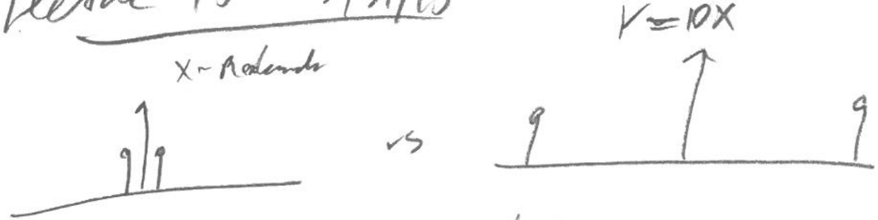
Math 241

Lecture 15 3/31/15



$E(X)=0,\ E(Y)=0$ but $X \overset{d}{\neq} Y$

Y is more dispersed than X about its center.

How to define dispersal? Need distance/loss/penalty function.

$e(x,\mu) = (x-\mu)^2$ squared error loss is a "normal" choice

but still arbitrary...

Given a realization of X, how far do we expect it to be from its expectation on avg? " $E[g(x)]$ (do proof) $= \sum g(x)p(x)$

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

for Rademacher

$= (-1-0)^2 \frac{1}{2} + (+1-0)^2 \frac{1}{2} = \boxed{1}$

expected loss has a special name: "Variance"

for Y

$= (-10-0)^2 \frac{1}{2} + (+10-0)^2 \frac{1}{2} = \boxed{100}$

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

let $X \sim Bern(p)$ $\mu = p$

$\overset{Var(X)=}{\sigma^2} = E(X-\mu)^2 = E(X-p)^2 = \sum_{x=0}^{1} (x-p)^2 p(x) = (0-p)^2 p(0) + (1-p)^2 p(1) = p^2(1-p) + (1-p)^2 p$

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

\$1 Bet on Black:

$X \sim \begin{cases} \$35 & \text{wp } \frac{1}{38} \\ -\$1 & \text{wp } \frac{37}{38} \end{cases} \quad \mu = -\$0.053$

$Var(X) = (-1 - -0.053)^2 \frac{37}{38} + (35 - -0.053)^2 \frac{1}{38}$

$= 33.207\ \$^2$ ← what's the units?

X is a discrete r.v. $X: \Omega \to \mathbb{R}$

$Supp(X) = \{x_1, x_2, \dots\}$

$\{\omega : X(\omega) = x_1\}$

$\to \{\omega : X(\omega) = x_2\}$

$\to \{\omega : X(\omega) = x_3\}$

$$\sum_{x \in Supp(X)} p(x) = 1 \quad \text{(Never proven)}$$

By def: $\displaystyle\int_\Omega dP(\omega) = 1$ defines a prob space

$\{\omega : X(\omega) = x_1\} \cap \{\omega : X(\omega) = x_2\} = \phi$

disjoint!

Otherwise $\exists\, \omega$ s.t. $X(\omega) = x_1$, $= x_2$

violates def of X!!

$\{\omega : X(\omega) = x_1\} \cup \dots = \Omega$

mutually exclusive, collectively exhaustive

$$= \int_{\{\omega : X(\omega) = x_1\}} dP(\omega) + \int_{\{\omega : X(\omega) = x_2\}} dP(\omega) + \dots = 1$$

$\in Supp(X)$

$$= P(X = x_1) + P(X = x_2) + \dots = 1$$

$$= \sum_{x \in Supp(X)} p(x) = 1 \checkmark$$

$$E[X] = \sum_{x \in Supp(X)} x\, p(x) \quad \text{not really ok def} \dots$$

$$E[X] = \int_\Omega X(\omega)\, dP(\omega) = \int_{\{\omega : X(\omega) = x_1\}} X(\omega)\, dP(\omega) + \int_{\{\omega : X(\omega) = x_2\}} X(\omega)\, dP(\omega) + \dots$$

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

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

$$E[g(X)] = \int_\Omega g(X(\omega))\, dP(\omega) = \int_{\{\omega : X(\omega) = x_1\}} g(X(\omega))\, dP(\omega) + \int_{\{\omega : X(\omega) = x_2\}} g(X(\omega))\, dP(\omega) + \dots$$

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

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

Can you interpret this unit? NO... we need to standardize the units

$\sigma := \sqrt{\sigma^2} = \sqrt{\text{Var}[X]} := \sqrt{33.207\$^2} = \$5.76$ the "standard deviation"

$SD[X] =$

$= SE[X]$

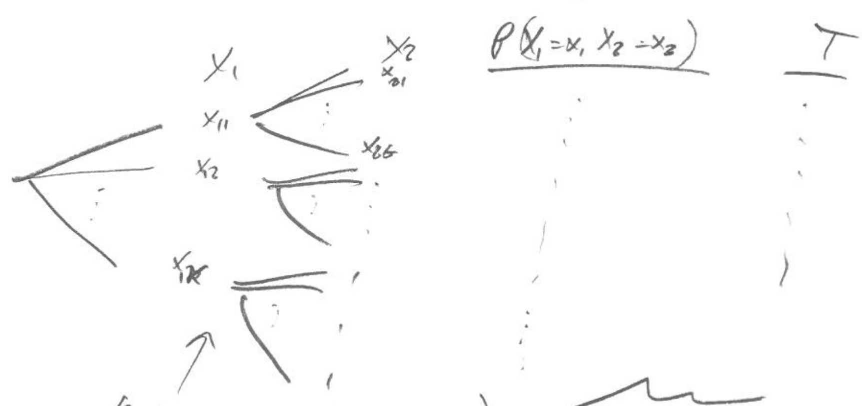
(deviation being another word for "variance".)

same units of the support of X.

= PAUSE =

Recall $T_2 = X_1 + X_2$

to get $p(t)$, need to traverse



$p(x_1, x_2)$ ← joint PMF (JMF)

add up all common t vals $\Rightarrow p(t)$

Now, we want $E[T_2] = \sum_{t \in Supp[T]} t\,p(t) = E[X_1 + X_2]$

If X_1, X_2 ind, $p(x_1, x_2) = p(x_1)\,p(x_2)$ JMF factors into the two "marginal" PMF's

→ Can we do anything? Yes...

$$E[g(X_1, X_2)] = \sum_{(x_1,x_2) \in Supp[X_1] \times Supp[X_2]} g(x_1,x_2)\,p(x_1,x_2) = \sum_{x_2}\sum_{x_1} g(x_1,x_2)\,p(x_1,x_2)$$

(this is a generalization of the previous proof)

$$E[X_1 + X_2] = \sum_{x_2}\sum_{x_1} (x_1 + x_2)\,p(x_1,x_2) = \sum_{x_2}\sum_{x_1} x_1\,p(x_1,x_2) + \sum_{x_2}\sum_{x_1} x_2\,p(x_1,x_2)$$

$$= \sum_{x_2}\sum_{x_1} x_1\,p(x_1)\,p(x_2) + \sum_{x_2}\sum_{x_1} x_2\,p(x_1)\,p(x_2)$$

$$= \underbrace{\sum p(x_2)}_{1}\ \underbrace{\sum x_1\,p(x_1)}_{E[X_1]} + \underbrace{\sum_{x_2} x_2\,p(x_2)}_{E[X_2]}\ \underbrace{\sum_{x_1} p(x_1)}_{1} = E[X_1] + E[X_2] \text{ if ind.}$$

What if not independent?

$$= \sum_{x_1} \sum_{x_2} x_1 \, p(x_1, x_2) + \sum_{x_2} \sum_{x_1} x_2 \, p(x_1, x_2)$$

$$= \sum_{x_1} x_1 \sum_{x_2} p(x_1, x_2) + \sum_{x_2} x_2 \sum_{x_1} p(x_1, x_2)$$

what are these?

Imagine JMF for X_1, X_2 $\quad Supp(X_1) = \{1, 7, 19\}$

$Supp(X_2) = \{5, 23, 88\}$

$p(x_1, x_2)$

X_2 \ X_1	1	7	19	$p(x_2)$
5	$\frac{1}{15}$	$\frac{1}{3}$	$\frac{2}{15}$	$\frac{2}{5}$
23	$\frac{1}{30}$	$\frac{1}{10}$	$\frac{1}{30}$	$\frac{1}{6}$
88	$\frac{1}{30}$	$\frac{1}{5}$	$\frac{1}{15}$	$\frac{13}{30}$
$p(x_1)$	$\frac{4}{30}$	$\frac{19}{30}$	$\frac{7}{30}$	1

Valid JMF? $\sum_{x_1} \sum_{x_2} p(x_1, x_2) = 1$

"Law of Tot Prob" Again!

$$P(X_2 = 5) = P(X_2 = 5 \ \& \ X_1 = 1) + P(X_2 = 5 \ \& \ X_1 = 7) + P(X_2 = 5 \ \& \ X_1 = 19)$$

$$= \frac{1}{15} + \frac{1}{3} + \frac{2}{15} = \boxed{\frac{2}{5}}$$

$$\Rightarrow P(X_2 = 5) = \sum_{x_1} p(x_1, 5)$$

this is called "marginalizing out" X_1

If continuous, integrating out

$$\int_{Supp(Y)} f(x,y)\,dy = g(x)$$

$$\to \sum_{x_1} x_1 \underbrace{\sum_{x_2} p(x_1, x_2)}_{p(x_1)} + \sum_{x_2} x_2 \underbrace{\sum_{x_1} p(x_1, x_2)}_{p(x_2)}$$

$$= \sum_{x_1} x_1 \, p(x_1) + \sum_{x_2} x_2 \, p(x_2) = E(X_1) + E(X_2)$$

$$\Rightarrow E[X_1 + X_2] = E[X_1] + E[X_2] \text{ for all discrete r.v.'s } X_1, X_2$$

general rule $\quad E[X_1 + \ldots + X_n] = E(X_1) + \ldots + E(X_n)$ AKA $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$

$X_1, \dots, X_n \overset{iid}{\sim} p(x)$

Remember $\bar{X} := \frac{X_1 + \dots + X_n}{n}$ and $E(aX) = a E(X)$

What is $E(\bar{X}) = E\left[\frac{X_1 + \dots + X_n}{n}\right] = E\left[\frac{1}{n}(X_1 + \dots + X_n)\right] = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{1}{n}(\overbrace{\mu + \dots + \mu}^{n}) = \boxed{\mu}$

the "average of the average r.v." is the expectation of the process!

This is a property called "unbiasedness." The sample avg is unbiased... what's next...?

Return to Var

rule just proved

$$Var(X) = E\left[(X-\mu)^2\right] = E\left[X^2 - 2X\mu + \mu^2\right] = E(X^2) - E[2X\mu] + E(\mu^2)$$
$$= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2$$
$$= E(X^2) - 2\mu\mu + \mu^2$$
$$\boxed{\sigma^2 = E(X^2) - \mu^2}$$

convenient formula

no rule for $E(X^2)$ ($g(x) = x^2$)

gotta do it manually...

Note:

$E(X)$	"first moment"	$E[X-\mu]$	"first central moment"
$E(X^2)$	"second moment"	$E[(X-\mu)^2]$	"second central moment" = (Variance)
$E(X^3)$	"third moment"	$E[\lvert(X-\mu)^3\rvert]$	"third central moment"
$\vdots$		$\vdots$	

HARD to compute... until end of class...

More rules...

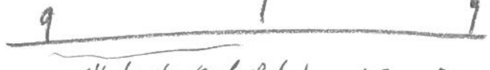
$Y = aX,\ a \in \mathbb{R}$

$E(Y) = aE(X)$

$Var(Y)$?

$$Var[Y] = E\left[(Y - E[Y])^2\right] = E\left[(aX - E(aX))^2\right]$$
$$= E\left[(aX - a\mu)^2\right] = E\left[a^2(X-\mu)^2\right] = a^2 E(X-\mu)^2 = a^2 Var(X)$$
$$\Rightarrow \boxed{Var(aX) = a^2 Var(X)}$$

⇒



all dots multiplied by 10 ⇒ Squared distance multiplied by 100 and std distance multiplied by 10.

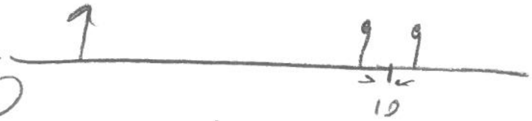
$$SD[aX] = \sqrt{Var(aX)} = \sqrt{a^2 Var(X)} = |a|\, SD(X)$$

$Y = X + c,\ c \in \mathbb{R}$ $E[X+c] = E(X) + c$

$$Var(Y) = E[(Y - E(Y))^2] = E[((X+c) - E(X+c))^2] = E[(X+c-\mu-c)^2] = E(X-\mu)^2 = Var(X)$$

makes sense $X \sim$ Rademacher, $Y = X + 10$

STOP



distances haven't changed!

$E[X_1 + X_2] = E[X_1] + E(X_2)$ what about $Var[X_1 + X_2]$?

$$= E\,(X_1 + X_2 - \underbrace{E(X_1 + X_2)}_{\mu_1 + \mu_2})^2 = E[(X_1 + X_2 - \mu_1 - \mu_2)^2]$$

$$= E[X_1^2 + X_2^2 + \mu_1^2 + \mu_2^2 + 2X_1X_2 - 2X_1\mu_1 - 2X_2\mu_1 - 2X_1\mu_2 - 2X_2\mu_2 + 2\mu_1\mu_2]$$

$$= E[X_1^2] + E[X_2^2] + \mu_1^2 + \mu_2^2 + 2E[X_1X_2] - 2\mu_1^2 - 2\mu_2\mu_1 - \cancel{2\mu_1\mu_2} - 2\mu_2^2 + \cancel{2\mu_1\mu_2}$$

$$= (E[X_1^2] - \mu_1^2) + (E[X_2^2] - \mu_2^2) + 2(E[X_1X_2] - \mu_1\mu_2)$$

$$= Var(X_1) + Var(X_2) + 2\,Cov[X_1, X_2]$$

$$E[X_1X_2] = \sum_{x_1}\sum_{x_2} x_1 x_2\, p(x_1, x_2)$$

$$= \sum_{x_1} x_1 \sum x_2\, p(x_1, x_2) \quad \text{Can't budge!}$$

But if X_1, X_2 iid $\Rightarrow p(x_1, x_2) = p(x_1)p(x_2)$

$$= \sum_{x_1} x_1 \sum x_2\, p(x_1)p(x_2)$$

$$= \underbrace{\sum_{x_1} x_1\, p(x_1)}_{E(X_1)}\ \underbrace{\sum x_2\, p(x_2)}_{E(X_2)} = \mu_1\mu_2$$

$$\Rightarrow Cov[X_1, X_2] = \mu_1\mu_2 - \mu_1\mu_2 = 0$$

$$\Rightarrow Var[X_1 + X_2] = Var[X_1] + Var(X_2)$$