variance:-

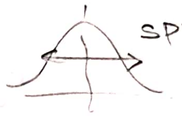
variance is approx. as- MIND THE "N-1".

$$\boxed{\text{variance} = s^2 = \frac{1}{N-1}\sum_{1}^{N}(\bar{x}-x_i)^2}$$

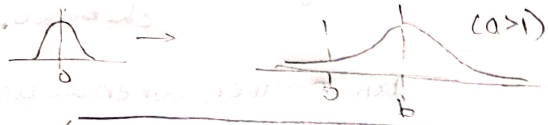
N-1 : very technical reason.

computed for Large N values.

$\sqrt{\text{variance}} = \sigma = S =$ standard deviation.


spread $\propto$ variance (or) std. deviation.

* $x_i \longrightarrow ax_i+b$



(a>1)

standard division $\times a$

practical application:-

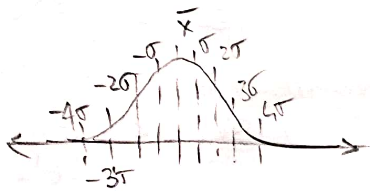
1) 100gm chips is told. make sure avg $\approx$ 100gm & σ is low.
2) definitⁿ of medical dieases:- "deficiency of RBC" "osteoporosis"- low bone density. ⋮ we have limits na!.

"Two sided-

→ Chebyshev's inequality:- suggests upper limit.

"The proportion of sample points; k or more than k (k>0) standard deviations away from the sample mean, is less than or equal to $\frac{1}{k^2}$" (proportion)

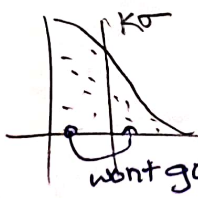
*Bcoz; if they are more than $\frac{N}{k^2}$; their squared sum contribution in σ^2 equⁿ would overshoot σ^2 itself.



Bounds are LOOSE, but NOT wrong.

$$\boxed{S_k = \{x_i : |x_i-\bar{x}| \geq k\sigma\} \quad \text{then } \frac{|S_k|}{N} \leq \frac{1}{k^2}}$$

* Gives lower bound on elements within $k\sigma$ & upper bound on those, out of $k\sigma$.



wont go anymore.

proof:

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

$$(n-1)\sigma^2 = \underbrace{\sum_{x_i\notin S_k}(x_i-\bar{x})^2}_{\text{near to mean}} + \underbrace{\sum_{x_i\in S_k}(x_i-\bar{x})^2}_{\text{far to mean}}$$

$$\therefore (n-1)\sigma^2 > \sum_{x_i\in S_k}\underbrace{(\bar{x}_i-x_i)^2}_{\geq k\sigma}$$

$$\therefore (n-1)\sigma^2 \geq k^2\sigma^2 |S_k|$$

$$\boxed{\frac{|S_k|}{N} \leq \frac{1}{k^2}} \Leftarrow \frac{1}{k^2} \cdot \frac{n-1}{nk^2} \geq \frac{|S_k|}{N}$$

101.

→ one sided chebysev's inequality: is stronger

chebyshev- cantelli inequality:-

"the proportion of sample points, k ore more than k s.d.s away from sample mean, & greater than the sample mean, is less than or equal to $\frac{1}{1+k^2}$"

$$\boxed{S_K = \{x : x_i - \bar{x} \geq K\sigma\} \qquad \frac{|S_K|}{N} \leq \frac{1}{1+K^2}}$$

Not an absolute value.

Stronger than 2-side chebysev.

but weaker, when we use this, to compute:

$S_K = \{x : |x_i - \bar{x}| \geq K\sigma\}$

$\frac{|S_K|}{N} \leq \frac{2}{1+K^2}$ weaker than $\frac{1}{K^2}$

Proof:-

$\{x_i\}_{i=1}^{N}$

let $y_i = x_i - \bar{x}$

$$\sum_{i=1}^{n} (y_i+b)^2 \geq \sum_{i: y_i \geq K\sigma} (y_i+b)^2 \quad b>0$$

$$\geq \sum_{y_i \geq K\sigma} (K\sigma+b)^2$$

$$\left(\sum_{i=1}^{n} y_i^2\right) + nb^2 \geq \quad = |S_K| \cdot (K\sigma+b)^2$$

$$(n-1)\cdot\sigma^2 + nb^2 \geq |S_K| \cdot (K\sigma+b)^2$$

$$\therefore |S_K| \leq \frac{(n-1)\sigma^2 + nb^2}{(K\sigma+b)^2}$$

$$\frac{|S_K|}{n} < \frac{\sigma^2+b^2}{(K\sigma+b)^2} \quad \text{for } b>0. \text{ any!..}$$

choose b; such that RHS minimised...

diff....

$$(K\sigma+b)^2(2b) - (K\sigma+b)(2)(\sigma^2+b^2)$$

$$(K\sigma+b)(2)[K\sigma b + \cancel{b^2} - \sigma^2 - \cancel{b^2}]$$

$$\therefore \boxed{b = \frac{\sigma}{K}}$$

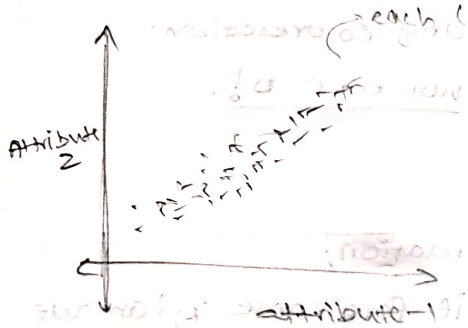
$$\therefore RHS = \frac{1}{1+K^2}$$

$$\therefore \boxed{\frac{|S_K|}{n} < \frac{1}{1+K^2}}$$

* co-relation b/w different attributes of same sample points:-

Eg: high fat intake → high heart diease.
high smokes → high cancer rates.

• use scatter plots:-



} we now, see a co-relation na! (able to suggest 'y' based on 'x' & viceversa).

co-relation coefficent:-

(x_i, y_i) be sample points.

σ_x, σ_y be standard deviations.

then.

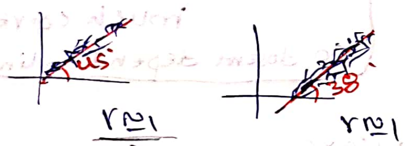
$$\boxed{r(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left(\sum(x_i-\bar{x})^2\right)\left(\sum(y_i-\bar{y})^2\right)}}} \in [-1,1]$$

cauchy schawrtz inequality

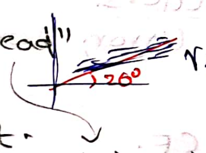
(or)

$$r(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1 \cdot \sigma_x \cdot \sigma_y}.$$

1) $r>0$: positive coreliation.

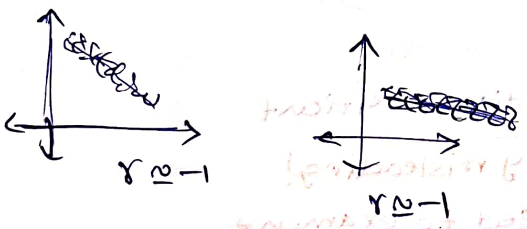


* depends on "spread" not angle, not intercept.



$(\vec{x_i}-\bar{x}\vec{u})\cdot(\vec{y_i}-\bar{y}\cdot\vec{u})$

2) $r<0$: negative correlation



see spread; for magnitude.

Not slope
Not intercept.

since
$x_i-\bar{x}$, $y_i-\bar{y}$ } intercept lost

$\frac{(\quad)}{(\quad)}$ } due to this division, slope lost.

∴ $\underbrace{x,\ ax+b}$
C·F = 1 if a>0
C·F = -1 if a<0

(PTO).

* take two vectors; in a n-dimension space.

$\vec{x} = x_1\cdot\hat{r}_1 + x_2\cdot\hat{r}_2 + \cdots x_n\cdot\hat{r}_n$

$\vec{x}-(\bar{x})\cdot\vec{u} = (x_1-\bar{x})\hat{r}_1 + \cdots (x_n-\bar{x})\hat{r}_n$

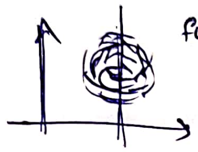
$||^{ly}$

$\vec{y}-(\bar{y})\cdot\vec{u} = (y_1-\bar{y})\hat{r}_1 + \cdots (y_n-\bar{y})\hat{r}_n$

∴ correlation factor = {angle's cosine} b/w the two "data" vectors.

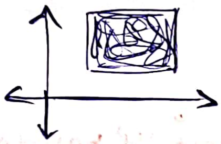
co-releation:- How strongly; I can suggest a unique 'y' for a unique 'x'.

r is defined & 0:-



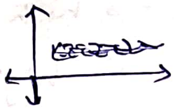
for an x; can't at all suggest a 'y'.

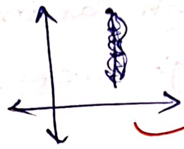
Numerat = 0
deno. ≠ 0.



⇒ we can say for sure; both are 0-correlated.
i.e. no relation b/w both.

r is undefined:- for x, i can say y, but for y, i can't say x.

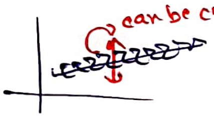




x = no. of ears & y = salary earned

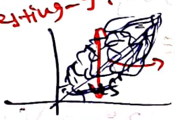
numerator = 0
denominator = 0

we can't say any co-relation. not even 0!.

* 

can be certain in suggesting y!

C.F ≃ 1



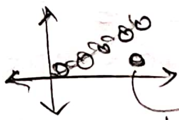
little uncertainity in suggesting y.

C.F > 0; ≠ 1
as spread more.

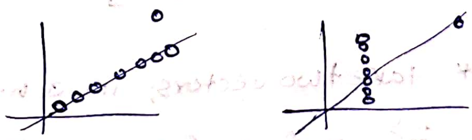
co-relation;
if for some x, i am able to "suggest" y; then I can say co-releation.
How strong I can say, that much co-releation.
So; dosent depend on line's slope.

* correlation coefficient — sensitive to outliers. (error, don't fit on graph).

- Ascombe's quartet.
 - graphically very different; but same CF.



r = 0.33
because of 1.



have same co-creleation coefficient
very misleading!
need to examine graph.

* $r_{uncentered}(x,y) = \dfrac{\Sigma x_i y_i}{\sqrt{(\Sigma x_i^2)(\Sigma y_i^2)}}$ — used SOMEwhere...

* co-oreleation dosent mean causation..

Discrete Probability:- ($11, 12^{th}$ topic).

* B is proper subset of A:

$B \subset A$ $\qquad B \neq A$ & $B \subseteq A$

$\therefore$ B can be NULL.

* Boole's inequality:-

$\boxed{P(AB) = P(A \cap B)}$

$$P(A \cup B) \leq P(A) + P(B)$$

$\therefore$

$$P(A_1 \cup A_2 \cup \cdots A_n) \leq \sum_{i=1}^{n} P(A_i)$$

** Bonferroni's inequality:-

$$P(A \cap B) \geq P(A) + P(B) - 1$$

(loose lower bound; $\because$ no other info, except $P(A)$, $P(B)$)

$\therefore$

$$P(A_1 \cap A_2 \cap \cdots A_n) \geq 1 - n + \sum_{i=1}^{n} P(A_i)$$

* conditional probability:-

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

$P(A)$; given B happened

notation

* joint probability:-

joint $P()$ of A, B $= \underline{P(AB) = P(A \cap B)}$ (Nothing!)

$= P(A,B)$

* Independent events:-

$$P(A|B) = P(A)$$
$$\Updownarrow$$
$$P(AB) = P(A) \cdot P(B)$$
$$\Updownarrow$$
$$P(B|A) = P(B)$$

implies they are dependent!

mutually exclusive events are not independent!

$P(A \cap B) = 0$
but neither $P(A) = 0$ or $P(B) = 0$
$\therefore P(AB) \neq P(A) \cdot P(B)$

for n-events:-

for any 'K'; $\boxed{K \leq n}$ events;

$$\boxed{P(A_1 A_2 \cdots A_K) = \prod_{i=1}^{K} P(A_i)}, K \leq n \quad \forall K$$

* only n-way independence doesn't imply events are independent.

* also A, B^c
A', B'
A', B

$\Big\}$ are independent.

Bayes's theorem:-

A,B happen

$$\boxed{P(B/A) = \frac{P(BA)}{P(A)} = \frac{P(A|B)\cdot P(B)}{\underbrace{P(A|B)\cdot P(B)}_{AB \text{ happen}} + \underbrace{P(A|B^c)\cdot P(B^c)}_{AB^c \text{ happen}}}}$$

Both

A happens.

B happens; given A happens.

$$\therefore \boxed{P(A) = P(A|B)\cdot P(B) + P(A|B^c)\cdot P(B^c)}$$

infact; in place of B, B^c; you can have ANY number of

- pairwise mutually exclusive
- exhaustive

events.

* Probability of B, given A $= \dfrac{\text{no. of A\&B events}}{\text{no. of A,B} + \text{no. of A,}B^c}$

$\boxed{\text{Simple!}}$

B^c could be

Event C

$\cup$

Event D

$\cup$

$\vdots$

So, more terms in denominator.

Random variable:-

- overview:-
  - discrete & continuous random variables
  - probability mass function (normal wala, for discrete)
    probability density function (PdF)
    cummilative distribution function (cdf)
  - joint & conditional pdf.
  - expectation operator. (a linear operator $E[ax+b] = a\cdot E[x] + E(b)$)
  - variance & co-variance
  - Markov's & chebyshev's inequality.
  - weak law of large numbers
  - Moment generating functns.

* $X, x \to$ R.V's value
↳ R.V. name

$P(X=x)$ is a probability mass function.

- continuous random variables:-

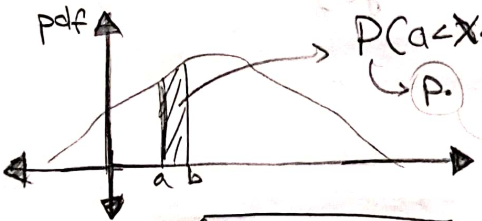
* $P(X=x)=0$ for any x.
(as, ∞'ly many x are there).

unlike discrete;
$P(X=x)=0$ dosent imply
$X=x$ dosen't occur.
It may occur!

define $P(X\le x)$ } exists!
(cdf.)

$$\boxed{F_X(x) = P(X\le x)}$$

now; $f_X(x) = \frac{d}{dx}(F_X(x))$ how much probability in interval 'dx'.
probability density function.
(pdf)



$P(a<X\le b) = \int_a^b f_X(x)dx$
↳ P. ; P. density
interval probability

$$\therefore \boxed{\int_{-\infty}^{\infty} f_X(x)dx = 1}$$

$$\boxed{F_X(x) = 1 \text{ as } x\to\infty}$$

weighted mean.

expectation value; not same as most probable value!

- (english misused).
- $E(X) \to$ might not exist; $P_X(x) = 1/x^2 \cdot K$
- $E(x) \to$ might not be a valid sample point.
  (dice throw; $E(X) = 3\cdot 5$)

let X be a random var.

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

let $g(X)$ be another random varial
inc. funct^n ; y
then

okay; if discrete, but cont.....

LOTUS: $\boxed{E(y) = \int_{-\infty}^{\infty} y\cdot f_X(x)dx}$

law of the unconscious statistisim.

> should hav been:

$$E(y) = \int_{-\infty}^{\infty} y\cdot f_Y(y)dy$$

Proof: $y=g(x)$ $F()$ ↳ cdf

$$\boxed{F_Y(y) = F_X(g^{-1}(y))}$$

common sense.

$\frac{d}{dx}$ on bothsides.

$$f_Y(y)\cdot\frac{dy}{dx} = f_X(g^{-1}(y))\cdot\frac{d}{dx}(g^{-1}(y))$$
x ; 1

$$\therefore \boxed{f_Y(y)dy = f_X(x)dx}$$

→ Expected value:

$E[\ ]$ → expectation.

- also called mean value of random variable.

$E(X) = \sum_{i=1}^{n} P(X=x_i)\cdot x_i$ for discrete.

$$\boxed{E(X) = \int_{-\infty}^{\infty} x\cdot(f_X(x)dx)}$$

NOTE: $E(X) \neq \text{mode}(X)$ } damn geezers were weak with english.
$E(X) = \text{mean}(X)$

- $E(X)$ might not even be a valid value of X.
  - might not even exist!.
    (as $\int_{-\infty}^{\infty} x\cdot f_X(x)dx \to \infty$).

* say; i have a new random variable, y
  $y = g(X)$.

(equal; if $g(x)$ is linear in x).

then; $E(y)$ need not equal $g(E(X))$

$$\boxed{E(y) = \int_{-\infty}^{\infty} g(x)\cdot f_X(x)\cdot dx}$$

LOTUS

might seem obvious
but; obvious - only for discrete x;

for continuous X;

$$E(y) = \int_{-\infty}^{\infty} y\cdot f_Y(y)\cdot dy$$

we can show both same;
(proof on before page.)

* $E(aX+b) = a\cdot E(X) + \underbrace{b}_{\text{equal to } E(b)}$

Linearity

an "OPERATOR"

* $E(g_1(X)+g_2(X)+\dots) = E(g_1(X))$
  $+ E(g_2(X))$
  $+ \ \vdots$

→ Markov's inequality:-

let $x>0$ be the possible values of X.

then

$$\boxed{P(X \geq a) \leq \frac{E(X)}{a}}$$

where $a>0$

very very loose.

proof:

$E(X) = \int_0^a x f_X(x)dx + \int_a^{\infty} x f_X(x)dx$

$\therefore E(X) \geq \int_a^{\infty} x f_X(x)dx$

$\therefore E(X) \geq \int_a^{\infty} a\cdot f(x)dx$

$$\therefore \boxed{P(X \geq a) \leq \frac{E[X]}{a}}$$

weak; usually RHS >> LHS.

* mea
(exp

med

F_X (

• varian

* als
var

* var.

* Va

INEQ
* if

P

* mean (expected value) $\longrightarrow$ minimizes the expectation of "Squared error" to 0.

~~minimizes~~ $\int_{-\infty}^{\infty} (x-c)^2 f_X(x)\,dx$

median $\rightarrow$ minimizes $\int_{-\infty}^{\infty} |x - x_0| \cdot f_X(x)\,dx$

$\boxed{F_X(\text{median}) = 0.5}$

cdf.

↳ a parameter; later called as median.

(do by differentiation)....

• variance;

$$\boxed{\sigma^2 = \text{var} = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f_X(x)\,dx.} \rightarrow \text{might not exist; as integral may go } \infty.$$

σ = positive sqrt(var)

* also

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

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

$$= E[x^2] - \mu^2$$

$$\star\ \text{var.} = E[x^2] - (E(x))^2$$

$$\star\ \text{var}[ax+b] = a^2 \cdot \text{var}(x)$$

$f_X(x) = \dfrac{k}{x\sqrt{x}}$ — supresses mean (x^1) but not variance (x^2)

cases;

i) mean × variance × } possible

(ii) mean ✓ variance × } possible

(iii) mean × variance ✓ } × ∵ mean is part of variance.

INEQUALITIES:

* if $x > 0$ then;

$$\boxed{P(X \geq a) \leq \frac{E[X]}{a}} \rightarrow \text{MARKOV's inequality - WEAK}$$

(E[X] → mean.)

often: RHS >>> LHS

Proof:

$$E(x) = \int x f_X(x)\,dx$$

$$= \int_0^a x \cdot f(x)\,dx + \int_a^{\infty} x \cdot f_X(x)\,dx$$

$$> \int_a^{\infty} x f_X(x)\,dx > \int_a^{\infty} a \cdot f_X(x)\,dx$$

$$\therefore \int_a^{\infty} f_X(x)\,dx \leq \frac{E(x)}{a}$$

→ chebyshev's inequality:-

$$\boxed{P\{|X-\mu| \ge k\} \le \frac{\sigma^2}{k^2}}$$

proof: take $Y = (X-\mu)^2$ ∴ $Y \ge 0$.

markov's on Y. with $a = k^2$

$\Rightarrow P\{(X-\mu)^2 \ge k^2\} \le \frac{\sigma^2}{k^2}$] def^n of variance; $\sigma^2 = E((X-\mu)^2)$

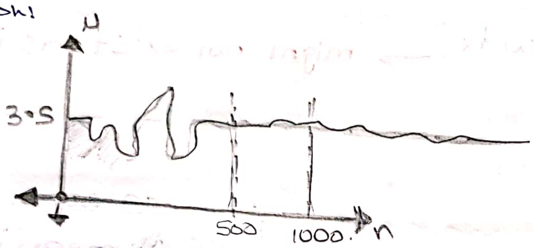
& hence;

$$\boxed{P\{|X-\mu| \ge k\sigma\} \le \frac{1}{k^2}}$$

→ weak law of large numbers:- Good proof, dammit!.

case: I rooll a die n times. what is the average value of die output?

wikipedia graph!



So; why is average 3.5 as n increases?

Ans: you know... 1 to 6 are equi-probable; so ... = $\frac{1}{6}(1+2+3+4+5+6) = 3.5$ da.

childish argument!

→ Manly arg:-

1) take 'n' dice.

2) X_i = random variable showing output on i^{th} die. → mean = μ = same for all.

so; associate each outcome with a random var.

3) so; now; we need:-

$E\left(\cancel{\frac{x_1+x_2+\cdots x_n}{n}}\right)$. } multiple random variables....

nope!

> we need $P\left(\left\{\frac{X_1+X_2+\cdots X_n}{N}\right\} = \mu\right) = 1$ (not expected value; we need to show; that mean ≡ it is unique value).

Strong law of large nos...

4) weak law of large numbers:-

let $X_1, X_2, \ldots X_n$ be a seq. of independent & identical distributed random variables; with mean μ (same for all); then for any $\varepsilon > 0$;

$$P\left\{\left|\frac{X_1+X_2+\cdots X_n}{n} - \mu\right| > \varepsilon\right\} \to 0 \text{ as } n \to \infty$$

Proof:- by chebyshevs; $\left(\because E\left(\frac{X_1+X_2+\cdots X_n}{n}\right) = \mu\right)$

$$P\{|\qquad| > \varepsilon\} \le \frac{var\left(\frac{X_1+X_2+\cdots X_n}{n}\right)}{\varepsilon^2}$$

$$\text{var}\left(\frac{x_1}{n}+\frac{x_2}{n}+\frac{x_3}{n}\cdots\frac{x_n}{n}\right) = \sum_{i=1}^{n} \text{var}\left(\frac{x_i}{n}\right) + \underbrace{\sum_{i=1}^{n}\sum_{j=i+1}^{n} \text{covar}\left(\frac{x_i}{n},\frac{x_j}{n}\right)}_{=0}$$

$$= \frac{1}{n^2}\cdot \underbrace{n\cdot\sigma^2}$$

$\because$ identical distribution.

$=0$ for independent events.

$\downarrow$ can have n^2 term, if not!

$\therefore P\{|\;|>\epsilon\} \le \frac{\sigma^2}{n\epsilon^2}$

$\therefore$ as $n \to \infty$;

$P\{||>\epsilon\} \to 0$ ✓

comments on weak law:

(i) assumption of identical σ is not necessary. but, should be finite.

(ii) Independance is NOT needed;

Just hav to be pair-wise uncorelated.

(i) covar( ) = 0

or

(ii) coeff. of corelation = 0.

5) Strong law of large numbers:-

$$P\left(\lim_{n\to\infty} \frac{x_1+x_2+\cdots x_n}{n} = \mu\right) = 1$$

(weak law; tends to 1) what shit kadha!

Note:

So; here; finally when we compute μ of 500 dice throws;

$\underbrace{1,1,1,1,1,1\cdots}_{\mu=1}$ $\underbrace{1,6,2,5,3,4,1,6,2,5,3,4,\cdots}_{\mu=3.5}$

$\underbrace{2,2,2,2,\cdots}_{\mu=2.}$ Both are possible values for mean; even after 500 tries

probability $=\left(\frac{1}{6}\right)^n$ ≈ 0 for large 'n'.

But! $P(\mu=3.5)$ is tending to 1.

Hence; we write a new random variable; using $x_1, x_2, x_3, \cdots$ (i.e. their mean)

& find $P(\text{new\_rand\_var} = \mu)$.

might not be high; for small n.

but $\to 1$; for $n\to\infty$.

$\because$ weak law of large numbers.

- (Incorrect) law of averages. (Gambler's fallacy)

coin tossed 20 times → 20 heads.

States. → ~~so; next time; higher probability for tails.~~

B.S.

next time also, same probability.

∵ 20 Heads + 1 Head   20 heads + 1 tail

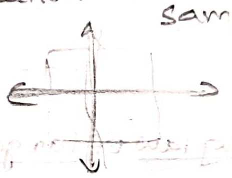
$\left(\frac{1}{2}\right)^{21}$ $\left(\frac{1}{2}\right)^{21}$

✓✓

Joint Pdfs:- Joint Cdfs:-

* For continuous r.v. X, Y.

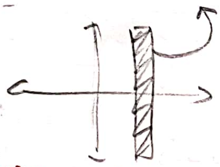
> $\boxed{F_{XY}(x,y) = P\{X \le x, Y \le y\}}$ - 2D plot of distribution of elements in sample space



also

> $F_X(x) = P\{X \le x, -\infty \le Y \le \infty\}$

same $F_Y(y)$



$= F_{XY}(x, \infty)$

"These" definitions are extended; to approach more variables.

→ joint pdf:-

for 1 variable; density over line;

for 2 variable; density over plane.

for probability in a region C:

$$P((x,y) \in C) = \iint_{(x,y)\in C} f_{XY}(x,y)\,dx\,dy. \checkmark$$

For discrete case:-

$$P(x_i) = \sum_{j=1}^{\infty} P(x_i, y_j)$$

now;

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

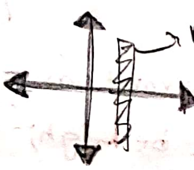
$$\therefore \boxed{f_{XY}(x,y) = \frac{d^2}{dx\,dy} F_{XY}(x,y)}$$

cool!

→ marginal pdf:-

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy.$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx$$



Independent Random Variables:-

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

"Joint pdf = product of margin pdfs"

& consequently,

$$F_{XY}(x,y) = F_X(x) \cdot F_Y(y)$$

capital means cummilative.

if n such random variables are indpendt then; any K (K ≤ n) such r.v.s should be "independent" i.e. joint pdf = product of marginl pdf.

- pairwise independent is necessary; not sufficient.

COVARIANCE:-

product

$$\boxed{\mathrm{cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y))}$$

dont write simply $E(X \cdot Y)$!

hence $\mathrm{cov}(X,X) = \mathrm{Var}(X) = E((X-\mu)^2)$

* $\mathrm{cov}(X,Y) = E(XY - \mu_X \cdot Y - \mu_Y \cdot X + \mu_X \mu_Y)$

$$\boxed{\mathrm{cov}(X,Y) = E(XY) - E(X) \cdot E(Y)}$$

properties:-

1) $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$
2) $\mathrm{cov}(aX,bY) = a \cdot b \cdot \mathrm{cov}(X,Y)$
3) coefficent of corelation:-

$$\boxed{r(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X) * \mathrm{Var}(Y)}}}$$

connect with stuff; in discrete statistics..

4) $\mathrm{cov}(X+Z,Y) = \mathrm{cov}(X,Y) + \mathrm{cov}(Z,Y)$

prove from,

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

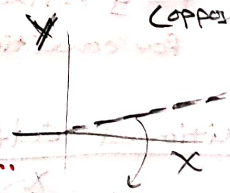
5) $\mathrm{cov}(\sum_i X_i, \sum_j Y_j) = \sum_i \sum_j \mathrm{cov}(X_i, Y_j)$ vokay!

6)

$$\boxed{\begin{array}{l} \mathrm{var}(X_1+X_2+\cdots X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \\ \quad\quad \| \\ \mathrm{covar} \quad\quad\quad\quad = \quad + \sum\sum_{i \neq j} \mathrm{covar}(X_i, X_j) \\ (X_1+X_2+\cdots, X_1+X_2+\cdots) \end{array}}$$

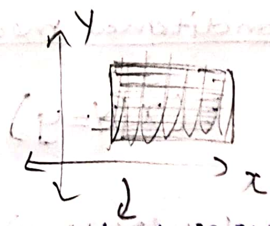
co-variance ≡ co-relatn.

co-variance × independent (opposite)



$r(X,Y) = 1$

∴ covariance is max. i.e. fully dependent (X,Y) are....



graph; if X,Y are independent corelation = 0.

cov = 0

* for independent events;

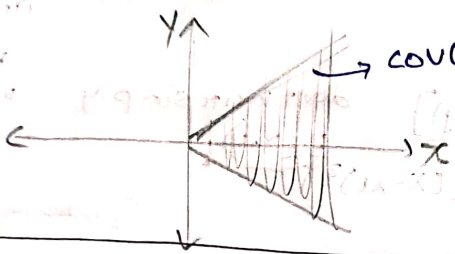
$E(XY) = E(X)\cdot E(Y)$

$\therefore$ $\boxed{cov(X,Y) = 0, \text{ for independent events}}$

'co'- variance (varying together) is 0.

but;

$\boxed{\text{if } cov(X,Y)=0 \text{ then independent is wrong.}}$

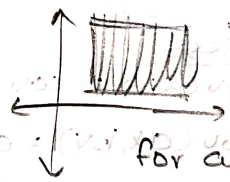


$cov(X,Y)=0$ ($\because$ for every $(X-\mu_X)$; we got $+y_0, -y_0$)

& X, Y are not independent

$\because$ $y \in [-x, x]$

(if X, Y are independent, then graph gotta be sq.)



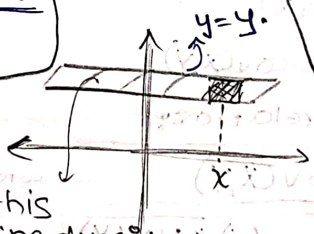
for any X; any Y is ok. & viceversa.

⟹ conditional pdf, cdf:-

joint pdf is $f_{XY}(x,y)$

$$\boxed{\underbrace{f_{X|Y}(x,y)}_{\text{a function in } x.} = \frac{f_{XY}(x,y)}{f_Y(y)}}$$

Y = y is fixed. for conditional.



this line density is $f_{X|Y}(x,y)$. proof vaguely

$\therefore$ $\boxed{\underbrace{f_{X|Y}(x,y)\cdot dx}_{\text{sample space is } (-\infty,\infty),\ (y,y+\delta y)\text{ needed is } (x,x+\delta x)(y,y+\delta y)} = \frac{f_{XY}(x,y)\cdot dx\cdot dy}{f_Y(y)dy}}$

points in $(x, x+\delta x)$ & $(y, y+\delta y)$

points in $(-\infty,\infty)$ & $(y, y+\delta y)$

• conditional cdf:

$$F_{X|Y}(x,y) = \int_{-\infty}^{x} f_{X|Y}(x,y)dx$$

x, not same r.a.

$$= \int_{-\infty}^{x} \frac{f_{XY}(z,y)dz}{f_Y(y)} \quad ✓$$

• conditional mean & variance:-

$$E(X|Y=y) = \int_{-\infty}^{\infty} x\cdot f_{X|Y}(x,y)dx$$

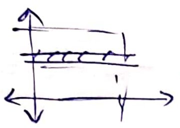
$$var(X|Y=y) = \int_{-\infty}^{\infty} (x - \underbrace{E(X|Y=y)}_{\text{take care!}})^2 \cdot f_{X|Y}(x,y)dx$$

not regular μ_X.
lol.....

Eg: $f_{XY}(x,y) = 2.4\,x(2-x-y)$ $\quad 0<x<1$, $0<y<1$

find conditional density of X given $Y=y$.

Sol)



$$f_{X|Y}(x,y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_{XY}(x,y)}{\int_0^1 f_{XY}(x,y)\,dx} = \frac{2.4\,x(2-x-y)}{(2.4)\left(1-\frac{1}{3}-\frac{y}{2}\right)}$$

y = parametric constant.

$$\therefore \boxed{f_{X|Y}(x,y) = \frac{6x(2-x-y)}{4-3y}}$$

MOMENT GENERATING FUNCTIONS:-

* **Moment** of random var. X of order n; is $E(X^n)$.

$\therefore$ $m_1 = E(X) =$ mean.

$m_2 = E(X^2)$ → 2nd order moment.

$$\boxed{m_i = E(X^i)} \longrightarrow i^{th} \text{ order moment.}$$

$E(X)$ → mass moment.

moment of inertia. → $(E(X^2))$

mass $\equiv$ probability mass.

* Moment generating function: (MGF)

function; to generate various order-moment of X.

$$\boxed{\phi_X(t) = E(e^{tX})}$$; t is a function parameter.

$$= E\left(1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \cdots\right)$$

$$\phi_X'(t) = E\left(\frac{X}{1!} + 2\cdot t\cdot\frac{X^2}{2!} + \cdots\right)$$

$\phi_X'(0) = E(X) =$ 1st order moment.

$\phi_X''(0) = E(X^2) =$ 2nd order moment.

$\vdots$

$$\boxed{\phi_X^{k}(0) = E(X^k)} = k^{th} \text{ order moment.}$$

↓

moment 'generating' funct^n $\phi_X(t)$.

The MGF is a compact way of encapsulating all ∞ moments of X.

Properties of MGF:-

1) if X,Y are independent,

$$\boxed{\phi_{X+Y}(t) = \phi_X(t)\cdot\phi_Y(t)}$$

$\iint e^{t(X+Y)}$. dec . . . . . fun fey

$= E[A\cdot B] = E(A)\cdot E[B]$

2) $$\boxed{\phi_{aX+b}(t) = e^{tb}\cdot\phi_X(at)}$$

3) let Z be X with probability p
& Y with " (1-p)

$\phi_Z(t) = p\cdot\phi_X(t) + (1-p)\phi_Y(t)$

→ Uniqueness:- (crazy!) $\phi_X(t)$ generates '∞' moments, which uniquely determi

* Pdf completely determines a random variable X.

Claim:

> A MGF $\phi(t)$ can uniquely determine a Pmf / Pdf & thus uniquely define a random variable X.

i) case of discrete:

$MGF = E(e^{tX}) = \Sigma P(x_i)\cdot e^{tx_i}$

e^{kx} all, are independent vectors kadha.

say; PMF of two random variables X, Y exist

& say $\phi_X(t) = \phi_Y(t)$

compare coeff. of $e^{t\cdot(c)}$ (this is bcoz; every e^{tx} is a base vector...)

$P(X=c) = P(Y=c)$ for any c

like if $a\cdot e^{k_1 x} = b\cdot e^{k_2 x}$ $\forall x$!
then $a=b$, $k_1=k_2$

∴ since $PMF_{X,Y}$ are equal;

$X \triangleq Y$

(ii) This uniqueness between MGF & pdf can be shown for continuous X too!

(formidable proof)

Proof : vague.

X Y

$f_X(c)$ $f_Y(c)$.

$\phi_X(t) = \phi_Y(t)$.

$$\underbrace{\int_{-\infty}^{\infty} e^{tx} f_X(x)dx}_{\text{like, do fourier transform for } MGF_X(t)} = \underbrace{\int_{-\infty}^{\infty} e^{ty} \cdot f_Y(y)dy}_{\text{do fourier transform for } \phi_Y(t)} \ldots$$

$$\sin(t) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)dx$$

find $f(x)$ from this fourier transform?...

y determining

X ✓

X-
random
variable.

...or...)

k_2x
$\forall x$
!

Families of Random variables.

discrete distribⁿ → | → continuous distribution.

Discrete distbⁿ:

1) **Bernoulie:-** X is head; for a coin toss. (X is 1 if head)

PMF: $P(X=1) = p$; $P(X=0) = 1-p$ — p is parameter. so - family. — thats it.

* $E(X) = p$
* $Var(X) = p(1-p)$ — mode ✓ median $[0,1]$; $\boxed{MGF = (1-p+pe^{t \cdot 1})}$

2) **binomial:**

n-coins; with 'p' for heads, for each coin.

PMF: $P(X=i)$ → i heads exact, among n coins.

$$= {}^{n}C_{i} \cdot p^{i} (1-p)^{n-i}$$

$\boxed{n, p}$ parameters.

E(X): let

X_i = bernoulie value for ith coin.

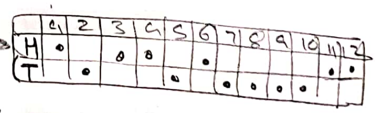
note that:

$$X = \Sigma X_i$$

$$E(X) = E(\Sigma X_i)$$

$$= \sum_{i=1}^{n} p$$

$$E(X) = np$$

$X \rightarrow$ 

$X = 6$ ✓

Var:

$$var(X) = var(\Sigma X_i) = \Sigma Var(X_i) = np(1-p)$$

∵ independence among X_is.

MGF: $= (1-p+p \cdot e^{t})^{n}$

$$\left(\because \phi_X(t) = \phi_{\Sigma X_i} t = \prod_{i=1}^{n} \phi_{X_i} t = (1-p+pe^{t})^{n} \right)$$

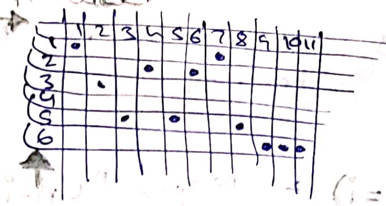
! only if independent.

3) Multinomial:-

instead of n-coins; we have n-dices.

- Here random variable X: is vector.

k^{th} dice.



X_i values.

$$P(X=[x_1,x_2,\dots x_n]) = \frac{n!}{x_1!\cdots x_n!}\, P_1^{x_1} \cdot P_2^{x_2} \cdots P_n^{x_n}$$

$(\because \Sigma x_i = n)$; $(\because \Sigma P_i = 1)$.

$X_i \rightarrow$ ith possible value of a outcome.

$X_{ij} \rightarrow$ the bernoullie variable; with $P_X(X_i) = P_i$ in, the jth dicethrow.

$\underline{E(X)}$: $= [E(X_1), E(X_2), \dots E(X_n)]$

$= [E(\Sigma X_{1j}), E(\Sigma X_{2j}), \dots]$

$= [n \cdot P_1, n \cdot P_2, \dots n \cdot p_k]$.

variance:

$$V(X_i) = Var\left(\sum_{j=1}^{n} X_{ij}\right)$$

$$= n \cdot P_i(1-P_i)$$

* for X; we write covariance matrix

$$= \begin{bmatrix} C(X_1,X_1) & C(X_1,X_2) & \cdots \\ C(X_2,X_1) & - & - \\ \vdots & \vdots & \end{bmatrix}$$

note that $C(X_i,X_i) = Var(X_i) = n \cdot P_i(1-P_i)$

$C(X_i,X_j) = -nP_iP_j$] proved in video.

okay proof ...

$\therefore$ covar matrix:-

symmetric,

diagonal positive, remaining negative.

MGF: = take only $P_1, P_2, \dots P_{k-1}$

$$= (P_1 \cdot e^{t_1} + P_2 \cdot e^{t_2} + \dots P_{k-1} \cdot e^{t_{k-1}} + 1 - P_1 - P_2 - \dots P_{k-1})^n$$

(we do with $t_1, t_2, \dots t_{k-1}$ to get the vectorcomponents).

4) Hypergeometric distribution:-

sampling without replacement.

Say N - good objects, M - bad objects ; total pick
i.e. $x=1$, $x=0$. n - objects.

$$P(x=i) = \frac{{}^{N}C_i \times {}^{M}C_{n-i}}{{}^{N+M}C_n}$$; X is hypergeometric r.v.

total; i

$[{}^aC_b = 0,$ if $b>a$, or $b<0.]$

* $P(X_i) = N/N+M$ $\left(\because P(X_1) = \frac{N}{N+M}\right.$

meaning; good object came in ith pick. - a bernoulie R.V.

$P(X_2) = P(X_2=1|X_1=1)\cdot P(X_1=1) + P(X_2=1|X_i=0)\cdot P(X_1=0)$

can tell directly. we've no data.

$= \frac{N}{N+M}$ (on simplificatn)

$\vdots$)

* $E(X)$:

$= E(X_1+X_2+\cdots X_n)$

$= \frac{n\cdot N}{N+m}$

$var(X_i)$:

$= \frac{NM}{(N+M)^2}$ (equal to $P_i\cdot 1-P_i$).

* $var(X) = Var(X_1+X_2+\cdots X_n)$

$= \sum_{i=1}^{n} Var(X_i) + \sum_i \sum_{j\neq i} covar(X_i, X_j)$

X_i, X_j are not independent

$= np(1-p)\left[1-\frac{n-1}{N+M-1}\right]$

covar(X_i,X_j):

$= E(X_i\cdot X_j) - E(X_i)\cdot E(X_j)$

$= P(X_iX_j=1)\cdot 1 + P(X_iX_j=0)\cdot 0 - E(X_i)\cdot E(X_j)$

$= P(X_i=1, X_j=1) - E(X_i)\cdot E(X_j)$

$= \frac{N-1}{N+M-1}\cdot\frac{N}{N+M} - \left(\frac{N}{N+M}\right)^2$

$= \frac{-NM}{(N+M-1)(N+M)^2}$

Geometric distribution:-

* probability that first heads occurs in the k^{th} trail

$$\boxed{f_X(k) = (1-p)^{k-1} \cdot p} \; ; \text{Here } P(X>k) = (1-p)^k$$

memory less

$$\boxed{P(X > s+t) = P(X>s) \cdot P(X>t)}$$

mean $= \frac{1}{p}$.

variance $= \frac{1-p}{p^2}$

II Continuous r.var. distribution:- 1) Gaussian:-

Parameters: mean $= \mu$

std.dev. $= \sigma$

$$\boxed{f_X(x) = \frac{1}{\sigma\cdot\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \rightarrow N(\mu,\sigma^2)$$

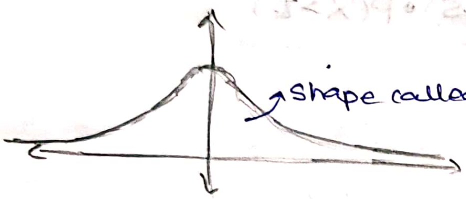
Not $N(\mu,\sigma)$

$$\boxed{\int_{-\infty}^{\infty} e^{-x^2}\cdot dx = \sqrt{\pi}}$$

proof by taking $\int_{-\infty}^{\infty} e^{-x^2}dx \cdot \int_{-\infty}^{\infty} e^{-y^2}dy$

(polar)

* $N(0,1) \rightarrow$ standard normal distribution.



shape called as Bell-curve.

mean $= \mu$.

variance $= \sigma^2$

Notes:-

$X \sim N(\mu,\sigma^2)$;

$aX+b \sim N(a\mu+b, a^2\sigma^2)$

CDF:

$$\phi(x) = F_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-z^2/2} dz \quad \text{for } N(0,1).$$

(we can't exactly integrate $\int_0^x e^{-x^2}dx$).

* error function $\boxed{\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \cdot \int_0^x e^{-x^2}dx}$

$\mathrm{erf}(x)$ as $x \to \infty = 1$.

$$\ast\ \phi(x) = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{0} e^{-z^2/2}dz + \int_0^x e^{-z^2/2}dz\right)$$

$$= \frac{1}{\sqrt{2\pi}}\left(\sqrt{2}\cdot\frac{\sqrt{\pi}}{2} + \sqrt{2}\cdot\int_0^{x/\sqrt{2}} e^{-z^2}dz\right)$$

$$\phi(x) = \frac{1}{2}\left(1+\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$$

$$\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left(1+\mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right)$$

] for mean $= \mu$; variance $= \sigma^2$.

* probability for R.V. to be with $\mu - n\sigma$ to $\mu + n\sigma$.

$n=1$	68.2%
$n=2$	95.4%
$n=3$	99.7%
$n=4$	99.9937%

MGF:

$$\phi_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

calculate $E(x)$ ✓

$E(x^2)$ ✓

Proof:

$$\int_{-\infty}^{\infty} e^{tx} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} \, dx.$$

$$e^{tx} \cdot e^{-x^2/2} \frac{1}{\sqrt{2\pi}}$$

$$e^{t^2/2} \cdot \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx$$

$$\sqrt{2\pi}$$

$$e^{t^2/2}$$

$$\frac{x-\mu}{\sigma} = K.$$

$$x = \sigma K + \mu.$$

$$\sigma \cdot dK.$$

$$e^{\mu}$$

$$e^{t(\sigma K + \mu)}$$

$$e^{t\mu} \cdot e^{t\sigma K} e^{-\frac{K^2}{2}} dK.$$

$$e^{\frac{(\sigma t)^2}{2}}$$

* Central Limit theorem:-

carry; n rounds of an experiment & random variables be

$x_1, x_2 \cdots$ are iid (identical, independent random var.)

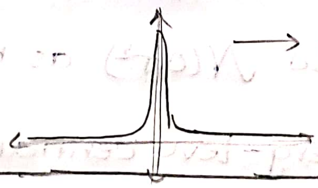
$x_1 \; x_2 \; x_3 \; \cdots \; x_n.$

mean $= \mu$;

(i) if $y = \frac{\sum x_i}{n} - \mu$:-

By weak law of large numbers;

as $n \to \infty$; $y \to 0$.



$\to$ Gauss: $N(0, \sigma \to 0)$

(ii) say $y = \sqrt{n}\left(\frac{\sum x_i}{n} - \mu\right)$:-

y itself will be a gaussian Random variable $N(0, \sigma^2)$ irrespective of the original distribution of x_i.

i.e. > for a single round; take n-experiments. (iir)

$X_1, X_2, \ldots X_n$ ↳ (μ, σ^2) each X_i

& take their mean $(= y)$

> do 't' such rounds

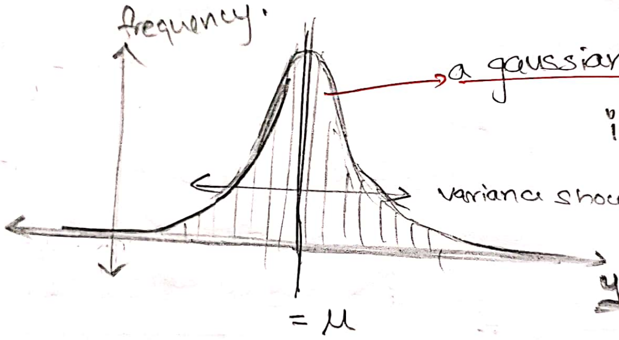
$$y_1 = \frac{(X_1)_1 + (X_2)_1 + \cdots (X_n)_1}{n}$$

$y_2 = \_\_\_\_\_\_$

⋮

$y_t = \_\_\_\_\_\_$

> now histogram of all such y:



a gaussian dist-!

i.e. Y is a gaussian RV.

variance should be $= \frac{\sigma^2}{n}$

* statement:-

consider $X_1, X_2, \ldots X_n$ to be a sequence of independent & identically distributed r.v.s with each, having

mean $= \mu$ $(\neq \infty)$

Variance $= \sigma^2$ $(\neq \infty)$ (no where talked about actual pdf (x_i)) so; CLT is powerful.

then the distribution of

$$Y_N = \sqrt{n}\left(\frac{\Sigma X_i}{n} - \mu\right)$$

converges to $N(0, \sigma^2)$ as $n \to \infty$.

Lindeberg-Levy central limit theorem.

* one version of CLT; requires only independence of $X_1, X_2 \ldots$

then $$Y_n = \left(\frac{\Sigma(X_i - \mu_i)}{\Sigma \sigma_i^2}\right)$$ is gaussian variant.

with some condn ( ) $\to 0$.

(out of portion?)

(i.e. can have different distributions too!)

* this, more general version of CLT is Lindeberg's CLT.
  - provides major motivation, of widespread use of gaussian distn.
  - errors in experiments are thus modelled gaussian.

* No disparity b/w CLT & Law of large numbers...

$\boxed{\text{gaussian } \mathcal{N}(\mu, \sigma)}$

* proof:-

consider $z = \left(\frac{\Sigma X_i - \mu n}{\sigma\sqrt{n}}\right)$

z is $\mathcal{N}(0,1)$ as $n \to \infty$.

• we do this by finding MGF(Z) as $n \to \infty$.

$$\phi_z(t) = \left(\phi_{x-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n$$

— use facts

i) $\phi_{x+y}(t) = \phi_x(t) \cdot \phi_y(t)$ for independent.

(ii) $\phi_{ax+b}(t) = e^{bt} \cdot \phi_x(at)$

we got to prove:

$$\lim_{n\to\infty} n \cdot \log\left(\phi_{\tilde{x}}\left(\frac{t}{\sigma\sqrt{n}}\right)\right) = \frac{t^2}{2}$$

Bcoz

MGF$(\mathcal{N}(0,1)) = e^{t^2/2}$

$n = \frac{1}{x^2}$ (substn) for lim. calculation.

& L-hopital rule 2-times. ✓.

$\therefore \boxed{\phi_z(t) = e^{t^2/2}}$

Now; By uniqueness of MGF;

$\because \phi_{\mathcal{N}(0,1)}(t) = e^{t^2/2}$

$\therefore$ $z \equiv \mathcal{N}(0,1)$

Hence proved CLT — simpler version.

• Gaussian Tail Bounds:

$CDF_x(X \geq x)$ ke liye, upper bound.

tail probability. not exact $P_x(X \geq x)$

$$P(X \geq x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt \leq \int_x^\infty \frac{t}{x\sqrt{2\pi}} e^{-t^2/2} dt$$

$\mathcal{N}(0,1)$ for $x > 0$

we get:

$$P(X \geq x) \leq \frac{e^{-x^2/2}}{x \cdot \sqrt{2\pi}}$$

* Binomial dist. to Gaussian dist:-

* for large n; binomial dist. approaches to Gaussian dist. with appropriate mean & var.

say n.

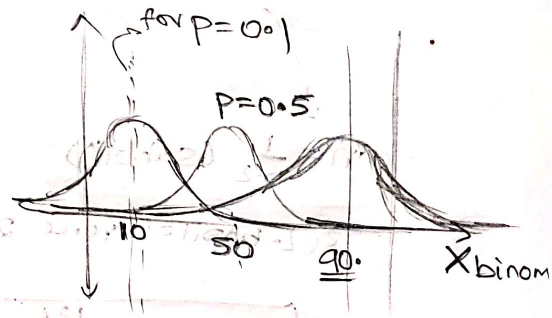
then $X_{binomial} = X_1 + X_2 + \cdots X_n$

Bernoullie...

$E(X_{bino}) = n \cdot p$.

$Var(X_{bino}) = n(p \cdot (1-p))$

$$Y = \frac{X_{bino} - n \cdot p}{\sqrt{n} \cdot \sqrt{(p)(1-p)}}$$



Y is $\mathcal{N}(0,1)$ for $n \to \infty$

$X_{bino} \sim Binomial(n, p)$

$\therefore X_{bino}$ is $\mathcal{N}(np, n^2 p(1-p))$

for sufficently Large n.

(50) ✓

Haha.

whether see from bino point of view (or) Gauss point of view

mean = np
Var. = np(1-p)
} won't change na

101.

I) Sample mean:-

consider 'n' i.i.r $X_1 X_2 \cdots X_r$ with μ, σ^2

then $\bar{X} = \frac{\Sigma X_i}{n}$ is a random variable. called the sample mean.

1) By Law of Large numbers:-

as $n \to \infty$ $\quad E(X) = \mu$

$P(X == \mu) \to 1.$ $\quad Var(X) = \frac{1}{n^2}(n \cdot \sigma^2) = \frac{\sigma^2}{n}.$

2) By CLT:-

for sufficiently big n;

X is a gaussian dist. (approximately)

$X \sim N(\mu, \frac{\sigma^2}{n})$

Note:-

HW-2 problem.

if $X_1, X_2, \cdots X_n$ were independent normal random (Gaussian) variables

* didn't say identical.

then $X = \frac{\Sigma X_i}{n}$ is also a normal (Gaussian) random variable.

(No need of CLT or $n \to \infty$).

Proof: By MGF:

$$\phi_X(t) = \phi_{\frac{X_1}{n} + \frac{X_2}{n} + \cdots \frac{X_n}{n}}(t)$$

$$= \phi_{X_1}(\tfrac{t}{n}) \cdot \phi_{X_2}(\tfrac{t}{n}) \cdots$$

$$= e^{\frac{t\mu_1}{n} + \frac{t^2}{2}\frac{\sigma_1^2}{n^2}} \cdot e^{t\mu_2 + \frac{t^2}{2}\sigma_2^2} \cdots$$

$$= e^{t\left(\frac{\Sigma \mu_i}{n}\right) + \frac{t^2}{2}\left(\frac{\Sigma \sigma_i^2}{n^2}\right)}$$

$$= e^{t(\mu_{net}) + t^2/2\left(\frac{\Sigma \sigma_i^2}{n^2}\right)}$$

By uniqueness theorem (1 MGF $\longleftrightarrow$ 1 pdf)

$$X \sim N(\mu_{net}, \sigma_{net}^2)$$

$$X \sim N\left(\frac{\Sigma \mu_i}{n}, \frac{\Sigma \sigma_i^2}{n^2}\right)$$

II) Sample variance:- (S^2) experimental variance.

$X_1, X_2, \cdots X_n$ are iirs.

then

$$S^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\cdot\bar{x}^2}{n-1}$$

now;

$E(S^2) = ?$ from a X_i; we are having n instances x_1, x_2, x_n

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

$$= \frac{1}{n-1}\left(\sum\left[\text{var}(x_i) + E(x_i)^2\right] - n\left[\text{var}(\bar{x}) + E(\bar{x})^2\right]\right)$$

their variance should be σ^2 na... variance of distribution.

$$= \frac{1}{n-1}\left(n\cdot\sigma^2 + n\cdot\mu^2 - n\frac{\sigma^2}{n} - n\mu^2\right)$$

$\therefore$ "expectation" of their variance should obviously be σ^2. lets see.

$$= \frac{\sigma^2\cdot(n-1)}{n-1}$$

$$= \sigma^2.$$

Note:

Here, we used $S^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n-1}$; rather than $S^2 = \frac{\sum(x_i - \bar{x})^2}{n}$

something related with unbiased estimator.

would give $E(S^2) = \frac{\sigma^2(n-1)}{n}$

$\therefore$

expected value of sample variance of n trials outcomes $\neq$ true variance of trials.

this is undesirable.

So; we multiply S^2 with $n/n-1$

$\boxed{\text{Bessel's correction}}$

* What about 'distribution' of sample variance. (i.e. S^2)?
· learn a new distribution.
Huff...

2) Chi-square distribution:-

- $z_1, z_2, \ldots z_n$ are independent, standard normal random vars, $\mathcal{N}(0,1)$

then

$$X = z_1^2 + z_2^2 + \cdots z_{n-1}^2 + z_n^2$$

(as $n \to \infty$; X should tend to gaussian. — CLT)

X is a chi-square random variable.
with 'n' degrees of freedom.
woah.

$$\boxed{X \sim \chi_n^2}$$ chi...

$$\boxed{f_X(x) = \frac{x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}}}{2^{n/2} \cdot \Gamma(n/2)}}$$

gamma function.

$$\Gamma(t) = \int_0^\infty x^{t-1} \cdot e^{-x} dx$$

$$\Gamma(n/2) = \frac{n}{2} \cdot \Gamma\left(\frac{n}{2}-1\right)$$

$$\Gamma(1/2) = \sqrt{\pi}.$$

if $n = 2k$;

$$\Gamma(n/2) = k!$$

* deriving for $n=1$.

$$X = z_1^2 \ ; \ z_1 \sim \mathcal{N}(0,1)$$

$$F_X(x) = P(z_1^2 \leq x)$$

$$= P(z_1 < \sqrt{x}) - P(z_1 < -\sqrt{x})$$

$$= F_{z_1}(\sqrt{x}) - F_{z_1}(-\sqrt{x})$$

$$\therefore f_X(x) = \frac{1}{2\sqrt{x}} \left(f_{z_1}(\sqrt{x}) + f_{z_1}(-\sqrt{x})\right)$$

$$= \frac{1}{2\sqrt{x}} \cdot \frac{e^{-x/2}}{\sqrt{2\pi}} \times 2$$

$$\therefore \boxed{f_X(x) = \frac{x^{-1/2} \cdot e^{-x/2}}{2^{1/2} \cdot \Gamma(1/2)}}$$

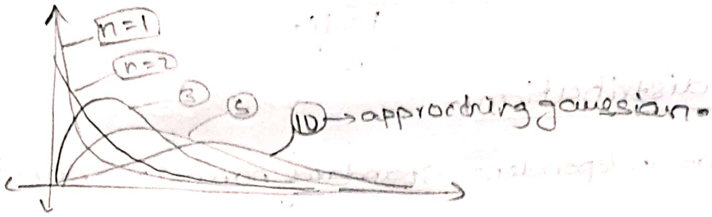
MGF; for X_n^2 is:-

$$\phi_{X_n^2}(t) = (1-2t)^{-n/2}$$

(do for $\phi_{X_1^2}(t)$ & multiply 'n' times)

Now; we can find, or "verify" $f_{X_n^2}(X)$

<plot of X for different n_i vs x>



- Additive property:-

$X = U_1 + U_2$

U_1: chi-sq. R.V. with n_1 deg. of freedom; U_2: n_2

then X is chi-sq. R.V. with (n_1+n_2) df.

---

* now; $S^2 = \dfrac{\sum x_i^2 - n(\bar{x})^2}{n-1} = \dfrac{\sum (x_i-\bar{x})^2}{n-1} = \dfrac{\sum (x_i-\mu)^2 - n(\bar{x}-\mu)^2}{n-1}$

(variance of $X = X+b$).

$$\therefore \underbrace{\sum\left(\frac{x_i-\mu}{\sigma}\right)^2}_{\chi^2_n} = \frac{\sum (x_i-\bar{x})^2}{\sigma^2} + \underbrace{\left(\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}\right)^2}_{\chi^2_1}$$

complete is a sum of 'n' standard Normal R.V. (x_i is normal R.V.)

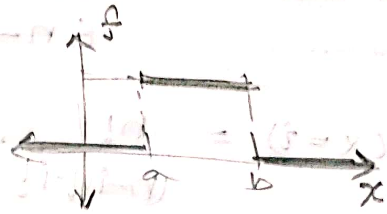
via CLT; this is std. gaussian R.V.

Both are independent? Yes!

$\therefore \dfrac{\sum (x_i-\bar{x})^2}{\sigma^2}$ is a χ^2_{n-1} R.V.

3) <u>Uniform distribution:-</u>

$$f_X(x) = \frac{1}{b-a}, \quad a<x<b$$
$$0 \quad , \text{otherwise.}$$



$$\star\ \mu(X) = \frac{b+a}{2}$$

$$\star\ Var(X) = \frac{(b-a)^2}{12}.$$

$$MGF = \int_a^b e^{tx}\cdot\frac{1}{b-a} = \frac{e^{bt}-e^{at}}{t(b-a)} \quad \text{for } t\neq 0.$$
$$= 1 \quad \text{for } t=0.$$

* <u>Application:-</u>

i) if we somehow bring uniform dist. R.V. X; then we can prepare samples for other distributions.
Like Gauss, poison, ...

ii) say;

$P(0)=0.3 \quad P(1)=0.3 \quad P(2)=0.4.$

now...

$X \sim$ uniform $\underbrace{[0,1]}$
the range of probability value.

if $(0\le X<0.3)$: sample value $=0$.

if $(0.3\le x\le 0.6)$: value $=1$.

if $(0.6\le x\le 1)$: value $=2$.

(iii) describe the working of randperm:-

- i.e. Select a subset of length K; from a set of length n.

Like;

bcoz; $\boxed{P(I_j \mid I_1, I_2, \ldots I_{j-1} \text{ are known}) = \frac{K-\sum^{j-1} I_i}{n-(j-1)}}$

now pick x from uniform $[0,1]$
if $\underline{x < P(I_j|\ldots)}$; $I_j = 1$
else —— $I_j = 0$

4) poission distribution:-

this is discrete R.V. case.

* A binomial dist. where $E(X)$ is fixed; & $n \to \infty$ is discussed as poission dist.

$$\text{Pmf}(X=i) = \frac{n!}{(n-i)!\cdot i!} \cdot (P)^i \cdot (1-p)^{n-i}$$

$\because n \to \infty$
& $E(X) = np = \lambda$.
$\therefore p = \frac{\lambda}{n}$.

$$P(X=i) = \frac{\lambda^i}{i!} \cdot \frac{n!}{(n-i)!\cdot n^i} \cdot \frac{\boxed{(\sim 1)^{n-i}}}{\left(1-\frac{\lambda}{n}\right)^{n-i}}$$ wait!

$$\simeq \frac{\lambda^i}{i!} \cdot \boxed{\left(1-\frac{\lambda}{n}\right)^n}$$

not 1.
this is $e^{-\lambda}$.

$$\boxed{P(X=i) = \frac{e^{-\lambda}\cdot\lambda^i}{i!}}$$

- sample size is large ($n \to \infty$) but $E(X)$ is finite. $= \lambda$.

poission Pmf
not Pdf

$\therefore \sum_{i=0}^{\infty} P(X=i) = 1$.

* $E(X) = \lambda$
* $Var(X) = \lambda$

think of binomial; with $np = \lambda$ & $p \to 0$

* Mgf: $\sum e^{-\lambda} \frac{(e^t\lambda)^i}{i!} = e^{-\lambda}(e^{e^t\lambda})$

$$\boxed{\phi_X(t) = e^{\lambda(e^t-1)}}$$

weirdest.

- for larger λ; $\text{Poission}(\lambda) \sim \mathcal{N}(\lambda, \lambda)$

* Say $z = X + Y$.

where X is poission (λ_1)

Y is poission (λ_2)

& both are independent.

then $\underline{Z = \text{poission}(\lambda_1 + \lambda_2)}$ proof by considering:-

$\phi_z(t) = \phi_x(t) \cdot \phi_y(t)$.

* $\frac{\text{PMF}(i+1)}{\text{PMF}(i)} = \frac{\lambda}{i+1}$

* if $X \sim$ poission (λ)

& $P(Y-|X=l) = \text{Binomial}(l, p)$ — n of Bino. dit.; p of Bino. dit

then

$Y \sim$ poission (λp)

thinning of poission rand. var. by a binomial.

$\text{Pmf}(y = x_0) = \sum P(X=l) \cdot X_{B(l,p)}(x_0)$

$P_i \quad e^{t \cdot P_i} \quad e^{t} \cdot e$

$P_i \cdot x_i \quad 1-P$

practical use:-

Points in a scene being imaged, send out photons at rate of λ (poisson).

of these; only a small frac. p, managed to enter the camera (binomial)

∴ effective rate captured by camera is λp (poission)

$= \sum_{l=x} \frac{e^{-\lambda} \lambda^l}{\cancel{l!}} \times \frac{\cancel{l!}}{(l-x)! \cdot x!} \cdot p^x (1-p)^{l-x}$

poisson eq:-

fixed no. of acc. per day....
accidents...

but police watches every 10 min.

5) Exponential distribution:-

* say; a process has 'λ' chances to succeed in unit time.

  $\therefore$ '$\lambda \cdot u$' chances; in 'u' time.

  poisson process

  let T denote the time; when first success occurs – waiting time.

* we hav

  T = a random variable.

$$F_T(T \le u) = 1 - \underbrace{P(T \ge u)}$$

meaning;
in time 'u';
'0' successes.

$= \text{Pmf}_{\text{poiss}(\lambda u)}(=0)$

$= e^{-\lambda u}$

$$\boxed{P(T>u) = e^{-\lambda u}}$$

imp. distinct feature of exp. rand. var.

$>$ $\therefore$ $\boxed{F_T(u) = 1 - e^{-\lambda u}}$ what the shit is going on here

$>$ $\therefore$ $\boxed{f_T(u) = \lambda \cdot e^{-\lambda u}}$ $u \ge 0$.

exponential random var.

* $E(T) = \dfrac{1}{\lambda}$ $\left(= \int_0^\infty t \cdot \lambda \cdot e^{-\lambda t} dt = \lambda \cdot \left[\cancel{\left[t \cdot \dfrac{e^{-\lambda t}}{-\lambda} \right]_0^\infty} + \left[\dfrac{e^{-\lambda t}}{-\lambda^2} \right]_0^\infty \right] \right.$

  $= \dfrac{1}{\lambda^2}$

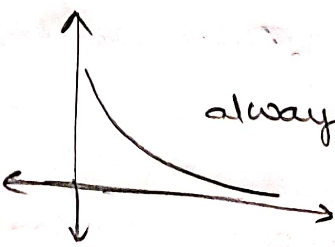
  $1/\lambda$

* $\text{Var}(T) = \dfrac{1}{\lambda^2}$

* $\text{MGF}(T)(t) = \left.\dfrac{\lambda}{\lambda - t}\right\}$ whatever.

  waw,

* mode is $u=0$



always exponential decay.

* median: $\int_0^{x_0} f_T(u) \cdot du = 1/2$

$$\boxed{x_0 = \frac{\ln 2}{\lambda}}$$

* This exp. R.V. is said to be memory less:-

$\forall s,u \geq 0 \quad P(T>s+u \mid T>u) = \underline{P(T>s)}$

This is not independence-101...

We haven't got 2 variables here.

also RHS $\neq P(T>s+u)$

proof:

$P(T>s+u \mid T>u)$

$$= \frac{P(T>s+u, T>u)}{P(T>u)}$$

$$= \frac{e^{-(s+u)\lambda}}{e^{-(u)\lambda}}$$

$$= e^{-s\lambda}$$

$$= P(T>s)$$

($\because$ exponential tail).

* if $x_1, x_2, \ldots x_n$ are exp. rand. var;

$X = \min(x_1, x_2, \ldots x_n)$ is also a exp. rand. var

$$\because \quad P(X>x_0) = P(X_1>x_0 \cap X_2>x_0 \cap \ldots)$$

$$= e^{-\lambda_1 x_0} \cdot e^{-\lambda_2 x_0} \cdots e^{-\lambda_n x_0}$$

$$\boxed{P(X>x_0) = e^{-(\lambda_{net})x_0}}$$

so,

$$F_X(x_0) = 1 - e^{-\lambda_{net} x_0}$$

$$\therefore f_X(x_0) = \lambda \cdot e^{-\lambda_{net} x_0}$$

Self:-

* to find $f_X(x)$ for some distribution; (methods)
Like $\min(X_1, X_2 \ldots)$
or just $X_1 + X_2 + \cdots$

(i) maybe multiply individual $f_{X_1}(x_1), f_{X_2}(x_2) \ldots$
(for independent variables & joint probability).

(ii) Find CDF first & get pdf from that.
$\equiv$ counting manually in case of continuous distributions

$y = \min(X_1, X_2, X_3, \ldots)$ or $y = \max(X_1, X_2 \ldots)$

then $P(Y \le y) = 1 - P(Y > y)$

$= 1 - P(X_1 > y) \cdot P(X_2 > y) \ldots$

MLE of uniform distribut^n.

$\therefore$ now; $\text{pdf} = \frac{\partial}{\partial y} P(Y \le y)$

discrete case.
$\shortparallel$
(iii) Start counting manually. (we do this for basic distributions) AND discrete r.v..

like $f(X = x) = {}^nC_x \cdot p^x \cdot (1-p)^{n-x}$
(binomial).

(iv) Look at MGF:
Third
(if MGF is of any one of known forms; the by uniqueness theorem; we can prove $f_X(x)$.

(used in CLT proof;
also used in proving
$X = x_1 + x_2 + \cdots x_n$
& each x_i is gaussian
then X is gaussian).

Self :-

* to find $E(x)$ & $Var(X)$ for some random var. X :-

(i) do the summation directly :-

$E(X) = \int x \cdot f_X(x)dx$ (or) $\sum x \cdot f_X(x)$ } used very less; in case of the standard distributions.

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

(ii) Try to write X as $X_1+X_2+X_3+\cdots$:-

now;

$E(X) = E(X_1) + E(X_2) + \cdots E(X_n)$ with or without independence.

$Var(X) = Var(X_1) + Var(X_2) + \cdots Var(X_n)$ with independence

$Var(X) = \sum Var(X_i) + \sum\sum Covar(X_i, X_j)$ without independence.

used in Binomial, multinomial, Geometric, Hypergeometric

where ever; we can have 1 trial consisting of 'n' events at a time..

(iii) from MGF :-

$E(X)$ ✓

$var(X) = E(X^2) - (E(X))^2$ ✓ ✓

Eg: n people have hats, in a party; group all hats & take one each.

Y = no. of people who got their own hat.

Find $E(Y)$ & $Var(Y)$.

Sol) $f_Y(y) = \dfrac{{}^nC_y\left((n-y)! - {}^{n-y}C_1 \cdot (n-y-1)! + {}^{n-y}C_2 (n-y-2)! \cdots\right)}{n!}$

derangments....

Soodhi formula.

can't use this for $E(y)$.

now; $Y = X_1 + X_2 + \cdots X_n$; here each (X_i, X_j) is not independent.

where X_1 is bernoulli with $p = 1/n$

$\therefore E(Y) = \sum E(X_i) = 1$.

$var(Y) = \sum var(X_i) + \sum covar(X_i, X_j)$

$= n \cdot \frac{1}{n}(1-\frac{1}{n}) + n(n-1)\left(covar(X_1, X_2)\right)$

$= E(X_1 X_2) - E(X_1)\cdot E(X_2)$

$= \frac{1}{n} \times \frac{1}{n-1} - \frac{1}{(n)^2}$

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

$= 1$

PARAMETER ESTIMATION:-

* let $x_1, x_2, --- x_n$ be a random sample from a distribution F_θ; where we don't know the parameters θ. (But we do know the family of F_θ)

Eg: • poisson; with unknown parameter λ. , we need to estimate this; with knowledge of $x_1, x_2, -- x_n$.

• normal dist.; with unknown mean & σ^2.

* In probability theory; we take that the parameters are known; whereas in Statistics theory; the opp. is true; we use observed data to inference on parameter

→ Maximum Likelihood Estimators:- (of a parameter)
ML estimator.

* Also called 'point estimate'; since we give a single value for θ, instead of a range (confidence interval) more statistics like.

Note:
Any statistic; used to estimate the value of a parameter is called estimator. There are many kinds of estimators $\bar{\theta} = 2$; irrespective of sample is an estimator (very bad one).
- observed value of estimator is estimate.

* Say; our sample is $x_1, x_2, x_3 --- x_n$. then;

$$f(x_1, x_2, --- x_n) = f(x_1) \cdot f(x_2) \cdots f(x_n)$$ x_i, x_j all are independent.

now; i) we know the distribution; but not the parameter θ.

ii) we choose a $\theta = \bar{\theta}$ such that the chances of $x_1, x_2, -- x_n$ is maximum!

iii) i.e. we find $\theta = \bar{\theta}$ such that $f(x_1, x_2, -- x_n)$ is maximum.

∴ $x_1, x_2, -- x_n$ is "maximu likely"

Hence $\boxed{\text{MLE of } \theta}$

→ lets calculate parameters for some distribution; by MLE :-

1) Bernoullie:-

say samples were $x_1, x_2, x_3, \ldots x_n$ (x_i, $x_i \in \{0,1\}$)

let p be the parameter of the bernoullie dist.

then $f(x_1) = p^{x_1}(1-p)^{1-x_1}$ true for $x_1 = 0$ or $x_1 = 1$

$\therefore f(x_1, x_2, \ldots x_n) = p^{x_1+x_2+\cdots} \cdot (1-p)^{n-x_1-x_2-\cdots}$

$$\boxed{\begin{array}{l} \text{maximize } f(x_1, x_2 \ldots x_n) \\ \therefore \text{ maximize } \log(f(x_1, x_2, \ldots x_n)) \end{array}}$$ important na.

$\therefore \log(f(x_1 x_2 \ldots x_n)) = (x_1+x_2+\cdots x_n)\log p + (n-x_1-x_2-\cdots)\log(1-p)$

$\frac{\partial}{\partial p} = 0$

$\Rightarrow \frac{1}{p}(\Sigma x_i) - \frac{(n-\Sigma x_i)}{1-p} = 0$

$\therefore (1-p)(\Sigma x_i) = pn - p\Sigma x_i$

$\therefore \boxed{p = \frac{\Sigma x_i}{n}}$

this is MLE for p.

2) Poission:-

$f(x_i) = \frac{e^{-\lambda}\lambda^i}{i!}$

$\therefore f(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!}$

$\log(f(x_1 \text{ to } x_n)) = n \cdot \log(e^{-\lambda}) + (\Sigma x_i) \cdot \log\lambda - \{\text{const.}\}$

$\frac{\partial}{\partial \lambda} = 0$

$\Rightarrow 0 = n(-1) + \frac{\Sigma x_i}{\lambda}$

$\therefore \boxed{\lambda = \frac{\Sigma x_i}{n}}$

MLE of λ.

3) MLE for gaussian:-

consider $\mathcal{N}(\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$f(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\therefore \; f(x_1 \text{ to } x_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\therefore \; \log(f(x_1 \text{ to } x_n)) = \sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2} - n\log(\sigma) + \{\text{const}\}$$

• now; we hav to maximize wrt 2 variables.

1) $\frac{\partial \log f}{\partial \mu} = 0$

$$\Sigma(x_i-\mu) = 0$$

$$\therefore \; \underbrace{\boxed{\mu = \frac{\Sigma x_i}{n}}}_{\text{MLE for } \mu}$$

2) $\frac{\partial \log f}{\partial \sigma} = 0$

$$\Sigma -\frac{(x_i-\mu)^2}{2}\frac{(-2)}{\sigma^3} - \frac{n}{\sigma} = 0$$

$$\therefore \; \underbrace{\boxed{\sigma^2 = \frac{\Sigma(x_i-\mu)^2}{n}}}_{\text{MLE for } \sigma^2}$$

we see that:-

MLE(mean) = mean($x_1 x_2 - x_n$)

MLE(var) = var($x_1 x_2 -- x_n$)

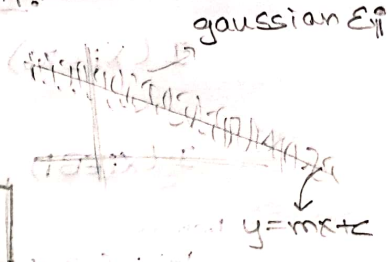
for other distribution to!

Bernoulie,
poission.

actual solving dosent involve seeing max. Likely hood at all!

→ ML for least square line fitting:-

Linear regression:-



* values are pairs (x_i, y_i)

our distribution is:

• not a probability distribution.

$$\boxed{y_i = mx_i + c + \varepsilon_i \quad \text{from } \mathcal{N}(0,\sigma^2)}$$

• we know x_i – accurately.
• we have noisy y_i.

we need to determine m, c

even though ε_i are iid; y_i are only independent

$$y_i \in \mathcal{N}(mx_i + c, \sigma^2)$$

$$\therefore \quad P(y_i ; x_i, m, c) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(y_i - mx_i - c)^2}{2\sigma^2}}$$

σ is given?

$$\therefore \quad \prod_{i=1}^{n} (P(y_i, x_i, m, c)) = \prod_{i=1}^{n} \frac{e^{-\frac{(y_i - mx_i - c)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$\therefore \quad \log(f) = \sum_{i=1}^{n} -\frac{(y_i - mx_i - c)^2}{2\sigma^2} + \{\text{constant}\} \text{ in } \sigma.$$

* we need MLE for m, c:- (is σ given?)

$$\frac{\partial}{\partial m} = 0 \quad \Rightarrow \quad \Sigma\, x_i (y_i - mx_i - c) = 0$$

$$\Rightarrow \boxed{m(\Sigma x_i^2) + c(\Sigma x_i) = \Sigma x_i y_i}$$

$$\frac{\partial}{\partial c} = 0 \quad \Rightarrow \quad \Sigma(y_i - mx_i - c) = 0$$

$$\boxed{c = \left(\frac{\Sigma y_i}{n}\right) - m\left(\frac{\Sigma x_i}{n}\right)}$$

solve simultaneously.

$$\therefore \quad \boxed{\bar{m} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \qquad \boxed{\bar{c} = \bar{y} - \bar{m}\bar{x}}$$

where $\bar{x} = \frac{\Sigma x_i}{n}$

$\bar{y} = \frac{\Sigma y_i}{n}$

* indicator function:-

$I(X_i \le x_0)$ is a bernoullie rand. var; with parameter = CDF(x_0)

$I(X_i \in B_i)$ is a bernoullie; with $p =$ CDF(X_2)$-$CDF(X_1)

where $B_i \equiv [x_1, x_2]$

& ∴ if $X = \{x_1, x_2, x_3, ---\}$

& $N_{x_0} = \sum_{i=1}^{n} I(X_i \le x_0)$

then N_{x_0} is a binomial r.v.

* Now; ML estimators are Random variables. why? Bcoz it is a "function" of samples from an underlying dist.

- the ML estimator has its own pdf
  mean
  variance.
  (When you calculate the value of MLE for a dataset, this is a sample from pdf(MLE))

→ Bias; variance; mean square deviation of an estimator:- not just MLE.

$X_1, X_2, \ldots X_n$ are r.v. (iid) from a distribution with parameter θ.

- let $\hat{\theta}$ be an estimator of θ.
  (How to decide, whether good or bad estimator?)

evaluate $(\hat{\theta}-\theta)^2$ (square deviation)
But it is R.V.
So; evaluate $E((\hat{\theta}-\theta)^2)$

* $E[(\hat{\theta}-\theta)^2]$ is called mean squared error of estimator.
  θ → true value! (we desire low MSE estimators).

* if $E[\hat{\theta}]=\theta$, unbiased
  else, biased.

$(E(\hat{\theta})-\theta) \longrightarrow$ Bias of the estimator.

Eg: unbiased:-
- MLE of mean for a gaussian sample.
- MLE of variance for a gaussian sample (when mean is known).

biased:
- MLE of variance for a gaussian sample. (when mean is unknown)
- interval of uniform dist.
  $\hat{\theta} = \max(X_1, X_2, X_3 \ldots)$
  $E(\hat{\theta}) = \int n x^{n-1}/\theta^n = n/_{n+1} \cdot \theta$

* variance of estimator = $E[(\hat{\theta} - E(\hat{\theta}))^2]$.
need.... not the true mean of sample!

*

$$MSE(\hat{\theta}) = \underbrace{E[(\hat{\theta}-E(\hat{\theta}))^2]}_{\text{variance}} + \underbrace{(E(\hat{\theta})-\theta)^2}_{\text{square bias}}.$$

* A biased estimator may have lower MSE; owing to its low variance.

(so; saying unbiased better than biased is B.S.)

- also; if the MSE is not going down as the 'n' increases; then estimator is undesirable.

Eg: let $x_1, x_2, x_3 \cdots$ be r.v. of dist. of true parameter θ.

if $\hat{\theta} = x_1$

now; $E(\hat{\theta}) = \theta$.

but $[Var(\hat{\theta}) = \sigma^2 \text{(of dist)}]$
and! this dosen't go down! with n.

→ Estimator consistency:-

let θ be the parameter of a dist.

& $\hat{\theta}$ be a value of estimator of θ.

• we say estimator is (asymptotically) consistent if

$$\lim_{n \to \infty} \overset{\text{probability}}{P}(|\hat{\theta}-\theta| > \varepsilon) = 0 \quad \text{for any } \varepsilon > 0.$$

MSE→0 as $n \to \infty$.

probability is zero...

like; $\theta = 50$

& for 1 sample set $\hat{\theta}$ turns out to be 5

then for 10^8 samplesets; $\hat{\theta}$ turns out to be 50.

So; probability wise, fine.

Eg: for a distribution; with true mean θ;

Sample be $X_1, X_2, X_3, \dots X_n$

take two estimators; $\theta' = 1$ (irrespective of sample set)

$\theta'' = X_1$

then:

θ':

biased

variance = 0

MSE is high

| inconsistent estimator

θ'':

unbiased $(E(\theta'') = \theta)$

variance is high

MSE is high

| in consistent estimator.

$\boxed{\text{an estimator can be unbiased \& still be inconsistent.}}$

weak law of large nos. won't apply here; bcoz θ'' is not $\dfrac{X_1 + X_2 + X_3 + \cdots}{n}$ form.

→ Motivation for MLE:-

hard facts.

- MLE is a consistent estimator.

  (as long as true values won't change with n; sample count)

- no consistent estimator; can achieve a lower asymptotic MSE than MLE.

Bias & variance of various MLEstimators:-

i) μ, σ^2 estimators for Gaussian:

$$\hat{\mu} = \frac{\Sigma x_i}{n}$$

- $E(\hat{\mu}) = \frac{n \cdot \mu}{n} = \mu$ $\boxed{\text{unbiased}}$
- $var(\hat{\mu}) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$ } variance decreasing with n.
- $MSE(\hat{\mu}) = \frac{\sigma^2}{n}$ ($\because$ bias = 0).

$$\hat{\sigma^2} = \frac{\Sigma (x_i - \hat{\mu})^2}{N}$$

if μ is known beforehand:-

$$E(\hat{\sigma^2}) = \frac{1}{N} \Sigma E((x_i-\mu)^2)$$
$$= E((x_i-\mu)^2)$$
$$= \sigma^2$$

if μ not known:-

$$E(\hat{\sigma^2}) = \frac{1}{N} \cdot \Sigma E((x_i - \hat{\mu})^2)$$
$$= \frac{1}{n}\left[\Sigma E(x_i^2) + \Sigma E(\hat{\mu}^2) - 2\Sigma E(x_i \hat{\mu})\right]$$

$$\boxed{E(\hat{\sigma^2}) = \sigma^2\left(1-\frac{1}{n}\right)}$$

Biased.

But; bias ↓ as n↑.

2) θ estimator; for uniform $[0,\theta]$:-

$$\hat{\theta} = \max(x_1, x_2, \ldots)$$

what is the distribution of $\hat{\theta}$?

$$P(\hat{\theta} \leq x) = P(\max(x_1, x_2 \ldots) \leq x)$$
$$= P(x_1 \leq x) \cdot P(x_2 \leq x) \cdot \ldots$$
$$= \frac{x}{\theta} \cdot \frac{x}{\theta} \ldots$$

$$P(\hat{\theta} \leq x) = \frac{x^n}{\theta^n}$$

$$\therefore \boxed{f_{\hat{\theta}}(x) = \frac{n x^{n-1}}{\theta^n}} \rightarrow \boxed{E(\hat{\theta}) = \frac{n}{n+1}\theta.}$$

for $x \in [0,\theta]$
$= 0$ otherwise.

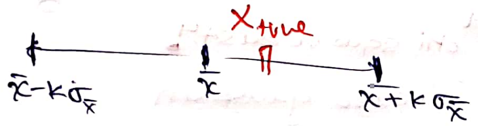
→ <u>confidence intervals:-</u>

→ θ_{true} might not be $\hat{\theta}$ in most of the cases. (point estimation).

So; we say;

$\theta_{true} \in [\hat{\theta} - c), \hat{\theta} + c)]$ with 99% probability.

confidence intervals.



Eg: 1) <u>ML estimate of Gaussian mean:-</u>

$X_1 X_2 \dots X_n$ are gaussian iids

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots X_n}{n}$$

now;

$$\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$$] true enough; if X_i are not gaussian also. (CLT).

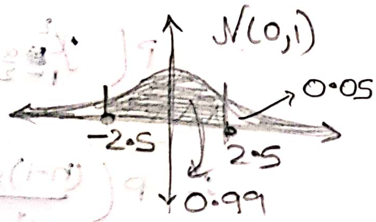
$$\therefore \quad P\left(-2.5 \le \sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right) \le 2.5\right) \simeq 0.99$$

if we don't know σ_{true}; approximate with $\sigma_{dataset}$.

$$\therefore \quad P\left(\bar{x} - 2.5\underbrace{\frac{\sigma}{\sqrt{n}}}_{\sigma_{\bar{x}}} \le \mu \le \bar{x} + \frac{2.5\sigma}{\sqrt{n}}\right) \simeq 0.99$$

$\therefore$ μ (true value) lies in $\left[\bar{x} - \frac{2.5\sigma}{\sqrt{n}}, \bar{x} + \frac{2.5\sigma}{\sqrt{n}}\right]$

with 99% confidence.

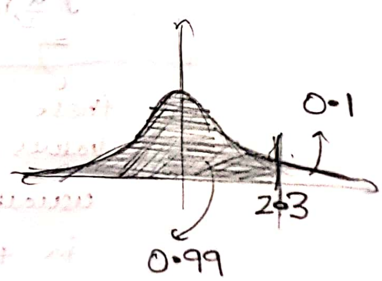


also;

$$P\left(\sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right) \le 2.3\right) \simeq 0.99$$

$$\therefore P\left(\bar{x} - \frac{2.3\sigma}{\sqrt{n}} \le \mu\right) \simeq 0.99.$$

one sided confidence....



μ lies right of ( ) with 99% confidence.

$\therefore Z_{0.05} = 2.5$

$Z_{0.1} = 2.3$ for $N(0,1)$

→ 2. for Variance's MLEstimator :-

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

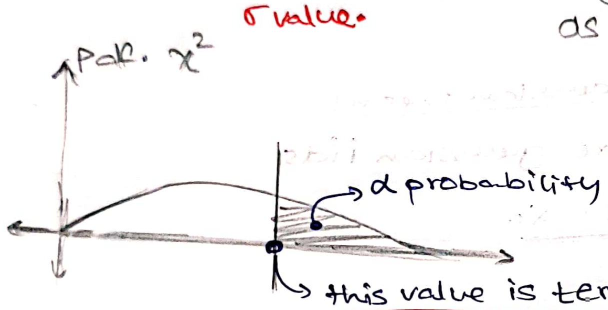
& as we've already seen;

$$(n-1)\left(\frac{S^2}{\sigma^2}\right) \sim \chi^2_{n-1}$$

chi square distn.

(also gaussian as n increases).

true σ value.



this value is termed Z_α.

or

$\chi^2_{\alpha, n-1}$.

$$\boxed{P(X \geq Z_\alpha) = \alpha}$$

∴ now;

$$P\left((n-1)\frac{S^2}{\sigma^2} \geq \chi^2_{\alpha/2, n-1}\right) \simeq \frac{\alpha}{2}$$

$$P\left((n-1)\frac{S^2}{\sigma^2} \geq \chi^2_{1-\frac{\alpha}{2}, n-1}\right) \simeq 1 - \frac{\alpha}{2}$$

$$\therefore \quad P\left(\chi^2_{1-\frac{\alpha}{2}, n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}, n-1}\right) \simeq 1 - \alpha.$$

$$\therefore \quad \boxed{P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right) \simeq 1 - \alpha}$$

these values are usually avaliable in tabular manner.

confidence interval.