9/6/ 6. Central Limit Theorems

& Inequalities

Central Limit Theorem:

Let $\{X_n\}$ be a sequence of R.V's which are independent, identically distributed with finite variance $[0 < \underbrace{Var(X_n)}_{\sigma^2} < \infty]$, & common mean μ.

Let $S_n = X_1 + X_2 + \cdots + X_n$, then $\forall x \in R$

$$\underset{n\to\infty}{Lt} P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right\} = \underset{n\to\infty}{Lt} P\left\{\frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \le x\right\}$$

$E(S_n) = E(X_1 + X_2 + \cdots + X_n) = n\mu$

$Var(S_n) = Var(X_1 + X_2 + \cdots + X_n) = n\sigma^2$

$S.D(S_n) = \sigma\sqrt{n}$

$$\underset{n\to\infty}{Lt} P\left\{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

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

$E(\bar{x}) = \frac{1}{n}\left(E(X_1) + E(X_2) + \cdots + E(X_n)\right) = \frac{n\mu}{n} = \mu$

$Var(\bar{X}) = \cancel{V E(X_1)}\ Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$

$= \frac{1}{n^2}\left(Var(X_1) + Var(X_2) + \cdots + Var(X_n)\right)$

$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

Scanned with CamScanner

If $n \geq 30$ central limit theorem suits well.

Ex: Let $X_1, X_2, \ldots X_n$ be iid RV's with

independent & identically distributed

common $B(\alpha, \beta)$ distribution then

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

then
$$\frac{S_n - n\left(\alpha/(\alpha+\beta)\right)}{\sqrt{n \cdot (\alpha+\beta)^2/(\alpha+\beta+1)}} \longrightarrow Z \sim N(0,1)$$

Ex: A bank serves customers standing in a queue one by one. Suppose that X_i for customer i is the service time with $E(X_i) = 2$, $Var(X_i) = 1$

Assume that services are independent.

Let Y be the total time bank spends on 50 service customers

Find $P(90 < Y < 110)$

sol
$$Y = X_1 + X_2 + \cdots + X_{50}$$

$$\mu = 2, \quad \sigma = 1$$

$$\frac{Y - n\mu}{\sigma\sqrt{n}} = \frac{Y - 50(2)}{(1)\sqrt{50}} = \frac{Y - 100}{\sqrt{50}}$$

$P(90 < Y < 1\cancel{1}0) = P(\text{total service time for 50 customers is b/w 90 \& 100})$

$$P\left(\frac{90-100}{\sqrt{50}} < Z < \frac{1\cancel{1}0-100}{\sqrt{50}}\right)$$

$$P\left(\frac{-10}{5\sqrt{2}} < Z < \frac{10}{5\sqrt{2}}\right)$$

$$P(-1.41 < Z < 1.41)$$

$$= \phi(1.41) - \phi(-1.41)$$

$$= \cancel{0.5 - 0.07927}$$

$$= 0.92073 - 0.07927$$

$P(90 < Y < 110) = 0.84146$

Ex: $Y \sim \text{Bernoulli } \cancel{\text{Binomial}}\ (n = 20, p = \frac{1}{2})$

$E(X) = p = \frac{1}{2}$

$\sigma^2 = pq = \frac{1}{4}$

$n\mu = np = 10$

$\sqrt{n}\,\sigma = \sqrt{20} \cdot \frac{1}{2} = \sqrt{5}$

$$P(8 < Y < 10) = P\left(\frac{8-10}{\sqrt{5}} < Z < \frac{10-10}{\sqrt{5}}\right)$$

$$= P\left(\frac{-2}{\sqrt{5}} < Z < 0\right)$$

$$= \phi(0) - \phi(-0.89)$$

$$= 0.3145$$

Cachy shwarz Inequality:

X & Y are two Random Variables then

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

Ex: $E(X) = 1$, $E(Y) = 2$ $Var(X) = 4$, $Var(Y) = 1$

$$E(X^2) = (E(X))^2 + Var(X)$$
$$= 1+4$$
$$= 5$$

$$E(Y^2) = (2)^2 + 1$$
$$= 5$$

$$(E(XY))^2 \leq 5\cdot 5$$
$$(E(XY))^2 \leq 25$$
$$E(XY) \leq 5$$

Proof: Statement continues

Equality holds

$(E(XY))^2 = E(X^2)\cdot E(Y^2)$ iff $X = aY$ i.e.,

X is a scalar multiple of Y.

Let us define R.V $U = (X - sY)^2$

$$E(U) \geq 0$$

Consider

$$g(s) = E((X-sY)^2) = E(X^2 - 2sXY + s^2Y^2)$$

$$= E(X^2) - 2s\,E(XY) + s^2 E(Y^2)$$

$$g(s) = s^2 E(Y^2) - 2s\,E(XY) + E(X^2)$$

$$= s^2\left(\sqrt{E(Y^2)}\right)^2 - 2\cdot s\cdot \frac{E(XY)\sqrt{E(Y^2)}}{\sqrt{E(Y^2)}} + \frac{(E(XY))^2}{E(Y^2)}$$

$$+ E(X^2) - \frac{(E(XY))^2}{E(Y^2)}$$

$$g(s) = \left(s\sqrt{E(Y^2)} - \frac{E(XY)}{\sqrt{E(Y^2)}}\right)^2 + E(X^2) - \frac{(E(XY))^2}{E(Y^2)}$$

$g(s)$ is positive if

$$E(X^2) - \frac{(E(XY))^2}{E(Y^2)} \geq 0$$

$$\therefore E(X^2)\cdot E(Y^2) - (E(XY))^2 \geq 0$$

$$\therefore E(X^2)\cdot E(Y^2) \geq (E(XY))^2$$

$$(E(XY))^2 \leq E(X^2)\cdot E(Y^2)$$ [Hence proved]

Scanned with CamScanner

9/6/21

Cauchy schwaz's Inequality

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

→ Correlation coefficient $|\rho| \leq 1$

Proof Let X & Y be two Random Variables with means μ_1 & μ_2

Let $U = X-\mu_1$
$V = Y-\mu_2$

$$(E(UV))^2 \leq E(U^2) \cdot E(V^2)$$

$$\left(E((X-\mu_1)(Y-\mu_2))\right)^2 \leq \left(E(X-\mu_1)^2\right)\left(E(Y-\mu_2)^2\right)$$

$$\left(E(XY) - \mu_2 E(X) - \mu_1 E(Y) + \mu_1\mu_2\right)^2$$

$$\leq E(X-\mu_1)^2 \cdot E(Y-\mu_2)^2$$

$$\left(E(XY) - E(X) \cdot E(Y)\right)^2 \leq E(X-\mu_1)^2 \cdot E(Y-\mu_2)^2$$

$$(Cov(X,Y))^2 \leq \sigma_x^2 \sigma_y^2$$

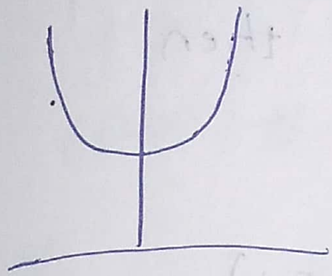
$$\left(\frac{Cov(X,Y)}{\sigma_x \sigma_y}\right)^2 \leq 1$$

$$|\rho_{xy}| \leq 1$$

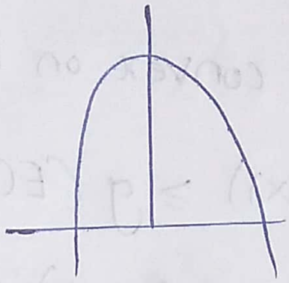
Scanned with CamScanner

~~Jensen's~~

Convex



Concave



Convex:

A function $g: R \to R$ is convex on $[a,b]$ if for $x, y \in [a,b]$ and each $\lambda \in [0,1]$ we have $g(\lambda x_1 + (1-\lambda)\cancel{y}x_2) \leq \lambda g(x_1) + (1-\lambda) g(\cancel{y}x_2)$

↓ Function value

↳ Interpolation

Concave:

$$g(\lambda x_1 + (1-\lambda)x_2) \geq \lambda g(x_1) + (1-\lambda) g(x_2)$$

if $f''(x) > 0$ it is convex

$f''(x) < 0$ it is concave

Ex: x^2, e^x → convex function

$\log x, -x^2, -e^x$ → convex function.

Jensen's Inequality:

Suppose X is a R.V $P(a \le x \le b) = 1$

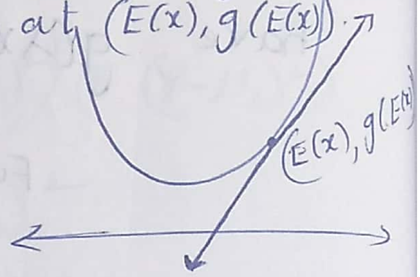
if $g: R \to R$, convex on $[a,b]$ then

$$E(g(x)) \ge g(E(x))$$

For concave $E(g(x)) \le g(E(x))$

Proof:

Let $L(x) = Ax + B$ which is tangent to the given curve such that the tangent line meets the graph at $(E(x), g(E(x)))$.



$g(x) \ge L(x)$, $x \in [a,b]$

$$E(g(x)) \ge E(L(x))$$

$$\ge E(Ax+B)$$

$$\ge A E(x) + B$$

$$\ge L(E(x))$$

$$E(g(x)) \ge g(E(x))$$

$L(E(x)) = g(E(x))$ since it is point common to curve & tangent.

Ex:

$g(x) = x^2$

$g'(x) = 2x$

$g''(x) = 2 > 0$

convex $E(g(x)) \ge g(E(x))$

$$E(x^2) \ge (E(x))^2$$

Ex 2 $g(x) = -x^2$

$g'(x) = -2x$

$g''(x) = -2 < 0$ (Concave)

$E(-x^2) \leq -(E(X))^2$

Ex 3

$g(x) = e^x$ $g'(x) = e^x$ $g''(x) = e^x > 0$

Convex $E(e^x) \geq (e^{E(x)})$

Law of Large Numbers: Sample size is Large

Trials are independent.

Weak Law of Large Numbers:

Let $X_1, X_2, \ldots X_n$ be the R.V's (R.V X_i's are iid)

then $\bar{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ will converge

to $\mu = E(X_i)$ as $n \to \infty$

$\bar{X} = \mu = E(X_i)$

Ex:

No. of Tosses	No. of Heads	Probability of Heads
4	1	25%
100	64	64%
1000	582	58.2%
10000	4989	49.89%

T H
0 1

$E(X) = \frac{1}{2}$

As no. of trials increases sample avg tends to actual expectation.

sample avg → very small +ve real num

$$\underset{n\to\infty}{Lt}\ P\big((\bar{X}-\mu) > \epsilon\big) = 0$$

$\mu \downarrow E(X)$

$\bar{X}$ & μ gets closer as $n\to\infty$

(or)

Proof: $\underset{n\to\infty}{Lt}\ P\big((\bar{X}-\mu) < \epsilon\big) = 1$

By using chebyshev's

$$P\big((\bar{X}-\mu) > \epsilon\big) \le \frac{\sigma^2}{n\epsilon^2}$$

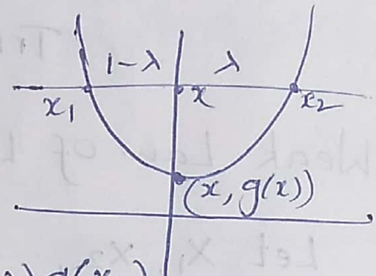
$$= 0 \quad \text{as } n\to\infty$$

11/6/21

Jensen's Inequality:

Any point on the line x_1, x_2 is given as $x = (1-\lambda)x_2 + \lambda x_1$

$$x = \lambda x_1 + (1-\lambda)x_2$$



$$g(\lambda x_1 + (1-\lambda)x_2) \le \lambda g(x_1) + (1-\lambda) g(x_2)$$

$$g(E(X)) \le E(g(X)) \rightarrow \text{Convex}$$

$$g(E(X)) \ge E(g(X)) \longrightarrow \text{Concave}$$

Strong Law of Large Numbers:

Let $X_1, X_2, \ldots X_n$ be iid R.V

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}, \quad E(X_i) = \mu$$

$$P\Big(\underset{n\to\infty}{Lt}\big((\bar{X}_n - E(X)) = 0\big)\Big) = 1$$

$\underset{n\to\infty}{Lt}\ (\bar{X} - E(X)) = 0$ or $\underset{n\to\infty}{Lt}\ \bar{X} = E(\bar{X})$ with probability 1, almost sure event

Conditions:

→ R.V's should be independent

→ $E(X_i) = \mu$ (finite)

strong Law implies weak Law but the converse is not true.

Ex: $X_i \sim$ Bernoulli's with p

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \qquad E(X) = P$$

$$\underset{n\to\infty}{Lt}\ \bar{x} \to P$$

or $\underset{n\to\infty}{Lt} (\bar{x} - P) = 0$ with probability 1

$$P\left(\underset{n\to\infty}{Lt}\left(\bar{x} - E(x) = 0\right)\right) = 1$$

→ Weak & strong Law ~~does~~ are not applicable with distribution having infinite mean or mean doesn't exist

Central Limit Theorem:

Ex: Let X_i ; $i = 1, 2, \dots 10$ which are iid R.V's each being uniformly distributed over $(0,1)$

Find $P(S > 7) = ?$ $\qquad P(X_1 + X_2 \dots + X_{10} > 7) = ?$

$E(X_i) = \frac{1}{2}$

$Var(X) = \frac{1}{12}$

Sol Mean $E(S) = E(X_1 + X_2 + \dots + X_{10}) = ~~\bot~~$

$= n\,E(X) = 10 \cdot \frac{1}{2}$

$= 5$

$Var(S) = Var(X_1 + X_2 + \dots + X_{10}) = Var X_1 + Var X_2 + \dots + Var X_{10}$

$= n\sigma^2 = \frac{10}{12} \implies \sigma_S = \sqrt{\frac{10}{12}}$

$$P(S>7) = P\left(\frac{S-5}{\sqrt{\frac{10}{12}}} > \frac{7-5}{\sqrt{\frac{10}{12}}}\right)$$

$$= P(Z > 2.19)$$

$$= 1 - P(Z < 2.19)$$

$$= 1 - \phi(2.19)$$

$$P(S>7) = 0.0143$$