Last time: Independence of r.v.
Expectations + joint distributions

\( \Theta \rightarrow E[X+Y] = E[X] + E[Y] \)

\( \times \times \rightarrow E[g(X) \cdot h(Y)] = E[g(X)] \cdot E[h(Y)] \)
when \( X \) and \( Y \) are independent.

Today: Covariance

Special case of \( \times \times \): \( E[X \cdot Y] = E[X] \cdot E[Y] \)
when \( X \) and \( Y \) are indep.
So $E[XY] - EEX \cdot EY$ is zero for indep. r.v.
We can use it as an indication of whether $X$ and $Y$ are indep.

**Definition:** The covariance of $X$ and $Y$ is

$$\text{Cov}(X, Y) = E[XY] - EEX \cdot EY.$$ 

We call $X$ and $Y$ uncorrelated if $\text{Cov}(X, Y) = 0.$

**Remarks:**

(*) independent $\Rightarrow$ uncorrelated
uncorrelated $\not\Rightarrow$ independent
If \( \text{Cov}(X,Y) > 0 \), we say that \( X \) and \( Y \) are **positively correlated**.

\[ \rightarrow \text{the larger} \ X \text{ is, the larger we expect} \ Y \text{ to be.} \]

If \( \text{Cov}(X,Y) < 0 \), we say that \( X \) and \( Y \) are **negatively correlated**.

\[ \rightarrow \text{the larger} \ X \text{ is, the smaller we expect} \ Y \text{ to be.} \]
Formulas:

\[ \text{Cov}(X, X) = \text{Var}(X) \]

\[ \text{Cov}(X, Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] \]

\[ \text{Cov}(aX+b, Y) = a \cdot \text{Cov}(X, Y) \]

\[ \text{Cov}(X, aY+b) = a \cdot \text{Cov}(X, Y) \]

\[ \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y) \]

In particular, \( \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) \) if \( X \) and \( Y \) are uncorrelated.
Example: Draw balls from an urn with 3 white and 2 black balls.

- $X = \#$ of draws until the 1st black.
- $Y = \#$ of draws between 1st and 2nd black.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$P(X = x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{3}{10}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10}$</td>
<td>0</td>
<td>$\frac{3}{10}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10}$</td>
<td>0</td>
<td>0</td>
<td>$\frac{3}{10}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{10}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{10}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$P(Y = y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\frac{4}{10}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{3}{10}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{2}{10}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{1}{10}$</td>
</tr>
</tbody>
</table>
Compute $E[XY]$ and $EX \cdot EY$.

$EX \cdot EY = \left( 1 \cdot \frac{y}{10} + 2 \cdot \frac{3}{10} + 3 \cdot \frac{2}{10} + 4 \cdot \frac{1}{10} \right)^2 = 4.$

$E[XY] = \frac{1}{10} \left( 1 \cdot 1 + 1 \cdot 2 + 1 \cdot 3 + 1 \cdot 4 + 2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3 + 3 \cdot 1 + 3 \cdot 2 + 4 \cdot 1 \right) = 3 \frac{1}{2}.$

$\Rightarrow Cov(X,Y) = E[XY] - EX \cdot EY = \square - \frac{1}{2}.$

$X$ and $Y$ are negatively correlated.

$E g(X,Y) = \sum_{x,y} g(x,y) \cdot P(X=x, Y=y)$
Example: Let \((X,Y)\) be the coordinates of a uniformly chosen point in the unit disk. We have seen that \(X\) and \(Y\) are not independent. Are they uncorrelated?

**Sol:** Joint p.d.f. \(f(x,y) = \left\{ \begin{array}{ll} \frac{1}{\pi} & \text{if } x^2+y^2 \leq 1 \\ 0 & \text{otherwise} \end{array} \right. \)

p.d.f.'s: \(f_X(x) = f_Y(y) = \left\{ \begin{array}{ll} \frac{2}{\pi \sqrt{1-x^2}} & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1 \end{array} \right. \)

\[ EX=EX = \int_{-1}^{1} x \cdot \frac{2}{\pi \sqrt{1-x^2}} \, dx = 0. \]
\[ E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x,y) \, dx \, dy \]

\[ = \iint xy \cdot \frac{1}{\pi} \, dx \, dy \]
\[ \{ x^2 + y^2 \leq 1 \} \]

\[ = \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} y \cdot dy \cdot dx = 0. \]

\[ \Rightarrow \text{Cov}(X,Y) = 0 - 0 = 0. \quad \rightarrow \quad X \text{ and } Y \text{ are uncorrelated.} \]
We have seen that the sign of $\text{Cov}(X, Y)$ has an intuitive interpretation.

On the other hand, the absolute value of $\text{Cov}(X, Y)$ does not have an immediate interpretation. This is because $\text{Cov}(aX, bY) = ab \cdot \text{Cov}(X, Y)$

For example: $X =$ time until next visit to hospital

$Y =$ time until your bike brakes fail.
We believe that $\text{Cov}(X,Y) > 0$.

If we measure time in days, we get some number $\text{Cov}(X,Y)$.

If we measure time in weeks,

then $X' = \frac{1}{7} X$

$Y' = \frac{1}{7} Y$ \rightarrow $\text{Cov}(X',Y') = \frac{1}{7^2} \cdot \text{Cov}(X,Y)$. 
Last time: Covariance

\[ \text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}X \cdot \mathbb{E}Y \]

Today: Correlation coefficient

Conditional probability distributions

Correlation coefficient:

Definition: The correlation coefficient of two random variables \( X \) and \( Y \) is

\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)} \]
Remark:

\[ \rho(aX, bY) = \frac{\text{Cov}(aX, bY)}{\sigma(aX) \cdot \sigma(bY)} \]

\[ = \frac{ab \cdot \text{Cov}(X, Y)}{a \sigma(X) \cdot b \sigma(Y)} = \rho(X, Y) \]

Example: Let \( X \sim \mathcal{N}(0, 2) \), \( Z \sim \mathcal{N}(0, 1) \) be independent.

Define \( Y = X + E \), \( E \sim \mathcal{N}(0, 1) \).
\[ \text{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y] = E[XY] \]

\[ = E[X(X + \varepsilon Z)] = E[X^2] + \varepsilon E[X \varepsilon Z] = E[X^2] + \varepsilon E[X] \cdot E[Z] = 2. \]

\( X, Z \) are indep.

\[ \Rightarrow \text{Var}(X) = 2, \quad \text{Var}(Y) = \text{Var}(X) + \text{Var}(\varepsilon Z) = 2 + \varepsilon^2. \]

\[ \Rightarrow \rho(X, Y) = \frac{2}{\sqrt{2(2 + \varepsilon^2)}} = \frac{1}{\sqrt{1 + \frac{\varepsilon^2}{2}}}. \]
Fact:

(i) $|\rho(x, y)| \leq 1$

(ii) If $\rho(x, y) = \pm 1$, then

$$y = ax + b \text{ for some } a, b \in \mathbb{R}$$

and $\text{sign}(a) = \text{sign} \rho(x, y)$
Conditional probability distributions

Discrete case: let $X, Y$ be two discrete r.v.

**Definition:** The conditional p.m.f. of $X$ given $Y$ is

$$P_{X|Y}(x | y) = \frac{P(X=x | Y=y)}{P(Y=y)}$$

**Example:** An urn with 3 white, 2 black balls.

$$P(\text{2nd black on last draw} | \text{1st black on 2nd draw}) = \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{3}$$

$$P_{X|Y}(2 | 3) = \frac{1/10}{3/10} = \frac{1}{3}.$$
Continuous case: let $X, Y$ jointly continuous.

**Definition:** The conditional p.d.f. of $X$ given $Y$ is

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)}$$

\footnote{We interpret $\frac{0}{0}$ as $0$ here.}
Example: Let \((X,Y)\) be the coordinates of a uniformly chosen point in the unit disk.

Joint p.d.f.: \(f(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}\)

Marginal: \(f_Y(y) = \begin{cases} \frac{2}{\pi} \sqrt{1-y^2} & \text{if } |y| \leq 1 \\ 0 & \text{otherwise} \end{cases}\)

\[ \Rightarrow f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{2\sqrt{1-y^2}} & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{if } x^2 + y^2 > 1 \end{cases} \]
Conditional expectations

Motivation: Suppose that $X$ is a "detailed" r.v., and that $Y$ is a "less detailed" r.v.

$\Rightarrow$ Assume you can only observe the outcome of $Y$.

How can we predict/estimate $X$ from the observation of $Y$?

$\Rightarrow$ We want to approximate $X$ by a r.v. of the form $g(Y)$. 
Conditional expectation

We have two r.v. $X$ and $Y$.
We would like to approximate $X$ by a r.v. of the form $g(Y)$.

$\Rightarrow$ Let's call $X - g(Y)$ the approximation error.

$\Rightarrow$ A good approximation should satisfy:

1. $\mathbb{E} [X - g(Y)] = 0$.
2. $\text{Var}(X - g(Y))$ should be minimal among all possible choices for $g$. 
Fact: The unique function \( g \) that satisfies (1) + (2) is given by

\[
g(y) = \begin{cases} 
\sum_x \times P_{X \mid Y}(x \mid y) & \text{discrete case} \\
\int_{-\infty}^{\infty} \times f_{X \mid Y}(x \mid y) \, dx & \text{cont. case}
\end{cases}
\]

We call \( g(Y) \) the \underline{condition expectation} of \( X \) given \( Y \).

We write \( g(y) = \mathbb{E}[X \mid Y = y] \)

\[
g(Y) = \mathbb{E}[X \mid Y]
\]
Proof that (1) holds: (discrete case)

\[ E \ g(Y) = \sum_{y} g(y) \cdot P(Y=y) \]

\[ = \sum_{y} \sum_{x} P_x(y|x) \cdot P(Y=y) \]

\[ = P(X=x|Y=y) \]

\[ = \frac{P(X=x, Y=y)}{P(Y=y)} \]

\[ = \sum_{x} \sum_{y} P(X=x, Y=y) \]

\[ = \sum_{x} P(X=x) \cdot \sum_{y} P(X=x, Y=y) = EX. \]
Remarks:

For fixed $y$,

$P_{X|Y}(x|y)$ is like a p.m.f for $X$.

$f_{X|Y}(x|y)$ is like a p.d.f for $X$,

and the condition expectation of $X$ given $Y=y$ is just the "usual" expectation of this cond. p.m.f./p.d.f instead of the original p.m.f./p.d.f.
\((\ast)\) \(g(Y)\) is a random variable.

That is, \(E[X \mid Y]\) is a r.v.

Moreover, we have just shown that

\[ E[g(Y)] = E[E[X \mid Y]] = EX \]

\(\rightarrow\) The expectation of the cond. expectation is the expectation.
Example: Let \((X,Y)\) be a uniform point in the triangle \[ T = \{(x,y) : 0 \leq x \leq y, 0 \leq y \leq 2\}. \]

Compute \(E[X|Y=1]\).

Sol: Intuition: given \(Y=1\), \(X\) behaves like a uniform point in \([0,1]\), since no point there is preferred over any other.

Thus, we expect that \(E[X|Y=1] = E[\text{Unif}[0,1]] = \frac{1}{2}\).
Let's show this by a calculation.

We need $f_{X|Y}(x|y)$ for $y=1$.

Joint p.d.f.: $f(x,y) = \begin{cases} \frac{1}{11} = \frac{1}{2} & \text{if } (x,y) \in T \\ 0 & \text{otherwise} \end{cases}$

Marginal: $f_Y(y) = \int_{-\infty}^{\infty} f(x,y) \, dx = \begin{cases} \frac{y}{2} & \text{if } 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$

Cond. p.d.f.: $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{y} & \text{if } (x,y) \in T \\ 0 & \text{otherwise} \end{cases}$
Observation: for fixed $0 < y < 2$, the cond. pdf. of $X$ given $Y = y$ is constant on $[0, y]$ and 0 outside this interval.

Thus, it describes the pdf. of a $\text{Unif}[0, y]$ r.v.

$$\Rightarrow \quad \mathbb{E}[X | Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) \, dx$$

$$\mathbb{E}[X | Y = y] = \frac{y}{2} \quad \text{if } y = 1,$$

then

$$\mathbb{E}[X | Y = 1] = \int_{0}^{1} x \cdot \frac{1}{y} \, dx = \frac{1}{y} \cdot \frac{x^2}{2} \bigg|_{0}^{1} = \frac{y}{2}.$$