Stat 134 lec 37

Last time

Sec 6.4 Covariance and the variance of sums

Let X, Y be RVs

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

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

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

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$$

Covariance has bilinearity properties:

$$\text{Cov}(X, 2Y) = 2\,\text{Cov}(X,Y)$$
$$\text{Cov}(X, Y+Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z)$$

If $X_1, \ldots, X_n$ are exchangeable (i.e. i.d.) (identically distributed)

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

If we draw n tickets, $X_1, X_2, \ldots, X_n$ from a box with N tickets, with variance σ^2,

$$\text{Cov}(X_1, X_2) = -\frac{\sigma^2}{N-1}$$

and $$\text{Var}(X_1 + \cdots + X_n) = n\sigma^2\left[\frac{N-n}{N-1}\right]$$

(correction factor)

Today

1. review concept test from last time.
2. Sec 6.4 Correlation

Stat 134

Wednesday April 24 2019

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

   a λ

   b λ^2

   c $1/\lambda^2$

   d none of the above

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

Discuss with your neighbor for 1 minute how you did this.

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
c	Find the Cov(T_1, T_3 - T_1), use Cov properties and Var(T_1) to find Cov(T_1, T_3)

$0 = \text{Cov}(T_1, T_3 - T_1) = \text{Cov}(T_1, T_3) - \text{Var}(T_1)$ ($\text{Var}(T_1) = \frac{1}{\lambda^2}$)

$\Rightarrow \text{Cov}(T_1, T_3) = 1/\lambda^2$

Sec 6.4 Correlation

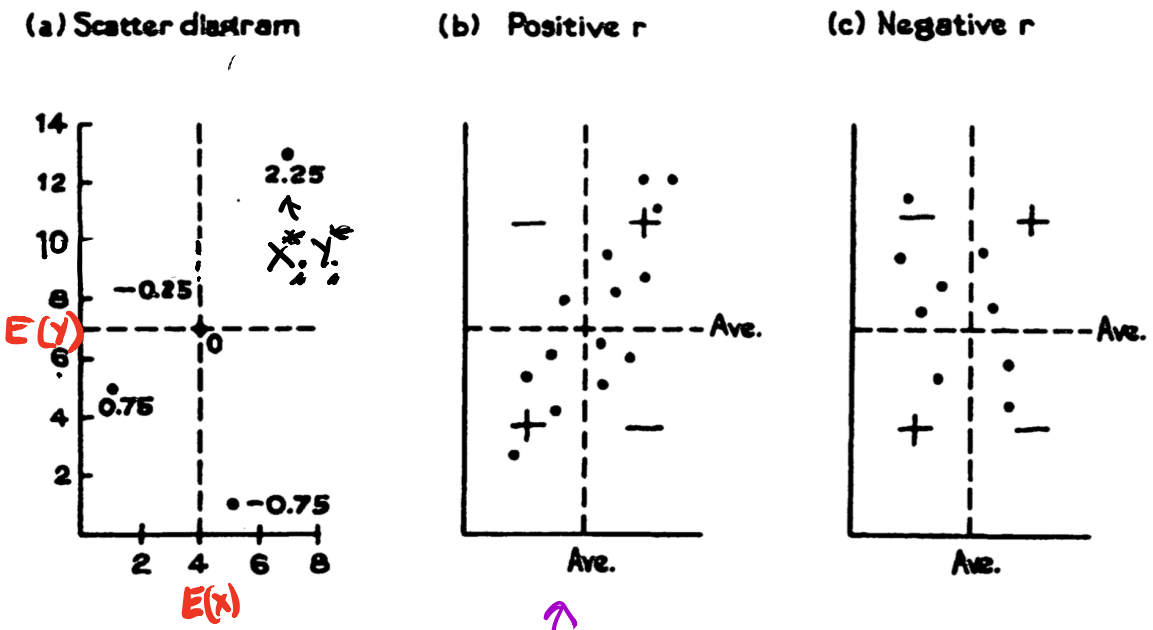
$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

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

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

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 Consider a Poisson process.

$T_r \sim \text{Gamma}(r, \lambda)$ be the r^{th} arrival time.

we showed $\text{Cov}(T_1, T_3) = \frac{1}{\lambda^2}$.

(recall $\text{Var}(T_r) = \frac{r}{\lambda^2}$)

$$\text{Corr}(T_1, T_3) = \frac{\text{Cov}(T_1, T_3)}{SD(T_1)\,SD(T_3)} = \frac{\frac{1}{\lambda^2}}{\frac{1}{\lambda} \cdot \frac{\sqrt{3}}{\lambda}} = \frac{1}{\sqrt{3}}$$

makes sense positively correlated.

Ex Suppose the sum of k exchangeable (i.e identically distributed) RVs is a constant

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

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

soln

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

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

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

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

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

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_i = # green at i^{th} person.

$X_i \sim Hyper(n=10, G=20, N=90)$ for $i=1,2,\ldots,9$.

Imagine randomly laying out 90 marbles flat in a line

If you pull out 2nd 10 marbles without looking at other marbles $P(X_2 = k) = \frac{\binom{20}{k}\binom{70}{10-k}}{\binom{90}{10}}$

$\binom{90}{10}$

Sum of exchangable X_i is a constant.

2. 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.
Find $Corr(X_1, X_2)$.

$X_1 + \dots + X_9 = 20$

$\Rightarrow Cov(X_1, X_2) = \frac{-1}{9-1} = \frac{-1}{8}$

a $-1/8$

b $-1/9$

c $-1/10$

d none of the above

If asked to find $Cov(X_1, X_2)$

$Cov(X_1, X_2) = Corr(X_1, X_2) SD(X_1) SD(X_2) = Cov(X_1, X_2) Var(X_1)$

$SD(X_1) = SD(X_2)$ since X_1, X_2 i.d.

$X_1 \sim Hyper(10, 20, 90)$

$\Rightarrow Var(X_1) = n \frac{G}{N} \frac{N-G}{N} \left[\frac{N-n}{N-1}\right]$

$= 10 \left(\frac{20}{90}\right)\left(\frac{70}{90}\right)\left(\frac{90-10}{90-1}\right)$

$\Rightarrow Cov(X_1, X_2) = \frac{-1}{8} \cdot 10 \left(\frac{20}{90}\right)\left(\frac{70}{90}\right)\left(\frac{90-10}{90-1}\right) = \frac{-1400}{7209}$

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.

ex 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) $-1 \le corr(X, Y) \le 1$

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 $-1 \le Corr(X^*, Y^*) \le 1$.

$$E(X^*) = 0 = E(Y^*)$$
$$SD(X^*) = 1 = SD(Y^*)$$
$$E(X^{*2}) = 1 = E(Y^{*2})$$

} Since X^*, Y^* are standard units.

$$E(X^{*2}) = Var(X^*) + (E(X^*))^2$$
($Var(X^*) = 1$, $E(X^*) = 0$)

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

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

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

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

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

Show $\text{Corr}(x,y) \leq 1$ by examining $E\left((x^* - y^*)^2\right)$:

$$(x^* - y^*)^2 \geq 0$$

so $E\left((x^* - y^*)^2\right) \geq 0$

$$E(x^{*2} + y^{*2} - 2x^* y^*) \geq 0$$

$$1 + 1 - 2E(x^* y^*) \geq 0$$

$$E(x^* y^*) \leq 1$$

$$\Rightarrow \boxed{\text{Corr}(x,y) \leq 1}$$

This finishes the proof. □