recap

Variance

how far the set of possible values is spread out from the expected value.

$$Var(X) = E[(X-\mu)^2] \quad \text{where } \mu = E[X]$$

$$Var(X) = E[X^2] - E[X]^2$$

if X, Y, Z independent ... $Var(X+Y+Z) = Var(X) + Var(Y) + Var(Z)$

always ... $Var(cX) = c^2 Var(X)$

$Var(X+c) = Var(X)$

Concentration Inequalities

Markov's : (X is nonnegative)

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

Chebyshev's : $P(|X-\mu| \geq c) \leq \frac{Var(X)}{c^2}$

$\mu = E[X]$

follows from applying Markov's on $Y = (X - E[X])^2$

WLLN

if we observe an outcome several times, the average of the observations converges very close to $E[X]$

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0$$

as $n \to \infty$

recap
Intuitive Markov's

suppose we have a field of flowers, avg height = 4 ft

$P(\text{height of a flower} \geq 10 \text{ ft}) \leq \frac{4}{10}$ by Markov's

Assume this is not true...

Total height of n flowers = $4n$

if more than 4/10 of flowers had height ≥ 10 ft,

= # of tall flowers

the total height of the tall flowers $> \frac{4}{10} \cdot n \cdot 10$.

total height of this subpopulation already greater than total height of population. not possible if all heights nonnegative.

* avg isn't the exact same as expected value, so only treat this exercise as a way to see the general intuition behind Markov's.

recap **Using Chebyshev to Estimate Bias of a Coin**

Estimate bias $\underbrace{p}_{\text{true mean}}$ empirically : $\underbrace{\hat{p} = S_n/n}_{\text{sample mean}}$

S_n = # of heads
n = tosses

By LLN, p should be close to $\hat{p}$ if n is large ← but how large should n be?

Say we want to estimate p with 95% confidence and tolerate error of 0.1
we ask... how many times should I toss st. there is at least a 95% chance that our observed $\hat{p}$ is within 0.1 of p.

$$S_n = X_1 + X_2 + \dots + X_n \qquad X_i = \begin{cases} 1 & \text{if toss } i = H \\ 0 & \text{otherwise} \end{cases}$$

$$E[\hat{p}] = \frac{1}{n}\left(E[X_1] + \dots E[X_n]\right) = \frac{1}{n} \cdot n \cdot p = p$$

Recall Chebyshev's : $P[|X - \mu| \geq \alpha] \leq \dfrac{Var(X)}{\alpha^2}$ Set X to be $\hat{p}$.

recap

$S_n = X_1 + X_2 + \dots + X_n$ $\quad X_i = \begin{cases} 1 & \text{if toss } i = H \\ 0 & \text{otherwise} \end{cases}$

$E[\hat{p}] = \frac{1}{n}(E[X_1] + \dots E[X_n]) = \frac{1}{n} \cdot n \cdot p = p$

Recall Chebyshev's : $P[|X - \mu| \geq \alpha] \leq \frac{Var(X)}{\alpha^2}$ set X to be $\hat{p}$.

$Var(X+Y) = Var(X) + Var(Y)$ if independent

$Var(cX) = c^2 Var(X)$

need $Var(\hat{p})$

$\hat{p}$ is a scaled sum of n indicator variables so

$$Var(\hat{p}) = \frac{1}{n^2} Var(S_n) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X_i) = \frac{p(1-p)}{n}$$

Cool fact. $p(1-p) \leq 1/4$ so we usually use $1/4$.

Plug into Chebyshev's

$$P[|\hat{p} - p| \geq 0.1] \leq \frac{p(1-p)}{n} \cdot \frac{1}{0.1^2}$$

probability that we DO deviate a lot.

$$\frac{p(1-p)}{0.01n} \leq 0.05$$

95% confidence

$$\frac{1}{4} \cdot \frac{1}{0.01} \cdot \frac{1}{0.05} \leq n$$