Stat 134 lec 37

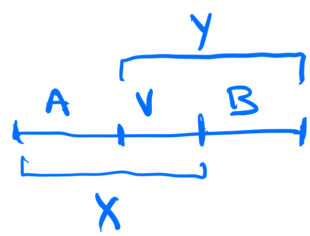
Warmup 8:00 – 8:10 AM

Toss a fair coin 30 times.

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

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

Find $Cov(X,Y)$



$$\left.\begin{array}{l} A \sim Bin(10, 1/2) \\ V \sim Bin(10, 1/2) \\ B \sim Bin(10, 1/2) \end{array}\right\} \text{ indep.} \qquad \begin{array}{l} X = A + V \\ Y = V + B \end{array}$$

$$Cov(X,Y) = Cov(A+V, V+B) = Cov(V,V)$$

$$= Var(V) = 10\left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)$$

$$= \frac{10}{4}$$

Last time

Sec 6.4 Covariance and the variance of sum

Let X, Y be RVs

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

Covariance relates to correlation between X, Y as we will see today.

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

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

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

$Cov(X_i, X_j) = Cov(X_1, X_2)$ and

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

Today

0. Sec 3.6 Exchangeable RVs,
1. review concept test from last time.
2. Sec 6.4 Correlation

0

Sec 3.6

<u>def</u> RVs $X_1, .., X_n$ are exchangeable RVs if

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

ex $X_1, \ldots, X_n$ iid

$P(x_i, x_j) = P(X_i) P(X_j) \overset{\text{i.d.}}{=} P(x_1) P(x_2) = P(x_1, x_2)$

<u>Note</u> if (X_i, X_j) and (x_1, x_2) have the same joint

then $g(X_i, X_j)$ and $g(x_1, x_2)$ have the same joint for any function g.

x_1, x_2 exchangeable $\Rightarrow X_i \overset{D}{=} X_1$ (i.e $X_1, .., X_n$ are identically distributed).

ex Flip coin 6 times

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

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

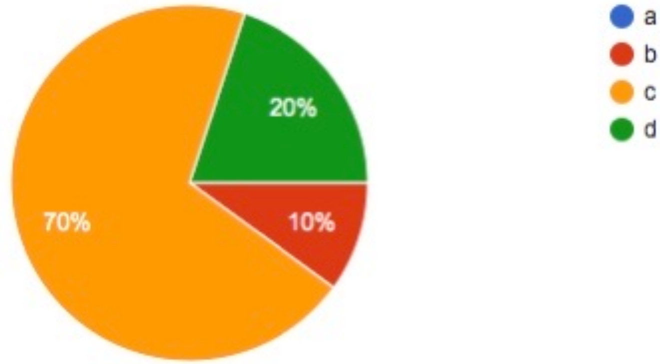
$P(I_2 = 1, I_3 = 1) = 0$ but $P(I_2 = 1, I_4 = 1) = P(I_2) P(I_4)$
$= (q^2 p)^2$

So I_2, I_3, I_4, I_5 are not exchangeable.

ex Let $X_1, .., X_n$ be a random sample without replacement

Then $X_1, .., X_n$ are exchangeable.

1



Stat 1

Wednesday A

1. Consider a Poisson(λ) process. Let $T_r \sim$ gamma(r, λ) be the rth arrival time. Cov(T_1, T_3) equals:

   Recall $\quad Var(T_r) = \frac{r}{\lambda^2}$

   a λ

   b λ^2

   c $1/\lambda^2$

   d none of the above

d	Each arrival in a poisson process is independent, therefore T1 and T3 are independent which means the covariant must be zero.

c	Break T3 into T3 = T1 + W2 + W3. Then W2, W3 are independent of T1. By bilinearity of covariance and covariance of independent variables is 0, we can simplify to Var(T1) so we get 1/lambda^2

$$Cov(T_1, T_1 + W_2 + W_3) = \underbrace{Cov(T_1, T_1)}_{= Var(T_1) = 1/\lambda^2} + \underbrace{Cov(T_1, W_2 + W_3)}_{=0}$$

ex

Suppose in a batch of 12 raffle tickets, we know there are 5 winners. I look at the 1st - 8th cards, and you look at the 5th - 12th cards (so we both observe cards 5-8). Let X be the number of winning tickets I see, and Y be the number of winning tickets you see. Find $Cov(X, Y)$. (Hint: use indicators.)

$p = \frac{5}{12}$

let $I_j = \begin{cases} 1 & \text{if ticket } j \text{ is winner} \\ 0 & \text{else} \end{cases}$

$X = I_1 + \cdots + I_5 + \cdots + I_8$

$Y = \quad I_5 + \cdots + I_8 + \cdots + I_{12}$

$Cov(X,Y) = \underline{4}\; Cov(I_5, I_5) + \underline{4(8) + 4(7)}\; Cov(I_1, I_{12})$

$Cov(I_5, I_5) = Var(I_5) = \frac{5}{12} \cdot \frac{7}{12}$

$Cov(I_1, I_{12}) = E(I_1 I_{12}) - E(I_1)E(I_{12})$

$E(I_1 I_{12}) = \frac{5}{12} \cdot \frac{4}{11}$, $\quad E(I_1)E(I_{12}) = \left(\frac{5}{12}\right)^2$

2

Sec 6.4 Correlation

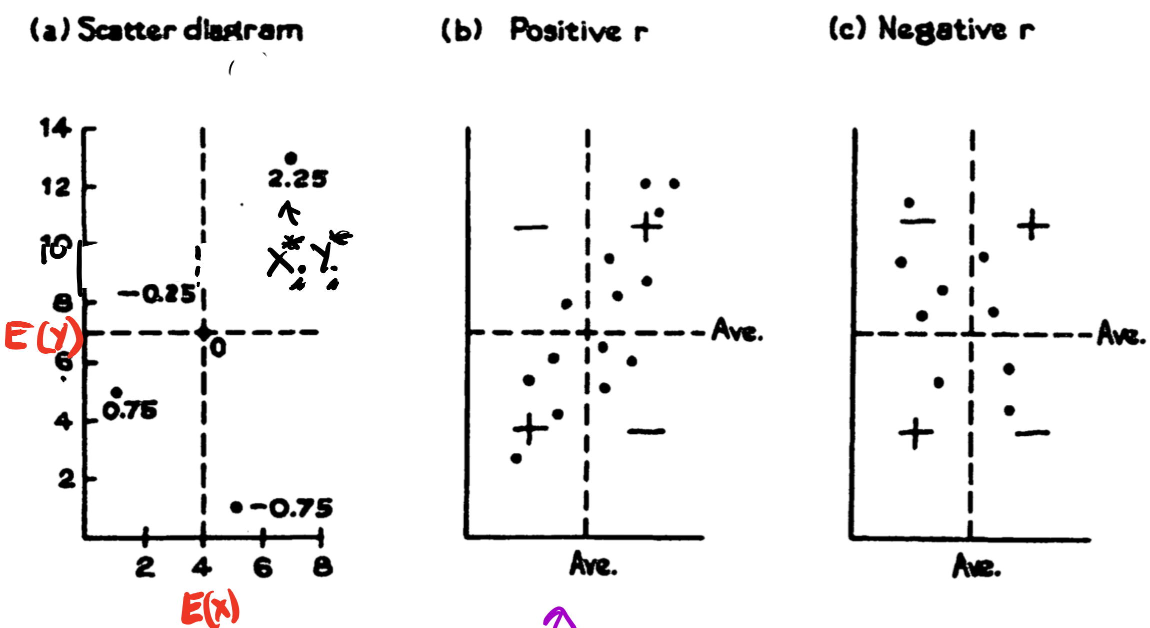
$$\mathrm{Cov}(X,Y) = E\big((X-\mu_X)(Y-\mu_Y)\big)$$

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

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

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

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

ex Suppose the sum of k exchangeable RVs is a constant

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

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

soln

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

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

Finish...

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

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

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

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

Properties of correlation

① 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.

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

Proof

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

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

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

□

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

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

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

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

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.

Appendix

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

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

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

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

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

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

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

This finishes the proof. $\square$