Warmup 1:00 - 1:10 pm

Let (X,Y) be bivariate normal. Then (2X+3Y+4, 6X-Y-4) is bivariate normal.

a true

b false

c not enough info to decide

Show $a(2x+3y+4) + b(6x-y-4)$ is normal for all a, b

$= \underbrace{(2a+6b)X + (3a-4b)Y}_{\text{normal}} + \underbrace{4a-4b}_{\text{constant}}$ is normal since (X,Y) bivariate normal.

$\Rightarrow (2x+3y+4,\ 6x-y-4)$ is BV.

Julio Alejo

5:53pm

Can we go over 3.2.13(c)? I could use a little refresher on tail sums.

Reply

13. Suppose a fair die is rolled ten times. Find numerical values for the expectations of each of the following random variables:

c) the maximum number in the first five rolls;

Sec 3.2 Tail Sum formula for expectation

Suppose X is a count $0, 1, 2, 3, \dots$

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) + \dots$$

$$= 1 \cdot \underbrace{P(X=1)}_{p_1} + 2 \cdot \underbrace{P(X=2)}_{p_2} + \dots$$

$$= \begin{array}{cccc} P(X \geq 1) & P(X \geq 2) & P(X \geq 3) & \\ p_1 & & & \\ p_2 & p_2 & & \\ p_3 & p_3 & p_3 & \\ p_4 & p_4 & p_4 & p_4 \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

$$= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots$$ Tail Sum Formula

This is useful when it is easy to find $P(X \geq k)$.

ex A fair die is rolled 5 times.

Let $X = \max(X_1, \ldots, X_5)$.

Find $P(X \geq k)$ $\quad k = 1,2,3,4,5,6$

$$
\begin{aligned}
P(X \geq k) &= 1 - P(X < k) \\
&= 1 - P(X_1 < k, X_2 < k, \ldots, X_5 < k) \\
&= 1 - P(X_1 < k) P(X_2 < k) \cdots P(X_5 < k) \\
&= 1 - P(X_1 < k)^5 \\
&= 1 - \left(\frac{k-1}{6}\right)^5
\end{aligned}
$$

so $E(X) = \underset{1}{\underset{\shortparallel}{P(X \geq 1)}} + \underset{1-\left(\frac{1}{6}\right)^5}{\underset{\shortparallel}{P(X \geq 2)}} + \cdots + \underset{1-\left(\frac{5}{6}\right)^5}{\underset{\shortparallel}{P(X \geq 6)}} + \underset{0}{\underset{\shortparallel}{P(X \geq 7)}} + \cdots$

$$= 6 - \left(\frac{1}{6}\right)^5 \left[1^5 + 2^5 + 3^5 + 4^5 + 5^5\right] = 5.43$$

① Sec 6.5 Corr, Regression and bivariate normal

Def^n (Standard Bivariate Normal Distribution)

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

$$Y = \rho X + \sqrt{1-\rho^2}\, Z \sim N(0,1)$$

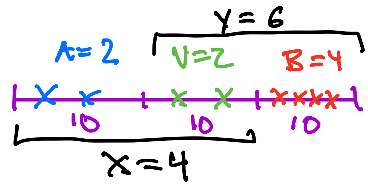
$$\text{Corr}(X,Y) = \rho$$

$$\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{SD(X)SD(Y)}$$

ex. Toss a fair coin 30 times

X = # heads first 20

Y = # heads 2nd 20



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

$X = A + V$

$Y = V + B$

a) Find $Corr(X, Y)$?

$$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

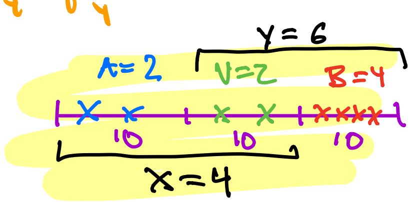
$$Cov(X,Y) = Cov(A+V, V+B) = Cov(V,V)$$

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

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

$$Corr(X,Y) = \frac{\frac{10}{4}}{20/4} = 1/2$$

$20/4$ — $SD(X)SD(Y) = \sqrt{\frac{20}{4}} \cdot \sqrt{\frac{20}{4}}$



b) Find $E(Y|X) = E(V+B|X)$

$$= E(V|X) + E(B|X)$$

$$E(B|X) = E(B) = 10 \cdot 1/2 = 5$$

What distribution is $V|X$?

hint use Bayes rule to find

$$P(V=v|X=x) = \frac{P(V=v, X=x)}{P(X=x)}$$

$$= \frac{P(V=v, A=x-v)}{P(X=x)} = \frac{P(V=v)P(A=x-v)}{P(X=x)}$$

$$= \frac{\binom{10}{v}\cancel{\left(\frac{1}{2}\right)^{10}}\binom{10}{x-v}\cancel{\left(\frac{1}{2}\right)^{10}}}{\binom{20}{x}\cancel{\left(\frac{1}{2}\right)^{20}}}$$

$$V|X \sim HG(N=20, G=10, n=X)$$

$$E(V|X) = n \cdot \frac{G}{N} = 10 \cdot \frac{X}{20} = \frac{1}{2}X$$

$$E(Y|X) = E(V|X) + \underbrace{E(B|X)}_{=5} = \frac{1}{2}X + 5$$

Note that (X,Y) is approximately Bivariate normal:

$$X \sim Bin(20, \tfrac{1}{2}) \approx N(\overset{20\cdot 1/2}{10}, \overset{20\cdot 1/2\cdot 1/2}{5})$$
$$Y \sim Bin(20, \tfrac{1}{2}) \approx N(10, 5)$$
$$Corr(X,Y) = \tfrac{1}{2}$$

$$(X,Y) \sim BV(10, 10, 5, 5, \tfrac{1}{2}) \quad \text{since}$$

$$aX + bY = a(A+V) + b(V+B)$$
$$= aA + (a+b)V + bB$$

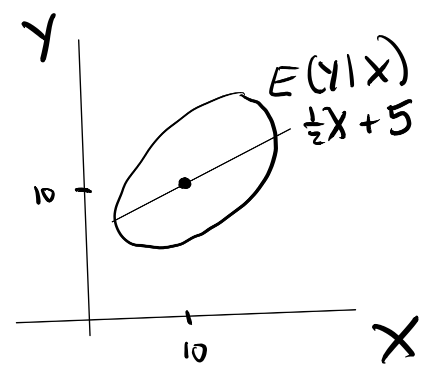
is normal since A, V, B are indep normals.

eqn of regression line $E(Y|X)$:

$$\frac{\hat{Y} - \mu_y}{\sigma_y} = \frac{1}{2}\,\frac{X - \mu_x}{\sigma_x}$$

$$\frac{\hat{y} - 10}{\sqrt{5}} = \frac{1}{2}\left(\frac{X - 10}{\sqrt{5}}\right)$$

$$\boxed{\hat{y} = \tfrac{1}{2}X + 5}$$



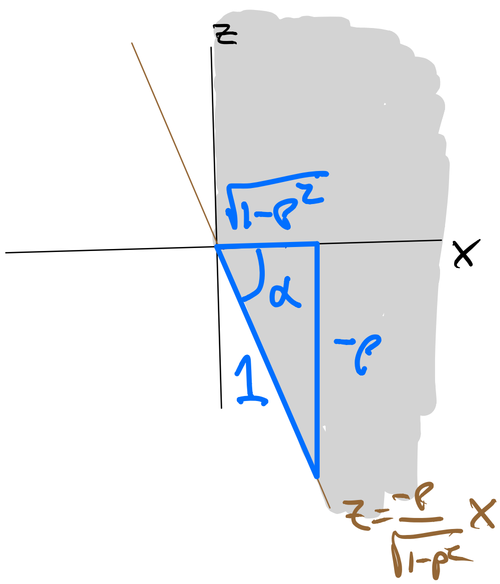
eg: X, Y std bivariate normal, $\rho > 0$

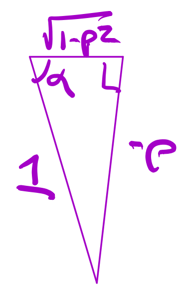
Find $P(X>0, Y>0)$

$$P(X>0, Y>0) = P(X>0, \rho X + \sqrt{1-\rho^2}\, Z > 0)$$

$$= P\left(X>0, Z > \frac{-\rho}{\sqrt{1-\rho^2}} X\right)$$

note X and Z are uncorrelated so joint (X,Z) is a symmetric bell over x,z plane.





$$\tan\alpha = \frac{-\rho}{\sqrt{1-\rho^2}}$$

$z = \frac{-\rho}{\sqrt{1-\rho^2}} x$ $\quad \alpha = \tan^{-1}\left(\frac{-\rho}{\sqrt{1-\rho^2}}\right)$

$$P\left(X>0, Z > \frac{-\rho}{\sqrt{1-\rho^2}} X\right) = \frac{90 + |\alpha|}{360} = \boxed{\frac{90 + \left|\tan^{-1}\left(\frac{-\rho}{\sqrt{1-\rho^2}}\right)\right|}{360}}$$

10. Suppose we have a collection of n i.i.d variables, $X_1, X_2, \ldots, X_n \sim \mathcal{N}(0, 1)$. We are interested in the relationship between a single data point X_1 and the sample mean, $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.

(a) Find $Corr(X_1, \overline{X})$. Use this to identify the joint distribution of $(X_1, \overline{X})$. Be sure to state all parameters. Write your answer as a simplified fraction.

(b) Find $\mathbb{P}(\overline{X} > 0 \mid X_1 > 0)$. (Use the conditional probability rule.)

a)
$$\mathrm{Cov}(X_1, \bar{X}) = \sum_{i=1}^{n} \mathrm{Cov}(X_1, \tfrac{1}{n}X_i)$$
$$= \mathrm{cov}(X_1, \tfrac{1}{n}X_1)$$
$$= \frac{1}{n}\mathrm{Var}(X_1) = \boxed{\frac{1}{n}}$$

$$\mathrm{Corr}(X_1, \bar{X}) = \frac{\frac{1}{n}}{1 \cdot \left(\frac{1}{\sqrt{n}}\right)} = \boxed{\frac{1}{\sqrt{n}}}$$

$X_1 \sim N(0,1)$

$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$

$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\underbrace{\mathrm{Var}(X_1, \ldots, X_n)}_{= n\mathrm{Var}(X_1)} = \frac{\overbrace{\mathrm{Var}(X_1)}^{=1}}{n} = \frac{1}{n}$

$SD(\bar{X}) = \frac{1}{\sqrt{n}}$.

$\bar{X} \sim N(0, \frac{1}{n})$

$(X_1, \bar{X})$ is BV? $aX_1 + b\bar{X} = (a + \frac{b}{n})X_1 + \frac{b}{n}X_2 + \cdots + \frac{b}{n}X_n$

$\Rightarrow$ normal.

b) $P(\bar{X} > 0 \mid X_1 > 0) = \dfrac{P(\bar{X} > 0, X_1 > 0)}{\underbrace{P(X_1 > 0)}_{= 1/2}}$

Find $P(\bar{X} > 0, X_1 > 0)$

$(\bar{X}, X_1)$ BV $\Leftrightarrow$ $(\sqrt{n}\,\bar{X}, X_1)$ std BV

$$\underbrace{\sqrt{n}\,\bar{X}}_{\text{std units}} = \underbrace{\rho}_{= \frac{1}{\sqrt{n}}} X_1 + \underbrace{\sqrt{1-\rho^2}}_{= \sqrt{\frac{n-1}{n}}}\, Z$$

$$\Rightarrow \bar{X} = \frac{1}{n} X_1 + \frac{1}{\sqrt{n}}\sqrt{\frac{n-1}{n}}\, Z$$

$$P(\bar{X} > 0, X_1 > 0) = P\left(\frac{1}{n}X_1 + \frac{1}{\sqrt{n}}\sqrt{\frac{n-1}{n}}\, Z > 0,\ X_1 > 0\right)$$

$$= P\left(Z > \frac{-\frac{1}{n}X_1}{\frac{1}{\sqrt{n}}\sqrt{\frac{n-1}{n}}},\ X_1 > 0\right)$$

$$= P\left(Z > \frac{-X_1}{\sqrt{n-1}},\ X_1 > 0\right)$$

From this point we can copy from answer above with $\dfrac{-\rho}{\sqrt{1-\rho^2}} = \dfrac{-1}{\sqrt{n-1}}$

$$P\left(Z > \frac{-X_1}{\sqrt{n-1}},\ X_1 > 0\right) = \frac{90 + \left|\tan^{-1}\left(\frac{-1}{\sqrt{n-1}}\right)\right|}{360}$$

....

Then

$$\boxed{P(\bar{x} > 0 \mid x_1 > 0) \simeq 2 \cdot \frac{90 + \left|\tan^{-1}\left(\frac{-1}{\sqrt{n-1}}\right)\right|}{360}}$$