Stat 134 lec 39

## Warmup 1:00–1:10

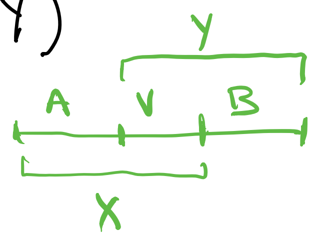
Toss a fair coin 30 times.

Let $X$ = # heads in the first 20 tosses

$Y$ = # heads in the last 20 tosses.

Find $Cov(X,Y)$

Hint:




$A \sim Bin(10, 1/2)$
$V \sim Bin(10, 1/2)$
$B \sim Bin(10, 1/2)$
$\}$ indep.

$X = A + V$
$Y = V + B$

$$Cov(X,Y) = Cov(A+V, V+B) = Cov(V,V)$$
$$= Var(V) = 10\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)$$
$$= \frac{10}{4}$$

## Last time

### Sec 3.6

defn RVs' $X_1, \dots, X_n$ are exchangeable RVs if

$$(X_i, X_j) \overset{D}{=} (X_1, X_2) \quad \text{(i.e. same joint distribution)}$$

ex Two cards drawn without replacement from a deck of cards is exchangeable.

* See appendix to notes for example of identically distributed RVs not exchangeable

### Sec 6.4 Covariance and the variance of sum

Let $X, Y$ be RVs

$$\text{Cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$
$$= E(XY) - E(X)E(Y)$$

If $X, Y$ are independent, $\text{Cov}(X,Y)=0$, and

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + \underbrace{2\text{Cov}(X,Y)}_{\overset{\text{"}}{0}}$$

If $X_1, \dots, X_n$ are exchangeable then

$\text{Cov}(X_i, X_j) = \text{Cov}(X_1, X_2)$ and

$$\text{Var}(X_1 + \dots + X_n) = n\,\text{Var}(X_1) + n(n-1)\,\text{Cov}(X_1, X_2)$$

## Today

1. Sec 6.4 Correlation
2. Sec 6.5 Bivariate Normal

2

## Sec 6.4 Correlation

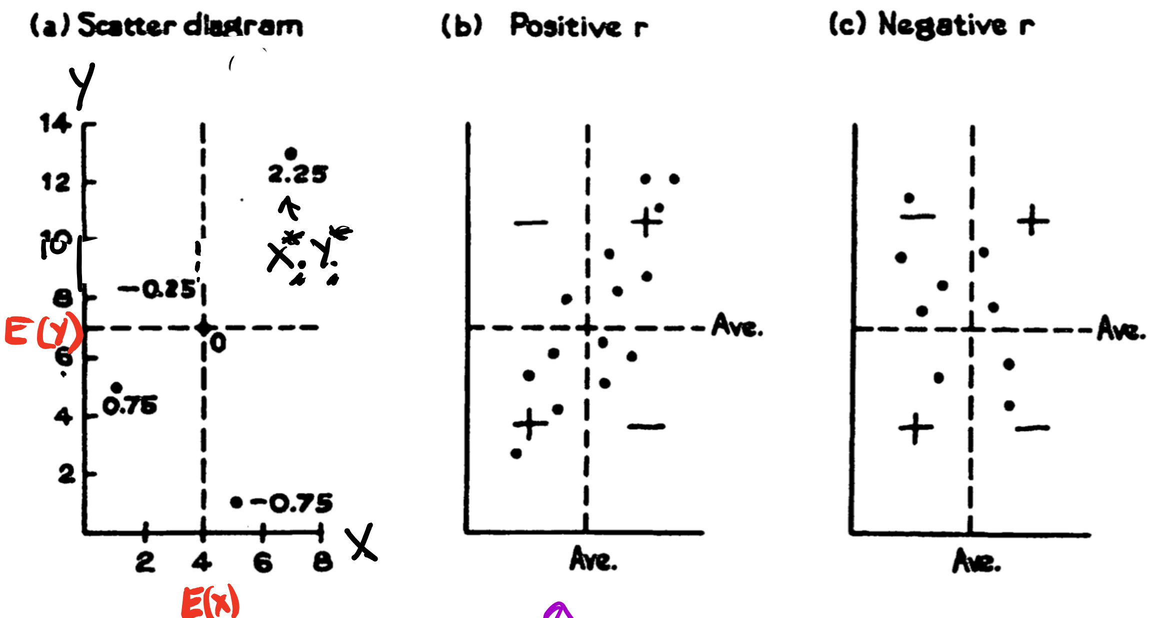
$$Cov(X,Y) = E\left((X-\mu_x)(Y-\mu_y)\right)$$

$$r = Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = E\left(\left(\frac{X-\mu_x}{SD_x}\right)\left(\frac{Y-\mu_y}{SD_y}\right)\right)$$

$$= E(X^* Y^*)$$

X, Y in standard units

How the correlation coefficient works.




This will have a positive correlation since more of the points are in the 1st and 3rd quadrants

ex

Let $X$ = # heads in the first 20 tosses
$Y$ = # heads in the last 20 tosses.

Find $\text{Corr}(X,Y)$

Note we showed $\text{Cov}(X,Y) = \frac{10}{4}$

$$\text{Corr}(X,Y) = \frac{10/4}{\underset{\underset{20\cdot\frac{1}{2}\cdot\frac{1}{2}}{\parallel}}{SD(X)\,SD(Y)}} = \frac{1}{2}$$

$\underline{\text{ex}}$ Suppose the sum of $k$ exchangeable RVs is a constant

$$N_1 + N_2 + \dots + N_k = c$$

Find $\text{Corr}(N_1, N_2)$.

$\underline{\text{soln}}$

$$N_1 + N_2 + \dots + N_k = c$$

$$\Rightarrow \quad \text{Var}(N_1 + \dots + N_k) = 0$$

Finish...

$$\Rightarrow \quad k\,\text{Var}(N_1) + k(k-1)\,\text{Cov}(N_1, N_2) = 0$$

$$\Rightarrow \quad \text{Cov}(N_1, N_2) = \frac{-\text{Var}(N_1)}{k-1}$$

$$\Rightarrow \quad \text{Corr}(N_1, N_2) = \frac{\overset{\frac{-\text{Var}(N_1)}{k-1}}{\overset{\parallel}{\text{Cov}(N_1, N_2)}}}{\underbrace{SD(N_1)\,SD(N_2)}_{\overset{\parallel}{\text{Var}(N_1)}}} = \boxed{-\frac{1}{k-1}}$$

tinyurl: http://tinyurl.com/dec4-pt1

**Stat 134**

**Friday April 26 2019**

1. An urn contains 90 marbles, of which there are 20 green, 25 black, and 45 red marbles. Alice randomly picks 10 marbles without replacement, and Bob randomly picks another 10 marbles without replacement. Let $X_1$ be the number of green marbles that Alice has and $X_2$ the number of green marbles that Bob has.
   To find $Corr(X_1, X_2)$ is
   $$X_1 + X_2 + \cdots + X_9 = 20$$
   a true identity that is useful? Explain.

**a** yes

**b** no

**c** not enough info to decide

$X_1, \ldots, X_9$ are exchangeable being random draws without replacement from an urn,

From above, $Corr(X_1, X_2) = \frac{-1}{k-1} = \frac{-1}{8}$

ex

Show $\text{Corr}(aX+b, cY+d) = \frac{ac}{|ac|}\text{Corr}(X,Y)$

$$\text{Corr}(aX+b, cY+d) = \frac{\text{Cov}(aX+b, cY+d)}{SD(aX+b)SD(cY+d)}$$

$$= \frac{ac\,\text{Cov}(X,Y)}{|a||c|SD(X)SD(Y)}$$

$$= \frac{ac \quad \text{Cov}(X,Y)}{|a||c| \; SD(X) \; SD(Y)} = \frac{ac}{|ac|}\text{Corr}(X,Y)$$

## Properties of correlation

(a) Correlation is invariant to change of scale except possibly by a sign.

(i.e $|\text{corr}(X,Y)| = |\text{corr}(aX+b, cX+d)|$

for constants $a, b, c, d$.

ex Correlation between Boston and NYC temperatures is the same whether temps in $°C$ or $°F = 1.8°C + 32$

Hence

$\text{Corr}(X,Y) = \text{Corr}(X^*, Y^*)$ since

$SD(X) > 0$ and $SD(Y) > 0$.

(2) $-1 \le \text{corr}(X, Y) \le 1$

**Proof**

Correlation is invariant if you convert $X, Y$ to standard units $X^*, Y^*$ since $SD(X) > 0$
$SD(Y) > 0$.

So we show that $-1 \le \text{Corr}(X^*, Y^*) \le 1$

$$\left.\begin{aligned} E(X^*) &= 0 = E(Y^*) \\ SD(X^*) &= 1 = SD(Y^*) \\ E(X^{*2}) &= 1 = E(Y^{*2}) \end{aligned}\right\} \text{ Since } X^*, Y^* \text{ are standard units.}$$

$E(X^{*2}) = \underbrace{\text{var}(X^*)}_{=1} + \underbrace{(E(X^*))^2}_{=0}$

$$(X^* + Y^*)^2 \ge 0$$

so $E((X^* + Y^*)^2) \ge 0$

$$E(X^{*2} + Y^{*2} + 2X^*Y^*) \ge 0$$

$$1 + 1 + 2E(X^*Y^*) \ge 0$$

$$E(X^*Y^*) \ge -1$$

$$\Rightarrow \quad -1 \le E(X^*Y^*)$$

$$\boxed{-1 \le \text{Corr}(X, Y)}$$

Similarly can show $\text{Corr}(X, Y) \le 1$ — See appendix.

ex

let $X, Z$ iid $N(0,1)$,

$Y = \rho X + \sqrt{1-\rho^2}\, Z$ where $-1 \le \rho \le 1$.

① What distribution is $Y$?

$Y$ is a linear combination of independent normals and hence normal.

$$E(Y) = E(\rho X + \sqrt{1-\rho^2}\, Z) = \rho \underbrace{E(X)}_{0} + \sqrt{1-\rho^2}\, \underbrace{E(Z)}_{0}$$

$= 0$

$$Var(Y) = Var(\rho X + \sqrt{1-\rho^2}\, Z) = \rho^2 \underbrace{Var(X)}_{1} + (1-\rho^2) \underbrace{Var(Z)}_{1}$$

$= 1$

$\Rightarrow Y \sim N(0,1)$

② What is $Corr(X, Y)$

$$Cov(X,Y) = Cov(X, \rho X + \sqrt{1-\rho^2}\, Z)$$

$$= \rho \underbrace{Var(X)}_{1} = \rho$$

$$Corr(X,Y) = \frac{\overbrace{Cov(X,Y)}^{\rho}}{\underbrace{SD(X)}_{1}\underbrace{SD(Y)}_{1}} = \boxed{\rho}$$

(2) Sec 6.5 <u>Bivariate Normal</u>

<u>Defn</u> (Standard Bivariate Normal Distribution)

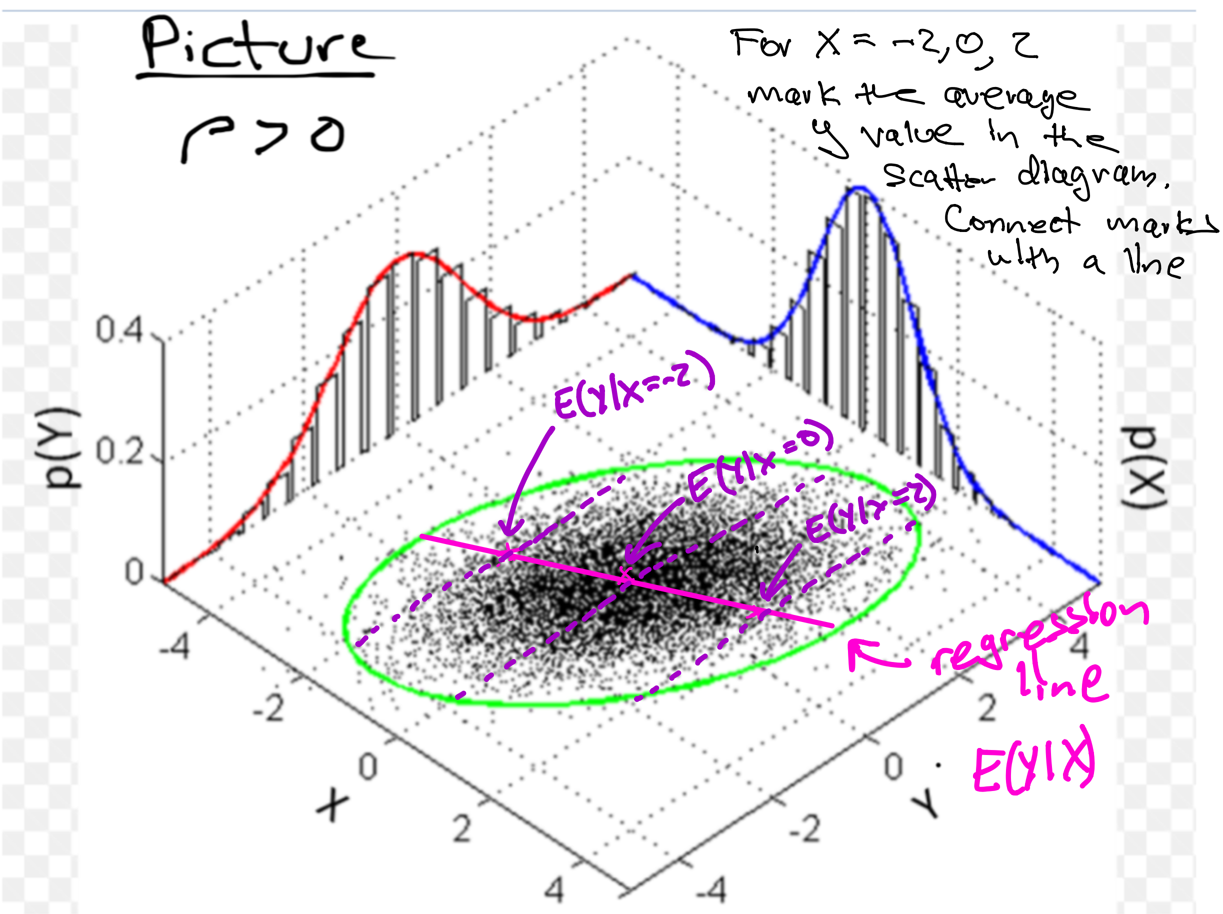
let $X, Z$ iid $N(0,1)$, $-1 \le \rho \le 1$

$$Y = \rho X + \sqrt{1-\rho^2}\, Z$$

We call the joint distribution $(X,Y)$ the standard bivariate normal with $\text{corr}(X,Y) = \rho$

written $(X,Y) \sim BV(0, 0, 1, 1, \rho)$

$\mu_X$ $\mu_Y$ $\sigma_X^2$ $\sigma_Y^2$



let, $X, Z \overset{iid}{\sim} N(0,1)$, $-1 \le \rho \le 1$

$$y = \rho X + \sqrt{1-\rho^2}\, Z$$

Find the "regression line" $\hat{y} = E(y|X)$

$$\hat{y} = E(y|X) = E(\rho X + \sqrt{1-\rho^2}\, Z | X)$$

$$= E(\rho X | X) + E(\sqrt{1-\rho^2}\, Z | X)$$

$$= \rho X + \sqrt{1-\rho^2}\, \underbrace{E(Z)}_{=0}$$

$$= \rho X$$

Find the conditional variance $Var(y|X)$

$$Var(y|X) = Var(\rho X + \sqrt{1-\rho^2}\, Z | X)$$

$$= Var(\rho X | X) + Var(\sqrt{1-\rho^2}\, Z | X) \quad \text{since } X, Z \text{ indep.}$$

$$= \rho^2 \underbrace{Var(X|X)}_{=0} + (1-\rho^2) \underbrace{Var(Z|X)}_{= Var(Z) = 1}$$

since $X, Z$ indep

$$= 1 - \rho^2$$

# Appendix

## Example of identically distributed RVs not exchangable

ex Flip coin 6 times

$q^2 p$

$I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ flip is start of a run of 1 head} \\ 0 & \text{else} \end{cases}$

T H T _ _ _

$I_2, I_3, I_4, I_5$ are identically distributed,

Is joint of $(I_2, I_3)$ the same as $(I_2, I_4)$ ?

No

$P(I_2=1, I_3=1) = 0$ but $P(I_2=1, I_4=1) = P(I_2)P(I_4)$
$= (q^2 p)^2$

so $I_2, I_3, I_4, I_5$ are not exchangeable.

# Appendix

Show $\text{Corr}(X,Y) \le 1$ by examining $E\left((X^* - Y^*)^2\right)$:

$$(X^* - Y^*)^2 \ge 0$$

so $E((X^* - Y^*)^2) \ge 0$

$$E(X^{*2} + Y^{*2} - 2X^*Y^*) \ge 0$$

$$1 + 1 - 2E(X^*Y^*) \ge 0$$

$$E(X^*Y^*) \le 1$$

$$\Rightarrow \boxed{\text{Corr}(X,Y) \le 1}$$

This finishes the proof.