05.12.2022

YZV 231E

Probability Theory & Stats

Week 11

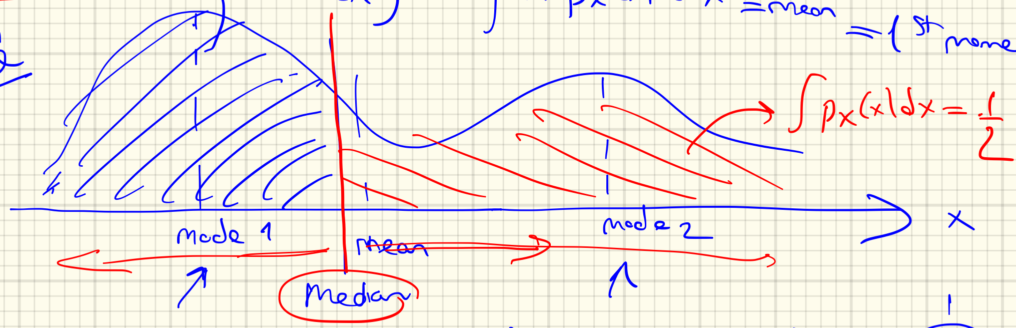
Gü.

Recap:

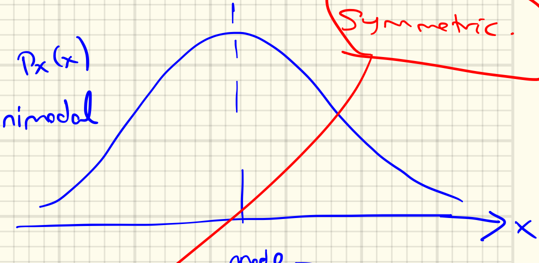
$E(X) = \int x \, p_X(x) \, dx$ = mean = 1st moment.

moments of prob. densities

$p_X(x)$ multimodal pdf



mode = $\arg\max_x p_X(x)$



Symmetric.

mode = mean = median

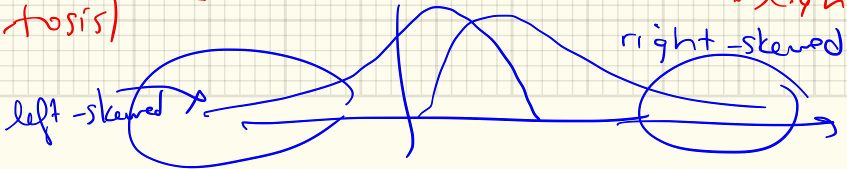
$Var(X) \propto$ 2nd moment

$E[(X-\mu)^2] = E[X^2] - (E[X])^2$

3rd moment measures symmetry. (skewness.)

4th moment : comparing tails — heavy, Gaussian, light

(kurtosis)

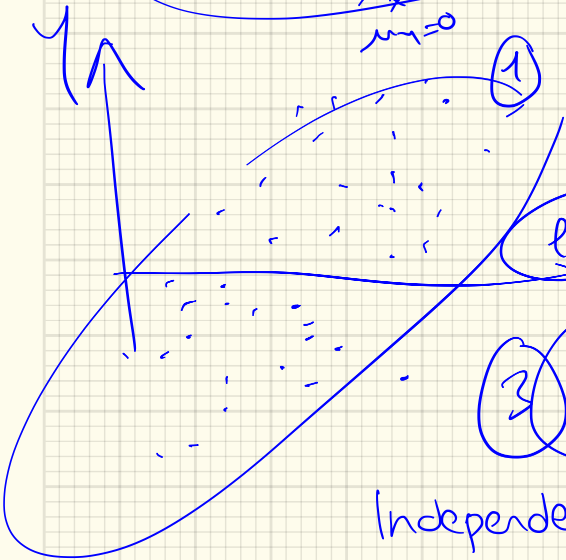


Covariance : $X_1, X_2, \dots, X_n$

$$E[X \cdot Y] \longrightarrow E[(X-\mu_x)(Y-\mu_y)]$$

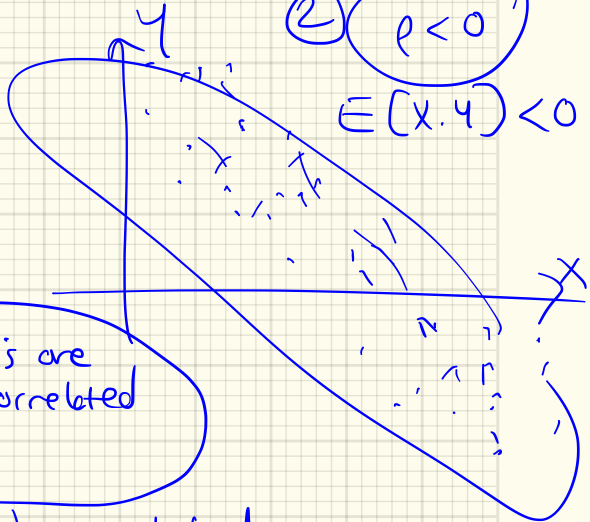
$\mu_x = 0$

$\mu_y = 0$



① Covariance $\updownarrow$ ρ : correlation coeff.

$\rho > 0$



② $\rho < 0$

$E[X \cdot Y] < 0$

③ $\rho = 0$: R.V.'s are Uncorrelated

Independence $\longrightarrow$ Uncorrelated.

$\cancel{\longleftarrow}$

Multivariate Gaussian pdf: joint pdf for N r.v.s.

$\mathcal{N}(\underline{\mu}, \underline{\underline{C}})$

random vector. $\underline{X} = [X_1, X_2, \ldots, X_N]^T$

$E[\underline{X}] = \underline{\mu} = [\mu_1 \ \mu_2 \ \ldots \ \mu_N]^T$.

Joint pdf for $\underline{X}$:

$$\boxed{P_{\underline{X}}(\underline{x})} = \frac{1}{(2\pi)^{N/2} \det^{1/2}(\underline{\underline{C}})} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{\underline{C}}^{-1} (\underline{x}-\underline{\mu})\right]$$

Covariance Matrix

$$\underline{\underline{C}} = \begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & \ldots & Cov(X_1,X_N) \\ Cov(X_2,X_1) & Var(X_2) & & \\ & & \ddots & \\ Cov(X_N,X_1) & \ldots & & Var(X_N) \end{bmatrix}$$

symmetric

positive-definite

$\det(\underline{\underline{C}}) > 0$

$\hookrightarrow$ then it is invertible

If $\underline{\underline{C}}$ is diagonal

then the r.v.s $\underline{X}^T = [x_1 \ x_2 \ldots x_N]$

are uncorrelated.

$$\underline{\underline{C}} = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \ \sigma_N^2 \end{bmatrix}$$

$$\det \underline{\underline{C}} = \prod_{i=1}^{N} \sigma_i^2$$

insert into $P_{\underline{X}}(\underline{x})$

$$\underline{\underline{C}} = \underline{\underline{U}} \, \Sigma \, \underline{\underline{V}}^T$$

$[v_1 \ldots v_n]$

$\begin{bmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \ d_n \end{bmatrix}$

EV)

$\boxed{\text{Show this}}$: $\underline{\underline{C}}^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \ldots, \frac{1}{\sigma_N^2}\right)$

Inverse Covariance matrix $= \begin{bmatrix} 1/\sigma_1^2 & & 0 \\ & 1/\sigma_2^2 & \\ 0 & & \ddots \\ & & & 1/\sigma_N^2 \end{bmatrix}$

$\det \underline{\underline{C}} =$

$$p_{\underline{X}}(\underline{x}) = \frac{1}{\prod_{i=1}^{N} \sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu_i)^2/\sigma_i^2\right]$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2}(x_i-\mu_i)^2/\sigma_i^2\right)$$

$$\boxed{p_{\underline{X}}(\underline{x}) = \prod_{i=1}^{N} p_{\underline{X}}(x_i)}, \quad \boxed{X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)}, \; i=1,\ldots,N$$

$\equiv \mathrm{cov}(X_i, X_j) = 0$ for $i \neq j$.

$\therefore$ For multivariate Gaussian, uncorrelated r.v.s $\longrightarrow$ independent r.v.s $X_1, \ldots, X_N$

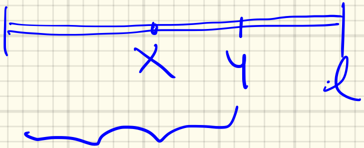
$\equiv \underline{\underline{C}}$ is diagonal.

Conditional Expectations : X, Y : two given r.v.s :

Given the value y of an r.v.

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x \, p_{X|Y}(x|y) \, dx \qquad \Big(\sum_x x \, p_{X|Y}\Big)$$

Recall stick-breaking ex. i) break uniformly at Y ;
ii) break " " X .



$E[X \mid Y = y] = \dfrac{y}{2}$ ← known y

$$\boxed{\underbrace{E[X \mid Y]}_{g(Y)} = \frac{Y}{2}}$$ ← Y .. r.v.

$g(Y)$: fn. of Y.

★ → $E[X \mid Y]$ is considered as an r.v.

$E[E[X \mid Y]]$ Law of Iterated Expectations →

$$E[E[X|Y]] = \sum_y \underbrace{E[X|Y=y]}_{g(y)} \cdot P_Y(y)$$

$$E[g(Y)]$$

Recall Total Expectation Thm.
taking weighted average of all possible scenarios.

$$= \sum_y \sum_x x \underbrace{P_{X|Y}(x|y) \cdot P_Y(y)}$$

$$\sum_x x \sum_y \underbrace{P_{X,Y}(x,y)}$$

$$\sum_x x \; P_X(x)$$

$$\boxed{\underbrace{E[E[X|Y]]}_{\text{expectation of a conditional expectation}} = \underbrace{E[X]}_{\text{unconditional expectation.}}}$$

expectation of a conditional expectation = unconditional expectation.

e.g. in the stick ex.

$$E[E[X|Y]] = E[X] = \frac{\ell}{4}$$

Conditional Variances : $\mathrm{Var}(X \mid Y)$ is an r.v. } consider all y's

~~$\mathrm{Var}(X \mid Y = y)$~~ for a specific y value.

$$\mathrm{Var}(X \mid Y=y) = E\Big[\big(X - \underbrace{E[X \mid Y=y]}_{\text{conditional exp.}}\big)^2 \;\Big|\; Y=y\Big]$$

Law of Total Variance :

Total Variance:

$$\mathrm{Var}(X) = \underbrace{E[\mathrm{Var}(X \mid Y)]}_{1} + \underbrace{\mathrm{Var}(E[X \mid Y])}_{2}$$

1

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

$$\mathrm{Var}(X \mid Y) = E[X^2 \mid Y] - (E[X \mid Y])^2$$

: valid for a condit. universe

$$E[\mathrm{Var}(X \mid Y)] = \underbrace{E[E[X^2 \mid Y]]}_{E[X^2]} - \cancel{E[(E[X \mid Y])^2]}$$

cancels w/ 1st term from 2 next page

$$\operatorname{Var}(X) = E[X^2] - (E[X])^2$$

2 $\operatorname{Var}(\underbrace{E[X|Y]}) = +E\left[(E[X|Y])^2\right] - (E[\underbrace{E[X|Y]}_{(E[X])}])^2$

$\Rightarrow$ 1 + 2 $= E[X^2] - (E[X])^2 = \operatorname{Var}(X)$

$$\Rightarrow \boxed{\operatorname{Var}(X) = E[\underbrace{\operatorname{Var}(X|Y)}] + \underbrace{\operatorname{Var}(E[X|Y])}}$$

Total Variance Law

$\longrightarrow$ ex:

Ex: A class of 30 students, w/ 2 sections.

They take a quiz.

2 r.v.s of interest:
$$\begin{cases} X: \text{ quiz score} \\ Y: \text{ section number} \end{cases}$$

$Y=1$ (Section 1) 10 students

$Y=2$ (Section 2) 20 students

Given quiz statistics.

Quiz average in Section 1: 90

" " " 2: 60

$E[X] = ?$

$Y=1$: $\frac{1}{10}\sum_{i=1}^{10} x_i = 90$

$Y=2$: $\frac{1}{20}\sum_{i=11}^{30} x_i = 60$

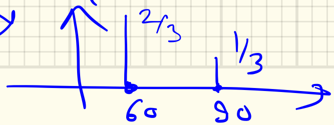
$$E[X] = \frac{1}{30}\sum_{i=1}^{30} x_i = \frac{90 \cdot 10 + 60 \cdot 20}{30} = 70$$

$E[X \mid Y=1] = 90$

$E[X \mid Y=2] = 60$

$E[X \mid Y]$ is a random variable
$$= \begin{cases} 90 & \text{w.p. } \frac{1}{3} \\ 60 & \text{w.p. } \frac{2}{3} \end{cases}$$

pmf:



$$E[E[X \mid Y]] = 90 \cdot \frac{1}{3} + 60 \cdot \frac{2}{3} = 70 = E[X] = \checkmark$$

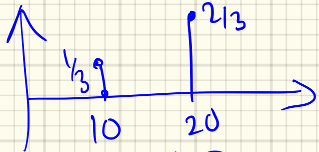
$$\mathrm{Var}(\underbrace{E[X|Y]}_{\text{r.v.}}) = \frac{1}{3}(90 - \underset{\text{mean}}{70})^2 + \frac{2}{3}(60 - \underset{\text{mean}}{70})^2 = 200$$

Say, we are given variances in each section:

$$\begin{cases} \mathrm{Var}(X|Y=1) = 10 \longrightarrow \dfrac{1}{10}\displaystyle\sum_{i=1}^{10}(x_i - 90)^2 = 10 \\ \mathrm{Var}(X|Y=2) = 20 \longrightarrow \dfrac{1}{20}\displaystyle\sum_{i=11}^{30}(x_i - 60)^2 = 20 \end{cases}$$

$$\mathrm{Var}(X|Y) \text{ is an r.v.} = \begin{cases} 10 & \text{w.p. } \frac{1}{3} \\ 20 & \text{w.p. } \frac{2}{3} \end{cases}$$

w/ pmf

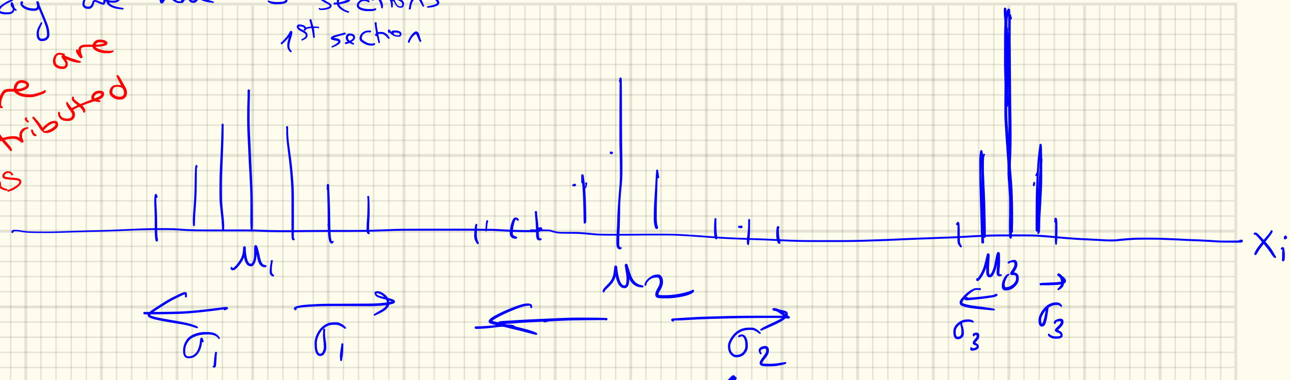


$$E[\mathrm{Var}(X|Y)] = \frac{1}{3}\cdot 10 + \frac{2}{3}\cdot 20 = \frac{50}{3}$$

$$\underset{\text{Total Variance}}{\mathrm{Var}(X)} = \underbrace{E[\mathrm{Var}(X|Y)]}_{\text{AVERAGE variability WITHIN SECTIONS (1)}} + \underbrace{\mathrm{Var}(E[X|Y])}_{\text{VARIABILITY BETWEEN SECTIONS. (2)}} = \frac{50}{3} + 200$$

Say we have 3 sections

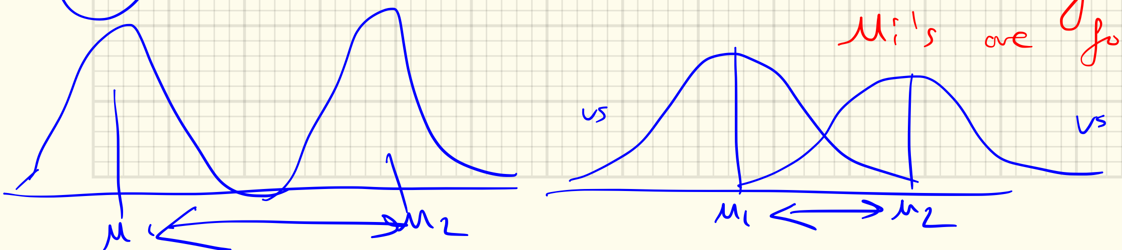
Scores are distributed as pmf.



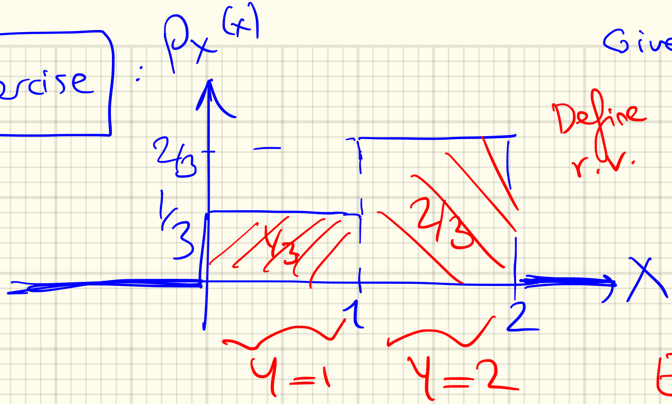
(1) Average of σ_i^2 : Average of variances within each section.

$E[Var(X|Y)]$

(2) $Var(E[X|Y])$: Variability of (μ_1, μ_2, μ_3) = how widely spread the μ_i's are for each section



Exercise:



Given a cont. r.v. X w/ pdf:

Define r.v. $Y = \begin{cases} 1, & 0 \le x \le 1 \quad \text{w.p. } \frac{1}{3} \\ 2, & 1 \le x \le 2 \quad \text{w.p. } \frac{2}{3} \end{cases}$

$\downarrow$

Y is a fn. X.

$E[X \mid Y=1] =$

$E[X \mid Y=2] =$

$Var(X \mid Y=1) =$

$Var(X \mid Y=2) =$

$E(X) = ?$

$Var(E[X \mid Y]) = ?$

Total var. law: $Var(X) = \ldots$

Note: $\underline{Var(X) \text{ directly calculated}}$ from the mathematical expression.

$Var(X) = ?$

Sum of a Random number of independent r.v.s.

N : # stores visited ≥ 0 integer r.v.
X_i : money spent in store i ;

X_i : i.i.d. (indep. identically distrib)
* X_i & N are independent.

N : is a non-negative integer r.v.

$Y = X_1 + X_2 + \cdots + X_N$: sum of money you spent.

Q. $E[Y] = ?$ ✓ $\quad Var(Y) = ?$

$$E[Y \mid N=n] = E[(X_1 + \cdots + X_n) \mid N=n]$$ due indep.
$$= E[X_1 + \cdots + X_n]$$ linearity of exp.
$$= E[X_1] + \cdots E[X_n] = n \cdot E[X]$$

r.v.
$$E[Y \mid N] = N \cdot E[X]$$

b/c it's an r.v. take its E[·].

a number

$$E[E[Y \mid N]] = E[Y] = E[N \, E[X]] = E[N] . E(X).$$

use Total Variance Law :

$$Var(Y) = \underbrace{E[Var(Y|N)]}_{①} + \underbrace{Var(E[Y|N])}_{②}$$

① $Var(Y|N=n) = n \cdot Var(X)$

$\tilde{Y} = X_1 + \cdots + X_n$ due : variance of a sum of n indep. r.v.s . = sum of their variances .

$Var(Y|N) = N \cdot Var(X)$ is an r.v.

$\rightarrow E[Var(Y|N)] = E[N] \cdot Var(X)$.

② $E[Y|N] = N \cdot \underbrace{E[X]}_{\text{a number}}$ (found on prev page)

$\underset{r.v.}{Z} = \overset{const}{a} \underset{r.v.}{N}$

$Var(Z) = a^2 Var(N)$

$Var(E[Y|N]) = (E[X])^2 \cdot Var(N)$

$\rightarrow Var(Y) = E[N] \underbrace{Var(X)}_{} + \underbrace{(E[X])^2 \cdot Var(N)}_{\text{Variability between}}$

~~average of within variability~~ ✓

Variability between .

LIMIT THEOREMS

→ 1) Weak Law of Large Numbers (WLLN) (Averages)

2) Central Limit Theorem (CLT) (Distributions)

1 Want to come up w/ an expected value.

$X_1, X_2, \dots, X_n$; i.i.d.

Sample mean: $\underbrace{M_n}_{} = \dfrac{X_1 + \dots + X_n}{n}$; $n \to \infty$.

Is my sample mean representative?

$\underline{M_n}$ is a random variable.

$M_n \xrightarrow[n\to\infty]{} E[X]$?

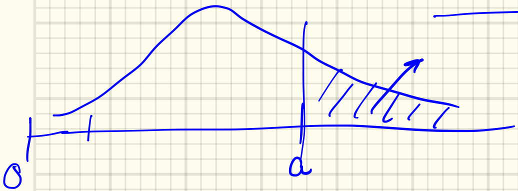
In what sense? ↗ in prob
↘ in distrib.

We'll first study a few tools in order to study these convergences.

Markov Inepuality:

- We have an r.v. X, $X \geqslant 0$; it's discrete r.v. (assume)

$$E[X] = \sum_x \underbrace{x}_{\geqslant 0} \cdot \underbrace{p_X(x)}_{\geqslant 0} \geqslant \sum_{x \geqslant a} x \cdot p_X(x)$$



$$\geqslant \sum_{x \geqslant a} a \cdot p_X(x)$$

$$= a \sum_{x \geqslant a} p_X(x)$$

$$= a\, P(X \geqslant a)$$

Markov Inepuality

$$E[X] \geqslant a \cdot P(X \geqslant a)$$

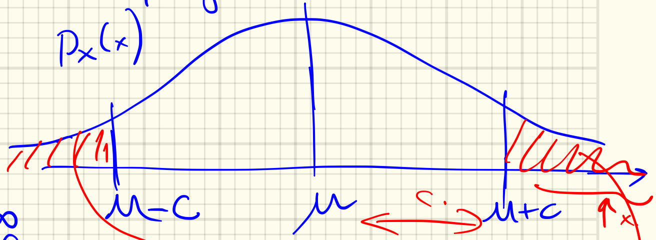
smallness of expected value $\longleftrightarrow$ smallness of probabilities.

rewrite its centered version.

$$E[(X-\mu)^2] \geq a^2 P((X-\mu)^2 \geq a^2)$$

$$\boxed{Var(X) \geq a^2 P((X-\mu)^2 \geq a^2)}$$

Another derivation: X an r.v. w/ finite mean μ & var. σ^2:



$$\sigma^2 = \int_{-\infty}^{\infty} \underbrace{(x-\mu)^2}_{\geq 0} p_x(x)\,dx$$

$$\geq \int_{-\infty}^{\mu-c} \underbrace{(x-\mu)^2}_{c^2} p_x(x)\,dx + \int_{\mu+c}^{+\infty} \underbrace{(x-\mu)^2}_{c^2} p_x(x)\,dx$$

$$\geq c^2 \int_{-\infty}^{\mu-c} p_x(x)\,dx + c^2 \int_{\mu+c}^{\infty} p_x(x)\,dx$$

$$= c^2 P(|X-\mu| \geq c)$$

$$\boxed{\sigma^2 \geq c^2 P(|X-\mu| \geq c)}$$

Tchebysheff Inequality: Require μ & σ^2 of the r.v.

$$\boxed{P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}}$$

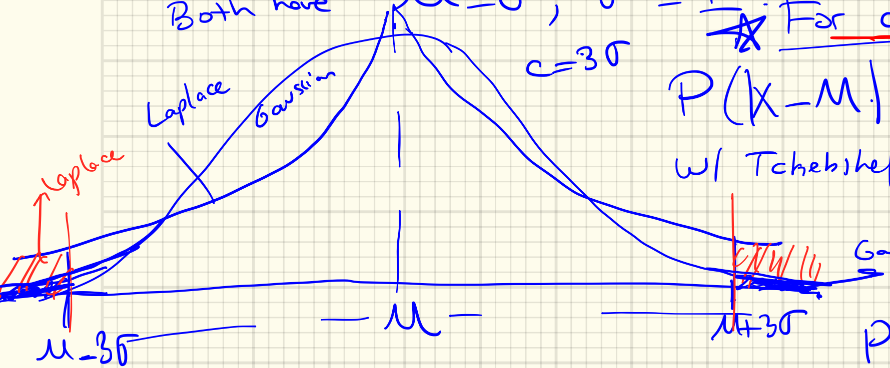
let $c = k\sigma$

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

Ex: $X \sim N(0,1)$, $X \sim \text{Laplace}\left(\frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}\right) \sim \text{Laplace}(\underset{E(X)}{0}, \underset{Var(X)}{1})$

Both have $\mu = 0$, $\sigma^2 = 1$

$c = 3\sigma$



★ For any distrib w/ mean μ & var σ^2

$$P(|X-\mu| > 3\sigma) \leq \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9} = 0.111$$

w/ Tchebyshelf inep.

very conservative bound

For Gaussian, we can calculate an exact value:

$$P(|X-\mu| \geq 3\sigma) = 1 - \int_{-3\sigma}^{3\sigma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2\sigma^2} dx$$

$$= \Phi(\mu-3\sigma) + 1 - \Phi(\mu+3\sigma) = \boxed{0.002}$$

very tight bound.

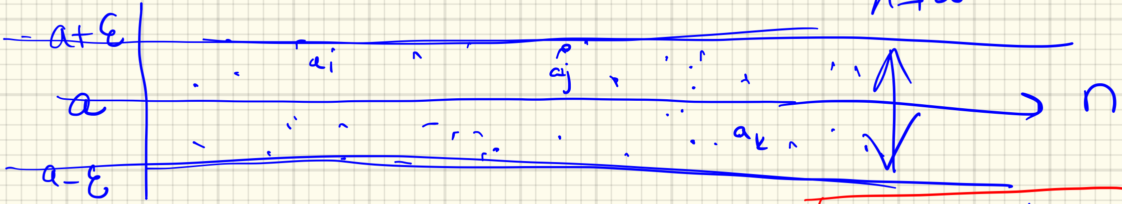
Φ: normal cdf

Convergence:

Given a seq. $a_1, a_2, \ldots a_n \longrightarrow a$

$$\lim_{n\to\infty} a_n = a.$$

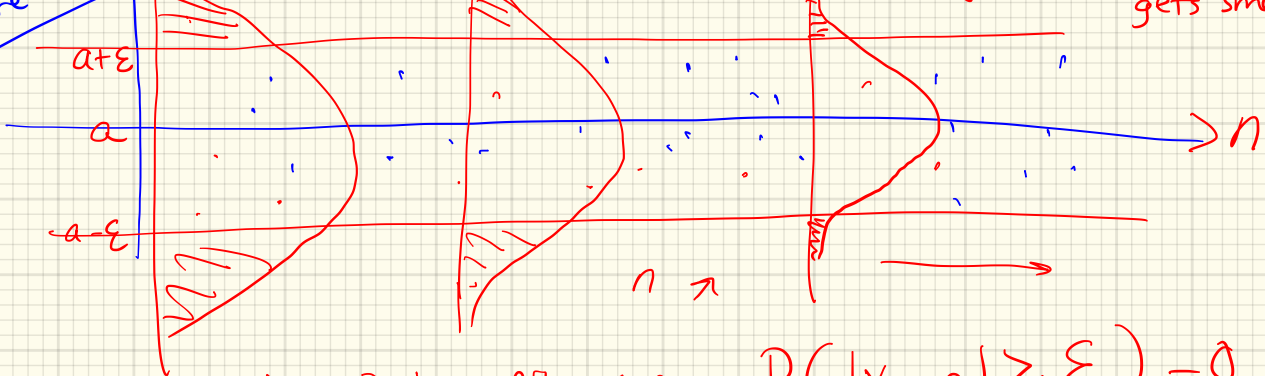
Calculus 101



$$\forall \varepsilon > 0,\ \exists\, n_0 \text{ s.t. } \forall n \geq n_0 \quad \boxed{|a_n - a| \leq \varepsilon.}$$

Now, a_n's are r.v.s

as $n \nearrow$ prob. of falling outside $a \pm \varepsilon$ gets smaller.



Convergence in Probability:

$$\lim_{n\to\infty} P(|X_n - a| \geq \varepsilon) = 0$$

Def (Convergence in Probability)

Sequence of r.vs X_n converges in probability to a number a,

For every $\varepsilon > 0$, $\lim_{n \to \infty} P(|X_n - a| \geq \varepsilon) = 0$

$\equiv$ tail probabilities go to 0 as $n \uparrow \infty$.

Weak Law of Large Numbers (WLLN)

$X_1, X_2, \ldots$ i.i.d. rvs w/ finite mean μ, finite variance σ^2.

Sample mean: r.v. $M_n = \frac{X_1 + \cdots + X_n}{n}$

$$E[M_n] = \frac{E[X_1] + \cdots + E[X_n]}{n} = \frac{n \cdot \mu}{n} = \mu.$$

How big is the variance?

$$\underline{Var(M_n)} = Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2} \cdot n\sigma^2 = \boxed{\frac{\sigma^2}{n}}$$

→ Use Tchebyshev ineq:

$$P\left(|M_n - \mu| \geq \varepsilon\right) \leq \frac{Var(M_n)}{\varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

$$\Rightarrow \quad M_n \xrightarrow[\text{convergence in prob.}]{} \mu.$$

WLLN: $\underbrace{\text{Sample mean}}_{M_n}$ $\underbrace{\text{converges}}_{\text{(in prob.)}}$ to the true mean μ. as $n \to \infty$.

Typical Ex: $\boxed{\text{POLLING}}$: What fraction of the population prefers "something"?

"Soğanlı menemen" vs "Soğansız menemen"

Eggs w/ onions — Eggs w/ onions.

$\boxed{\text{80 million}}$ out of

30 million,

$\boxed{P = \frac{3}{8}}$

437K voted
50.6% soğanlı
49.4% soğansız.

p : fraction of population that prefers smt. → Predict p

ith person randomly polled, $X_i = \begin{cases} p, & \text{Yes (soğanlı)} \\ (1-p), & \text{No (soğansız)} \end{cases}$

Bernouilli r.v. w.p. p

prediction of fraction : $M_n = \dfrac{X_1 + \dots + X_n}{n} \xrightarrow{?} p$

Goal: w/ 95% confidence ≤ 1% error.

Tell me p within 1% error.

(eg. if $p = 0.45$ answer within $(0.44, 0.46)$.

desired accuracy

$$P(|M_n - p| \geq 0.01) \leq 0.05$$

1 - confidence.

desired confidence.

specifications → Specs

Q. How large n, sample size, you should use to satisfy the specs given by the pollsters?

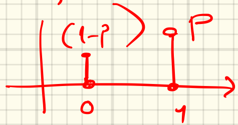
accuracy & confidence.

Note: Importance of sampling uniformly,

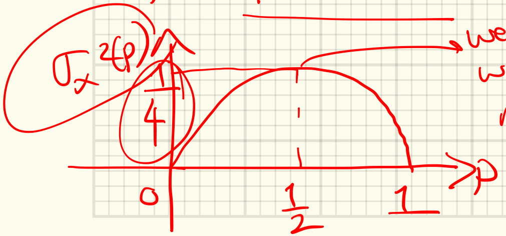
eg. don't sample only from the relatives of a candidate in voting polls !!

Tchebyshelff: $P(|M_n - p| \geq 0.01) \leq \dfrac{\sigma_{M_n}^2}{(0.01)^2} = \dfrac{\sigma_x^2}{n\cdot(0.01)^2} \leq 0.05$

σ_x^2: variance of a Bernouilli r.v.



$\sigma_x^2 = p(1-p)$.



we'll use the worst case σ_x^2

max value $= \frac{1}{4}$

$$\frac{\sigma_x^2}{n\cdot(0.01)^2} \leq 0.05 \qquad \sigma_x^2 = 1/4$$

$$\frac{1}{4n\,10^{-4}} \leq 5\cdot 10^{-2} \Rightarrow n \geq \frac{1}{4\cdot 10^{-4}\cdot 5\cdot 10^{-2}} = 50K$$

If $n = 50{,}000$ (conservative estimate thru Tchebyshev) w/ 1% error

$$P(|M_n - p| \geq 0.01) \leq 0.05$$

50K is not a very practical #. ! What to do?

You can make it less conservative: allow 3% error,

$$P(|M_n - p| \geq 0.03) < 0.05 \rightarrow \frac{1}{(0.03)^2} \rightarrow$$

instead of 1% save you a factor ~10

→ 5K people.