# Probability and Statistics I

## 21. Expected Values, Covariance, Correlation

# 5.2 Expected Values, Covariance, and Correlation

# 5.2.1 Expected Values

## Expected Values for Jointly Distributed RVs

Suppose that $X$ and $Y$ are jointly distributed discrete random variables with joint pmf $p(x, y)$. The expected value of any function of $X$ and $Y$, $h(X, y)$, is

$$E[h(X, Y)] = \sum_x \sum_y h(x, y) p(x, y)$$

## Expected Values for Jointly Distributed RVs

Suppose that $X$ and $Y$ are jointly distributed continuous random variables with pdf $f(x, y)$. The expected value of any function of $X$ and $Y$, $h(X, y)$, is

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f(x, y) \; dxdy$$

**Example 21.1 ctd**

The simplest possible joint distribution is that for two Bernoulli random variables. Suppose that $X$ and $Y$ take the values 0 and 1 according to the following joint pmf:

| $x$ | 0 | | 1 | |
|---|---|---|---|---|
| $y$ | 0 | 1 | 0 | 1 |
| $p(x,y)$ | $p_{00}$ | $p_{01}$ | $p_{10}$ | $p_{11}$. |

a) What is the expected value of $XY$?

# 5.2.2 Covariance and Correlation

**Covariance**

The covariance between two random variables $X$ and $Y$ is

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_Y)]$$

## Covariance

The covariance between two discrete random variables $X$ and $Y$ is

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_x)(Y - \mu_Y)] \\ &= \sum_x \sum_y h(x, y)p(x, y).\end{aligned}$$

The covariance between two continuous random variables $X$ and $Y$ is

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - \mu_x)(Y - \mu_Y)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x, y)f(x, y)\ dxdy.\end{aligned}$$

### Shortcut Formula for Covariances

The covariance of any two random variables $X$ and $Y$ is

$$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

It is often more efficient to use the shortcut formula

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

## Correlation

The correlation (coefficient) of $X$ and $Y$ is

$$\mathrm{Corr}(X, Y) = \rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

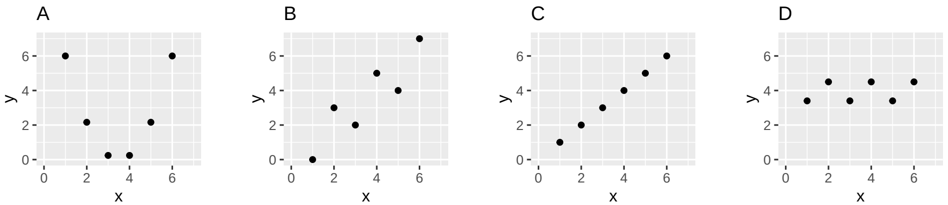
## Correlation

Correlation measures the strength of the *linear* relationship between two variables.

- $\text{Corr}(X, Y) = -1$: $Y = aX + b$ for some $a < 0$.
- $-1 < \text{Corr}(X, Y) < 0$: $Y \approx aX + b$ for some $a < 0$.
- $\text{Corr}(X, Y) = 0$: the best fitting line has slope 0.
- $0 < \text{Corr}(X, Y) < 1$: $Y \approx aX + b$ for some $a <> 0$.
- $\text{Corr}(X, Y) = 1$: $Y = aX + b$ for some $a > 0$.

## Example 21.1

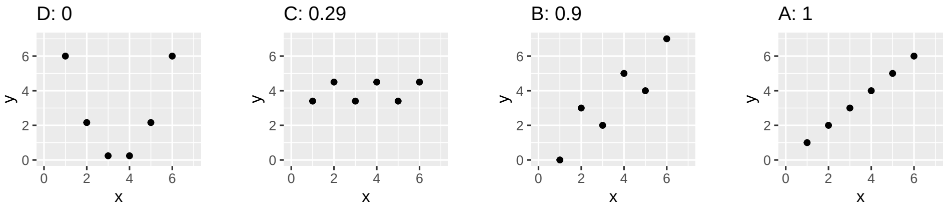
Each of the following plots represents the joint pmf of two random variables, $X$ and $Y$. The points, $(x, y)$ represent the possible values of $(X, Y)$. The distribution places equal probability, $1/6$, at each point.




Order the plots according to their correlation.

## Example 21.1 ctd

Each of the following plots represents the joint pmf of two random variables, $X$ and $Y$. The points, $(x, y)$ represent the possible values of $(X, Y)$. The distribution places equal probability, $1/6$, at each point.




Order the plots according to their correlation.

## Example 21.1 ctd

The simplest possible joint distribution is that for two Bernoulli random variables. Suppose that $X$ and $Y$ take the values 0 and 1 according to the following joint pmf:

| $x$ | 0 | | 1 | |
|---|---|---|---|---|
| $y$ | 0 | 1 | 0 | 1 |
| $p(x, y)$ | $p_{00}$ | $p_{01}$ | $p_{10}$ | $p_{11}$. |

1. What is the expected value of $XY$?
2. What are the covariance and correlation of $X$ and $Y$?

## 5.2.3 Sums of Random Variables

## Sums of Random Variables

Let $X_1, X_2, \ldots$ be any sequence of random variables and $c_1, c_2, \ldots$ be any sequence of constants. Then

$$E\left[\sum c_k X_k\right] = \sum c_k E(X_k).$$

The sum may either be finite or infinite.

In the case of two random variables

$$E(aX + bY) = aE(X) + bE(Y).$$

### Variance of a Sum of Random Variables

Let $X$ and $Y$ be any two random variables and $a, b \in \Re$. Then

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

If $X$ and $Y$ are independent then $\text{Cov}(X, Y) = 0$[1] In this case

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y).$$

[1] The reverse is **not** true in general.

## Example 21.1 ctd

The simplest possible joint distribution is that for two Bernoulli random variables. Suppose that $X$ and $Y$ take the values 0 and 1 according to the following joint pmf:

| $x$ | 0 | | 1 | |
|---|---|---|---|---|
| $y$ | 0 | 1 | 0 | 1 |
| $p(x,y)$ | $p_{00}$ | $p_{01}$ | $p_{10}$ | $p_{11}$. |

1. What is the expected value of $XY$?
2. What are the covariance and correlation of $X$ and $Y$?
3. What are the mean and variance of $Z = 2X + 4Y$?

# 5.2.4 Products of Independent Random Variables

## Products of Independent Random Variables

Let $X_1, X_2, \ldots$ be any set of independent random variables. Then

$$E\left[\prod X_k\right] = \prod E(X_k).$$

In the case of two independent random variables

$$E(XY) = E(X)E(Y).$$

## Example 21.1 ctd

The simplest possible joint distribution is that for two Bernoulli random variables. Suppose that $X$ and $Y$ take the values 0 and 1 according to the following joint pmf:

| $x$ | 0 | | 1 | |
|---|---|---|---|---|
| $y$ | 0 | 1 | 0 | 1 |
| $p(x,y)$ | $p_{00}$ | $p_{01}$ | $p_{10}$ | $p_{11}$. |

1. What is the expected value of $XY$?
2. What are the covariance and correlation of $X$ and $Y$?
3. What are the mean and variance of $Z = 2X + 4Y$?
4. Under what conditions are $X$ and $Y$ independent? What is are the mean $XY$ in this case?

# Questions?

## Exercise 21.1

Consider rolling two fair, three-sided die. Let $S$ denote the sum of the values showing on the two die and $D$ the absolute value of the difference. E.g., if one die shows the value 1 and the second shows the value 2 then $S = 3$ and $D = 1$, regardless of which was thrown first.

a) Construct a table showing the joint pmf of $S$ and $D$.
b) Compute the marginal pmf of both $S$ and $D$.
c) Compute the expected value and variance of $S$ and $D$.
d) Compute the covariance and correlation of $S$ and $D$.
e) Are $S$ and $D$ independent? Justify your answer.