# STAT 345/445 Lecture 16

## Multiple Random Variables

Covariance and correlation – Section 4.5
Beyond $n = 2$ – Section 4.6

# Covariance and Correlation

## Definition

Let $(X, Y)$ be a random vector with $E(X) = \mu_X$, $\mathrm{Var}(X) = \sigma_X^2$, $E(Y) = \mu_Y$, $\mathrm{Var}(Y) = \sigma_Y^2$. We define

$$\mathrm{Cov}(X, Y) = E\left(\overbrace{(X - \mu_X)(Y - \mu_Y)}^{g(X,Y)}\right)$$

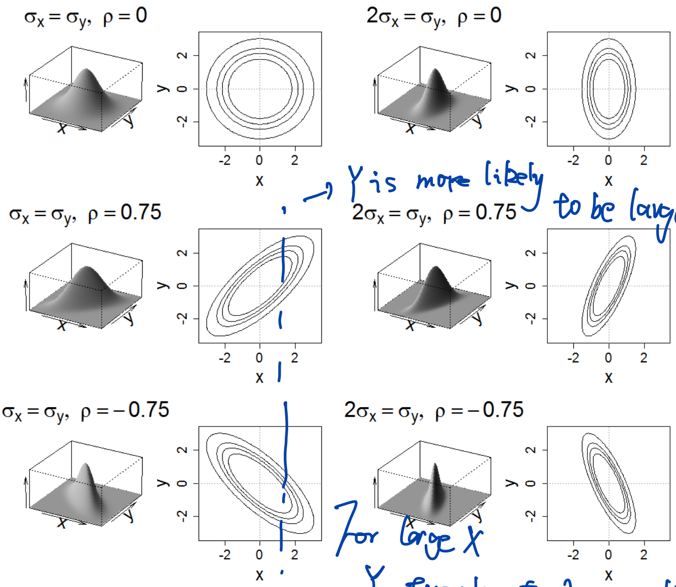
$$= \iint (x-\mu_x)(y-\mu_y) f(x,y)\, dx\, dy$$

$$\mathrm{Cor}(X, Y) = \rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

- Measures of the strength of a *linear* relationship between two random variables

Note:

$$\mathrm{Cov}(X, X) = E\left(\overbrace{(X - \mu_X)(X - \mu_X)}^{= (X-\mu_X)^2}\right) = \mathrm{Var}(X)$$




Y is more likely to be large when X is large

For large X Y tends to be small

From "Analysis of Neural Data" by Kass, Eden, and Brown

# Covariance

Similar to $V(X) = E(X^2) - \mu_X^2$

**Theorem**

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

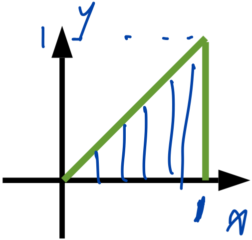
Proof... $\mathrm{Cov}(X, Y) = E((X-\mu_X)(Y-\mu_Y))$

$$= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y)$$

$$= E(XY) - \mu_Y \underbrace{E(X)}_{=\mu_X} - \mu_X \underbrace{E(Y)}_{=\mu_Y} + \mu_X\mu_Y$$

$$= E(XY) - \mu_X\mu_Y$$

# Continuous example 2




$$f(x,y) = \begin{cases} 8xy & \text{if } 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = 4x^3 \; I_{[0,1]}(x)$$

$$f_Y(y) = 4(y - y^3) \; I_{[0,1]}(y)$$

- Find the covariance and correlation for this pdf
- First find $\mu_X$ and $\mu_Y$:

$$\mu_X = \int_0^1 x \; 4x^3 \; dx =$$

$$\mu_Y = \int_0^1 y \; 4(y - y^3) \; dy =$$

# Continuous example 2

$$f(x,y) = \begin{cases} 8xy & \text{if } 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = 4x^3 \; I_{[0,1]}(x)$$

$$f_Y(y) = 4(y - y^3) \; I_{[0,1]}(y)$$

- Find the covariance and correlation for this pdf
- First find $\mu_X$ and $\mu_Y$:

$$\mu_X = \int_0^1 x \, 4x^3 \, dx = \left. \frac{4x^5}{5} \right|_0^1 = \frac{4}{5}$$

$$\mu_Y = \int_0^1 y \, 4(y - y^3) \, dy =$$

# Continuous example 2

$$f(x,y) = \begin{cases} 8xy & \text{if } 0 < y < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_X(x) = 4x^3 \; I_{[0,1]}(x)$$

$$f_Y(y) = 4(y - y^3) \; I_{[0,1]}(y)$$

- Find the covariance and correlation for this pdf
- First find $\mu_X$ and $\mu_Y$:

$$\mu_X = \int_0^1 x \, 4x^3 \; dx = \left. \frac{4x^5}{5} \right|_0^1 = \frac{4}{5}$$

$$\mu_Y = \int_0^1 y \, 4(y - y^3) \; dy = \left. \frac{4y^3}{3} - \frac{4y^5}{5} \right|_0^1 = \frac{8}{15}$$

# Continuous example 2 - continued

$x \in (0,1)$
$y \in (0,x)$

Next:

$$E(XY) = \iint xy\, f(x,y)\,dx\,dy = \iint xy \cdot 8xy\,dx\,dy$$

$$= \iint 8x^2y^2\,dx\,dy = \int_0^1 8x^2 \int_0^x y^2\,dy\,dx$$

$$= \int_0^1 8x^2 \cdot \left[\tfrac{1}{3}x^3 - 0\right] dx$$

$$= \int_0^1 \frac{8}{3}x^5\,dx$$

$$= \frac{8}{18}x^6\Big|_0^1 = \frac{8}{18} = \frac{4}{9}$$

positive. Large or small covariance?

units of cov = unit of X · unit of Y

$\Rightarrow \mathrm{Cov}(X, Y) =$

$= \frac{4}{9} - \frac{4}{5} \cdot \frac{8}{15} = \frac{4}{225}$

# Continuous example 2 - continued

To find the correlation we need the marginal variances:

$$\sigma_X^2 = \int_0^1 x^2\, 4x^3\, dx - \frac{4^2}{5^2} = \frac{2}{75}$$

$$\sigma_Y^2 = \int_0^1 y^2\, 4(y - y^3)\, dy - \frac{8^2}{15^2} =$$

$$\Rightarrow \mathrm{Cor}(X, Y) =$$

# Continuous example 2 - continued

To find the correlation we need the marginal variances:

$$\sigma_X^2 = \int_0^1 x^2\, 4x^3\, dx - \frac{4^2}{5^2} = \left.\frac{4x^6}{6}\right|_0^1 - \frac{4^2}{5^2} = \frac{4}{6} - \frac{4^2}{5^2}$$
$$= \frac{2}{75} = 0.02666667$$
$$\sigma_Y^2 = \int_0^1 y^2\, 4(y - y^3)\, dy - \frac{8^2}{15^2} =$$

$$\Rightarrow \mathrm{Cor}(X, Y) =$$

# Continuous example 2 - continued

To find the correlation we need the marginal variances:

$$\sigma_X^2 = \int_0^1 x^2\, 4x^3\, dx - \frac{4^2}{5^2} = \left.\frac{4x^6}{6}\right|_0^1 - \frac{4^2}{5^2} = \frac{4}{6} - \frac{4^2}{5^2}$$
$$= \frac{2}{75} = 0.02666667$$
$$\sigma_Y^2 = \int_0^1 y^2\, 4(y - y^3)\, dy - \frac{8^2}{15^2} = \left.\frac{4y^4}{4} - \frac{4y^6}{6}\right|_0^1 - \frac{8^2}{15^2}$$
$$= \frac{1}{3} - \frac{64}{225} = \frac{11}{225} = 0.04888889$$

$$\Rightarrow \mathrm{Cor}(X, Y) = \frac{0.01778}{\sqrt{\frac{2}{75}} \cdot \sqrt{\frac{11}{225}}} = \cdots$$

# 3 coins example

| x | y: 0 | 1 | 2 | 3 | $f_X(x)$ |
|---|---|---|---|---|---|
| 0 | $\frac{1}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | 0 | $\frac{1}{2}$ |
| 1 | 0 | $\frac{1}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | $\frac{1}{2}$ |
| $f(y) =$ | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ | |

- We know that the marginal distributions are

$\mu_Y = 3 \cdot 0.5 = 1.5$ $\qquad$ $\mu_X = 0.5$

$$Y \sim \text{Binomial}(3, 0.5) \qquad \text{and} \qquad X \sim \text{Bernoulli}(0.5)$$

$\sigma_Y = \sqrt{3 \cdot 0.5 \cdot 0.5} = \frac{1}{2}\sqrt{3}$ $\qquad$ $\sigma_X = \sqrt{\frac{1}{4}} = \frac{1}{2}$

- Find the covariance and correlation for this pmf

$E(XY) = \sum_{\text{all } x,y} xy\, f(x,y) = \sum_{x=0}^{1} \sum_{y=0}^{3} xy\, f(x,y)$

$$= 0 \cdot 0 \cdot \frac{1}{8} + 0 \cdot 1 \cdot \frac{2}{8} + 0 \cdot 2 \cdot \frac{1}{8} + 0 \cdot 3 \cdot 0 +$$

$$1 \cdot 0 \cdot 0 + 1 \cdot 1 \cdot \frac{1}{8} + 1 \cdot 2 \cdot \frac{2}{8} + 1 \cdot 3 \cdot \frac{1}{8} = \frac{8}{8} = 1$$

$$\Rightarrow \quad Cov(x, y) = E(XY) - \mu_x \mu_y$$

$$= 1 - 1.5 \cdot 0.5 = 0.25$$

$$Cor(x, y) = \frac{Cov(x, y)}{\sigma_x \cdot \sigma_Y} = \frac{0.25}{\frac{\sqrt{3}}{2} \cdot \frac{1}{2}} = \frac{1}{\sqrt{3}}$$

$$= 0.577$$

# Variance and covariance

## Theorem

Let $(X, Y)$ be a random vector and let $a, b$ be constants. Then

$$\mathrm{Var}(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X, Y)$$

Proof... $= E\left( \left(aX+bY - E(aX+bY)\right)^2 \right)$

$= E\left( (aX+bY - a\mu_X - b\mu_Y)^2 \right)$

$= E\left(a^2(X-\mu_X)^2\right) + E\left(b^2(Y-\mu_Y)\right) + E\left(2ab(X-\mu_X)(Y-\mu_Y)\right)$

$= a^2 V(X) + b^2 V(Y) + 2ab\,Cov(X, Y)$

# Correlation

**Theorem**

Let $(X, Y)$ be a random vector and let $\rho_{XY} = \mathrm{Cor}(X, Y)$. Then

(a) $-1 \leq \rho_{XY} \leq 1$

(b) $|\rho_{XY}| = 1$ if and only if there exists a constant $a \neq 0$ and $b$ such that $P(Y = aX + b) = 1$

- What part (b) tells us:
  - $\rho_{XY} = 1$ or $\rho_{XY} = -1$ can only happen if there is an *exact* linear relationship between $X$ and $Y$
  - $\rho_{XY} = 1 \Leftrightarrow a > 0$ and $\rho_{XY} = -1 \Leftrightarrow a < 0$
    - sign of correlation = sign of the slope
- Proof of part (a) ...

Proof (a): $-1 \leq \rho_{XY} \leq 1$

$Var\left(\frac{1}{\sigma_X} X + \frac{1}{\sigma_Y} Y\right) \geq 0$, (Variance is always $\geq 0$)

$\Rightarrow \frac{1}{\sigma_X^2} V(X) + \frac{1}{\sigma_Y^2} V(Y) + 2 \frac{1}{\sigma_X} \frac{1}{\sigma_Y} Cov(X, Y) \geq 0$

$\Rightarrow 1 + 1 + 2\rho_{XY} \geq 0 \Rightarrow \rho_{XY} \geq -1$

$$Var\left(\frac{1}{\sigma_x}X-\frac{1}{\sigma_Y}Y\right)\geq 0 \Rightarrow \frac{1}{\sigma_x^2}V(X)+\left(-\frac{1}{\sigma_Y}\right)^2V(Y)$$

$$+2\frac{1}{\sigma_x}\left(-\frac{1}{\sigma_Y}\right)cov(X,Y)\geq 0$$

$$\Leftarrow$$

$$1+1-2\rho_{XY}\geq 0 \Rightarrow \rho_{XY}\leq 1$$

# Covariance and independence

**Theorem**

If $X$ and $Y$ are independent random variables then

$$\mathrm{Cov}(X, Y) = 0 \qquad \text{and} \qquad \rho_{XY} = 0$$

Recall: If independ

$E(g(X)h(Y))$

$= E(g(X)) \cdot E(g(Y))$

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

$= E(X) \cdot E(Y) - \mu_X\mu_Y = \mu_X\mu_Y - \mu_X\mu_Y$

$= 0$

- But: We can have $\mathrm{Cov}(X, Y) = 0$ even when $X$ and $Y$ are not independent (see example in book)

Soln: CounterExample: $\mu_X = 0$ $\mu_Y = 0$

$E(XY) \neq 0$

# Multivariate random vectors

- $n$-dimensional random vector

$$f(\underline{x}) = f(x_1, x_2, \ldots, x_n)$$

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)$$

bold

- Most things follow naturally from the $n = 2$ setting
- Example: $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5)$
  - (joint) marginal distribution of $(X_1, X_3, X_4)$:

$$f_{134}(x_1, x_3, x_4) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4, x_5)\, dx_2\, dx_5$$

  - (joint) conditional distribution of $(X_1, X_3)$ given $(X_2, X_4, X_5)$:

$$f(x_1, x_3 \mid x_2, x_4, x_5) = \frac{f(x_1, x_2, x_3, x_4, x_5)}{f(x_2, x_4, x_5)}$$

# Mutually independent random variables

- There are some generalizations for $n > 2$ to take note of.

## Definition

Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a random vector with joint pdf/pmf $f(x_1, x_2, \dots, x_n)$ and marginal pdfs/pmfs $f_1(x_1), f_2(x_2), \dots, f_n(x_n)$.

$X_1, X_2, \dots, X_n$ are called **mutually independent** random variables if

$$f(\underline{x}) = f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n) \qquad \forall \mathbf{x} \in \mathbb{R}^n$$

$\Rightarrow \operatorname{Cov}(X_i, X_j) = 0, \ \forall i \neq j.$

- If $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are random vectors with joint pdf/pmf $f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ and (joint) marginal pdfs/pmfs $f_1(\mathbf{x}_1), f_2(\mathbf{x}_2), \dots, f_n(\mathbf{x}_n)$ we say that $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are **mutually independent random vectors** if

$$f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \prod_{i=1}^{n} f_i(\mathbf{x}_i) \qquad \forall (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$$

# Mutually independent random variables

## Theorem

Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables. Then for any functions $g_1(\cdot) \ldots, g_n(\cdot)$

(i) $E(\overbrace{g_1(X_1)}^{X_1 \text{ only}}\overbrace{g_2(X_2)}^{X_2 \text{ only}} \cdots g_n(X_n)) = E(g_1(X_1))E(g_2(X_2)) \cdots E(g_n(X_n))$

(ii) $M_{X_1+X_2+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t) = (M_{X_1}(t))^n$, if all $X_i$ has same distribution

(iii) $M_{b+a_1X_1+a_2X_2+\cdots+a_nX_n}(t) = e^{tb}M_{X_1}(a_1t)M_{X_2}(a_2t) \cdots M_{X_n}(a_nt)$

(iv) $g_1(X_1), g_2(X_2), \ldots, g_n(X_n)$ are mutually independent

# Mutually independent random variables

**Theorem**

$X_1, X_2, \ldots, X_n$ are mutually independent random variables if and only if the joint pdf/pmf can be written as

$$f(x_1, x_2, \ldots, x_n) = g_1(x_1)g_2(x_2)\cdots g_n(x_n)$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

# Example

- Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables where
$$X_i \sim N(\mu_i, \sigma_i^2) \qquad i = 1, 2, \ldots, n$$
Find the distribution of
$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + b$$

Know: $M_{X_i} = e^{t\mu_i + t^2\sigma_i^2/2}$ for $i = 1, \ldots, n$.

$Y = b + a_1X_1 + \cdots + a_nX_n \Rightarrow M_Y(t) = e^{tb}\prod_{i=1}^{n} M_{X_i}(a_i t) = \rightarrow$

$= e^{tb}\prod_{i=1}^{n} e^{a_i t\mu_i + a_i^2 t^2 \sigma_i^2/2}$

$= \exp\left(tb + \sum_{i=1}^{n} a_i t\mu_i + \sum_{i=1}^{n} a_i^2 t^2 \sigma_i^2/2\right)$

$$= \exp\left(t\underbrace{(b + a_i \sum \mu_i)}_{\text{mean}} + t^2 \underbrace{a_i^2 \sum_{i=1}^{n} \sigma_i^2}_{\text{variance}} /2\right)$$

$$= \text{mgf of } N\left(b + \sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

distribution of $Y$

# Multivariate transformations

continuous case, one-to-one transformation

$\mathbb{R}^n \to \mathbb{R}^n$

- Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector with joint pdf $f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$

$\underline{g}(\underline{x}) = (g_1(\underline{x}), \ldots, g_n(\underline{x}))$ was to be one-to-one

- Want the joint pdf of $U_1 = g_1(\mathbf{X}), U_2 = g_2(\mathbf{X}), \ldots, U_n = g_n(\mathbf{X})$
- Find the inverse functions:

$$\begin{aligned} u_1 &= g_1(x_1, x_2, \ldots, x_n) \\ &\vdots \\ u_n &= g_n(x_1, x_2, \ldots, x_n) \end{aligned} \quad \Rightarrow \quad \begin{aligned} x_1 &= h_1(u_1, u_2, \ldots, u_n) \\ &\vdots \\ x_n &= h_n(u_1, u_2, \ldots, u_n) \end{aligned}$$

# Multivariate transformations

## continuous case, one-to-one transformation

- Then the joint pdf of **U** is

absolute value

$$f_{\mathbf{U}}(\mathbf{u}) = f_{\mathbf{X}}(\underbrace{h_1(\mathbf{u})}_{x_1}, \underbrace{h_2(\mathbf{u})}_{x_2}, \ldots, h_n(\mathbf{u}))\ |J|$$

 where $J$ is the Jacobian:

$$J = \det\left(\begin{bmatrix} \frac{\partial h_1(\mathbf{u})}{\partial u_1} & \cdots & \frac{\partial h_1(\mathbf{u})}{\partial u_n} \\ \frac{\partial h_2(\mathbf{u})}{\partial u_1} & \cdots & \frac{\partial h_2(\mathbf{u})}{\partial u_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n(\mathbf{u})}{\partial u_1} & \cdots & \frac{\partial h_n(\mathbf{u})}{\partial u_n} \end{bmatrix}\right)$$

n×n dim.