AMAT 362 - PROBABILITY for STATISTICS

Lecture 9 : Mean, Median & Mode

μ or $\bar{x}$ = $E(X)$

med ← 50th percentile

most probable value = $\underset{x}{\arg\max}\ p(x)$

⋮

Lecture 12 : INTRO TO VARIANCE & STANDARD DEVIATION

→ $E(XY) = E(X)E(Y)$

only for independent X & Y

→ $Var(X+Y) = Var(X) + Var(Y)$

MOTIVATION

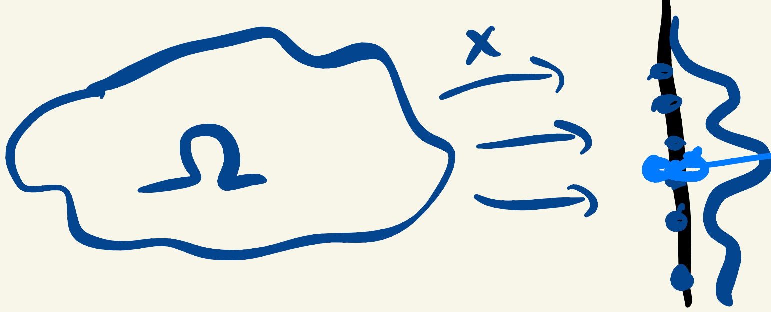
↓

PREDICTION

Recall If X is a RV, then for g a function

$$E(g(X)) = \sum_{x} g(x) P(X=x)$$

Expected Loss for different functions



Pick a value b

Let $L(x,b)$ = money lost if predict b and outcome is x

3 Different Loss Functions

$$L(x,b) = \begin{cases} 0 & \text{if } x = b \\ \$ & \text{if } x \neq b \end{cases}$$

I. All or Nothing Loss

$$L(x,b) = \begin{cases} 0 & x = b \\ 1 & x \neq b \end{cases}$$

BIG IDEA

Toggle b to minimize $E[L(x,b)]$

$$E[L(x,b)] = 0 \cdot P(X=b) + 1\, P(X \neq b)$$
$$= 1 - P(X=b)$$

Minimize by picking b = Most probable outcome $= \text{Mode}(X)$

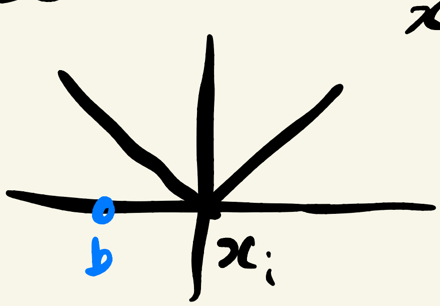
II. Penalty = Distance

$$L(x,b) = |x - b|$$

MINIMIE $E[L(x,b)]$ by taking derivative wrt b & set to 0.

$$E[L(x,b)] = \sum_{x_i} |x_i - b|\, P(X = x_i)$$

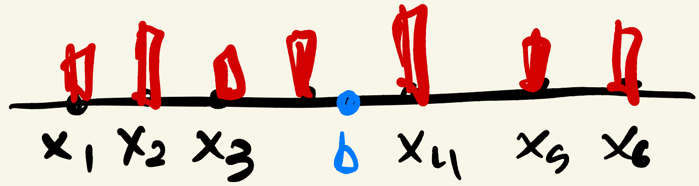
$$\frac{d}{db} E[L(x,b)] = \sum_{x_i} \frac{d}{db}\Big[|x_i - b|\Big] P(X = x_i)$$



Recall $\frac{d}{db}|x_i - b| = \begin{cases} -1 & b < x_i \\ 1 & b > x_i \end{cases}$

$$\Rightarrow \sum_{x_i < b} (+1) P(X = x_i) - \sum_{x_i > b} P(X = x_i)$$

If $x_k < b < x_{k+1}$



$$= \underbrace{\sum_{x_i \le x_k} P(X = x_i)} - \sum_{x_i > x_k} P(X = x_i)$$

$$= P(X \le x_k) \underbrace{- P(X > x_k)}$$

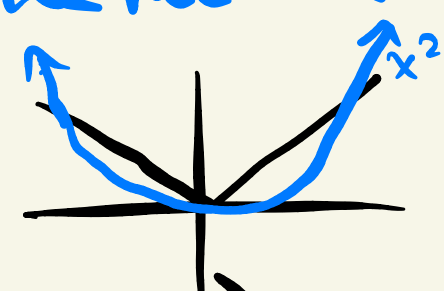
$$= \qquad - (1 - P(X \le x_k))$$

$$\frac{\partial}{\partial b} E = 2P(X \le x_k) - 1 \underset{\text{To minimize}}{=} 0$$

choose $x_k < b < x_{k+1}$ s.t. $P(X \le x_k) = \frac{1}{2}$

Choose b to be the median

III Loss = Squared Distance



$$L(x, b) = (x - b)^2$$

$$E[L(x,b)] = E((X-b)^2) = E(X^2 - 2bX + b^2)$$

$$= \underbrace{E(X^2) - 2bE(X) + b^2}$$

Take deriv wrt b

$$\frac{d}{db} \searrow -2E(X) + 2b \underset{\text{To minimize}}{=} 0 \Rightarrow b = E(X) = \mu$$

Guessing $b = \mu = E(X)$ is best!

UPSHOT: For squared error predicting $b = E(X)$ is best.

Q: What is $E[(X-\mu)^2]$?

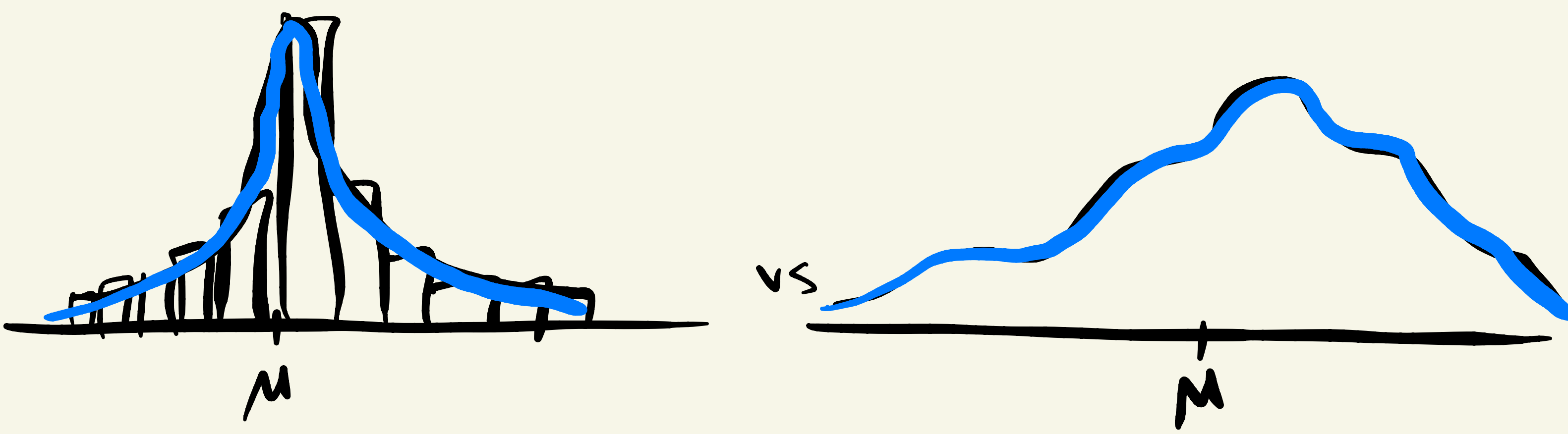
VARIANCE & STANDARD DEVIATION

Def If X is a RV then

the VARIANCE of X $\boxed{VAR(X) = E((X-\mu)^2)}$

the STANDARD DEVIATION $\boxed{SD(X) = \sqrt{VAR(X)}}$

$\sigma_X = SD(X)$ $\qquad$ $\sigma_X^2 = Var(X)$



LOW VARIANCE $\qquad$ HI VARIANCE

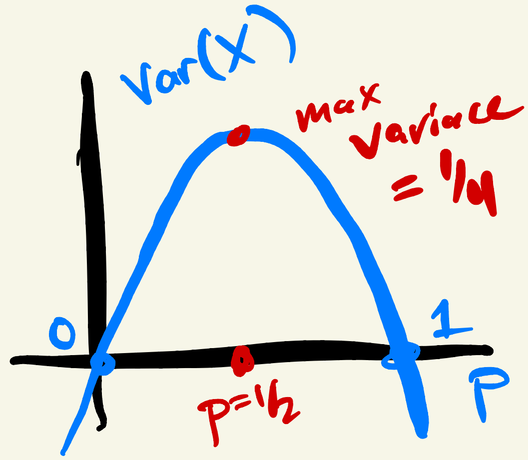
ALT. FORMULA: $\boxed{VAR(X) = E(X^2) - E(X)^2}$

$$E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2$$

$$\Longrightarrow \; = E(X^2) - \mu^2$$

Indicator (Bernoulli) R.V.

$$X = \begin{cases} 1 & \text{w/ prob } p \\ 0 & \text{w/ prob } q = 1-p \end{cases}$$



$$E(X) = 0 \cdot q + 1 \cdot p = p$$

$$E(X^2) = 0^2 \cdot q + 1^2 \cdot p = p$$

$$\Rightarrow Var(X) = E(X^2) - E(X)^2 = p - p^2$$
$$= p(1-p)$$
$$= pq$$

Ex Flipping a biased coin $\quad p(\text{Heads}) = p$, $\quad p(\text{Tails}) = q$

BINOMIAL (= Sum of Independent Identically Distributed indicator RVs)

$$X = X_1 + X_2 + \dots + X_n \quad \text{where } X_j = \begin{cases} 1 & \text{if } j\text{th try is a success, prob} = p \\ 0 & \text{o.w., prob} = q \end{cases}$$

$$E(X) = np$$

$$E(X^2) = ??? = \sum_{k=0}^{n} k^2 \cdot \underbrace{P(X=k)}_{\binom{n}{k} p^k q^{n-k}}$$

By next page
$$\boxed{Var(X) = npq}$$

$$\boxed{\text{Rmk } E(X^k) = k\text{th moment of } X}$$

TWO IMPORTANT FACTS

1) **If** X & Y are independent R.V.s
($i.e.$ $P(X=x, Y=y) = P(X=x)P(Y=y)$)
$(P(A \text{ and } B) = P(A)P(B))$

then $E[X \cdot Y] = E[X]E[Y]$

$$\sum_{x,y}{}' x \cdot y\, P(X=x, Y=y) \overset{\text{By indep}}{=} \sum_{x,y}{}' xy\, P(X=x)P(Y=y)$$
$$= \sum{}' x P(X=x) \cdot y P(Y=y)$$
$$= \underbrace{\left(\sum_x{}' x P(X=x)\right)}_{E(X)} \underbrace{\left(\sum_y{}' y P(Y=y)\right)}_{E(Y)}$$

2) **If** X & Y are independent RVs
then $Var(X+Y) = Var(X) + Var(Y)$

Application

$X = X_1 + \cdots + X_n$ sum of IID Bernoulli = Binomial

$$Var(X) = \boxed{Var(Binom(n,p)) = npq}$$
$$= Var(X_1) + \cdots + Var(X_n)$$
$$= pq + \cdots + pq$$
$$= npq$$