Stat 134 lec 40

Plan

Friday 4/27 — Finish 6.5
Practice final on website

Monday 4/28 — Adam has extended OH
10 – 12 SLC
Class time review.
Review 6.5 in section
(have 6.5 suggested problems)
for final

Wed, Fri TBA.

Sec 6.5 <u>Bivariate normal</u>

Last time:

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

we call the joint distribution X, Y the standard bivariate normal w/ correlation ρ

<u>Properties</u>

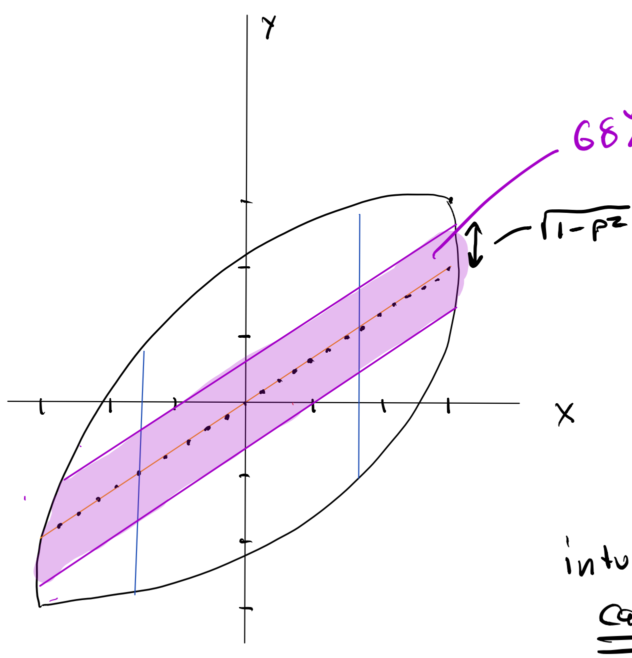
1) $Y \sim N(0,1)$

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

3) $Y \mid X=x \sim N(\rho x, 1-\rho^2)$

SD$(Y \mid X=x)$

So 68% of data is within $\sqrt{1-\rho^2}$ of regression line.

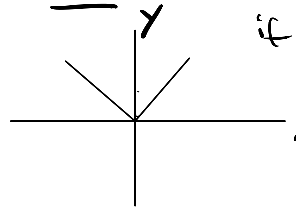


4) X, Y indep iff $\rho = 0$

but see next page for proof.

intuitively:

<u>can't</u> have scatter diagram



if x, y bivariate normal.

Theorem

If X, Y are standard bivariate w/ $corr(X,Y) = \rho$

X, Y are independent iff $\rho = 0$

Pf/

Recall if $W \sim N(\mu, \sigma^2)$, $f_W(w) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(w-\mu)^2}$

multiplication rule for density says

$$f(x,y) = f_X(x) f_{Y|X=x}(y)$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}\cdot\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(y-\rho x)^2}$$
$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2}{2} - \frac{1}{2(1-\rho^2)}(y-\rho x)^2}$$

Algebra

$$-\frac{x^2}{2} - \frac{1}{2(1-\rho^2)}(y-\rho x)^2 = \frac{-x^2(1-\rho^2) - y^2 + 2\rho xy - \rho^2 x^2}{2(1-\rho^2)}$$
$$= \frac{-x^2 + x^2\rho^2 - y^2 + 2\rho xy - \rho^2 x^2}{2(1-\rho^2)}$$
$$= -\frac{(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)}$$

so $f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)}$

is joint density for bivariate normal.

If $\rho = 0$ then $f(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$
$$= f_X(x) f_Y(y).$$
$\Rightarrow X, Y$ indep.

Conversely, if X, Y indep, $\rho = 0$ for any RV X, Y $\square$

<u>Theorem</u> Let $X, Y \sim N(0,1)$ and $\mathrm{Corr}(X,Y) = \rho$.
X, Y are standard bivariate normal correlation ρ
iff $aX + bY$ is normal for every constant a, b.

<u>Proof</u> (sketch)

$\Rightarrow$: Suppose X, Y std bivariate normal.
(i.e. $Y = \rho X + \sqrt{1-\rho^2}\, Z$ for X, Z iid $N(0,1)$)

$$aX + bY = aX + b\left(\rho X + \sqrt{1-\rho^2}\, Z\right)$$
$$= (a + b\rho) X + b\sqrt{1-\rho^2}\, Z$$

is normal since X, Z are independent normals.

$\Leftarrow$: Suppose $aX + bY$ is normal for any a, b.

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

Z is normal ? — yes by assumption.

$E(Z) = 0$

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

$$= \frac{1-\rho^2}{1-\rho^2} = 1$$

It remains to show that X and Z are independent. This is surprising since X is in the definition of Z! You can check that $Corr(X,Z)=0$ but this doesn't prove independence. The proof is beyond the scope of this class. The proof involves

1. Use the fact that $aX+bY$ is normal to find the joint density $f_{X,Y}$ (you need moment generating functions for this which you will learn about in Stat 135 or Stat 150).

2. Use the bivariate version of the change of variable rule to show that
$f_{X,Z}(x,z) = \frac{1}{2\pi} e^{-(x^2+z^2)}$. It follows then that
$f_{X,Z}(x,z) = f_X(x) f_Z(z)$ so X and Z are independent.

So finally we have, $X, Z \overset{iid}{\sim} N(0,1)$ and
$Y = \rho X + \sqrt{1-\rho^2}\, Z \Rightarrow X, Y$ are bivariate normal.

$\square$

Example of $X, Y \sim N(0,1)$ whose joint <u>isnt</u> std bivariate normal.

<u>ex</u> $X \sim N(0,1)$, $W = \begin{cases} 1 & \text{w/ prob } 1/2 \\ -1 & \text{w/ prob } 1/2 \end{cases}$

indep RVs.

let $Y = WX = \begin{cases} X & \text{if } W = 1 \\ -X & \text{if } W = -1 \end{cases}$

Next we show that $Y \sim N(0,1)$.

Note by change of variable rule that $-X \sim N(0,1)$.

Then
$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(Y \le y \mid W=1)P(W=1) + P(Y \le y \mid W=-1)P(W=-1) \\
&= \tfrac{1}{2}P(X \le y \mid W=1) + \tfrac{1}{2}P(-X \le y \mid W=-1) \\
&= \tfrac{1}{2}P(X \le y) + \tfrac{1}{2}P(-X \le y) \quad \text{since } X, W \text{ indep.} \\
&= \tfrac{1}{2}\Phi(y) + \tfrac{1}{2}\Phi(y) \quad \text{since } X, -X \sim N(0,1) \\
&= \Phi(y).
\end{aligned}
$$

Thus $Y \sim N(0,1)$.

Then $X + Y = \begin{cases} 2X & \text{with prob } 1/2 \\ 0 & \text{with prob } 1/2 \end{cases}$

is a mixed distribution, <u>not</u> normal.

So X, Y normal but $X+Y$ not normal.

Furthermore $\mathrm{Cor}(X,Y) = 0$ since

$E(YX) = E(WX^2) = \underset{\overset{\shortparallel}{0}}{E(W)}E(X^2)$ by indep of W and X.

but X, Y are clearly dependent contradicting property of bivariate normal.

bivariate normal X, Y with correlation ρ

Def^n

RVs X, Y are said to have the bivariate normal distribution with parameters $(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$ if

$$X^* = \frac{X - \mu_X}{\sigma_X}, \quad Y^* = \frac{Y - \mu_Y}{\sigma_Y} \quad \text{have}$$

the std bivariate normal dist w/ correlation ρ.

Note

From lec 38 we saw $\mathrm{Corr}(2X+3, 4Y-5) = \mathrm{Corr}(X, Y)$

$$\rho = \mathrm{Corr}(X^*, Y^*) = \mathrm{Corr}\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right)$$

$$= \mathrm{Corr}\left(\frac{1}{\sigma_X}X - \frac{\mu_X}{\sigma_X}, \frac{1}{\sigma_Y}Y - \frac{\mu_Y}{\sigma_Y}\right)$$

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

Thm

X, Y is bivariate normal w/ corr ρ

iff $aX + bY$ is normal for any $a, b \in \mathbb{R}$,

Thm

X, Y is bivariate normal w/ corr ρ

iff there exists $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

such that

$$X = c_1 Z_1 + c_2 Z_2$$

$$Y = c_3 Z_1 + c_4 Z_2$$

for some constants c_1, c_2, c_3, c_4.

Proof/ let $Z_1 = X^*$

$$Z_2 = \frac{-\rho}{\sqrt{1-\rho^2}} X^* + \frac{1}{\sqrt{1-\rho^2}} Y^*$$

Check $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

$$X = \sigma_X Z_1 + \mu_X$$

$$Y = \sigma_Y \left(\rho Z_1 + \sqrt{1-\rho^2}\, Z_2\right) + \mu_Y$$

Stat 134

Friday April 27 2018

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

   a true

   b false

$a(2x+3y+4)+b(6x-y-4)$

is a linear combo of

x, y hence normal.

Note

we see from this that

if x, y are bivariate normal that

$c_1 X + c_2 Y + c_3$, $c_4 X + c_5 Y + c_6$

are bivariate normal, in

other words linear combinations

of bivariate normal is bivariate

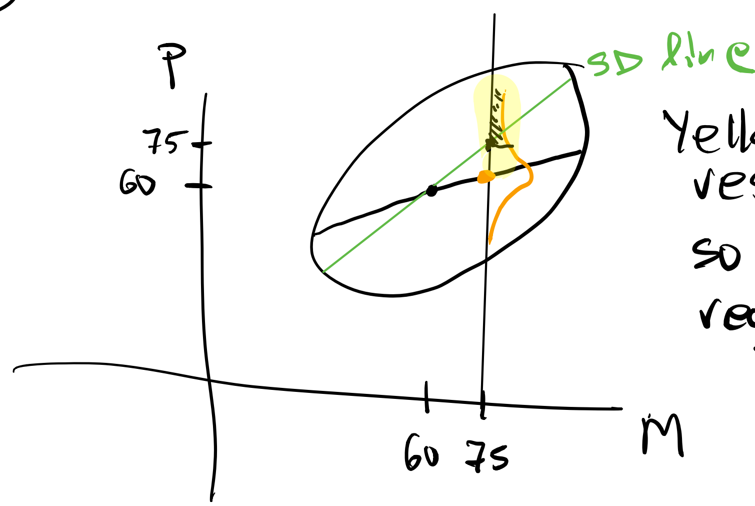
normal.

2. A test score in Math and Physics is bivariate normal, $\rho > 0$. The average is 60 on both tests and the SDs are the same. Of students scoring 75 on the Math test:

   a about half scored over 75 on Physics

   b more than half scored over 75 on Physics

   c less than half scored over 75 on Physics

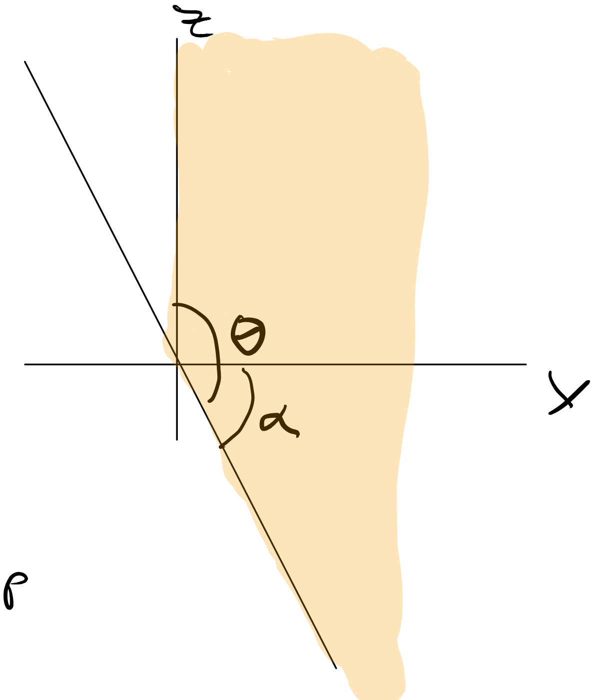


Yellow shaded region is 50%
so black shaded region < 50%

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

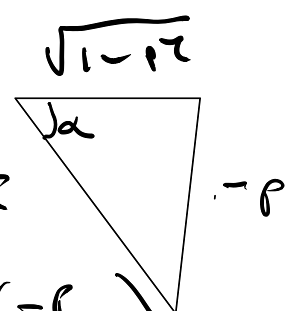
$$P(X>0,\ Y>0) = P\left(X>0,\ \rho X + \sqrt{1-\rho^2}\,Z > 0\right)$$

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



$$= \frac{|\theta^\circ|}{360}$$

$$|\theta| = 90^\circ + |\alpha|$$



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

$$\alpha = \arctan\left(\frac{-\rho}{\sqrt{1-\rho^2}}\right)$$

$$= \boxed{\frac{90^\circ + \left|\arctan\left(\frac{-\rho}{\sqrt{1-\rho^2}}\right)\right|}{360}}$$

Another example

ex M = midterm score

F = final score

M and F have bivariate $(66, 73, 10^2, 8^2, 0.6)$

let overall score $S = .3M + .7F$

What is joint distribution of F and S?

Need 5 param: $\mu_F = 73$

$\sigma_F^2 = 8^2$

$\mu_S = 0.3(66) + (.7)(73) = 70.9$

σ_S^2 — ? — derived below

corr (F, S) — ? — derived below

$$\text{Cov}(M, F) = \text{corr}(M, F)\,\sigma_M \sigma_F$$
$$= .6(10)(8) = 48$$

$$\text{Cov}(F, S) = \text{Cov}(F, S) = \text{Cov}(F, .3M + .7F)$$
$$= .3\,\text{Cov}(F, M) + .7\,\text{Cov}(F, F)$$
$$= .3(48) + .7(8^2)$$
$$= \boxed{59.2}$$

$$\mathrm{Var}(S) = \mathrm{Var}(0.3M + 0.7F)$$

$$= 0.3^2\,\mathrm{Var}(M) + (0.7)^2\,\mathrm{Var}\,F + 2\mathrm{Cov}(0.3M, 0.7F)$$

$$= 0.3^2(10^2) + 0.7^2(8^2) + 2(0.3)(0.7)(48)$$

$$= 60.5$$

$$\mathrm{Corr}(F, S) = \frac{\mathrm{Cov}(F, S)}{SD(F)\,SD(S)} = \frac{59.2}{\sqrt{60.2}} = 0.95$$