November 3, 2016

$W \sim \begin{cases} 7 \text{ min wp. } 0.7 \\ 12 \text{ min wp. } 0.3 \end{cases}$ – Custom built r.v. It's not Bern, Hyper, ...

$E[W] = (7 \cdot 0.7) + (12 \cdot 0.3) = 8.5$ min – Long term property of a r.v.

Uber charges \$0.40/min

$B = \$0.40 \times W$ ← Charge per minute. This is a new r.v.

$B \sim \begin{cases} \$2.80 \text{ wp } 0.7 \\ \$4.80 \text{ wp } 0.3 \end{cases}$ If you take the cab many times, you will, on avg spend \$3.12 per ride

$E[B] = (\$2.80 \times 0.7) + (\$4.80 \times 0.3) = \$3.12$

$E[B] = \underbrace{\$0.40 \times E[W]}$

Does this relationship hold, generally?

r.v. X, r.v. $Y = aX$. Does $E[Y] = E[X]$?

$E[X] := \int_\Omega X(\omega) P(d\omega)$

If X is discrete, then $Supp[X] = \{x_1, x_2, x_3, \ldots\}$

$\Omega = A_1 \cup A_2 \cup A_3 \cup \ldots \Rightarrow X(\omega) \in Supp[X] \; \forall \omega \in \Omega$ by def.

$= \int_{\{\omega : X(\omega) = x_1\}} X(\omega) P(d\omega) + \int_{\{\omega : X(\omega) = x_2\}} X(\omega) P(d\omega) + \ldots$

$= x_1 \int P(d\omega) + x_2 \int P(d\omega) + \ldots$

$= x_1 P(X = x_1) + x_2 P(X = x_2) + \ldots$

$= E[X] = \sum_{x \in Supp[X]} x\, p(x)$

$E[g(X)] := \int_\Omega g(X(\omega)) P(d\omega)$

$= \int_{\{\omega : X(\omega) = x_1\}} g(X(\omega)) + \int_{\{\omega : X(\omega) = x_2\}} g(X(\omega)) P(d\omega) + \ldots$

$= g(x_1) P(X = x_1) + g(x_2) P(X = x_2) + \ldots$

☆ $\boxed{E[g(X)] := \sum_{x \in Supp[X]} g(x)\, p(x)}$

If $Y = \underbrace{aX}_{g(x)}$,

$$E[Y] = \sum_{x \in Supp[X]} aX\, p(x) = a \sum_{x \in Supp[X]} x\, p(x) = a E[X]$$

☆ so $E[aX] = aE[X]$

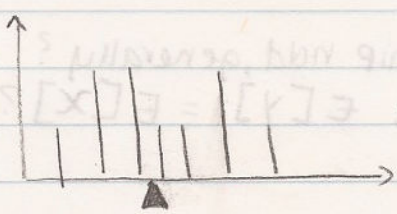
> Base Fare = \$3

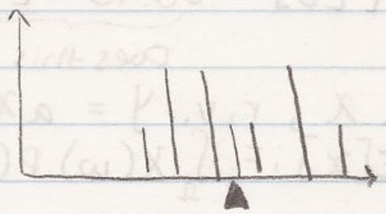
Calculate r.v. for base fare + price per mile.

$T = B + \$3$

new r.v., a function of B.

$E[T] = E[B] + \$3 = \$6.12$





when you add a constant, everything shifts to the right by constant c.

$$Y = X + c$$

$$E[Y] : E[X+c] = \sum_{x \in Supp[X]} (x+c)\, p(x) = \sum x\, p(x) + \sum c\, p(x)$$

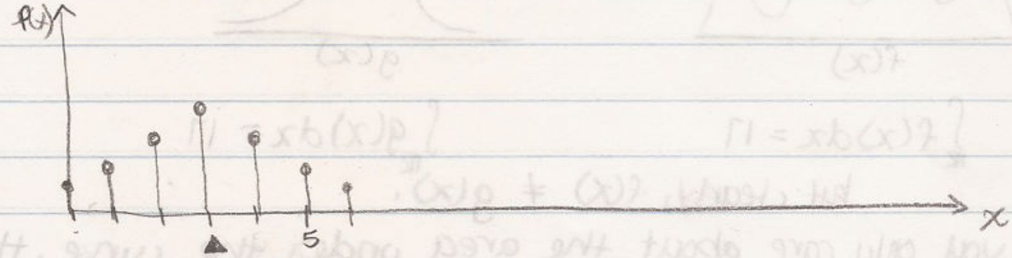
$$= E[X] + c \underbrace{\sum p(x)}_{=1} = E[X] + c$$

$$Y = aX + c \Rightarrow E[Y] = aE[X] + c$$

so $E[aX + c] = aE[X] + c$

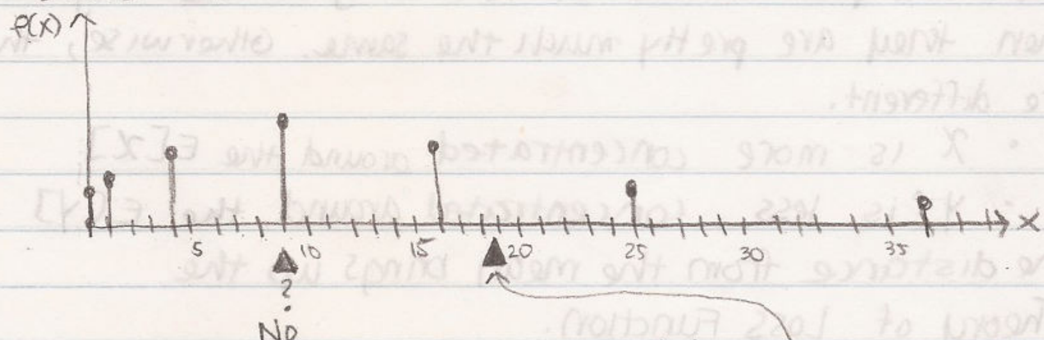
> $X \sim \text{Binom}(6, \frac{1}{2})$

$E[X] = 6 \cdot \frac{1}{2} = 3$



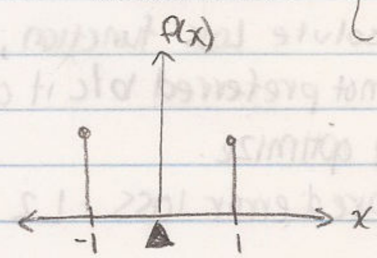
> $Y = X^2$

$E[X] = ?$



$$E[Y] = \sum x^2 p(x) = \sum x^2 \binom{6}{x}\left(\frac{1}{2^6}\right) = 17.5$$

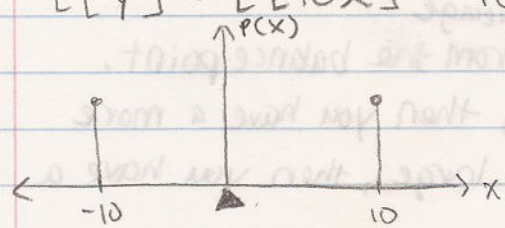
> $X \sim \text{Rademacher} := \begin{cases} 1 & \text{wp } \frac{1}{2} \\ -1 & \text{wp } \frac{1}{2} \end{cases}$



$E[X] = (1 \cdot \frac{1}{2}) - (1 \cdot \frac{1}{2}) = 0$

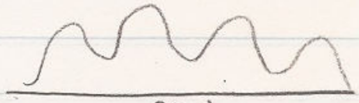
$Y = 10X$

$E[Y] = E[10X] = 10 \cdot E[X] = 10 \cdot 0 = 0$

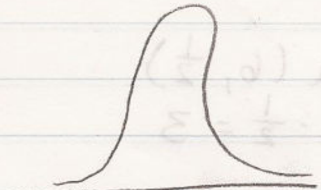


X and Y are not the same r.v.'s. They are not equal in distribution. PMF's are not the same, but if you did a lot of X and Y experiments, on average, you'd get 0 for both.

This is similar to ...



f(x)



g(x)

$\int_{\mathbb{R}} f(x)dx = 17$ $\qquad$ $\int_{\mathbb{R}} g(x)dx = 17$

but clearly, $f(x) \neq g(x)$.

If you only care about the area under the curve, then $f(x)$ is the same as $g(x)$. Similarly, if you only care about the probabilities of X and Y in the long run, then they are pretty much the same. Otherwise, they are different.

- X is more concentrated around the $E[X]$
- Y is less concentrated around the $E[Y]$

The distance from the mean brings up the Theory of Loss Function.

something that comes up expected value

$e(x, \mu) = x - \mu$ } Not a loss function b/c you can't have a cost of $-\$1$.

$e(x, \mu) = |x - \mu|$ } This is a loss function! "Absolute Loss function, L1 loss" But not preferred b/c it can only optimize.

$e(x, \mu) = (x-\mu)^2$ } "Squared error loss, L2 loss"

New random variable

> $L := (X - \mu)^2$

$E[L]$ tells us how far we are on average from the balance point.

If it is something small, then you have a more concentrated r.v. If it's large, then you have a less concentrated r.v.

> Variance of a r.v. X

$$\ast \quad \underset{\text{sigma}}{\sigma^2} := \underset{\text{variance}}{\mathrm{Var}[X]} = \underset{\text{Expectation}}{E}[L] = E[(X-\mu)^2] = \sum_{x \in \mathrm{Supp}[X]} (x-m)^2 \, p(x)$$

loss function: $L = (X-\mu)^2$

> Finding the variance of

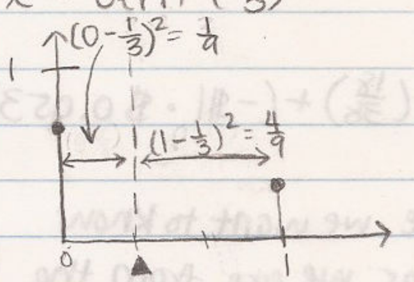
$$X \sim \text{Rademacher} := \begin{cases} 1 & \text{wp } \frac{1}{2} \\ -1 & \text{wp } \frac{1}{2} \end{cases} \quad \text{(supp[X])}$$

$$\mathrm{Var}[X] = ((-1)-0)^2\left(\tfrac{1}{2}\right) + (1-0)^2\left(\tfrac{1}{2}\right)$$
$$= 1 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2}$$
$$= 1$$

> Finding the variance of

$$Y = 10X$$
$$\mathrm{Var}[Y] = ((-10)-0)^2\left(\tfrac{1}{2}\right) + (10-0)^2\left(\tfrac{1}{2}\right)$$
$$= 100 \cdot \tfrac{1}{2} + 100 \cdot \tfrac{1}{2}$$
$$= 100$$ ← The "artistic" difference, the concentration.

> $X \sim \mathrm{Bern}\left(\frac{1}{3}\right)$



What's the long run average?

$$E[X] = \tfrac{1}{3}$$

How concentrated are we around the expectation?

$$\mathrm{Var}[X] = \left(1-\tfrac{1}{3}\right)^2\left(\tfrac{1}{3}\right) + \left(0-\tfrac{1}{3}\right)^2\left(\tfrac{2}{3}\right)$$
$$= \tfrac{4}{9} \cdot \tfrac{1}{3} + \tfrac{1}{9} \cdot \tfrac{2}{3}$$
$$= \tfrac{6}{27} = \tfrac{2}{9} = 0.259$$

> $X \sim \mathrm{Bern}(p)$

$$E[X] = p$$
$$\mathrm{Var}[X] = (0-p)^2(1-p) + (1-p)^2(p)$$
$$= p^2(1-p) + p(1-p)^2$$
$$= (1-p)(p^2 + (1-p)p)$$
$$\mathrm{Var}[X] = p(1-p)$$

> Roulette: Bet on lucky #7

$$X_7 \sim \begin{cases} \$35 & wp\ \frac{1}{38} \\ -\$1 & wp\ \frac{37}{38} \end{cases}$$

$\mu = -\$0.053$

This is less concentrated b/c you would have to go all the way to $x=35$. Variance should be bigger.

$Var[X_7] = (\$35 \cdot -\$0.053)^2(\frac{1}{38}) + (-\$1 \cdot \$0.053)^2(\frac{37}{38})$

$= 33.207\ \$^2$

$\sqrt{Var[X_7]} = \$5.29$

> Bet on black

$$X_B \sim \begin{cases} \$1 & wp\ \frac{18}{38} \\ -\$1 & wp\ \frac{20}{38} \end{cases}$$

$\mu = -\$0.053$

This is more concentrated

$Var[X_B] = (\$1 \cdot \$0.053)^2(\frac{18}{38}) + (-\$1 \cdot \$0.053)^2(\frac{20}{38})$

$= 0.997\ \$^2$

because we want to know how far we are from the balance point, but it makes no sense so we find the square root!

$\sqrt{Var[X_B]} = \$1.00$

$$\underbrace{\sigma := SE[x] := \sqrt{Var[x]}}= \sqrt{\sigma^2}$$

"Standard Error" } Makes a variance into a standard unit.
"Standard Deviation }

Law of large # $\left\{ \begin{array}{l} \overline{X_7} \longrightarrow -\$0.053 \\ \overline{X_B} \longrightarrow -\$0.053 \end{array} \right.$

Which one gets to \$0.053 faster? x_B, because the smaller the variance, the more concentrated it is. The more concentrated, the faster it'll get to the \$0.053.