

Lectures prepared by:
Elchanan Mossel
Yelena Shvets

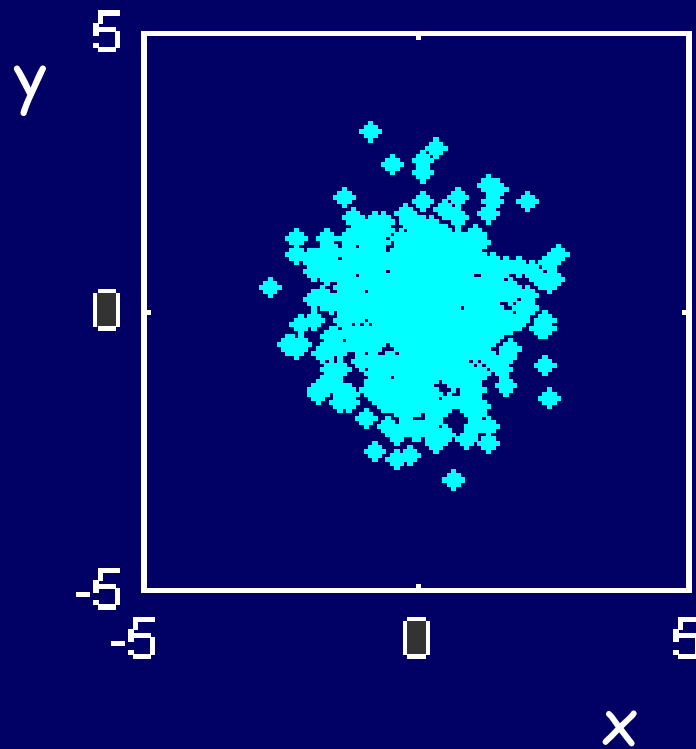
Follows Jim Pitman's
book:
Probability
Sections 6.5

Bivariate Normal

Let (X, Y) be independent Normal variables.

$$\text{The joint density: } f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2+y^2)}$$

$\rho = 0;$



Bivariate Normal

Question: What is the joint distribution of (X, Y)

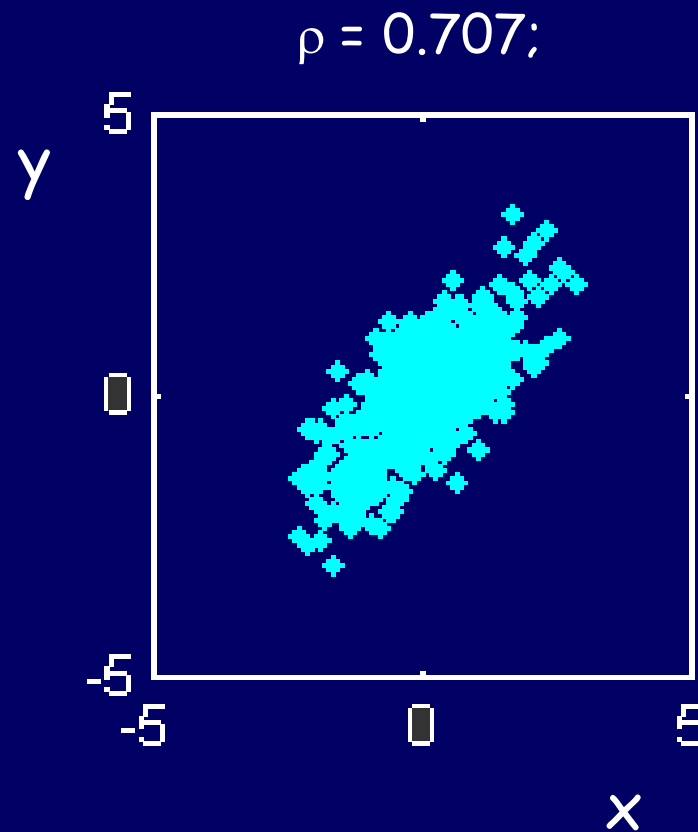
Where $X = \text{child height}$ and $Y = \text{parent height}$?

- We expect X and Y to be Gaussian
- However, X and Y are not independent:

$$\rho(X, Y) > 0.$$

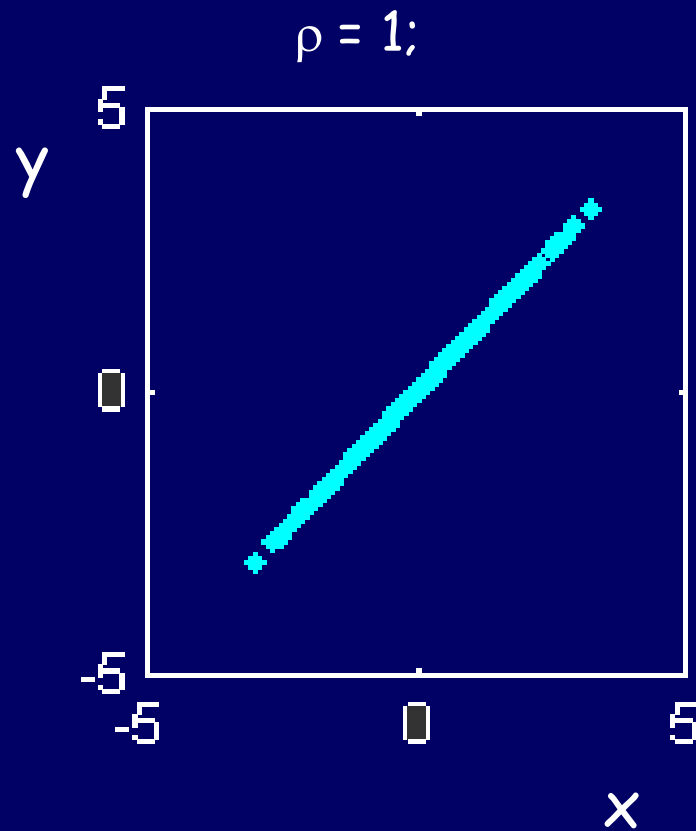
Bivariate Normal

Intuitive sample of Gaussian (X, Y) with $\rho = 0.7$



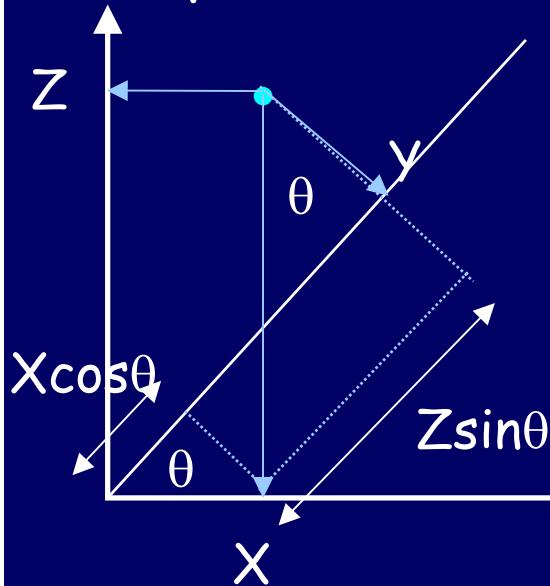
Bivariate Normal

Intuitive sample of X, Y with $\rho(X, Y) = 1$.



Construction of Bivariate Normal

Construction of correlated Gaussians: Let $X, Z \sim N(0,1)$ be independent.

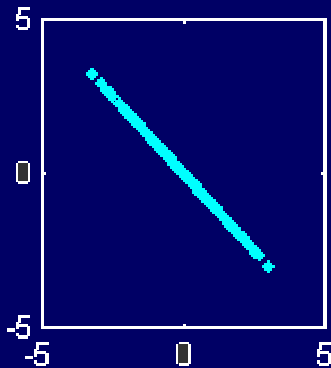


• Let: $Y = X \cos \theta + Z \sin \theta$;

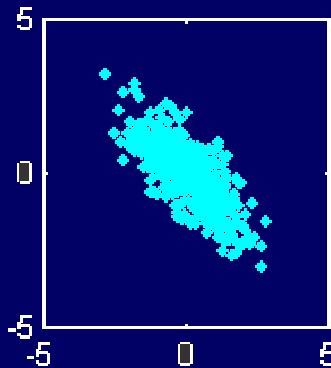
- Then: $E(Y) = E(X) = E(Z) = 0$
 $\text{Var}(Y) = \cos^2 \theta + \sin^2 \theta = 1$
 $\text{SD}(Y) = \text{SD}(X) = \text{SD}(Z) = 1$

$$\rho(X, Y) = E(XY) = E(X^2) \cos \theta + E(XZ) \sin \theta = \cos \theta$$

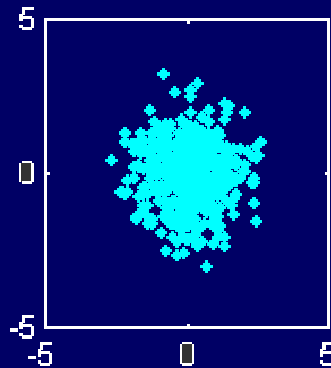
$$\rho = -1; \theta = \pi$$



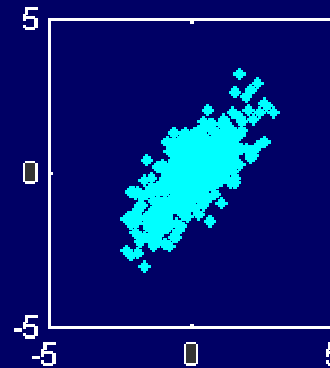
$$\rho = -0.707; \theta = 3\pi/4$$



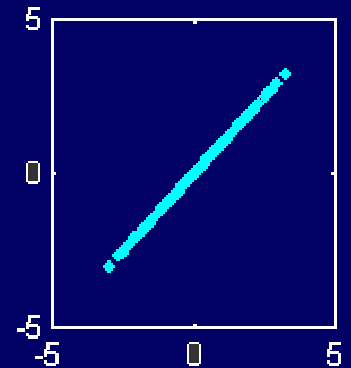
$$\rho = 0; \theta = \pi/2$$



$$\rho = 0.707; \theta = \pi/4$$



$$\rho = 1; \theta = 0$$



Standard Bivariate Normal

Def: We say that X and Y have **standard bivariate normal distribution** with correlation ρ if and only if

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

Where X and Z are independent $N(0,1)$.

Claim: If (X,Y) is ρ correlated bivariate normal then:

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

Marginals:

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

Conditionals:

$$X|Y=y \sim N(\rho y, 1-\rho^2); \quad Y|X=x \sim N(\rho x, 1-\rho^2)$$

Independence:

X, Y are independent if and only if $\rho = 0$

Symmetry:

(Y, X) is ρ correlated bivariate normal.

Bivariate Normal Distribution

Definition: We say that the random variables U and V have bivariate normal distribution with parameters μ_U , μ_V , σ^2_U , σ^2_V and ρ if and only if the standardized variables

$$X = (U - \mu_U) / \sigma_U \quad Y = (V - \mu_V) / \sigma_V$$

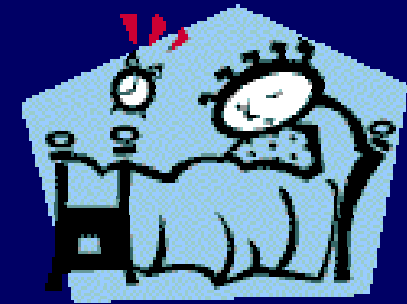
have standard bivariate normal distribution with correlation ρ .

Errors in Voting Machines



There are two candidates in a certain election: Mr.B and Mr.K. We use a voting machine to count the votes. Suppose that the voting machine at a particular polling station is somewhat capricious - it flips the votes with probability ε .

A voting Problem



- Consider a vote between two candidates where:

- At the morning of the vote:
Each voter tosses a coin and votes according to the outcome.

- Assume that the winner of the vote is the candidate with the majority of votes.

- In other words let $V_i \in \{\pm 1\}$ be the vote of voter i . So

- if $V = \sum_i V_i > 0$ then $+1$ wins;
- otherwise -1 wins.



A mathematical model of **voting machines**

Which voting schemes are more **robust** against noise?

Simplest model of noise: The voting machine flips each vote independently with probability ϵ .

Intended vote



Registered vote

prob ϵ



+1

prob $1-\epsilon$



-1

prob ϵ



-1

prob $1-\epsilon$



+1

On Errors in voting machines

Question: Let $V_i =$ intended vote i .

$W_i =$ registered vote i .

What is dist of $V = \sum_{i=1}^n V_i$, $W = \sum_{i=1}^n W_i$ for large n ?

Answer: $V \sim N(0,n)$; $W \sim N(0,n)$.

Question: What is $P[V > 0]$? $P[W > 0]$?

Answer: $\sim \frac{1}{2}$.

Question: What is the probability that machine errors flipped the elections outcome?

Answer: $P[V > 0, W < 0] + P[V < 0, W > 0]$?
 $= 1 - 2 P[V > 0, W > 0]$.

On Errors in voting machines

Answer continued: Take $(X, Y) = (V, W)/n^{1/2}$.

Then (X, Y) is bivariate-normal where $SD(X) = SD(Y) = 1$ and

$$\rho = \rho(X, Y) = \rho(V, W)/n = \rho(V_i, W_i) = 1 - 2\varepsilon.$$

Need to find $1 - 2P[X > 0, Y > 0]$.

Answer continued: Let $\rho = \cos \theta$. Then we need to find:

$$\begin{aligned} P[X > 0, Y > 0] &= P[X \cos 0 + Z \sin 0 > 0, X \cos \theta + Z \sin \theta > 0] = \\ &= P[0 < \text{Arg}(X, Z) < \pi, -\theta < \text{Arg}(X, Z) < \pi - \theta] \\ &= P[0 < \text{Arg}(X, Z) < \pi - \theta] = \frac{1}{2}((\pi - \theta)/\pi) \\ &= \frac{1}{2}(1 - (\arccos \rho)/\pi). \end{aligned}$$

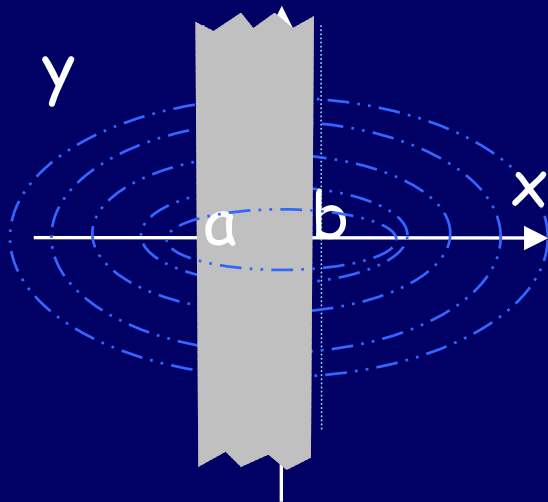
$$\begin{aligned} \text{So } 1 - 2P[X > 0, Y > 0] &= (\arccos \rho)/\pi \\ &= (\arccos(1 - 2\varepsilon))/\pi \end{aligned}$$



Conditional Expectation given an interval

Suppose that (X, Y) has the Standard Bivariate Normal Distribution with correlation ρ .

Question: For $a < b$, what is the $E(Y|a < X < b)$?



Solution:

$$\begin{aligned} E(Y|a < X < b) &= \frac{\int_a^b \int_{-\infty}^{\infty} y f(x, y) dy dx}{\int_a^b \int_{-\infty}^{\infty} f(x, y) dy dx} \\ &= \frac{\int_a^b \int_{-\infty}^{\infty} y f_y(y|X = x) f_X(x) dy dx}{\int_a^b f_X(x) dx} \\ &= \frac{\int_a^b E(Y|X = x) f_X(x) dy dx}{\int_a^b f_X(x) dx}. \end{aligned}$$

Conditional Expectation given an interval

Solution: $E(Y | a < X < b) = \int_a^b E(Y | X = x) f_X(x) / \int_a^b f_X(x) dx$

We know that $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $\int_a^b f_X(x) dx = \Phi(b) - \Phi(a)$.

Since $Y = \rho X + \sqrt{1 - \rho^2} Z$, where X & Z are independent

We have: $E(Y|X=x) = E(\rho X + \sqrt{1 - \rho^2} Z | X=x) = \rho x$

$$\begin{aligned} E(Y|a < X < b) &= \frac{\rho}{\sqrt{2\pi}} \int_a^b x e^{-\frac{x^2}{2}} dx / (\Phi(b) - \Phi(a)) \\ &= \frac{\rho}{\sqrt{2\pi}} [e^{-\frac{a^2}{2}} - e^{-\frac{b^2}{2}}] / (\Phi(b) - \Phi(a)) \end{aligned}$$