* Inequalities :-

① Bonferroni's inequality :-

$$\sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

② Boole's inequality :-

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq 1 - \sum_{i=1}^{n} P(\bar{A_i}).$$

③ Markov's inequality :-

$$P\{X \geq a\} \leq \frac{E(X)}{a}$$ $X \rightarrow$ non-negative R.V

④ Chebyshew's inequality :- This inequality is used when PDF is unknown.

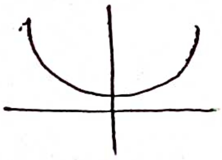
$$P\{|X-\mu| \geq K\} \leq \frac{\sigma^2}{K^2}$$

⑤ Cauchy-shwarz inequality :-

$$(E(XY))^2 \leq E(X^2).E(Y^2).$$

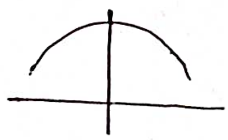
⑥ Jensen's inequality :-

- $f''(x) > 0 \rightarrow$ convex.



$\Rightarrow g(E(x)) \leq E(g(x))$

$f''(x) < 0 \rightarrow$ concave.



$\Rightarrow g(E(x)) \geq E(g(x))$.

⑦ Central limit theorem :-

$X_1, X_2, \ldots X_n \Rightarrow$ iid. (independent identically distributed R.V's).

$S_n = X_1 + X_2 + \ldots X_n$.

As $\underset{n\to\infty}{Lt} \dfrac{S_n - E(S_n)}{S.D(S_n)} \Rightarrow$ follow standard Normal dist. (Z).

⑧ Strong law of large no.s :-

$X_1, X_2, X_3, \ldots X_n \Rightarrow$ iid. R.V's.

$\Rightarrow$ As $\underset{n\to\infty}{Lt} \bar{X} = E(X)$. where $\bar{X}$ = Sample mean, $E(X)$ = population mean.

$\Rightarrow$ Its probability = 1 (Sure event).

⑨ Weak law of large no.s :-

As $\underset{n\to\infty}{Lt} P(\bar{X} - \mu > \epsilon) \Longrightarrow 0$. where $\bar{X}$ = Sample mean. μ = population mean.

* Discrete distributions :-

① Degenerate distribution :-

$$P\{X=k\} = \begin{cases} 1 \end{cases} \text{ and '0' otherwise.}$$

$E(X^l) = k^l \Rightarrow E(X) = k$

$V(X) = 0.$

② Uniform distribution :-

$X_1, X_2, \dots, X_n \Rightarrow$ 'n' R.V's

PMF, $P\{X = x_i\} = \frac{1}{n}$.

$E(X) = \frac{n+1}{2}$, $V(X) = \frac{n^2-1}{12}$.

③ Bernouli's distribution :-

One 1 trial.

p = success, q = failure.

$E(X) = p$, $V(X) = pq$.

④ Binomial distribution :-

'n' trials.

success $= p$, failure $= q$.

$E(X) = np$, $V(X) = npq$.

PMF, $\Rightarrow$ $p\{X=k\} = {}^nC_k \cdot q^{n-k} \cdot p^k$.

⑤ Poison distribution :-

PMF, $P\{X=x\} = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$

$E(X) = V(X) = \lambda$.

⑥ Geometric distribution :-

$(n-1)$ failures, '1' success.

PMF, $P\{X=n\} = (1-p)^{n-1} \cdot p$

$= q^{n-1} \cdot p$

$\Rightarrow E(X) = \frac{1}{p}$, $V(X) = \frac{q}{p^2}$.

* Continuous distributions :-

① Uniform distribution :-

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

$E(X) = \frac{b+a}{2}$, $V(X) = \frac{(b-a)^2}{12}$.

② Gamma distribution :-

PDF, $f(y) = \begin{cases} \frac{y^{\alpha-1} \cdot e^{-y/\beta}}{\Gamma(\alpha) \cdot \beta^\alpha} & ; 0 < y < \infty \\ 0 & ; \text{otherwise.} \end{cases}$

$\Rightarrow \Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$

$= (\alpha-1)!$

$E(X) = \alpha\beta$, $V(X) = \alpha\beta^2$

③ Exponential distribution :-

Special case of gamma dist. $\alpha = 1$.

PDF, $f(x) = \begin{cases} \frac{e^{-y/\beta}}{\beta} & ; 0 < x < \infty \\ 0 & ; \text{otherwise.} \end{cases}$

$E(X) = \beta$, $V(x) = \beta^2$.

④ Beta distribution :-

PDF, $f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & ; 0 < x < 1 \\ 0 & ; \text{otherwise.} \end{cases}$

$B(\alpha,\beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

$E(X) = \frac{\alpha}{\alpha+\beta}$, $V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

⑤ Normal distribution :-

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

~~Mean~~ $\Rightarrow$ Mean $= \mu$

Variance $= \sigma^2$

⑥ Standard Normal dist :-

PDF, $f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}$; $-\infty < Z < \infty$

Mean $= 0$, Variance $= 1$, $Z = \frac{X-\mu}{\sigma}$.

(*) Negative Binomial distribution :-

Here, we find probability of 'x' failure before r^{th} success.

PMF, $P\{X=x\} = {}^{x+r-1}C_x \cdot p^r \cdot q^x$

p = success

q = failure.

$\Rightarrow E(X) = \frac{qr}{p}$

$V(X) = \frac{qr}{p^2}$.

(*) Hyper geometric distribution :-

PMF,

$$P\{X=x\} = \frac{{}^{M}C_x \cdot {}^{N-M}C_{n-x}}{{}^{N}C_n}.$$

n = Sample, N = Total no. of things.

M = No. of things marked

$\Rightarrow E(X) = \frac{nM}{N}$, ~~$V(X) = nM$~~

$$V(X) = \frac{nM\,(N-M)(N-n)}{N^2(N-1)}.$$

(*) Sampling distributions :-

(1) Chi-square distribution :-

Special case of gamma dist.

$\alpha = \frac{n}{2}$, $\beta = 2$.

PDF,

$$f(x) = \begin{cases} \dfrac{e^{-x/2} \cdot x^{n/2-1}}{\Gamma(n/2) \cdot 2^{n/2}} & ; 0 < x < \infty \\ 0 & ; \text{otherwise}. \end{cases}$$

$X \sim \chi^2(n)$, n = degrees of freedom = No. of R.V's

$\Rightarrow E(X^k) = \frac{2^k}{\Gamma(\frac{n}{2})} \cdot \Gamma(\frac{n}{2}+k)$.

$E(X) = n$, $V(X) = 2n$

(2) T-distribution :-

$T = \frac{X}{\sqrt{\frac{Y}{n}}}$; $X \sim N(0,1)$, $Y \sim \chi^2(n)$.

PDF,

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \cdot \left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)} \quad ; -\infty < t < \infty.$$

$\Rightarrow E(T) = 0$, $V(T) = \frac{n}{n-2}$; $n > 2$.

$\Rightarrow$ Cauchy's distribution :-

$n = 1$ (degrees of freedom)

$$\boxed{f(t) = \frac{1}{\pi(1+t^2)}}$$

- Note :-

$$\boxed{t_{n,1-\alpha} = -t_{n,\alpha}}$$

(3) F-distribution :-

$F = \frac{X/m}{Y/n}$; $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$.

PDF,

$$g(f) = \begin{cases} \Gamma(\frac{m+n}{2}) \cdot (\frac{m}{n}) \cdot (\frac{m}{n} \cdot f)^{\frac{m}{2}-1} \cdot (1+\frac{m}{n}f)^{-\left(\frac{m+n}{2}\right)} & ; f \geq 0. \\ 0 & ; f < 0 \end{cases}$$

$\Rightarrow E(F) = \frac{n}{n-2}$ | $V(F) = \frac{(2n^2)(m+n-2)}{m(n-4)(n-2)^2}$.

Note :-

(1) $F_{m,n,1-\alpha} = \frac{1}{F_{n,m,\alpha}}$.

(2) $X \sim F(m,n) \Rightarrow \frac{1}{X} \sim F(n,m)$

(3) $F = T^2$, $m = 1$.

$F(1,n) = t^2(n)$.

Properties :-

① Normal distribution :-

→ Bell-shaped curve.

→ Symmetric about mean.

→ Curve is peak at mean.

→ Mean = median = mode. ($\because$ Symmetric Curve).

② Chi-square distribution :-

→ Suppose $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi^2(n)$.

→ It is sum of squares of Normal dist.

→ Curve is non-symmetric, skewed right.

→ Curve is different for each 'n'.

→ Always +ve.

→ For $n>30$, it follows normal dist.
For $n<30$, it is chi-square dist.

③ T-distribution :-

→ Symmetric about mean, $\mu=0$.
Mean = Median = Mode = 0.

→ As $n \to \infty$, function value (or) PDF $\to 0$.

→ As $n \to \infty$, it is close to normal dist.

→ As $t \to \infty$ (or) $t \to -\infty$,
tails of t-dist. slower than tails of Normal dist.

→ Probability of t-dist. > Probability of Z-dist.

④ F-distribution :-

→ Not symmetric curve.

→ Different curves for different pairs of m, n.

→ 'F' is always +ve.

→ As degrees of freedom ↑, it approaches normal dist.

→ Mean depends on 'n' variable.
Mean changes if 'n' changes.
It remains same if 'm' changes.

⑤ Covariance :- (cov(X,Y))

$\text{cov}(X,Y) = E(XY) - E(X)\cdot E(Y)$.

→ $\text{cov}(X,Y) > 0 \Rightarrow$ +vely related.
$\text{cov}(X,Y) < 0 \Rightarrow$ -vely related.
$\text{cov}(X,Y) = 0 \Rightarrow$ no relation.

→ $\text{cov}(X,Y)$ lies b/w $-\infty$ to ∞.

→ If $\text{cov}(X,Y)$ is high $\Rightarrow$ stronger the relation
If $\text{cov}(X,Y)$ is low $\Rightarrow$ weaker the relation.

→ $\text{cov}(ax+b, cx+d) = ac\,\text{cov}(X,Y)$.

→ $\text{cov}(X,X) = \text{Var}(X)$

→ If X, Y independent $\Rightarrow \text{cov}(X,Y) = 0$
Converse need not to be true.

⑥ Correlation coefficient :- (ρ_{XY}).

$$\rho_{XY} = \frac{\text{cov}(X,Y)}{S.D(X)\,.\,S.D(Y)}.$$

→ 'ρ' lies b/w -1 to 1.

→ If $\rho = 0 \Rightarrow \text{cov}(X,Y) = 0 \Rightarrow$ no relation.

→ If ρ = +ve → +vely related
ρ = -ve → -vely related.

→ $0 < |\rho| < 0.29 \Rightarrow$ weakly related.
$0.30 < |\rho| < 0.49 \Rightarrow$ moderately "
$0.5 < |\rho| < 1 \Rightarrow$ Strongly "

Mode = 3 Median − 2 Mean

Skewness :- (β_1) → lack of symmetry

It is 3rd standardized non-central moment.

$$\Rightarrow \beta_1 = E(Z^3) = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right)$$

$$\beta_1 = \frac{\mu_3}{\sigma^3}$$

$$\Rightarrow \beta_1 = \frac{3(\text{Mean}-\text{median})}{\sigma} = \frac{\text{Mean}-\text{mode}}{\sigma}$$

→ if $\beta_1 > 0$ → +vely skewed

if $\beta_1 < 0$ → −vely skewed.

if $\beta_1 = 0$ → symmetric.

Kurtosis :- (β_2) ⇒ Measure of peakness.

It is standardized 4th non-central moment.

$$\beta_2 = E(Z^4) = E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right).$$

$$= \frac{\mu_4}{(\mu_2)^2}.$$

⇒ if $\beta_2 = 3$ → mesokurtic (normal)

if $\beta_2 < 3$ → Platykurtic (less kurtosis)

if $\beta_2 > 3$ → Leptokurtic (high kurtosis).

→ Extral Kurtosis = $\underbrace{\beta_2}_{\text{actual Kurtosis}} - 3$.

Mode :-

Value which is more likely to occur.

→ In discrete-case, Mode = x (with highest probability).

→ In continuous-case, Mode = 'x' value for which $f'(x) = 0$.

Median :- Middle value.

→ In discrete-case, Median = x (at which $P(x) = \frac{1}{2}$).

→ In continuous-case, Median = x. $\left(\Rightarrow \int_{-\infty}^{x} f(x) = 1/2\right)$.

MGF :-

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

→ nth derivative of mgf at $t=0$ is μ_n'.

Characteristic (c.f) function :-

$$\phi_x(t) = E(e^{itx}).$$

→ nth derivative of $\phi_x(t)$ at $t=0$ is $(i)^n \cdot \mu_n'$.

→ Sometimes mgf may not exists but c.f always exists.

Random Process / Stochastic process :-

In Random Process, Random variable is function of time.

$$X = f(t).$$

⇒ We find required statistic like ~~mean~~, expectation, variance in terms of time. Then, we can easily find them at the given instant of time 't'.

Jointly distributed R.V's :-

→ Joint cdf,

$$F(x,y) = P(X \le x, Y \le y) \rightarrow \text{discrete}$$

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)\,dx\,dy \rightarrow \text{continuous}.$$

→ Marginal PMF of X :- $P(X=x)$

Collection (or) Sum of all Probabilities of Y w.r.t one X-value.

→ Marginal PMF of Y :- $P(Y=y)$

Collection (or) Sum of all Probabilities of 'X' w.r.t one Y-value.

→ **Conditional PMF :-**

$$P\{X=x_i/Y=y_i\} = \frac{P\{X=x_i, Y=y_i\}}{P\{Y=y_i\}}.$$

→ **Marginal PDF of 'X' :-**

$$f(X) = \int_{y\text{-int}} f(x,y)\cdot dy.$$

→ **Marginal PDF of 'Y' :-**

$$f(Y) = \int_{x\text{-int}} f(x,y)\cdot dx.$$

→ **Conditional PDF :-**

$$f(X/Y) = \frac{f(X,Y)}{f(Y)}.$$

• $$f(Y/X) = \frac{f(X,Y)}{f(X)}$$

→ **Joint independent R.V's :-**

Discrete, $P\{X=x, Y=y\} = P\{X=x\}.P\{Y=y\}$

Continuous, $f(X,Y) = f(X).f(Y)$.

Mutually independent :-

$f(X_1, X_2, \dots X_n) = f(X_1).f(X_2)\cdots f(X_n)$.

⇒ $f(X_i, X_j) = f(X_i).f(X_j)$, $\forall i,j$

→ **Moment for Joint R.V's :-**

Non central moments	Central moments
$\mu_{ij}' = E(X^i Y^j)$	~~$\mu_{ij} = E((X-\mu)^j$~~ $\boxed{\mu_{ij} = E((X-E(X))^i(Y-E(Y))^j)}$
⇒ $\mu_{10}' = E(X)$ $\mu_{01}' = E(Y)$	⇒ $\mu_{10} = 0$; $\mu_{01} = 0$ $\mu_{20} = V(X)$ $\mu_{02} = V(Y)$ $\mu_{11} = COV(X,Y)$ $= E(XY) - E(X).E(Y)$

⊛ **Bivariate R.V's :- / Multivariate R.V's**

X_1, X_2

Mean matrix, $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

Covariance matrix, $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$

⇒ $X_1 + X_2$ ⇒ follows normal dist.

⇒ $E(X_1+X_2) = E(X_1) + E(X_2)$

$V(X_1+X_2) = V(X_1) + V(X_2) + 2COV(X_1, X_2)$

⇒ Conditional dist.,

$$E(X/Y=y) = \mu_1 + \rho.\frac{\sigma_1}{\sigma_2}(y-\mu_2)$$

$$E(Y/X=x) = \mu_2 + \rho.\frac{\sigma_2}{\sigma_1}(x-\mu_1).$$

$$Var(X/Y=y) = (1-\rho^2)\sigma_1^2$$

$$Var(Y/X=x) = (1-\rho^2).\sigma_2^2$$

⇒ $Var(aX_1+bX_2+cX_3) = A\Sigma A^T$

where, A = Co-efficient matrix (or) Vector.
$= [a \ b \ c]$

⊛ **Transformations :- (Functions of R.V's)**

* **Theorem :-** (Function of Single R.V).

⇒ Suppose PDF of X is given i.e., $f(x)$.

⇒ $Y = g(X)$ function of X.

⇒ If $g'(x) > 0$ (or) $g'(x) < 0$ $\forall x$.

then PDF of 'Y',

⇒ $$f(y) = f(g^{-1}(y)).\left|\frac{d}{dy}.g^{-1}(y)\right|$$

and, Y-intervals can be find by substituting X-intervals in given function.

(*) Theorem :- (Function of multiple R.V's).

⇒ Suppose $X_1, X_2, \dots X_n$ ⇒ 'n' R.V's.

⇒ Given, $Y_1, Y_2 \cdots Y_n$ as a function of 'X' R.V's.

⇒ First find each 'X' R.V in terms of 'Y' R.V's. by inverse mapping. by denoting $f(h_1), f(h_2), \dots f(h_n)$.

⇒ Next, find Jocobian matrix. ⇒ $J = \dfrac{\partial (X \text{ R.V's})}{\partial (\text{each 'Y' R.V})}$

⇒ Finally, continuous PDF of Y is.

$$f(y_1; y_2, \dots, y_n) = |J| \cdot f(h_1, h_2, \dots h_n).$$

(*) Distributions of sum of R.V's :-

$Z = X + Y$; X & Y independent R.V's.

Discrete :- PMF,

$$P(Z = X+Y) = \sum P(X=K) \cdot P(Y = Z-K).$$

Continuous :-

PDF, $f_{X+Y}(Z) = \int_{y=-\infty}^{\infty} f_X(Z-y) \cdot f_Y(y) \cdot dy$

CDF (or) convolution of X, Y,

$$\Rightarrow F_{X+Y}(Z) = \int_{y=-\infty}^{\infty} F_X(Z-y) \cdot f_Y(y) \cdot dy.$$

(*) Sampling distributions :-

(*) Population :- collection of large no. of objects for investigation.

(*) Sample :- Subset of population.

(*) Parameter :- Statistical measurement related to population ⇒ μ, σ^2

(*) Statistic :- Statistical measurement related to sample ⇒ $\bar{X}, s^2$

(*) Why we do sampling ?

a) To estimate the unknown parameters of population.

b) To check validity of a statement.

(*) Distributions of $\bar{X}$ & s^2 :-

Sample mean, $\bar{X} = \dfrac{X_1 + X_2 + \dots X_n}{n}$

Sample variance, $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

→ $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$

⇒ $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim t(n-1).$

(*) Statistical Inference :-

(*) Estimation :-

Estimation is to choose a good estimator to estimate unknown population parameter.

(*) Types of estimation :-

(1) Point estimation :- Single value as estimate

(2) Interval estimation :- Range of values as estimate. (or) Confidence interval.

(*) Properties of estimator :-

(1) Unbiased estimator :-

$T(X)$ → function of R.V's.

⇒ $T(X)$ is unbiased estimator if $E(T(X)) = g(\theta)$.

⇒ $T(X)$ is biased estimator if

$$E(T(X)) = \underbrace{g(\theta)}_{\text{Population parameter}} \pm \underbrace{b(\theta)}_{\text{bias part}}$$

② **Consistency** :- (weak law of large no's)

Estimator is consistent if.

As $\lim_{n\to\infty} T(x) \longrightarrow g(\theta)$

$\Rightarrow \lim_{n\to\infty} P(|T(x)-g(\theta)| > \epsilon) = 0.$

③ **Mean squared error** :-

$$MSE = Var(T(x_i))$$

$$= E\Big(T(x_i) - E(T(x_i))\Big)^2$$

like $T(x) - g(\theta)$

∴ If variance is smaller, it is a better estimator.

(*) **Methods of estimation** :-

① **Method of moments** :-

$x_1, x_2, \ldots x_n \Rightarrow$ given population.

⇒ Consider non-central moment $\mu_k' = \frac{1}{n}\sum (x_i)^k$

⇒ write $\mu_1', \mu_2', \ldots \mu_n'$ as function of parameters.

⇒ No. of equations req. is the no. of parameters.

⇒ After that, express each parameter as a function of non-central moments.

⇒ Thus, we can obtain the estimator.

② **Method of likelyhood estimation** :-

⇒ $L(\theta, x) = \prod_{i=1}^{n} f(x_i, \theta)$ (product of all PDF's)

⇒ Apply 'log' on both sides.

⇒ Differentiate 'L' w.r.t 'parameter θ' to find MLE estimator.

⇒ Thus, we can obtain the MLE.

(*) **Addition rule** :-

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

(*) **Monotone** :-

$A \subseteq B \Rightarrow P(A) \le P(B)$

(*) **Subtractive** :-

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

(*) **Confidence-interval** :-

① **For 'μ'** :- (mean)

Z-test

$$\left(\bar{x} \pm Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right).$$

T-test

$$\left(\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{\sigma}{\sqrt{n}}\right).$$

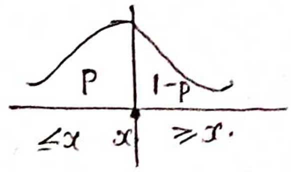
② **For 'σ^2'** :- (variance)

χ^2-test

$$\sigma^2 \in \left(\frac{s^2(n-1)}{\chi^2_{\alpha/2}}, \frac{s^2(n-1)}{\chi^2_{1-\frac{\alpha}{2}}}\right).$$

(*) **Quantiles** :-

→ These are used to divide given probability into segments.



→ p^{th} quantile,

$P\{X \le x\} \ge p \mid P\{X \ge x\} \ge 1-p.$

→ if $p = \frac{1}{2}$ then x = median.

→ Upper quartile, $p = \frac{3}{4}$; cdf $= \frac{3}{4}$

Lower quartile, $p = \frac{1}{4}$; cdf $= \frac{1}{4}$

(*) **Conditional Probability** :-

$$P(A \text{ given } B) \Rightarrow P(A/B) = \frac{P(A\cap B)}{P(B)}$$

(*) **Total Probability** :-

'A' can be happened through 'n' ways.

⇒ $P(A) = P(E_1) \cdot P\left(\frac{A}{E_1}\right) + P(E_2) \times P\left(\frac{A}{E_2}\right) + \ldots$

(*) **Baye's theorem** :-

$$P\left(\frac{A}{E_k}\right) = \frac{P(E_k) \cdot P(A/E_k)}{\sum_{i=1}^{n} P(E_i) \cdot P(A/E_i)}$$

A occured in 'n' diff. events.