July 27

correlation recap

uncorrelated $\not\Rightarrow$ independent

$\Leftarrow$

$\Rightarrow Y_1 = X$ corr $=1$

$X \sim U[-1, 1]$

$Y_2 = -X$ corr $= -1$

$\text{cov}(X, Y) = 0$

$Y = X^2$

useful ineq. & LLN

Ex ① fair coin, toss N time, probability of seeing more than $(\frac{3}{4}N)$ heads?

$\#H = X \sim \text{Binomial}(N, \frac{1}{2})$

$$\mathbb{P}(X \geq \tfrac{3}{4}N) = \sum_{k=(3/4)N}^{N} \binom{N}{k} \frac{1}{2^k} \quad \Leftarrow$$

Markov's Inequality

for a nonnegative r.v. X $\left(X(\omega) \geq 0, \forall \omega \in \Omega \right)$

$$\mathbb{P}(X > c) \leq \frac{\mathbb{E}(X)}{c} \quad \forall c > 0$$

proof: consider, $\forall a > 0$, r.v. $Y_a = \begin{cases} 0 & X < a \\ a & X \geq a \end{cases}$, $X \geq 0$

$$Y_a \leq X \Rightarrow \mathbb{E}(Y_a) \leq \mathbb{E}(X)$$

$$\Rightarrow 0 \cdot \mathbb{P}(X < a) + a \cdot \mathbb{P}(X \geq a) = a\mathbb{P}(X \geq a) \leq \mathbb{E}(X)$$

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

∎

$$\mathbb{P}(\# H > \tfrac{3}{4}N) \le \frac{\mathbb{E}(\# H)}{\tfrac{3}{4}N} = \frac{\tfrac{1}{2}N}{\tfrac{3}{4}N} = \frac{2}{3} \Leftarrow$$

Generalized Markov's Inequality

X is an Arbitrary r.v. w/ finite mean $|\mathbb{E}(X)| < \infty$

$\forall . c, r > 0$

$$\boxed{\mathbb{P}(|X| > c)} \le \frac{\mathbb{E}(|X|^r)}{c^r}$$

proof: indicator function $I(z) = \begin{cases} 1 & \text{if } z \text{ is true} \\ 0 & \text{if } z \text{ is false} \end{cases}$

$$\mathbb{E}(I(z)) = \mathbb{P}(z \text{ is true})$$

$\forall c, r > 0.$

$$|X|^r \geq |X|^r I\{|X| \geq c\} \geq c^r I\{|X| \geq c\}$$

$$\mathbb{E}(|X|^r) \geq \mathbb{E}(c^r I\{|X| \geq c\}) = c^r \mathbb{E}(I\{|X| \geq c\})$$

$$= c^r \mathbb{P}(|X| \geq c)$$

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}(|X|^r)}{c^r} \qquad \blacksquare$$

when $|x| \geq c$: $|x|^r \geq c^r$ and $I\{|x| \geq c\} = 1$

when $|x| < c$ $I\{|x| \geq c\} = 0$

so $|x|^r I\{|x| \geq c\} \geq c^r I\{|X| \geq c\}$

Chebyshev Inequality

If X is a r.v. $\mu = \mathbb{E}(X)$ $\sigma^2 = \text{var}(X)$ then

$$\mathbb{P}(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2} \quad \forall c > 0$$

proof. new r.v. $Y = (X-\mu)^2$ $Y \geq 0$

$$\mathbb{E}(Y) = \mathbb{E}((X-\mu)^2) = \text{var}(X) = \sigma^2$$

$$\mathbb{P}(Y \geq c^2) \leq \frac{\mathbb{E}(Y)}{c^2}$$

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

$$(X-\mu)^2 \geq c^2 \iff |X-\mu| \geq c \quad c>0$$

$\blacksquare$

$\mathbb{P}(\#H \geq \frac{3}{4}N)$ $\quad\quad X : \#H$ over N toss

$$\mathbb{P}(X \geq \frac{3}{4}N) = \mathbb{P}(X - \frac{N}{2} \geq \frac{N}{4})$$

$$\leq \mathbb{P}(|X - \frac{N}{2}| \geq \frac{N}{4})$$

$\mathbb{P}(X - \frac{N}{2} \geq \frac{N}{4})$

$\mathbb{P}(-X + \frac{N}{2} \geq \frac{N}{4})$

$$\leq \frac{\mathrm{var}(X)}{(\frac{N}{4})^2}$$

$$\mathrm{var}(X) = N \cdot p(1-p) = \frac{N}{4}.$$

$$\mathbb{P}(X \geq \frac{3}{4}N) \leq \frac{4}{N}$$

Ex ② unfair coin $\quad \mathbb{P}(X_i = H) = p$ unknown.

each toss X_i

$$\mathbb{E}(\#H) = \sum_{1}^{N} \mathbb{E}(X_i) = Np$$

$$\mathrm{var}(aX) = a^2 \mathrm{Var}(X)$$

find an "estimator" $\hat{p}$

$$\mathrm{var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$$

$$= \sigma^2 + \sigma^2 = 2\sigma^2$$

$$\hat{p} := \frac{1}{N}(X_1 + X_2 + \cdots + X_N)$$

$$\mathbb{E}(\hat{p}) = p \qquad \mathrm{var}(\hat{p}) = \mathrm{Var}\left(\frac{1}{N}(X_1 + \cdots + X_N)\right)$$

$$= \frac{1}{N^2} \cdot N\sigma^2 = \frac{\sigma^2}{N}$$

$$\mathbb{P}(|\hat{p} - p| \geq \varepsilon) \leq \frac{\mathrm{var}(\hat{p})}{\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2} \quad \Leftarrow \text{chebyshev.}$$

$$\boxed{\mathbb{P}(|\hat{p}-p| \geq \varepsilon) \leq \frac{\sigma^2}{N\varepsilon^2}}$$

to ensure $\mathbb{P}(|\hat{p}-p| \geq \varepsilon) \leq \delta$

$$\frac{\sigma^2}{N\varepsilon^2} \leq \delta \quad \Rightarrow \quad N \geq \frac{\sigma^2}{\varepsilon^2 \delta}$$

the weak law of large number.

a sequence of r.v. $X_1, X_2, \cdots, X_n$ are iid, $\mathbb{E}(X_i) = \mu$

$var(X_i) = \sigma^2$. we define sample mean as

$$M_n = \frac{X_1 + \cdots + X_n}{n} \qquad \mathbb{E}(M_n) = \mathbb{E}\left(\frac{\sum^n X_i}{n}\right) = \mu$$

$$Var(M_n) = \frac{Var(\sum_i X_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

WLLN says. $\forall \varepsilon > 0$

$$\mathbb{P}(|M_n - \mu| \geq \varepsilon) \to 0 \quad \text{as} \quad n \to \infty$$

$$\mathbb{P}(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$$

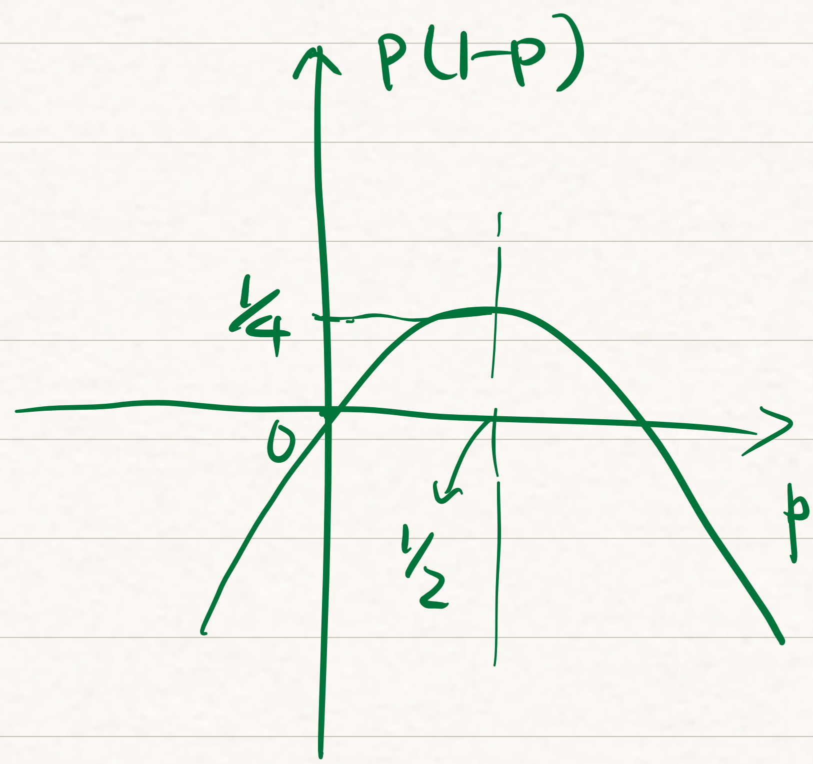
Ex ③ polling, p: fraction of people support A

interview "randomly" n voters.

M_n to be fraction of people out n voting A

$Var(X_i) = \sigma^2 = p(1-p)$ $\qquad var(M_n) = \frac{p(1-p)\cdot n}{n^2}$

$$\mathbb{P}(|M_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$$



$p(1-p) \leq \frac{1}{4}$

So. $\mathbb{P}(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$

$$P(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$$

$|M_n - p| \leq 0.01$

95% confident. of accuracy within 0.01 of p

how many people to interview

$$\mathbb{P}(|M_n - p| \geq 0.01) \leq \frac{1}{4n(0.01)^2} \leq 0.05 = 5\%$$

$n \geq 50,000$

$$\left(1 - \mathbb{P}(|M_n - p| \leq \varepsilon)\right) \leq \frac{1}{4n\varepsilon^2}$$