Lecture 19 Math 341 11/3/16

$W \sim \begin{cases} 7 \text{ min } w.p\ 0.7 \\ 12 \text{ min } w.p\ 0.3 \end{cases}$ $E[W] = 8.5 \text{ min}$

Uber Charges \$0.80/min

$B = \$0.80/\text{min} \cdot W$ } a transformation of a r.v.

$B \sim \begin{cases} \$2.80 \text{ w.p } 0.7 \\ \$4.80 \text{ w.p } 0.3 \end{cases}$ $E[B] = \$2.80 \cdot 0.7 + \$4.80 \cdot 0.3 = \$3.12$

figured out this PMF through common sense... Math 821 covers the theory to do this unconventionally

Is there an easier way?

Note that $B = g(W) = 0.8W$ ← B is a function of another variable. the function is simple: it's a "scale": $g(x) = a f(x)$ s.t. $a \in \mathbb{R}$ g is a scaling function.

Is $E(B)$ related to $E(W)$ through g?

Beyond scope of course... "FAKE PROOF"

$E(X) := \int_{\Omega} X(\omega) P(d\omega)$ A "Lebesgue" integral on a reasonable function

If X is discrete $Supp(X) = \{x_1, x_2, x_3, \dots\}$

$X(\omega) \in Supp(X)$ $\forall \omega \in \Omega$ by def.

Recall $\int_1^3 f(x)\,dx = \int_1^2 f(x)\,dx + \int_2^3 f(x)\,dx$

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

Motivates the def

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

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

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

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

$$= \sum_{x \in S_{g(X)}} g(x) p(x)$$

What if $Y = \underbrace{aX}_{g(X)}$ scaling just like $B = 0.4w$

$$E(Y) = E(aX) = \sum_{x \in S_{g(X)}} ax\, p(x) = a \sum_{x \in S_X} x\, p(x) = aE[X]$$

by def'n

So .. $E(B) = 0.4 E(w) = \$3.12$

there is a base fee ... $\$3$

$T = B + \$3$

$E(T)$?

$Y = X + c$

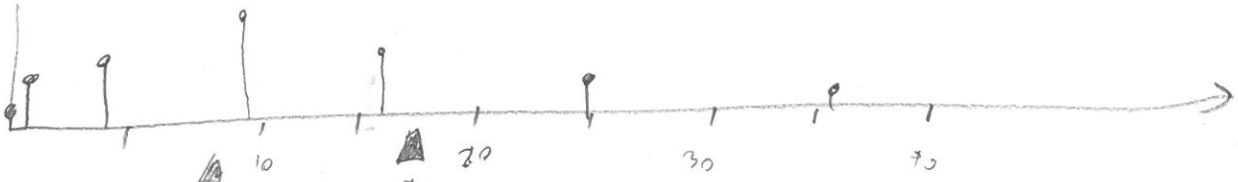
$$E(T) = E(X + c) = \sum_{x \in S_X} (x + c) p(x) = \sum x\, p(x) + \sum c\, p(x) = E(X) + c \sum p(x) = E(X) + c$$

$Y = aX + c \Rightarrow E(Y) = aE(X) + c$

$E(Y) = E(X) + \$3 = \$3.12 + 3 = \$6.12$

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

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



? No!

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

gotta do it...

Note: $E[g(X)] \neq g(E(X))$

However... $Quant[g(X), p] = g\left(Quant[X, p]\right)$

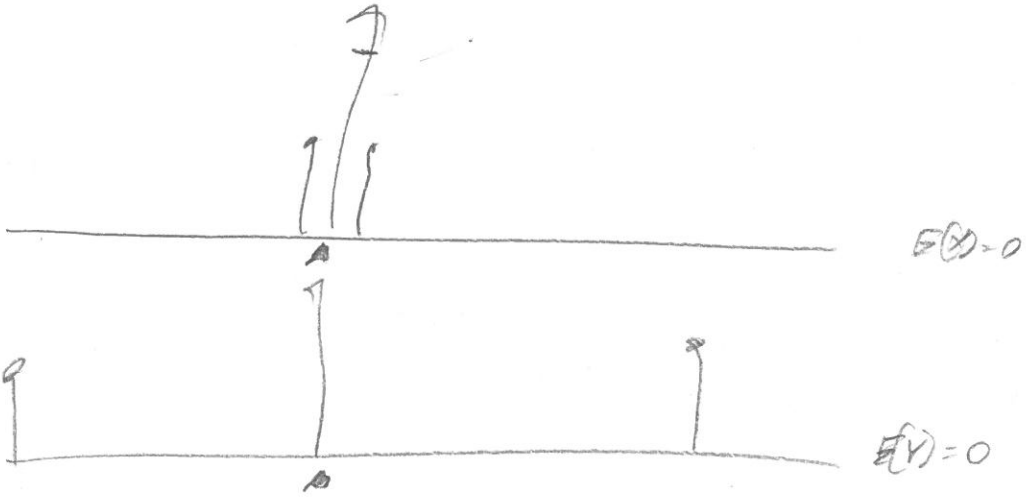
Why? $Q[X, p] = \min\{x : F(x) \geq p\}$

$Q[g(X), p] = \min\{g(x) : P(X \leq x) \geq p\}$

$P(g(X) \leq g(x)) \geq p$

$X \sim \text{Rademacher} := \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$ $\quad E(X) = (1)\cdot\frac{1}{2} + (-1)\cdot\frac{1}{2} = 0$

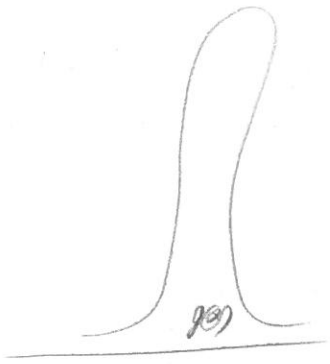
$Y = 10X \Rightarrow E(Y) = 10E(X) = 10 \cdot 0 = 0$



pmf's

$E(X) = E(Y) \not\Rightarrow P_X(x) = P_Y(y)$

just as shown as ...



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

$\int_{\mathbb{R}} g(x)\,dx = 17$ but $f(x) \neq g(x)$

Of course many different shapes can yield the same area!
f is more "diffuse" or "spread" or "varied" than g

We need a metric to capture part of this difference between X, Y. r.v.'s

How about something to do with how far the support is away from zero
In X, close to zero; in Y, far from zero. Y more "dispersed".

"error function", "loss function"

↓

$e(X, \mu) = X - \mu$ but it depends on if it's ⊕ or ⊖

$e(X, \mu) = |X - \mu|$ bad for taking derivative! Hard to prove things...

$e(X, \mu) = (X - \mu)^2$ "squared error loss" or "L2 error".

let $L := (X - \mu)^2$ (r.v.'s)

$E[L]$ is what? The expected sqd error distance from μ_X.

It answers "how far on avg is a realization of X from its own mean"?

$$E[L] = E[g(X)] = \sum_{x \in Supp(X)} g(x)\, p(x) = \sum_{x \in Supp(X)} (x - \mu)^2 p(x)$$

e.g.

$X \sim$ Rademacher $\Rightarrow \mu = E(X) = 0$

$$E[L] = ((-1) - (0))^2 p(-1) + ((1) - (0))^2 p(1) = 1 \cdot 0.5 + 1 \cdot 0.5 = \boxed{1}$$

$Y = 10X \Rightarrow \mu = 0$

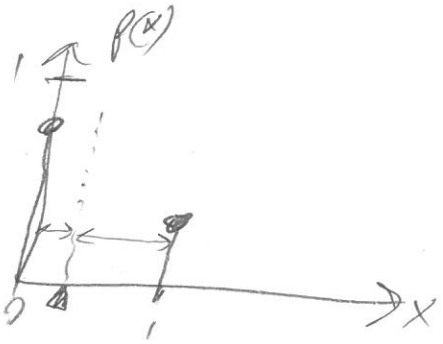
$$E[L] = (-10 - 0)^2 p(-1) + (10 - 0)^2 p(1) = 100 \cdot 0.5 + 100 \cdot 0.5 = \boxed{100}$$

we did our job.. This func so useful has its own name

Variance

$$\sigma^2 := Var(X) := E[(X - \mu)^2]$$

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



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

$$\text{Var}(X) = \left(0 - \tfrac{1}{3}\right)^2 \tfrac{2}{3} + \left(1 - \tfrac{1}{3}\right)^2 \tfrac{1}{3}$$

$$= \tfrac{1}{9} \cdot \tfrac{2}{3} + \tfrac{4}{9} \cdot \tfrac{1}{3} = \tfrac{6}{27} = \tfrac{2}{9} = .259$$

general formula?

$X \sim \text{Bern}(p)$

$$\text{Var}(X) = ((0) - (p))^2 (1-p) + ((1) - (p))^2 p$$

$$= p^2(1-p) + (1-p)^2 p$$

$$= (1-p)\left(p^2 + \underbrace{(1-p)}^{p-p^2} p\right)$$

$$= \boxed{p(1-p)}$$

Roulette: Bet on lucky #7

$$X_7 \sim \begin{cases} \$35 & \text{w.p. } \frac{1}{38} \\ -\$1 & \text{w.p. } \frac{37}{38} \end{cases} \qquad \mu = -\$0.053$$

$$\text{Var}(X_7) = \left(\$35 - -\$0.053\right)^2 \tfrac{1}{38} + \left(-\$1 - -\$.053\right)^2 \cdot \tfrac{37}{38}$$

$$= 33.207\ \$^2$$

" " Bet on Black

$$X_B \sim \begin{cases} \$1 & \text{w.p. } \frac{18}{38} \\ -\$1 & \text{w.p. } \frac{20}{38} \end{cases} \qquad \mu = -\$0.053$$

$$\text{Var}(X_B) = \left(\$1 - -\$0.053\right)^2 \tfrac{18}{38} + \left(-\$1 - -\$0.053\right)^2 \tfrac{20}{38}$$

$$= 0.997\ \$^2$$

$\bar{X}_7 \to \mu$

$\bar{X}_6 \to \mu$

which goes faster?

they are with lowest variance...

(we will prove why later)

Wait! $\$^2$ ... has no meaning!

Easier way to solve $\sqrt{\$^2} = \$$ ← nice and interpretable

let $\sigma := SE(X) := \sqrt{Var(X)} = \sqrt{\sigma^2}$

"Standard error" or "Standard deviation"

$\sigma_7 = \$5.79$, $\sigma_6 = \$1.00$

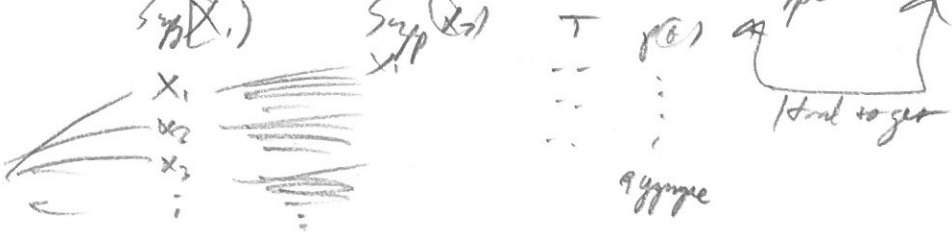
σ^2 uninterpretable

σ interpretable ... not so clear!

It is not an expectation... it is just a practical strategy to report spread.

But... it will be useful later...

$T_2 := X_1 + X_2$ $\quad E(T) = \sum_{t \in Sup(T)} t \cdot p(t)$



Is there a better way?

X_1, X_2 indep $p(x_1, x_2) = p(x_1) p(x_2)$