Stat 134 lec 38

Warmup 10:00-10:10

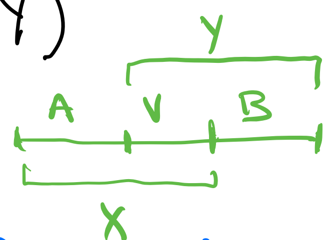
Toss a fair coin 30 times.

Let X = # heads in the first 20 tosses

Y = # heads in the last 20 tosses.

Find $Cov(X,Y)$

Hint:



$Cov(X,Y) = E(XY) - E(X)E(Y)$

$$\underbrace{Cov(A+V, V+B)}_{=Cov(X,Y)} = \underbrace{Cov(A,V)}_{=0} + \underbrace{Cov(A,B)}_{=0} + \underbrace{Cov(V,V)}_{=Var(V) = npq} + \underbrace{Cov(V,B)}_{=0}$$

$$npq: \quad n = 10, \quad p = 1/2, \quad q = 1/2$$

$$= \boxed{10/4}$$

Last time

Sec 3.6

def^n RVs' $X_1, \dots, X_n$ are exchangeable RVs if

$$(X_i, X_j) \overset{D}{=} (X_1, X_2)$$ (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

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

If X, Y are independent, $Cov(X,Y)=0$, and

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

Let $X_1, \dots, X_n$ be exchangeable

We wish to compute

$$Var\left(\sum_{i=1}^{n} X_i\right) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right) = Cov(X_1, X_1) + Cov(X_1, X_2)$$
$$+ \dots + Cov(X_n, X_n).$$

The variance-covariance matrix has all n^2 terms

$$
\begin{array}{c|c}
 & \begin{matrix} x_1 & x_2 & \cdots & x_n \end{matrix} \\
\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{matrix} & \begin{bmatrix} Cov(x_1,x_1) & Cov(x_1,x_2) & & \\ & Cov(x_2,x_2) & & \\ & & \ddots & \\ & & & Cov(x_n,x_n) \end{bmatrix}_{n \times n}
\end{array}
$$

Note $Cov(x_1, x_2) = Cov(x_i, x_j)$ since $x_1, \ldots, x_n$ are exchangable.

$$\Rightarrow \quad Var\left(\sum_{i=1}^{n} x_i\right) = \underbrace{n\,Var(x_1)}_{\text{diagonal}} + \underbrace{n(n-1)\,Cov(x_1, x_2)}_{\text{off diagonal}}$$

Today

1. Sec 6.4 Correlation

1

Sec 6.4 Correlation

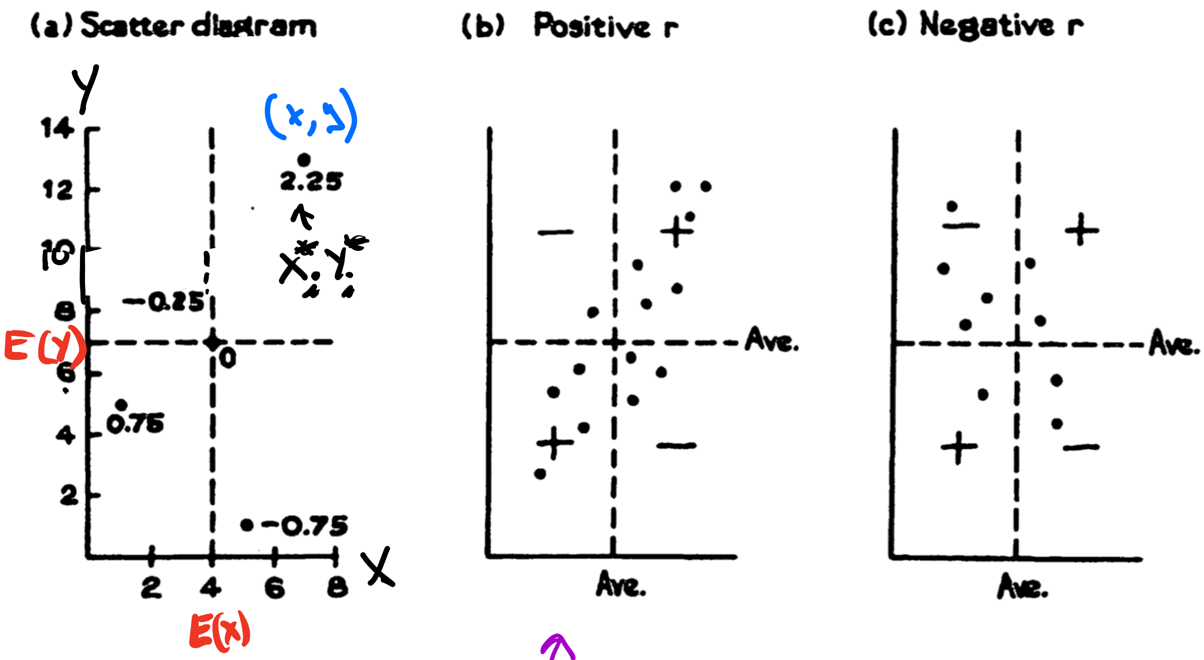
$$Cov(X,Y) = E\big((X-\mu_X)(Y-\mu_Y)\big)$$

$$r = Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)\,SD(Y)} = E\left(\left(\frac{X-\mu_X}{SD_X}\right)\left(\frac{Y-\mu_Y}{SD_Y}\right)\right)$$

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

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 $\mathrm{Corr}(X,Y)$

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

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{SD(X)\,SD(Y)} = \frac{10/4}{\sqrt{20\cdot\frac{1}{2}\cdot\frac{1}{2}}\cdot\sqrt{20\cdot\frac{1}{2}\cdot\frac{1}{2}}}$$

$$= \boxed{1/2}$$

ex Suppose the sum of k exchangeable RVs is a constant

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

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

soln

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

$$\Rightarrow \mathrm{Var}(N_1 + \cdots + N_k) = 0$$

Finish...

$$\mathrm{Var}(N_1 + \cdots + N_k) = k\,\mathrm{Var}(N_1) + k(k-1)\,\mathrm{Cov}(N_1, N_2)$$

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

$$\mathrm{Corr}(N_1, N_2) = \frac{\mathrm{Cov}(N_1, N_2)}{\underbrace{\mathrm{SD}(N_1)\,\mathrm{SD}(N_2)}_{= \mathrm{Var}(N_1)}} = \frac{\left(\frac{-\cancel{\mathrm{Var}(N_1)}}{k-1}\right)}{\cancel{\mathrm{Var}(N_1)}} = \boxed{\frac{-1}{k-1}}$$

* Assume all students take the same number of marbles.

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

$X_1, \dots, X_9$ are exchangeable RVs being random draws without replacement from an URN. From above

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

ex

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

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

recall

$\operatorname{Var}(aX) = a^2 \operatorname{Var}(X)$

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

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

Properties of correlation

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

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

$\operatorname{Corr}(X,Y) = \operatorname{Corr}(X^*, Y^*)$ 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 $E((X^* + Y^*)^2) \ge 0$

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

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

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

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

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

Note $E(X^*) = 0$, $E(Y^*) = 0$, so $E(X^*Y^*) = E(X^*Y^*) - E(X^*)E(Y^*) = \text{Corr}(X^*, Y^*)$

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

See appendix.

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

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

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

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

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

This finishes the proof.