Prof.
B.C. Giri

Random Experiment

An expirement E is called a random experiment if :-

(i) All possible outcomes of E are known in advance

(ii) It is impossible to predict which outcome will occur at a particular performance at E.

(iii) E can be repeated.

Q. If the experiment consists of flipping 2 coins, then what is the sample space/event space?

Soln. $\{(H,H), (H,T), (T,T), (T,H)\}$

Axiom 1 :- $0 \leq P(E) \leq 1$.

Axiom 2 :- $P(S) = 1$

Axiom 3 :- $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$.

Q. If two dice are rolled, what is the probability that the sum of the upturned face will equal 7?

Soln. 1/6

Q. If 3 balls are randomly drawn from 6W & 5 black ball. What is the probability that one of the drawn ball is White & other 2 black?

Soln. $\frac{{}^6C_1 \cdot {}^5C_2}{{}^{11}C_3}$.

② Conditional Probability

(1) Let A & B be any two events connected to a given random experiment E. Then the conditional probability of the event A on the hypothesis that the event B has occurred denoted by $P(A|B)$ is defined as :-

$$P(A|B) = \frac{P(AB)}{P(B)}$$

provided $P(B) \neq 0$

Q A coin is flipped. We assume that the sample space are equally likely, what is the conditional probability that both flips result in heads, given that the first flip does ?

Soln.

$$\frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}.$$

$E = \{(H,H)\}$ denotes the event that both flips land head head &

$F = \{(H,H), (H,T)\}$ the event that the first flips land head.

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{P(E \cap F)}{P(F)} = \frac{P(\{H,H\})}{P\{(H,H),(H,T)\}} = \frac{1}{2}.$$

Scanned with CamScanner

Q. In an exam 30% of the students failed in maths, 15% of the students failed in chemistry, 10% of the students failed both chemistry & maths. A student is selected at random. If he failed in chemistry, then what is the probability he passed.

Soln. $P(A|B) = \dfrac{P(A \cap B)}{\cancel{P(A)+}\,P(B)} = \dfrac{\frac{1}{10}}{\cancel{0.3+}\,0.15} = \dfrac{0.1}{0.15} = \dfrac{10}{15}$

$= \dfrac{2}{3}$.

$1 - P(A) = 1/3$.

Q. Two players A & B alternatively throw a pair of die. A wins if A throws 6 before B throws 7 before A throws 6. If A begins, then the probability that A wins is P

Soln.

$$\frac{5}{36}\left[1 + \left(\frac{31}{36}\cdot\frac{5}{6}\right) + \left(\frac{31}{36}\cdot\frac{5}{6}\right)^2 + \cdots\right]$$

$$= \frac{30}{61}$$

Mutually Exclusive events :- Two events connected to a given random experiment E are said to be mutually exclusive if A, B can never happen simultaneously, if $AB = 0$.

Example. Random experiment of throwing a die, the events multiple of 3 and a prime no. are not mutually exclusive.

But the events odd no. & event no are mutually exclusive.

Exhaustive set of Events

The collection of events $\{A_\alpha : \alpha \in I\}$ is exhaustive if and only if $\sum_{\alpha \in I} A_\alpha = S$. where I is an Index set & S is the corresponding event space.

Eg. The Random experiment of throwing a die, the collection of events $\{A_1, A_2, A_3\}$ is exhaustive where $A_1 = \{1, 3, 5\}$, $A_2 = \{2\}$, $A_3 = \{4, 5\}$.

Theorem :- If $A_1, A_2, \ldots, A_n$ be pairwise mutually exhaustive events one which certainly occur then

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B/A_i)$$

Problem!

A jar contains 2 white balls & 3 black ~~black~~ balls. The balls are drawn from the jar one by one and placed on the table in the order drawn. What is the probability that they are drawn in the order white, black, black, white, black?

$$\frac{2}{5} \times \frac{\cancel{3}}{\cancel{4}_2} \times \frac{\cancel{2}}{\cancel{3}} \times \frac{1}{\cancel{2}} \times 1$$

$$= \frac{1}{10}$$

Scanned with CamScanner

Baye's Theorem

Let $A_1, A_2, A_3, \ldots, A_n$ be n pairwise mutually exclusive events ~~connected~~ connected to a random experiment E where at least one of $A_1, A_2, A_3, \ldots, A_n$ is sure to happen. Let X be an arbitrary event connected to E, where $P(X) \neq 0$. Also let probabilities $P(X|A_1), P(X|A_2), \ldots, P(X|A_n)$ are all known.

Then,

$$\boxed{P(A_i|X) = \frac{P(X|A_i) \cdot P(A_i)}{\sum_{r=1}^{n} P(A_r)\, P(X|A_r)}}$$

Problem

(2) ~~Soln~~ From an urn containing 5 white & 5 black balls, 5 balls are transferred at random into an empty second urn from which one ball is drawn and it is found to be white. What is the probability that all balls transferred from the first urn are white?

Soln

E_0 = 5 black balls drawn from the 1st urn.

E_1 = 1 white & 4 black balls

E_2 = 2 W & 3 B

E_3 = 3 & 2 $\quad E_4$ = 4 & 1 $\quad E_5$ = 5 & 0.

Soln contd →

The chance that a doctor will diagnose a certain disease correctly is 60%. The chance that a patient will die by his treatment after correct diagnosis is 40%. and the chance of death by wrong diagnosis is 70%. A patient of the doctor who had the disease dies. What is the probability that the disease ~~was~~ was diagnosed correctly?

Scanned with CamScanner

Let X be the event one white drawn from the 2nd urn. ③

$P(E_0) = \frac{5C_5}{10C_5}$ ~~$\frac{1}{0 + 1/5 + 2}$~~

$= \frac{1}{252} = P(E_5).$

$P(E_1) = \frac{5C_1 \times 5C_4}{252} = \frac{25}{252} = P(E_4)$

$P(E_2) = \frac{5C_2 \times 5C_3}{252} = P(E_3). = \frac{25}{63}$

$$\frac{1 \times \frac{1}{252}}{0 \times \frac{1}{252} + \frac{1}{5} \times \frac{25}{252} + \frac{2}{5} \times \frac{25}{63} + \frac{3}{5} \times \frac{25}{252} + \frac{4}{5} \times \frac{25}{252} + 1 \times 1.}$$

$P(X|E_0) = 0$

~~P(E~~ $P(X|E_1) = 1/5$

$P(X|E_2) = 2/5$

$P(X|E_3) = 3/5$

$P(X|E_4) = 4/5$

$P(X|E_5) = 1$.

$= \frac{1}{126}$.

Soln ③.

$E_{correct diagnosis} = 0.6$ $E_{wrong diagnosis} = 0.4$.

$P(die | E_{correct diagnosis}) = 0.4$

$P(die | E_{wrong diagnosis}) = 0.7$

$$P(\cancel{E} E_{correct} | die) = \frac{\overset{P(E_{correct})}{\cancel{P(die)}} \cdot P(die|E_{correct})}{\cancel{P(die)} \cdot P(die|E_{correct}) P(E_{correct}) + P(die|E_{wrong\,dig}) \cdot P(E_{wrong})}$$

$$= \frac{0.6 \times 0.4}{0.6 \times 0.4 + 0.4 \times 0.7}$$

$$= \frac{0.24}{0.24 + 0.28}$$

$$= \frac{24}{52} = \frac{12}{26} = \frac{6}{13}.$$

Scanned with CamScanner

27/9/2019

Mathematics

Prof B.C. Giri

Q. Sum of 2 numbers is 2n. Find the probability that their product is not less than $\frac{3}{4}$ the maximum pro.

Soln.

~~Prob~~ Max Product = n^2.

$\therefore$ Product $\geq \left(n^2 \cdot \frac{3}{4}\right)$

$$x(2n-x) \geq \frac{3}{4}n^2$$

$$\Rightarrow 2nx - x^2 \geq \frac{3}{4}n^2$$

$$\Rightarrow x^2 - 2nx + \frac{3}{4}n^2 \leq 0$$

$$4x^2 - 8nx + 3n^2 \leq 0$$

$$\Rightarrow (2x-n)(2x-3n) \leq 0$$

$$\Rightarrow \frac{n}{2} \leq x \leq \frac{3n}{2}$$

$$\frac{3n}{2} - \frac{n}{2} = \cancel{\frac{2n}{2}}\, n$$

Ans: $\dfrac{\text{Favourable Range}}{\text{Total Range}} = \dfrac{n}{2n} = \dfrac{1}{2}$.

Q. 5 figure No. is formed by digits 0, 1, 2, 3, 4 without repetition. Probability that the number formed is divisable by 4.

Soln

_ _ _ 4 0 $\rightarrow$ $3 \times 2 \times 1$

3 2 $\rightarrow$ $2 \times 1 + 2 \times 1$ (0 cannot be in MSB)

1 2 $\rightarrow$ 4

0 4 $\rightarrow$ $3 \times 2 \times 1$

2 4 $\rightarrow$ 4

2 0 $\rightarrow$ 6.

Total = $4! \times 4$
= 96

$\therefore$ Probability = $\frac{5}{16}$

Q. A player tosses a coin. He scores 1pt for every head. & 2 pts for every tail turned up. He is to play on until his score reaches to n it P_n is the chance of getting exactly n. Find P_n.

Soln

1s	2s
n	0
n-1	1
n-4	2
⋮	
1	$2\left(\frac{n-1}{2}\right)$

$(n-1) \rightarrow$ head

$(n-2) \rightarrow$ tail.

$\therefore P_n = \frac{1}{2}(P_{n-1} + P_{n-2})$.

$\Rightarrow 2P_n = P_{n-1} + P_{n-2}$

$P_n = \frac{2}{3} + \left[(-1)^n\left(\frac{1}{2}\right)^n \cdot \frac{1}{3}\right]$

$P_1 = \frac{1}{2}$

$P_2 = \frac{3}{4}$.

Q. A Letter is known to come either from Tatanagar or from Kolkata. On the envelope, 2 consecutive letters are visible. The letters are 'TA'. Find the probability that the letters came from Tatanagar.

Soln. E_1 (From Tata...)

E_2 (From Kolkata)

E (TA $\rightarrow$ consecutive).

$P(E_1|E) = \frac{P(E) \cdot P(E|E_1)}{\sum P(E) \cdot P(E|E_i)}$

$= \frac{\frac{2}{8}}{\frac{2}{8} + \frac{1}{7}} = \frac{7}{11}$

Q. A & B are 2 weak students of Stat. Their chances of solving a problem correctly are $\frac{1}{8}$ & $1/8$ respectively. If their prob. of making a common error is $\frac{1}{505}$, & they obtain same ans. Find the prob. that their answer is correct.

Ans. 1/6.

E_1 (A & B both solve correctly)

E_2 (1 of them " ")

E_3 (Neither " ")

E (They got same ans)

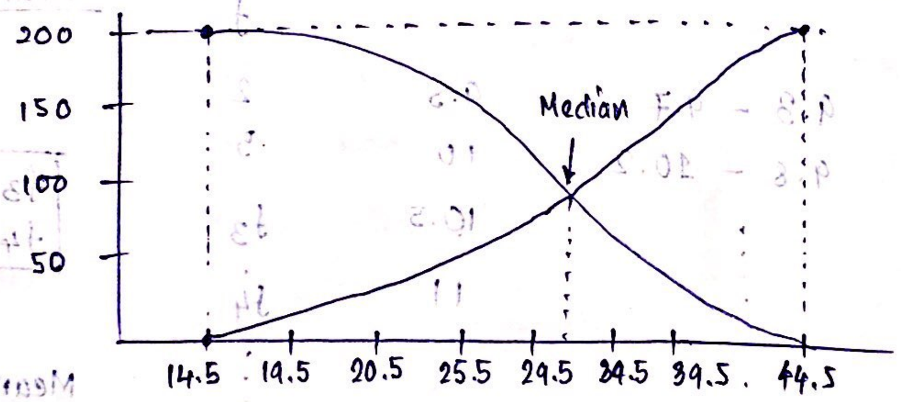
25/10/2019 **Statistics** Prof. BC Giri

①

- Histogram
  → Cumulative Frequency Distribution
- Ogive

	f		cumulative freq. (Less than type)	cdf greater than type
		14.5		200
15 - 19	37	19.5	37	
20 - 24	81	24.5	118	163
25 - 29	43	29.5	161	82
30 - 34	24	34.5	185	39
35 - 39	9	34.5	144	15
40 - 44	6	44.5	200	6
	200			0

Class Boundary



suppose $x_1, x_2, x_3, \ldots, x_n$ are n data
$\downarrow \quad \downarrow \quad \downarrow \qquad \downarrow$
$f_1 \quad f_2 \quad f_3 \qquad f_n$

$$\boxed{\text{Mean} = \frac{\sum x_i f_i}{\sum f_i}}$$

$A = \bar{x}$ ②

Group frequency distribution

$-f_1$

$-f_2$

$\vdots$

$-f_{n+1}$

$\dfrac{\sum x_i f_i}{\sum f_i}$

x

$y = \dfrac{x - A}{d}$

fy

$\dfrac{\sum fy}{y} = \bar{y}$

$y \times d = x - A$

$A = \bar{x}$

Missing frequencies

Mean = 11.09

Total = 60

		f
9.3 – 9.7	9.5	2
9.8 – 10.2	10	5
⋮	10.5	f_3
⋮	11	f_4
		⋮
		N = 60

$\boxed{f_3 = 12 \quad f_4 = 17}$

$$\text{Mean} = \frac{\sum x_i f_i}{f_i}$$

Median

$$\text{Median} = l_1 + \frac{\frac{N}{2} - F}{f_m} \times c$$

N → Total frequency

F → Cumulative frequency ~~of~~ the preceding median class

l_1 — lower limit of median class

f_m — frequency of median class

c — width.

Scanned with CamScanner

8.

15 - 25	4	4
25 - 35	11	15
35 - 45	19	34
45 - 55	14	48
55 - 65	0	48
65 - 75	2	50 = N.

Median, $= 35 + \dfrac{\frac{50}{2} - 15}{19} \times 10$

$= 40.26$

8. Missing freq. distribution. It is known that the total freq. is 1000 and median = 413.11. Estimate the frequencies.

300 - 325	5
325 - 350	14
350 - 375	80
375 - 400	p
400 - 425	325
425 - 450	?
450 - 475	88
475 - 500	9

$N = 1000$
$M = 413.11$

Mode $= l_1 + \left(\dfrac{d_1}{d_1 + d_2}\right) \times c$

diff b/w the

If d_1 is the largest frequency & the frequency of the class just preceding the modal class.
Similarly d_2 is the diff. b/w the largest frequency & the freq. of the class following the modal class.

Scanned with CamScanner

Measures of dispersion

$$\text{variance} = \sigma^2 = \left(\frac{\Sigma f_i x_i}{N}\right)^2 - \frac{\Sigma f_i x_i^2}{N}$$

$$\sigma^2 = \frac{1}{N}\Sigma f_i(x_i - \bar{x})^2$$

St. deviation = $\sqrt{\text{Variance}}$

$$\boxed{\text{Mean} - \text{Mode} = 3(\text{Mean} - \text{Median})}$$

Independent of change in origin but dependent of change in state

Q. In a group of 200 candidates, the mean & standard deviation of scores were found to be 40 & 15 resp. Later on it was discovered that the scores 43 & 35 were ~~mis re~~ misread as 34 & 53 resp. Find the correct mean & standard deviation for the data.

(Mean = 39.95, σ = 14.97)

Scanned with CamScanner

08/11/2019 Statistics & Probability Prof- B.C. Giri

1

Co-relation

Bi-variate data

$(x_1, y_1), \ldots, (x_n, y_n)$

$\rho = \frac{Cov(x,y)}{\sigma_x \sigma_y}$

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

$= \frac{1}{n}\sum(x_i - \bar{x})^2$

$= \frac{\sum x_i^2}{n} - \bar{x}^2$

$\sigma_y^2 = \frac{1}{n}\sum y_i^2 - \bar{y}^2$

$cov(x,y) = E\{(x-\bar{x})(y-\bar{y})\}$

$= \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$

$= \frac{1}{n}\left[\sum(x_i y_i - \bar{x} y_i - \bar{y} x_i + \bar{x}\bar{y})\right]$

$= \frac{1}{n}\left[\sum(x_i y_i)\right] - \frac{\bar{x}}{n}\sum y_i - \frac{\bar{y}}{n}\sum x_i + \frac{1}{n}\sum \bar{x}\bar{y}$

$= \frac{\sum(x_i y_i)}{n} - 2\bar{x}\bar{y} + \bar{x}\bar{y}$

$= \frac{\sum(x_i y_i)}{n} - \bar{x}\bar{y}$

** covariance is independent of the change of origin but depends on the change of scale.

$u_i = \frac{x_i - c}{d}$, $v_i = \frac{y_i - c'}{d'}$

$$\boxed{cov(x,y) = dd' \; Cov(u,v)}$$

S.D is always positive.

→ what about covariance ?

cov (x,y) may not always give a +ve result, may give negative value.

Scanned with CamScanner

ρ is independent of change in origin & scale.

$\boxed{-1 \le \rho \le 1}$

Proof

Hint let $u_i = \frac{x_i - \bar{x}}{\Sigma x}$

and $v_i = \frac{y_i - \bar{y}}{\Sigma y}$ then calculate

Σu_i^2, Σv_i^2 and $\Sigma u_i v_i$

then use the result

$(u_i + v_i)^2 \geq 0$

& $(u_i - v_i)^2 \geq 0$

8. Suppose that n pairs of values of 2 variables are $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$,

if the variances of $x, y, x-y$ are $\sigma_x^2, \sigma_y^2, \sigma_{(x-y)}^2$ then show that $\rho_{xy} = \frac{\sigma_x^2 \cdot \sigma_y^2 - \sigma_{(x-y)}^2}{2\sigma_x \cdot \sigma_y}$

~~Soln~~ ~~Hint: Find σ~~ Hint

Rewrite. ~~i) Find~~ $\sigma_{(x-y)}^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_x \cdot \sigma_y \cdot \rho_{xy}$

and put the ~~required~~ formula.

Let $u = x - y$
$u_i = x_i - y_i$

~~Σu_i^2~~ Then. $\boxed{\sigma_u^2 = \frac{1}{n}\Sigma(u_i - \bar{u})^2}$

~~Pipeline~~

~~to Completing~~

Calculate ρ for the following data :- setting c.

x =	65	66	67	67	68	69	70	72
y =	67	68	65	68	72	72	69	71

$$u = \frac{x-c}{d}$$

$$v = \frac{y-c'}{d'}$$

$\rho_{uv} = \rho_{xy}$.

$$\rho_{uv} = \frac{cov(u,v)}{\sigma_x \sigma_y} = \underline{0.603}$$ Ans.

Note

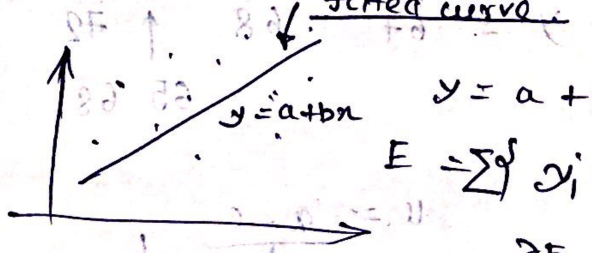
$cov(x,y) = 0$

x :	-3	-2	-1	0	1	2	3
y :	9	4	1	0	1	4	9

Even though $cov(x,y) = 0$ we cannot say all the time x & y are not related.

Regression

Predicting the value of one variable given value of another variable. ✓



$$y = a + bx$$

$$E = \sum \{ y_i - (a + b x_i) \}^2$$

$$\frac{\partial E}{\partial a} = 0$$

$$\frac{\partial E}{\partial b} = 0$$

We will get 2 regression lines.

They intersect at $(\bar{x}, \bar{y})$.

① Regression of y on x.

② Regression line of x on y.

$$\boxed{y - \bar{y} = \beta_{yx} (x - \bar{x})}$$

$$\boxed{x - \bar{x} = \beta_{xy} (y - \bar{y})}$$

Regression coefficients

$$\beta_{xy} = \rho_{xy} \cdot \frac{\sigma_x}{\sigma_y}$$

$$\beta_{yx} = \rho_{xy} \frac{\sigma_y}{\sigma_x}.$$

if θ be the angle b/w 2 regression lines then it can be shown that

$$\tan\theta = \frac{\rho^2 - 1}{\rho \cdot \left(\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x} \right)}$$

① if $\rho = \pm 1$.
→ 2 reg. line will coincide.

② if $\rho = 0$..
$\theta = 90°$. ($\perp$ to each other).

Scanned with CamScanner

Q. Find the eqn of line of regression of x on y

x:	1.0	1.5	2.0	2.5	3.0	3.5	4.0
y:	5.3	5.7	6.3	7.2	8.2	8.7	8.4

$$x-\bar{x} = b_{xy}(y-\bar{y}) = \rho_{xy}\cdot\frac{\sigma_x}{\sigma_y}(y-\bar{y}),$$

Binomial Distbution

$$P(X=r) = {}^nC_r \cdot p^r q^{n-r} \qquad r=0,1,2,\ldots,r.$$

Mean $= np$

Variame $= npq$.

Q. In a bombing attack there is a 50% chance that bomb will strike the target. 2 direct hits are required to destroy the target completely. How many bombs must be dropped to give a 99% or better of completely destroying the target. $(P \geq 0.99)$. Ans. 11.

15/11/2019. Statistics and Probability (Courtesy: Titir Adhikary). Prof. BC Giri

Chebyshev Inequality

Let X be a random value with mean μ and variance σ^2. Then for any $K \in Z^+$

$$P(|x - \mu| \geq K\sigma) \leq \frac{1}{K^2}$$

OR $$P(|x - \mu| < K\sigma) > 1 - \frac{1}{K^2}$$

Proof

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

$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \quad \text{where } f(x) = \text{pdf of } x$$

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

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

$$+ \int_{\mu+K\sigma}^{\infty} (x-\mu)^2 f(x)\,dx$$

Combining

$$|x-\mu| \geq K\sigma$$

$$\sigma^2 \geq K^2\sigma^2 \left[\int_{-\infty}^{\mu-K\sigma} f(x)\,dx + \int_{\mu+K\sigma}^{\infty} f(x)\,dx \right]$$

$$= K^2\sigma^2 \left[P(x \leq \mu - K\sigma) + P(x \geq \mu + K\sigma) \right]$$

$$= K^2\sigma^2 \left[P(|x-\mu| \geq K\sigma) \right]$$

$$\Rightarrow P(|x-\mu| \geq K\sigma) \leq \frac{1}{K^2}$$

Scanned with CamScanner

Applications of Chebychev's Inequality

A symmetric unbiased dice is thrown 600 times. Probability of finding the lower bound 80 ~ 120 sixes.

$$P(80 \leq \quad \leq 120) \geq 0.$$

Let s be the total no of sixes.

then $E(S) = 600 \times \frac{1}{6} = 100.$

$$\sigma(s) = 100 \times \frac{5}{6} \cancel{\times \frac{1}{6}} = \frac{500}{6}$$

$$\Rightarrow P\left(\mu - k\sigma \leq s \leq \mu + k\sigma\right) \geq 1 - \frac{1}{k^2}$$

$$\Rightarrow P\left(1 - k\sqrt{\frac{500}{6}} \leq s \leq 1 + k\sqrt{\frac{500}{6}}\right) \geq 1 - \frac{1}{k^2}$$

~~$k = 14$~~

Poisson Distribution

Poisson distribution can be approximated to binomial distribution under the following conditions:- No. of trials n is very large, Probability of success is too small that np is finite.

(Refer to Internet).

Normal Distribution

(Refer to Internet).