\,+}\,\sigma_y^2)/4$

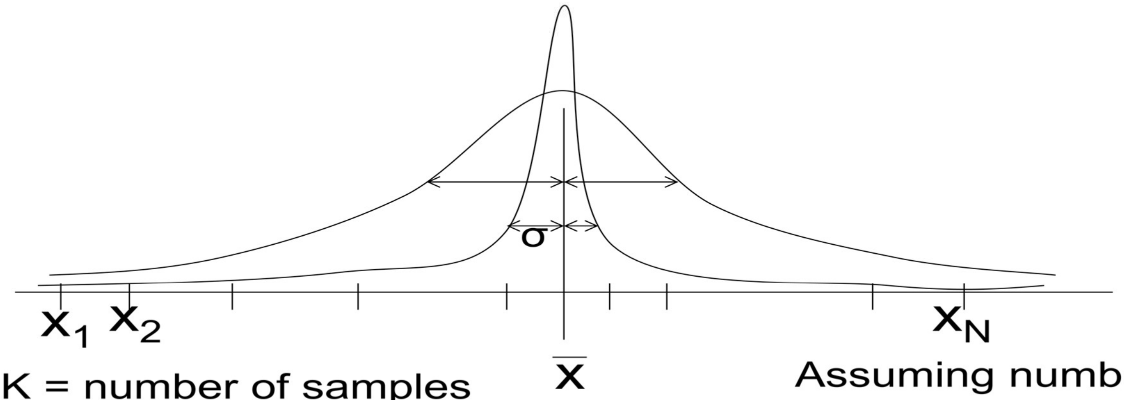
Taking k sample, $z = (x + x' + x'' + \ldots. + x''^{\ldots k})/k$

$\sigma_z^2 = (k\sigma_x^2\,)/k^2$

$\sigma_z^2 = \sigma_x^2\,)/k$

Basics of Sampling Theory

Normal Distribution curve



Assuming numbers are sorted

- K = number of samples
- $\overline{z}$ = sample mean
- as k increases, $\overline{z}$ comes closer to $\overline{x}$

Variance:-

In probability theory and statistics, the variance is a measure of how far a set of numbers is spread out. It is one of several descriptors of a probability distribution, describing how far the numbers lie from the mean (expected value). In particular, the variance is one of the moments of a distribution. In that context, it forms part of a systematic approach to distinguishing between probability distributions. While other such approaches have been developed, those based on moments are advantageous in terms of mathematical and computational simplicity.

The variance is a parameter describing in part either the actual probability distribution of an observed population of numbers, or the theoretical probability distribution of a sample (a not-fully-observed population) of numbers. In the latter case a sample of data from such a distribution can be used to construct an estimate of its variance: in the simplest cases this estimate can be the sample variance, defined below.

Examples: The variance of a random variable or distribution is the expectation, or mean, of the squared deviation of that variable from its expected value or mean. Thus the variance is a measure of the amount of variation of the values of that variable, taking account of all possible values and their probabilities or weightings (not just the extremes which give the range).

For example, a perfect six-sided die, when thrown, has expected value of

$$\frac{1}{6}(1+2+3+4+5+6)=3.5.$$

Its expected absolute deviation—the mean of the equally likely absolute deviations from the mean—is

$$\frac{1}{6}(|1-3.5|+|2-3.5|+|3-3.5|+|4-3.5|+|5-3.5|+|6-3.5|)=\frac{1}{6}(2.5+1.5+0.5+0.5+1.5+2.5)=1.5.$$

But its expected squared deviation—its variance (the mean of the equally likely squared deviations)—is

$$\frac{1}{6}(2.5^2+1.5^2+0.5^2+0.5^2+1.5^2+2.5^2)=17.5/6\approx 2.9.$$

As another example, if a coin is tossed twice, the number of heads is: 0 with probability 0.25, 1 with probability 0.5 and 2 with probability 0.25. Thus the expected value of the number of heads is:

$$0.25 \times 0 + 0.5 \times 1 + 0.25 \times 2 = 1,$$

and the variance is:

$$0.25 \times (0-1)^2 + 0.5 \times (1-1)^2 + 0.25 \times (2-1)^2 = 0.25 + 0 + 0.25 = 0.5.$$

Units of measurement

Unlike expected absolute deviation, the variance of a variable has units that are the square of the units of the variable itself. For example, a variable measured in inches will have a variance measured in square inches. For this reason, describing data sets via their standard deviation or root mean square deviation is often preferred over using the variance. In the dice example the standard deviation is $\sqrt{2.9} \approx 1.7$, slightly larger than the expected absolute deviation of 1.5.

The standard deviation and the expected absolute deviation can both be used as an indicator of the "spread" of a distribution. The standard deviation is more amenable to algebraic manipulation than the expected absolute deviation, and, together with variance and its generalization covariance, is used frequently in theoretical statistics; however the expected absolute deviation tends to be more robust as it is less sensitive to outliers arising from measurement anomalies or an unduly heavy-tailed distribution.

Estimating the variance

Real-world distributions such as the distribution of yesterday's rain throughout the day are typically not fully known, unlike the behavior of perfect dice or an ideal distribution such as the normal distribution, because it is impractical to account for every raindrop. Instead one estimates the mean and variance of the whole distribution as the computed mean and variance of a sample of n observations drawn suitably randomly from the whole sample space, in this example the set of all measurements of yesterday's rainfall in all available rain gauges.

This method of estimation is close to optimal, with the caveat that it underestimates the variance by a factor of $(n - 1) / n$. (For example, when $n = 1$ the variance of a single observation is obviously zero regardless of the true variance). This gives a bias which should be corrected for when n is small by multiplying by $n / (n - 1)$. If the mean is determined in some other way than from the same samples used to estimate the variance then this bias does not arise and the variance can safely be estimated as that of the samples.

To illustrate the relation between the population variance and the sample variance, suppose that in the (not entirely observed) population of numerical values, the value 1 occurs 1/3 of the time, the value 2 occurs 1/3 of the time, and the value 4 occurs 1/3 of the time. The population mean is $(1/3)[1 + 2 + 4] = 7/3$. The equally likely deviations from the population mean are $1 - 7/3$, $2 - 7/3$, and $4 - 7/3$. The population variance — the expected squared deviation from the mean 7/3 — is $(1/3)[(-4/3)^2 + (-1/3)^2 + (5/3)^2] = 14/9$. Now suppose for the sake of a simple example that we take a very small sample of $n = 2$ observations, and consider the nine equally likely possibilities for the set of numbers within that sample: (1, 1), (1, 2), (1,4), (2, 1), (2,2), (2, 4), (4,1), (4, 2), and (4, 4). For these nine possible samples, the sample variance of the two numbers is respectively 0, 1/4, 9/4, 1/4, 0, 4/4, 9/4, 4/4, and 0. With our plan to observe two values, we could end up computing any of these sample variances (and indeed if we hypothetically could observe a pair of numbers many times, we would compute each of these sample variances 1/9 of the time). So the expected value, over all possible samples that might be drawn from the population, of the computed sample variance is (1/9)[0 + 1/4 + 9/4 + 1/4 + 0 + 4/4 + 9/4 + 4/4 + 0] = 7/9. This value of 7/9 for the expected value of our sample variance computation is a substantial underestimate of the true population variance, which we computed as 14/9, because our sample size of just two observations was so small. But if we adjust for this downward bias by multiplying our computed sample variance, whichever it may be, by $n/(n-1) = 2/(2-1) = 2$, then our estimate of the population variance would be any one of 0, 1/2, 9/2, 1/2, 0, 4/2, 9/2, 4/2, and 0. The average of these is indeed the correct population variance of 14/9, so on average over all possible samples we would have the correct estimate of the population variance.

The variance of a real-valued random variable is its second central moment, and it also happens to be its second cumulant. Just as some distributions do not have a mean, some do not have a variance. The mean exists whenever the variance exists, but the converse is not necessarily true.

Definition

If a random variable X has the expected value (mean) $\mu = E[X]$, then the variance of X is given by:

$$\operatorname{Var}(X) = \operatorname{E}\left[(X-\mu)^2\right].$$

That is, the variance is the expected value of the squared difference between the variable's realization and the variable's mean. This definition encompasses random variables that are discrete, continuous, or neither (or mixed). It can be expanded as follows:

$$\begin{aligned}
\operatorname{Var}(X) &= \operatorname{E}\left[(X-\mu)^2\right] \\
&= \operatorname{E}\left[X^2 - 2\mu X + \mu^2\right] \\
&= \operatorname{E}\left[X^2\right] - 2\mu\,\operatorname{E}[X] + \mu^2 \\
&= \operatorname{E}\left[X^2\right] - 2\mu^2 + \mu^2 \\
&= \operatorname{E}\left[X^2\right] - \mu^2 \\
&= \operatorname{E}\left[X^2\right] - (\operatorname{E}[X])^2.
\end{aligned}$$

A mnemonic for the above expression is "mean of square minus square of mean". The variance of random variable X is typically designated as Var(X), σ_X^2, or simply σ^2 (pronounced "sigma squared").

Continuous random variable

If the random variable X is continuous with probability density function f(x), then the variance equals the second central moment, given by

$$\operatorname{Var}(X) = \int (x-\mu)^2 f(x)\,dx\,,$$

where $\mu$is the expected value,

$$\mu = \int x\,f(x)\,dx\,,$$

and where the integrals are definite integrals taken for x ranging over the range of X.

If a continuous distribution does not have an expected value, as is the case for the Cauchy distribution, it does not have a variance either. Many other distributions for which the expected value does exist also do not have a finite variance because the integral in the variance definition diverges. An example is a Pareto distribution whose index k satisfies $1 < k \leq 2$.

Discrete random variable

If the random variable X is discrete with probability mass function $x_1 \mapsto p_1, ..., x_n \mapsto p_n$, then

$$\operatorname{Var}(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2$$

where $\mu$is the expected value, i.e.

$$\mu = \sum_{i=1}^{n} p_i \cdot x_i$$
.

(When such a discrete weighted variance is specified by weights whose sum is not 1, then one divides by the sum of the weights.) That is, it is the expected value of the square of the deviation of X from its own mean. In plain language, it can be expressed as "The mean of the squares of the deviations of the data points from the average". It is thus the mean squared deviation.

Examples

Exponential distribution

The exponential distribution with parameter λ is a continuous distribution whose support is the semi-infinite interval [0,∞). Its probability density function is given by:

$$f(x) = \lambda e^{-\lambda x},$$

and it has expected value $\mu = \lambda^{-1}$. Therefore the variance is equal to:

$$\int_0^\infty f(x)(x-\mu)^2\,dx = \int_0^\infty \lambda e^{-\lambda x}(x-\lambda^{-1})^2\,dx = \lambda^{-2}.$$

So for an exponentially distributed random variable $\sigma^2 = \mu^2$.

Fair dice

A six-sided fair die can be modelled with a discrete random variable with outcomes 1 through 6, each with equal probability $\frac{1}{6}$. The expected value is $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$. Therefore the variance can be computed to be:

$$\begin{aligned}\sum_{i=1}^{6} \tfrac{1}{6}(i-3.5)^2 = \tfrac{1}{6}\sum_{i=1}^{6}(i-3.5)^2 &= \tfrac{1}{6}\left((-2.5)^2+(-1.5)^2+(-0.5)^2+0.5^2+1.5^2+2.5^2\right)\\ &= \tfrac{1}{6}\cdot 17.50 = \tfrac{35}{12} \approx 2.92.\end{aligned}$$

The general formula for the variance of the outcome X of a die of n sides is:

$$\begin{aligned}\sigma^2 = E(X^2) - (E(X))^2 &= \frac{1}{n}\sum_{i=1}^{n} i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} i\right)^2\\ &= \tfrac{1}{6}(n+1)(2n+1) - \tfrac{1}{4}(n+1)^2\\ &= \frac{n^2-1}{12}.\end{aligned}$$

Properties

Variance is non-negative because the squares are positive or zero.

$$\operatorname{Var}(X) \geq 0.$$

The variance of a constant random variable is zero, and if the variance of a variable in a data set is 0, then all the entries have the same value.

$$P(X = a) = 1 \Leftrightarrow \operatorname{Var}(X) = 0.$$

Variance is invariant with respect to changes in a location parameter. That is, if a constant is added to all values of the variable, the variance is unchanged.

$$\operatorname{Var}(X+a)=\operatorname{Var}(X).$$

If all values are scaled by a constant, the variance is scaled by the square of that constant.

$$\operatorname{Var}(aX)=a^2\operatorname{Var}(X).$$

The variance of a sum of two random variables is given by:

$$\operatorname{Var}(aX+bY)=a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)+2ab\operatorname{Cov}(X,Y),$$

$$\operatorname{Var}(X-Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)-2\operatorname{Cov}(X,Y),$$

In general we have for the sum of N random variables:

$$\operatorname{Var}\left(\sum_{i=1}^{N}X_i\right)=\sum_{i,j=1}^{N}\operatorname{Cov}(X_i,X_j)=\sum_{i=1}^{N}\operatorname{Var}(X_i)+\sum_{i\neq j}\operatorname{Cov}(X_i,X_j).$$

The variance of a finite sum of uncorrelated random variables is equal to the sum of their variances. This stems from the above identity and the fact that for uncorrelated variables the covariance is zero.

$$\operatorname{Cov}(X_i,X_j)=0\ (i\neq j)\Rightarrow\operatorname{Var}\left(\sum_{i=1}^{N}X_i\right)=\sum_{i=1}^{N}\operatorname{Var}(X_i).$$

These results lead to the variance of a linear combination as:

$$\begin{aligned}\operatorname{Var}\left(\sum_{i=1}^{N}a_iX_i\right)&=\sum_{i=1}^{N}\sum_{j=1}^{N}a_ia_j\operatorname{Cov}(X_i,X_j)\\&=\sum_{i=1}^{N}a_i^2\operatorname{Var}(X_i)+\sum_{i\neq j}a_ia_j\operatorname{Cov}(X_i,X_j)\\&=\sum_{i=1}^{N}a_i^2\operatorname{Var}(X_i)+2\sum_{i<j}a_ia_j\operatorname{Cov}(X_i,X_j).\end{aligned}$$

Suppose that the observations can be partitioned into equal-sized subgroups according to some second variable. Then the variance of the total group is equal to the mean of the variances of the subgroups plus the variance of the means of the subgroups. This property is known as variance decomposition or the law of total variance and plays an important role in the analysis of variance. For example, suppose that a group consists of a subgroup of men and an equally large subgroup of women. Suppose that the men have a mean height of 180 and that the variance of their heights is 100. Suppose that the women have a mean height of 160 and that the variance of their heights is 50. Then the mean of the variances is (100 + 50) / 2 = 75; the variance of the means is the variance of 180, 160 which is 100. Then, for the total group of men and women combined, the variance of the height will be 75 + 100 = 175. Note that this uses N for the denominator instead of N − 1.

In a more general case, if the subgroups have unequal sizes, then they must be weighted proportionally to their size in the computations of the means and variances. The formula is also valid with more than two groups, and even if the grouping variable is continuous.

This formula implies that the variance of the total group cannot be smaller than the mean of the variances of the subgroups. Note, however, that the total variance is not necessarily larger than the variances of the subgroups. In the above example, when the subgroups are analyzed separately, the variance is influenced only by the man-man differences and the woman-woman differences. If the two groups are combined, however, then the men-women differences enter into the variance also.

Many computational formulas for the variance are based on this equality: The variance is equal to the mean of the square minus the square of the mean:

$$\operatorname{Var}(X) = \operatorname{E}[X^2] - \operatorname{E}[X]^2.$$

For example, if we consider the numbers 1, 2, 3, 4 then the mean of the squares is (1 × 1 + 2 × 2 + 3 × 3 + 4 × 4) / 4 = 7.5. The regular mean of all four numbers is 2.5, so the square of the mean is 6.25. Therefore the variance is 7.5 − 6.25 = 1.25, which is indeed the same result obtained earlier with the definition formulas. Many pocket calculators use an algorithm that is based on this formula and that allows them to compute the variance while the data are entered, without storing all values in memory. The algorithm is to adjust only three variables when a new data value is entered: The number of data entered so far (n), the sum of the values so far (S), and the sum of the squared values so far (SS). For example, if the data are 1, 2, 3, 4, then after entering the first value, the algorithm would have n = 1, S = 1 and SS = 1. After entering the second value (2), it would have n = 2, S = 3 and SS = 5. When all data are entered, it would have n = 4, S = 10 and SS = 30. Next, the mean is computed as M = S / n, and finally the variance is computed as SS / n − M × M. In this example the outcome would be 30 / 4 − 2.5 × 2.5 = 7.5 − 6.25 = 1.25. If the unbiased sample estimate is to be computed, the outcome will be multiplied by n / (n − 1), which yields 1.667 in this example.

Properties, formal

Sum of uncorrelated variables (Bienaymé formula)
See also: Sum of normally distributed random variables

One reason for the use of the variance in preference to other measures of dispersion is that the variance of the sum (or the difference) of uncorrelated random variables is the sum of their variances:

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

This statement is called the Bienaymé formula.[1] and was discovered in 1853. It is often made with the stronger condition that the variables are independent, but uncorrelatedness suffices. So if all the variables have the same variance σ^2, then, since division by n is a linear transformation, this formula immediately implies that the variance of their mean is

$$\operatorname{Var}\left(\overline{X}\right) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}\left(X_i\right) = \frac{\sigma^2}{n}.$$

That is, the variance of the mean decreases when n increases. This formula for the variance of the mean is used in the definition of the standard error of the sample mean, which is used in the central limit theorem.

Product of independent variables

If two variables X and Y are independent, the variance of their product is given by[2][3]

Sum of correlated variables

In general, if the variables are correlated, then the variance of their sum is the sum of their covariances:

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j).$$

(Note: This by definition includes the variance of each variable, since $\operatorname{Cov}(X_i,X_i) = \operatorname{Var}(X_i)$.)

Here Cov is the covariance, which is zero for independent random variables (if it exists). The formula states that the variance of a sum is equal to the sum of all elements in the covariance matrix of the components. This formula is used in the theory of Cronbach's alpha in classical test theory.

So if the variables have equal variance σ^2 and the average correlation of distinct variables is ρ, then the variance of their mean is

$$\operatorname{Var}(\overline{X}) = \frac{\sigma^2}{n} + \frac{n-1}{n}\rho\sigma^2.$$

This implies that the variance of the mean increases with the average of the correlations. Moreover, if the variables have unit variance, for example if they are standardized, then this simplifies to

$$\operatorname{Var}(\overline{X}) = \frac{1}{n} + \frac{n-1}{n}\rho.$$

This formula is used in the Spearman–Brown prediction formula of classical test theory. This converges to ρ if n goes to infinity, provided that the average correlation remains constant or converges too. So for the variance of the mean of standardized variables with equal correlations or converging average correlation we have

$$\lim_{n\to\infty} \operatorname{Var}(\overline{X}) = \rho.$$

Therefore, the variance of the mean of a large number of standardized variables is approximately equal to their average correlation. This makes clear that the sample mean of correlated variables does generally not converge to the population mean, even though the Law of large numbers states that the sample mean will converge for independent variables.

[edit] Weighted sum of variables

The scaling property and the Bienaymé formula, along with this property from the covariance page: Cov(aX, bY) = ab Cov(X, Y) jointly imply that

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab\, \operatorname{Cov}(X, Y).$$

This implies that in a weighted sum of variables, the variable with the largest weight will have a disproportionally large weight in the variance of the total. For example, if X and Y are uncorrelated and the weight of X is two times the weight of Y, then the weight of the variance of X will be four times the weight of the variance of Y.

The expression above can be extended to a weighted sum of multiple variables:

$$\operatorname{Var}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \operatorname{Var}(X_i) + 2\sum_i \sum_{j>i} a_i a_j \operatorname{Cov}(X_i, X_j)$$

Decomposition

The general formula for variance decomposition or the law of total variance is: If X and Y are two random variables and the variance of X exists, then

$$\operatorname{Var}(X) = \operatorname{Var}(\operatorname{E}(X|Y)) + \operatorname{E}(\operatorname{Var}(X|Y)).$$

Here, $\operatorname{E}(X|Y)$ is the conditional expectation of X given Y, and $\operatorname{Var}(X|Y)$ is the conditional variance of X given Y. (A more intuitive explanation is that given a particular value of Y, then X follows a distribution with mean $\operatorname{E}(X|Y)$ and variance $\operatorname{Var}(X|Y)$. The above formula tells how to find $\operatorname{Var}(X)$ based on the distributions of these two quantities when Y is allowed to vary.) This formula is often applied in analysis of variance, where the corresponding formula is

$$MS_{\text{Total}} = MS_{\text{Between}} + MS_{\text{Within}};$$

here MS refers to the Mean of the Squares. It is also used in linear regression analysis, where the corresponding formula is

$$MS_{\text{Total}} = MS_{\text{Regression}} + MS_{\text{Residual}}.$$

This can also be derived from the additivity of variances, since the total (observed) score is the sum of the predicted score and the error score, where the latter two are uncorrelated.

Similar decompositions are possible for the sum of squared deviations (sum of squares, SS):

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}},$$

$$SS_{\text{Total}} = SS_{\text{Regression}} + SS_{\text{Residual}}.$$

Computational formula

Main article: computational formula for the variance

See also: algorithms for calculating variance

The computational formula for the variance follows in a straightforward manner from the linearity of expected values and the above definition:

$$\begin{aligned}\operatorname{Var}(X) &= \operatorname{E}(X^2 - 2\,X\,\operatorname{E}(X) + (\operatorname{E}(X))^2)\\ &= \operatorname{E}(X^2) - 2(\operatorname{E}(X))^2 + (\operatorname{E}(X))^2\\ &= \operatorname{E}(X^2) - (\operatorname{E}(X))^2.\end{aligned}$$

This is often used to calculate the variance in practice, although it suffers from catastrophic cancellation if the two components of the equation are similar in magnitude.

Characteristic property

The second moment of a random variable attains the minimum value when taken around the first moment (i.e., mean) of the random variable, i.e. $\mathrm{argmin}_m\, \mathrm{E}((X-m)^2) = \mathrm{E}(X)$. Conversely, if a continuous function φ satisfies $\mathrm{argmin}_m\, \mathrm{E}(\varphi(X-m)) = \mathrm{E}(X)$ for all random variables X, then it is necessarily of the form $\varphi(x) = ax^2 + b$, where a > 0. This also holds in the multidimensional case.[4]

Calculation from the CDF

The population variance for a non-negative random variable can be expressed in terms of the cumulative distribution function F using

$$2\int_0^\infty uH(u)\,du - \Big(\int_0^\infty H(u)\,du\Big)^2.$$

where H(u) = 1 − F(u) is the right tail function. This expression can be used to calculate the variance in situations where the CDF, but not the density, can be conveniently expressed.

Matrix notation for the variance of a linear combination

Let's define X as a column vector of n random variables $X_1, \ldots, X_n$, and c as a column vector of N scalars $c_1, \ldots, c_n$. Therefore $c^T X$ is a linear combination of these random variables, where c^T denotes the transpose of vector c. Let also be Σ the variance-covariance matrix of the vector X. The variance of $c^T X$ is given by[5]:

$$\mathrm{Var}(c^T X) = c^T \Sigma c.$$

Approximating the variance of a function

The delta method uses second-order Taylor expansions to approximate the variance of a function of one or more random variables: see Taylor expansions for the moments of functions of random variables. For example, the approximate variance of a function of one variable is given by

$$\mathrm{Var}\left[f(X)\right] \approx \left(f'(\mathrm{E}\left[X\right])\right)^2 \mathrm{Var}\left[X\right]$$

provided that f is twice differentiable and that the mean and variance of X are finite.

Population variance and sample variance

In general, the population variance of a finite population of size N is given by

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

where

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

is the population mean.

In many practical situations, the true variance of a population is not known a priori and must be computed somehow. When dealing with extremely large populations, it is not possible to count every object in the population.

A common task is to estimate the variance of a population from a sample. We take a sample with replacement of n values $y_1, ..., y_n$ from the population, where n < N, and estimate the variance on the basis of this sample. There are several good estimators. Two of them are well known:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) - \overline{y}^2,$$ and

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i^2 - \overline{y}^2\right) \\ &= \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{n}{n-1}\overline{y}^2 \end{aligned}$$

The first estimator, also known as the second central moment, is called the biased sample variance. The second estimator is called the unbiased sample variance. Either estimator may be simply referred to as the sample variance when the version can be determined by context. Here, $\overline{y}$ denotes the sample mean:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

The two estimators only differ slightly as can be seen, and for larger values of the sample size n the difference is negligible. While the first one may be seen as the variance of the sample considered as a population, the second one is the unbiased estimator of the population variance, meaning that its expected value $E[s^2]$ is equal to the true variance of the sampled random variable; the use of the term $n - 1$ is called Bessel's correction. In particular,

$$\mathrm{E}[s^2] = \sigma^2,$$

while, in contrast,

$$\mathrm{E}[s_n^2] = \frac{n-1}{n}\sigma^2.$$

The unbiased sample variance is a U-statistic for the function $f(x_1, x_2) = (x_1 - x_2)^2/2$, meaning that it is obtained by averaging a 2-sample statistic over 2-element subsets of the population.

Distribution of the sample variance

Being a function of random variables, the sample variance is itself a random variable, and it is natural to study its distribution. In the case that y_i are independent observations from a normal distribution, Cochran's theorem shows that s^2 follows a scaled chi-squared distribution.

$$(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

As a direct consequence, it follows that $E(s^2) = \sigma^2$.

If the y_i are independent and identically distributed, but not necessarily normally distributed, then

$$\mathrm{E}[s^2] = \sigma^2, \quad \mathrm{Var}[s^2] = \sigma^4\left(\frac{2}{n-1} + \frac{\kappa}{n}\right),$$

where κ is the excess kurtosis of the distribution. If the conditions of the law of large numbers hold, s^2 is a consistent estimator of σ^2.

Samuelson's inequality

Samuelson's inequality is a result that states, given that the sample mean and variance have been calculated from a particular sample, bounds on the values that individual values in the sample can take.[8] Values must lie within the limits $m \pm s(n - 1)^{1/2}$.

Relations with the harmonic and arithmetic means

It has been shown that for a sample of real numbers that

$$\mathrm{Var} \leq 2M(A - H)$$

where M is the maximum of the sample, A is the arithmetic mean, H is the harmonic mean of the sample and Var is the variance of the sample.

This bound has been improved on and it is known that variance is bounded by

$$\mathrm{Var} \leq \frac{M(A-H)(M-A)}{M-H}$$

$$\mathrm{Var} \geq \frac{m(A-H)(A-m)}{H-m}$$

where m is the minimum of the sample.

Generalizations If X is a vector-valued random variable, with values in $\mathbb{R}^n$, and thought of as a column vector, then the natural generalization of variance is $\mathrm{E}((X-\mu)(X-\mu)^{\mathrm{T}})$, where $\mu = \mathrm{E}(X)$ and X^{T} is the transpose of X, and so is a row vector. This variance is a positive semi-definite square matrix, commonly referred to as the covariance matrix.

If X is a complex-valued random variable, with values in $\mathbb{C}$, then its variance is $\mathrm{E}((X-\mu)(X-\mu)^{\dagger})$, where $X^{\dagger}$ is the conjugate transpose of X. This variance is also a positive semi-definite square matrix.

Introduction to Hypothesis Testing Significance testing is used to help make a judgment about a claim by addressing the question, Can the observed difference be attributed to chance? We break **up significance testing into three (or four) steps:**

Step A: Null and alternative hypotheses

The first step of hypothesis testing is to convert the research question into null and alterative hypotheses. We start with the null hypothesis (H_0). The null hypothesis is a claim of "no difference." The opposing hypothesis is the alternative hypothesis (H_1). The alternative hypothesis is a claim of "a difference in the population," and is the hypothesis the researcher often hopes to bolster. It is important to keep in mind that the null and alternative hypotheses reference population values, and not observed statistics.

Step B: Test statistic

We calculate a test statistic from the data. There are different types of test statistics. This chapter introduces the one-sample z-statistics. The z statistic will compare the observed sample mean to an expected population mean μ_0. Large test statistics indicate data are far from expected, providing evidence against the null hypothesis and in favor of the alternative hypothesis.

Step C: p Value and conclusion

The test statistic is converted to a conditional probability called a P-value. The P- value answers the question "If the null hypothesis were true, what is the probability of observing the current data or data that is more extreme?"

Small p values provide evidence against the null hypothesis because they say the observed data are unlikely when the null hypothesis is true. We apply the following conventions:

- When p value > .10 → the observed difference is "not significant"
- When p value ≤ .10 → the observed difference is "marginally significant"
- When p value ≤ .05 → the observed difference is "significant"
- When p value ≤ .01 → the observed difference is "highly significant"

Use of "significant" in this context means "the observed difference is not likely due to chance." It does not mean of "important" or "meaningful."

Step D: Decision (optional)

Alpha (α) is a probability threshold for a decision. If P ≤ α, we will reject the null hypothesis. Otherwise it will be retained for want of evidence.

Page

What is a Monte Carlo simulation:-

- In a Monte Carlo simulation we attempt to follow the `time dependence' of a model for which change, or growth, does not proceed in some rigorously predefined fashion (e.g. according to Newton's equations of motion) but rather in a stochastic manner which depends on a sequence of random numbers which is generated during the simulation.

Random Walk:

- Markov chain is a sequence of events with the condition that the probability of each succeeding event is uninfluenced by prior events
- Choosing from Probability Distribution: Any random variable has a probability distribution for its occurrence. We need to choose a random variable which mimics that probability distribution
- Best way to relate random number to a random variable is to use cumulative probability distribution and equating it to the random number

Random Numbers:

- Uniformly distributed numbers in [0,1]
- Most useful method for obtaining random numbers for computer use is a pseudo random number generator
- How random are these pseudo random numbers?

Application to Microscale Heat Transfer :-

- Boltzmann Transport Equation (BTE) for phonons best describes the heat flow in solid nonmetallic thin films
- difficult to solve analytically or even numerically using deterministic approaches
- alternative is to solve the BTE using stochastic or Monte Carlo techniques

Boltzmann Transport Equation for Particle Transport

Distribution Function of Particles: $f = f(\mathbf{r},\mathbf{p},t)$

--probability of particle occupation of momentum **p** at location **r** and time t

Equilibrium Distribution:

f_0, i.e. Fermi-Dirac for electrons, Bose-Einstein for phonons, Plank for photons, etc.

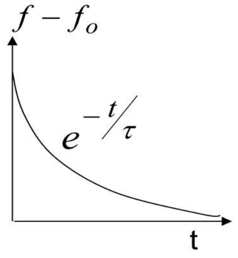
Non-equilibrium, e.g. in a high electric field or temperature gradient:

$$\boxed{\frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}} f + \mathbf{F}\cdot\nabla_{p} f = \left(\frac{\partial f}{\partial t}\right)_{scat}}$$

Relaxation Time Approximation

$$\left(\frac{\partial f}{\partial t}\right)_{scat} = \sum_{\mathbf{p}'}\left[\underbrace{W(\mathbf{p},\mathbf{p}')f(\mathbf{p}')}_{\mathbf{p}'\to\mathbf{p}} - \underbrace{W(\mathbf{p}',\mathbf{p})f(\mathbf{p})}_{\mathbf{p}\to\mathbf{p}'}\right] \approx \frac{f_o - f}{\tau(\mathbf{r},\mathbf{p})}$$

Relaxation time



Monte Carlo Solution Technique:-

- Phonons are drawn from the six individual stochastic spaces, including three wave-vector components and the three position vector components
- Phonons are then allowed to drift (or unrestrained motion) and scatter in time, and their statistics is collected at various points in time and space, and processed to extract the necessary information

Initial Conditions:-

- number of phonons per unit volume and polarization (p) is usually an extremely large number
- a scaling factor is used to simulate only a fraction of the phonons

- A series of random numbers properly distributed to match the equilibrium distribution are drawn to initialize the positions, frequencies, polarizations, and wavevectors of the ensemble of phonons chosen for the simulation

- Mazumdar and Majumdar developed a numerical scheme to obtain the number of phonons within the *i*th frequency interval D*w* as:

$$N_i = \langle n(\omega_{0,i}, LA)\rangle D(\omega_{0,i}, LA)\Delta\omega_i + 2\langle n(\omega_{0,i}, TA)\rangle D(\omega_{0,i}, TA)\Delta\omega_i$$

Monte Carlo Simulation of Silicon Nanowire Thermal Conductivity:-

- Boundary scattering play an important role in thermal resistance as the structure size decreases to nanoscale