Lecture 37

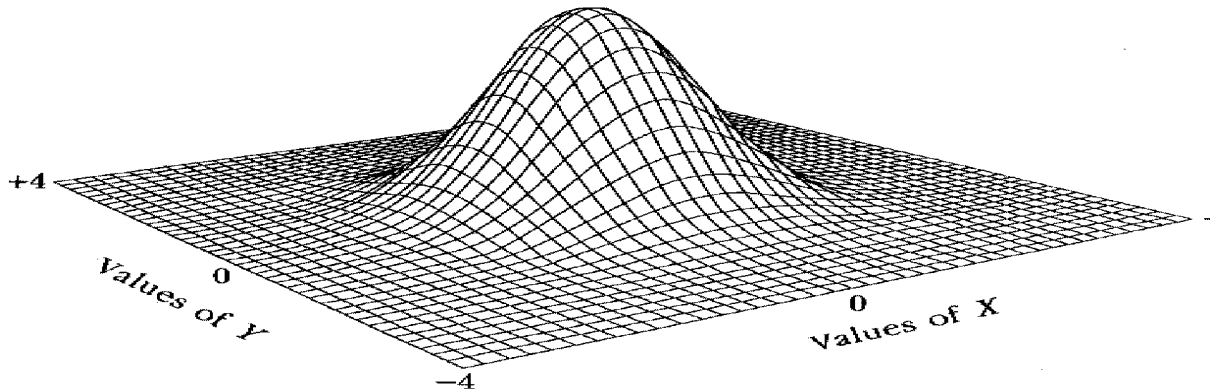
Regression to the mean

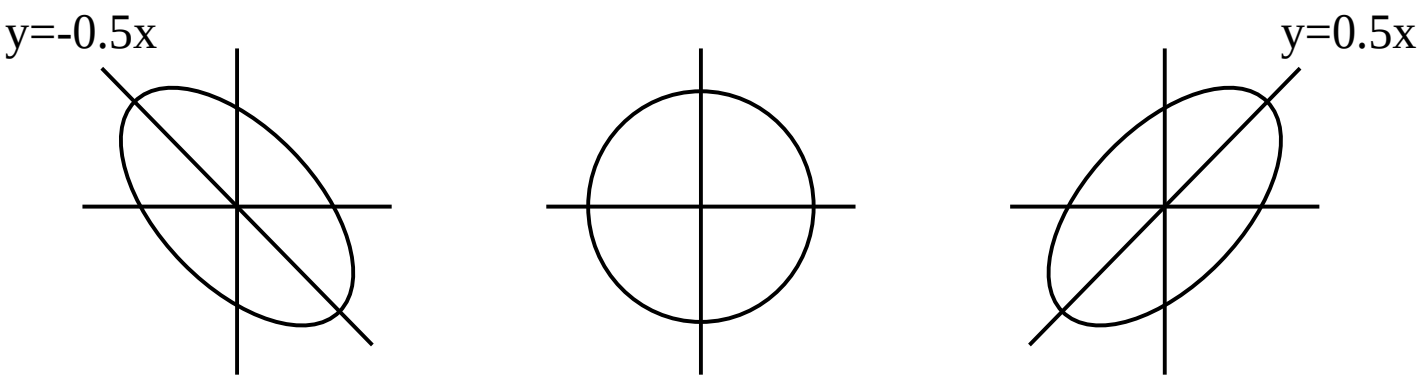
Computing some probabilities

Polar coordinates and simulation

Normal graphics

FIGURE 1. Perspective plot of the joint density of X and Y.





ρ=-0.5 ρ=0 ρ=0.5

General bivariate normal

- X and Y have a general normal distribution with parameters μ_X, μ_Y, σ_X^2, σ_Y^2 and ρ $\iff$ the normalised random variables

$$U = \frac{X - \mu_X}{\sigma_X} \qquad V = \frac{Y - \mu_Y}{\sigma_Y}$$

  have a standard bivariate normal distribution with correlation ρ where

$$\rho = Corr(U, V) = Corr(X, Y)$$

- Note:

$$X = \sigma_X U + \mu_X, \qquad Y = \sigma_Y V + \mu_Y$$

- Proofs by standard transformation calculations.

Computing some probabilities

- Take (X, Y) bivariate normal, means zero, unit variance, and correlation ρ.
- Find $P\{X > 0, Y > 0\}$
- Note, in general, this is equivalent to the probability that two correlated Gaussians are both greater than their expectations. (Independent of variances!)
- e.g. Both father and son are taller than average.
- If $\rho = 0$, then $P\{X > 0, Y > 0\} = 1/4$, either by product of marginals or computing a 2-dimensional integral.

- We reduce to the following problem:

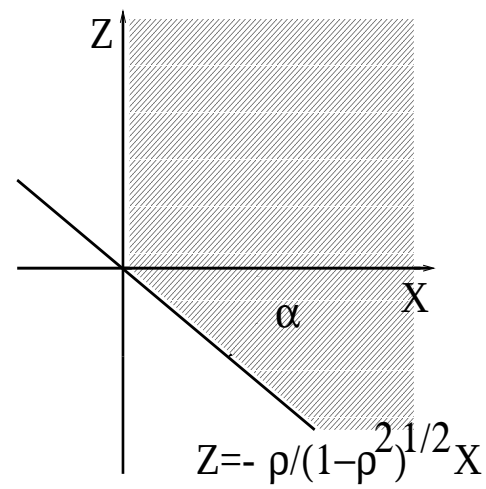
  - Take $Z \sim N(0,1)$, independent of X
  - Consider Y to be $Y = \rho X + \sqrt{1-\rho^2}Z$.
  - Recall: $Corr(X,Y) = \rho$.
  - We need to find

$$
\begin{aligned}
P\{X > 0,\ Y > 0\} &= P\{X > 0,\ \rho X + \sqrt{1-\rho^2}Z > 0\} \\
&= P\left\{X > 0,\ Z > \frac{-\rho X}{\sqrt{1-\rho^2}}\right\}
\end{aligned}
$$

- But this is a problem about uncorrelated, independent standard normals!

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

- Note that $\tan\alpha = -\rho/\sqrt{1-\rho^2}$
- Suppose $\rho = 0.5$. (Fathers and sons)
- $\Rightarrow\ \ \alpha = -30^o$
- $\Rightarrow\ \ P\{X > 0,\ Y > 0\} = 1/3$



Inverse problem

- $P\{X > 0,\ Y > 0\} = 3/8$. Find ρ

$$\frac{3}{8} = \frac{135^o}{360^o} \Rightarrow \alpha = -45^o \Rightarrow \frac{-\rho}{\sqrt{1-\rho^2}} = \tan\alpha = 1 \Rightarrow \rho = \frac{1}{\sqrt{2}}$$

Fathers and Sons

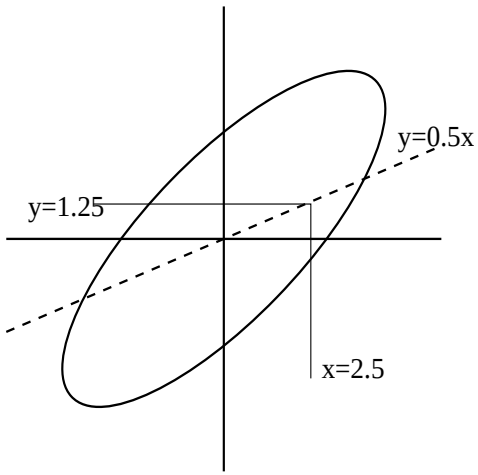
- Karl Pearson studied 1078 fathers and sons, and found the following re heights:
  - Fathers: Mean height=μ_F =175cm, σ_F = 6cm
  - Sons: Mean height=μ_S =178cm, σ_S = 6cm
  - Correlation between father and son heights: 0.5
- Predict the height of a son whose father is 190cm tall.
- To simplify, move to standard units, so that we want the height, in standard units, of a son whose father is

$$\frac{190 - 175}{6} = 2.5$$

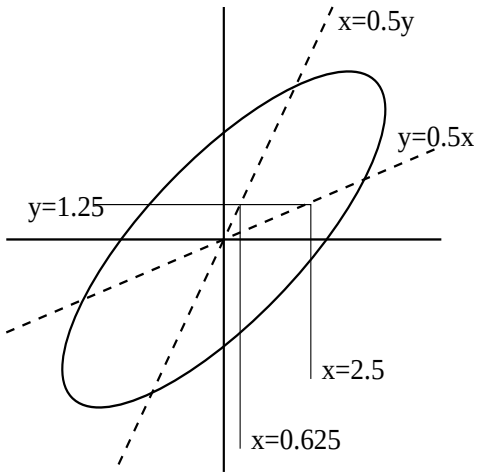
  standard units tall.
- Assume the joint distribution is bivariate normal.

Father $\longrightarrow$ son

- Set H_F for father's normalised height
- Set H_S for son's normalised height
- $E(H_S|H_F = h) \;=\; \rho\, h \;=\; 0.5\, h$
- $\Rightarrow$ prediction for son, in normalised units, is $0.5 \times 2.5 = 1.25$
- $\Rightarrow$ prediction for son, in regular units, is $178 + 1.25{\times}6 = 185.5$
- i.e. Tall, but not as tall as his father!



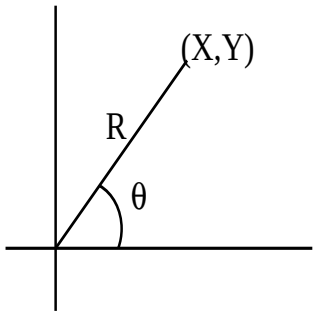
Son $\longrightarrow$ father



- Now look at a son of height 185.5cm and predict his father's height.
- $E(H_F|H_S = h) \;=\; \rho\, h \;=\; 0.5\, h$
- $\Rightarrow$ prediction for father, in normalised units, is $0.5 \times 1.25 = 0.625$
- $\Rightarrow$ prediction for father, in regular units, is $175 + 0.625{\times}6 = 182.5$
- i.e. The father has shrunk!
- Regression to the mean

Polar coordinates and simulation

- (X, Y) independent, standard, Normals.
- $R = \sqrt{X^2 + Y^2}$,



$$\theta = \begin{cases} \arctan(Y/X) & Y > 0 \\ \pi + \arctan(Y/X) & Y < 0 \end{cases}$$

$$X = R\cos\theta, \quad Y = R\sin\theta, \quad J = \det \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} = r$$

$$\begin{aligned} f_{XY}(x, y) &= \frac{1}{2\pi} e^{-(x^2+y^2)/2} \\ f_{R\theta}(r, \theta) &= f_{XY}(r\cos\theta, r\sin\theta) \cdot |J| \\ &= \frac{1}{2\pi}\ r\ e^{-r^2/2} \end{aligned}$$

Cont'd

$$f_{R\theta}(r,\theta) \quad = \quad \begin{cases} \frac{1}{2\pi}\ r\ e^{-r^2/2} & r > 0,\ \theta \in [0, 2\pi] \\ 0 & \text{otherwise} \end{cases}$$

- Marginal for R: (Rayleigh)

$$f_R(r) \ = \ \int_0^{2\pi} \frac{1}{2\pi}\ r\ e^{-r^2/2}\ d\theta \ = \ r\ e^{-r^2/2} \qquad r > 0$$

- Marginal for θ: $\theta \sim U[0, 2\pi]$.

- R and θ are independent

Simulating (X, Y) from independent uniforms (U, V):

- $\theta = 2\pi \times U$
- $F_R(r) \;=\; \int_0^r r\; e^{-r^2/2}\; dr \;=\; [-e^{-r^2/2}]_0^r = 1 - e^{-r^2/2}$
- $F_R^{-1}(v) \;=\; \sqrt{-2\ln(1-v)}, \quad v \in [0,1]$
- $\Rightarrow \sqrt{-2\ln(1-V)}$ has a Rayleigh distribtion
- We already know that

$$(R\cos\theta,\;\; R\sin\theta)$$

  is distributed like (X, Y): i.e. two independent normals

- In summary:

$$(X, Y) \;\sim\; \left(\sqrt{-2\ln(1-V)}\cdot\cos(2\pi U),\;\; \sqrt{-2\ln(1-V)}\cdot\sin(2\pi U)\right)$$