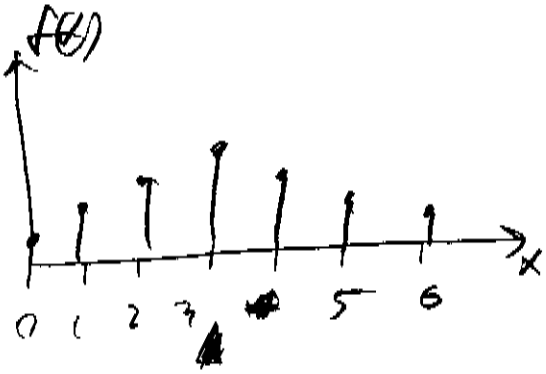
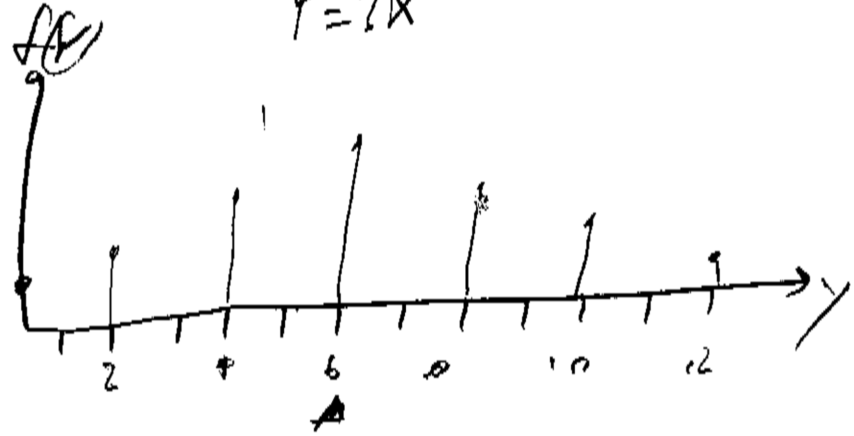
Lecture 14 10/30/14

$X \sim \text{Binom}(6, \frac{1}{2})$



$E(X) = 6 \cdot \frac{1}{2} = 3$

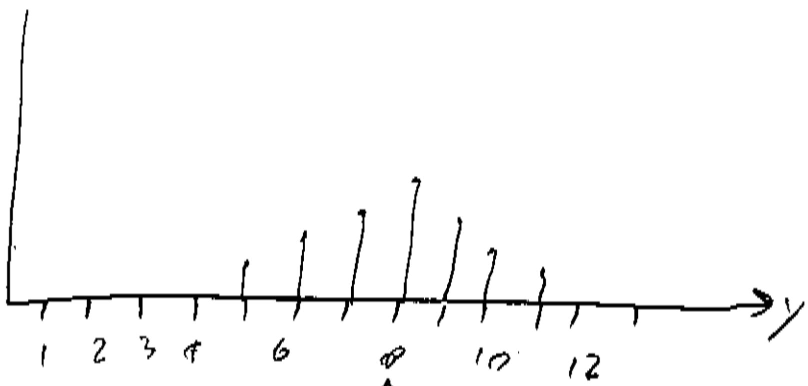
$Y = 2X$



$E(2X) = 2E(X) = 2 \cdot 6 \cdot \frac{1}{2} = 6$

Pivot is doubled since all chances are doubled

$Y = X+5$

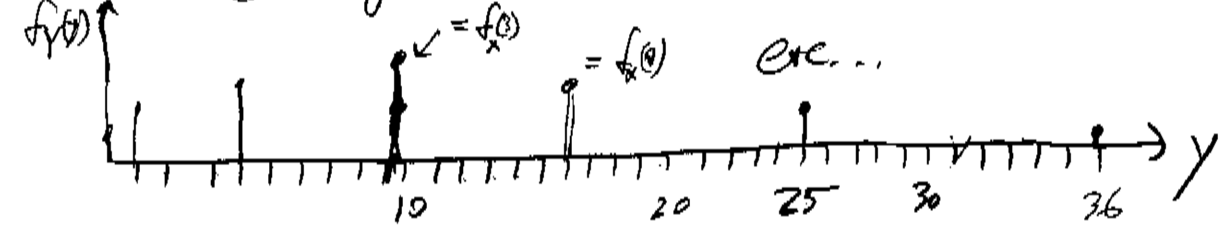


$E(X+5) = E(X)+5 = 3+5 = 8$

Everything shifts over by same amount, so does pivot...

How about $Y = g(X)$. g is an arbitrary function. Takes any $x \in Supp(X)$ and transforms it into $g(x)$. $Supp(Y) = \{g(x) : x \in Supp(X)\}$

Let $X \sim Bin(6, \frac{1}{2})$, $Y = g(X) = X^2$



Where does plot go? Not so simple, need to go by definition

$$E(Y) = \sum_{y \in Supp(Y)} y\, f_Y(y) \overset{f_Y(y) = f_X(x)}{=} \sum_{x \in Supp(X)} g(x) f_X(x) = \sum_{x \in Supp(X)} x^2 f_X(x) = \ldots \text{ need to compute } =$$

$$= 0 f(0) + 1 f(1) + 4 f(2) + 9 f(3) + 16 f(4) + 25 f(5) + 36 f(6)$$

$= 17.5$ no tricks, just a spreadsheet...

In general, $E(Y) = E[g(X)] = \sum_{x \in Supp(X)} g(x) f(x)$ No simple rule! Have to compute!

$E(X)$ is what's expected over many iid trials. It is a special "parameter" and it has a special name $\mu := E[X]$. Greek letters are used for the "truth" in the model. "Truth" and model are not "real." The greeks invented geometry & logic and basic math.

Sometimes we are also interested in how far away the x is from μ.

For example, $X \sim Rademacher$, $Y = 10X$

$E[X] = 0$, $E(Y) = 10E(X) = 10 \cdot 0 = 0$ ✓

$\mu_X = \mu_Y = 0$. But $X \overset{d}{\neq} Y$. Y is more "dispersed" than X.

Any given realization of Y is much further from μ_Y than any given realization of X in its distance from μ_X! How far do we expect a random X to be from its theological center? $e(X, \mu_X)$ where e is an error function.

What should we pick for e? ① e should be symmetric $e(3,4) = e(4,3)$, $e(x,y) = e(y,x)\ \forall x,y \in \mathbb{R}$ ② Negative error doesn't make sense so $e(x,y) \geq 0\ \forall x,y \in \mathbb{R}, x \neq y$
③ and if the values are the same, there should be no error $e(x,x) = 0\ \forall x \in \mathbb{R}$

Plenty of choices e.g. $e(x,\mu) := |x-\mu|$, $e(x,\mu) = (x-\mu)^2$, ... etc.

In Statistics, we choose $e(x,\mu) := (x-\mu)^2$ → "square loss" or "L2" error.
Since e is a diff. function ①

② $e(x,\mu) = (x-\mu)^2$ function appears all over the place notably e.g. in the density of the normal distribution.

③ makes ~~sense~~ that you should penalize errors further away more heavily
e.g. $e(3,4) = 1$, $e(3,5) = 4$, $e(3,13) = 100$. Being further away is worse.

→ But at the end of the day, it is our choice, and it is fundamentally arbitrary.

$$L := \overbrace{(X-\mu)^2}^{g(X)}$$

a function of X

squared error loss of X

$\sigma^2 := Var(X) :=$

$$E(L) = E\left[(X-\mu)^2\right] = \sum_{x \in Supp(X)} (x-\mu)^2 f(x)$$

special name: **Variance!** From the word "vary": How much does it vary from what you expect?

$$X \sim Bernoulli(0.3) \Rightarrow \mu = E(X) = 0.3$$

$$Var(X) = E\left[(X-0.3)^2\right] = \sum_{x=0}^{1} (x-0.3)^2 f(x) = (0-0.3)^2 f(0) + (1-0.3)^2 f(1)$$
$$= (0.3)^2\, 0.7 + (0.7)^2\, 0.3$$
$$= 0.09 \cdot 0.7 + 0.49 \cdot 0.3$$
$$= 0.063 + 0.147$$
$$= \boxed{0.21} \leftarrow \text{expected square error loss}$$

$X \sim Bernoulli(p)$ X has only one parameter $\Rightarrow$ σ^2 is a function of p and constants.

$$\sigma^2 := E(X-\mu)^2 = E(X-p)^2 = \sum_{x=0}^{1} (x-p)^2 f(x) = (0-p)^2 f(0) + (1-p)^2 f(1)$$
$$= p^2(1-p) + (1-2p+p^2)p$$
$$= \cancel{p^2} - \cancel{p^3} + p - 2p^2 + \cancel{p^3}$$
$$= p - p^2 = \boxed{p(1-p)}$$

$$X \sim \begin{cases} 35 & \text{w.p. } \frac{1}{38} \\ -1 & \text{w.p. } \frac{37}{38} \end{cases} \qquad \mu = -0.053$$

(from last time)

$$Var[X] = E(X-\mu)^2 = (35 - -0.053)^2 \frac{1}{38} + (-1 - -0.053)^2 \frac{37}{38}$$
$$= 33.207$$

What are the units? $L = (X-\mu)^2 \quad \underbrace{(\$ - \$)^2}_{\$} \Rightarrow \$^2$

Variance is in the underlying units squared... Not so useful...

Hard to think in $\$^2$, hr^2, mi^2, etc...

Need a way to "standardize" the units of dispersion.

$$\sigma := SD(X) := \sqrt{Var(X)}$$ the "standard deviation"

by taking the square root, $\sqrt{\$^2} = \$$, the units of $SD(X)$

is the same as the units of X, $Range(X, p)$, $E(X)$. $SD[X] = \sqrt{33.207\,\$^2} = \$5.76$

$\sigma^2 = Var[X] =$

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

$\Rightarrow E(X^2) = \sigma^2 + \mu^2$ a convenient formula for $g(X) = X^2$.

is called the 2nd moment

$E(X-\mu)^2$

is called the 2nd central moment

$E(X^{17})$ is the 17th moment

$E(X-\mu)^{17}$ " central moment

$E[X]$ is the first moment

$E[X-\mu]$ is the first central moment

Skew... moments etc...

~~Asymmetry? central moments: skewness~~ ~~a measure of symmetry~~

Moments are in general difficult to find e.g.

$X \sim \text{Binom}(n, p)$

$$\sigma^2 = \text{Var}(X) = E(X^2) - \underbrace{\mu^2}_{E(X)} \qquad \mu^2 = n^2p^2$$

$$\sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

$$x^2 \binom{n}{x} = x^2 \frac{n!}{x!(n-x)!} = x \frac{n!}{(x-1)!(n-x)!} = nx \frac{(n-1)!}{(x-1)!(n-x)!} = nx \binom{n-1}{x-1}$$

$$np \sum_{x=1}^{n} x \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x}$$

let $y = x-1 \Rightarrow x = y+1$

$$\Rightarrow np \sum_{y=0}^{n-1} (y+1) \binom{n-1}{y} p^y (1-p)^{n-y-1}$$

let $m = n-1$

$$np \sum_{y=0}^{m} (y+1) \binom{m}{y} p^y (1-p)^{m-y}$$

$$np\Big(\underbrace{\sum_{y=0}^{m} y \binom{m}{y} p^y (1-p)^{m-y}}_{\underbrace{E(X)}_{mp}} + \underbrace{\sum_{y=0}^{m} \binom{m}{y} p^y (1-p)^{m-y}}_{\underbrace{\sum f(x)}_{1}} \Big)$$

$$\Rightarrow np(mp+1) = np((n-1)p+1)$$
$$= np(np - p + 1)$$
$$= n^2p^2 - np^2 + np$$

$$\cancel{n^2p^2} - np^2 + np - \cancel{n^2p^2} = np - np^2 = \boxed{np(1-p)} = \sigma^2$$

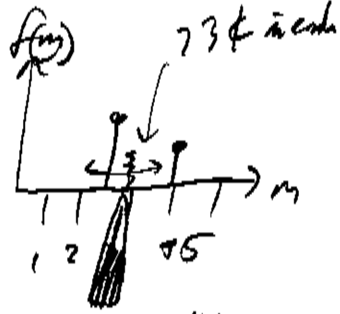
$$\Rightarrow \boxed{SD(X) = \sqrt{np(1-p)}} = \sigma$$

Easier way ... mgf's

Note $E[X^2] = \sigma^2 + \mu^2 = np(1-p) + n^2p^2 \neq \mu^2$ In general, $E[g(X)] \neq g(\mu)$ 7

Ans for question in class... next class...

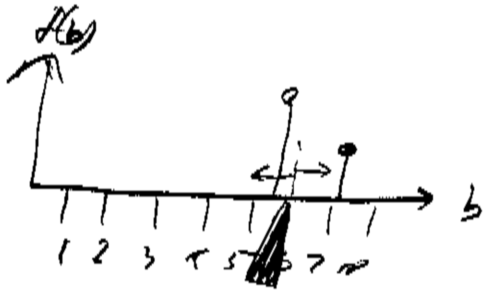
Back to Uber example $M = \$0.40 \cdot W$ $Supp(M) = \{2.40, 4.00\}$



$E(M) = \$2.88$

$Var[M] = 0.538 \$^2 \Rightarrow SD(M) = \$0.733$

$B = \$3 + M \Rightarrow E(B) = \$5.88$



$Var(B)$? What about dispersion?

$= Var(3+M)$ It makes sense that $Var[3+M] = Var(M)$ dispersion doesn't change on shift

Proof $\sigma^2 = Var[X]$

$$Var(X+c) = E(X+c-\mu)^2 = E(X+c)^2 - (E(X+c))^2$$
$$= E[X^2 + 2cX + c^2] - (\mu + c)^2$$
$$= EX^2 + 2c\mu + c^2 - (\mu^2 + 2\mu c + c^2)$$
$$= \underbrace{EX^2 - \mu^2}_{\sigma^2} \cancel{+2c\mu} \cancel{+c^2} \cancel{-2\mu c} \cancel{-c^2}$$

Should make common sense $\Rightarrow$ $SD[X+c] = SD(X)$ just take square's.

$$Var[W] = (6-7.2)^2 \cdot 0.7 + (10-7.2)^2 \cdot 0.3 = 3.36 \Rightarrow SD[W] = 1.83$$

Note: $0.4 \cdot 1.83 = 0.73$