Joint distributions: extensions

Chrysafis Vogiatzis

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 12

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

Last time..

Notation	Name	Calculation
$f_{XY}(x,y)$	Joint pmf	$P(X=x, Y=y) = f_{XY}(x,y)$ $P\left(\begin{array}{c}\alpha_1 \leq X \leq \beta_1 \\ \alpha_2 \leq Y \leq \beta_2\end{array}\right) = \sum_{x=\alpha_1}^{\beta_1} \sum_{y=\alpha_2}^{\beta_2} f_{XY}(x,y)$
	Joint pdf	$P(X=x, Y=y) = 0$ $P\left(\begin{array}{c}\alpha_1 \leq X \leq \beta_1 \\ \alpha_2 \leq Y \leq \beta_2\end{array}\right) = \int_{\alpha_1}^{\beta_1} \int_{\alpha_2}^{\beta_2} f_{XY}(x,y)dydx$
$f_X(x)$ or $f_Y(y)$	Marginal pmf	$f_X(x) = \sum_y f_{XY}(x,y)$ $P(\alpha_1 \leq X \leq \beta_1) = \sum_{x=\alpha_1}^{\beta_1} f_X(x)$
	Marginal pdf	$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy$ $P(\alpha_1 \leq X \leq \beta_1) = \int_{\alpha_1}^{\beta_1} f_X(x)dx$
$f_{X\|Y}(x)$ or $f_{Y\|X}(y)$	Conditional pmf	$f_{X\|Y}(x) = f_{XY}(x,y)/f_Y(y)$ $P(\alpha_1 \leq X \leq \beta_1 \| Y=y) = \sum_{x=\alpha_1}^{\beta_1} f_X(x)/f_Y(y)$
	Conditional pdf	$P(\alpha_1 \leq X \leq \beta_1 \| Y=y) = \int_{\alpha_1}^{\beta_1} f_X(x)dx/f_Y(y)$

Expectations and variances

We define two types of expectations and variances:

1. $E[X], E[Y]$ and $Var[X], Var[Y]$ for the marginal distribution.
2. $E[X|y], E[Y|x]$ and $Var[X|y], Var[Y|x]$ for the conditional distribution.

	Discrete	Continuous	
	$E[X] = \sum_x x f_X(x)$	$= \int_{-\infty}^{+\infty} x f_X(x) dx$	$= \mu_X$
	$Var[X] = \sum_x x^2 f_X(x) - \mu_X^2$	$= \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu_X^2$	$= \sigma_X^2$
	$E[X\|y] = \sum_x x f_{X\|y}(x)$	$= \int_{-\infty}^{+\infty} x f_{X\|y}(x) dx$	$= \mu_{X\|y}$
	$Var[X\|y] = \sum_x x^2 f_{X\|y}(x) - \mu_{X\|y}^2$	$= \int_{-\infty}^{+\infty} x^2 f_X\|y(x) dx - \mu_{X\|y}^2$	$= \sigma_{X\|y}^2$

ILLINOIS

Expectations and variances

We define two types of expectations and variances:

1. $E[X], E[Y]$ and $Var[X], Var[Y]$ for the marginal distribution.
2. $E[X|y], E[Y|x]$ and $Var[X|y], Var[Y|x]$ for the conditional distribution.

Discrete	Continuous	
$E[X] = \sum_x x f_X(x)$	$= \int_{-\infty}^{+\infty} x f_X(x) dx$	$= \mu_X$
$Var[X] = \sum_x x^2 f_X(x) - \mu_X^2$	$= \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu_X^2$	$= \sigma_X^2$
$E[X\|y] = \sum_x x f_{X\|y}(x)$	$= \int_{-\infty}^{+\infty} x f_{X\|y}(x) dx$	$= \mu_{X\|y}$
$Var[X\|y] = \sum_x x^2 f_{X\|y}(x) - \mu_{X\|y}^2$	$= \int_{-\infty}^{+\infty} x^2 f_X\|y(x) dx - \mu_{X\|y}^2$	$= \sigma_{X\|y}^2$

Expectations and variances

We define two types of expectations and variances:

1. $E[X], E[Y]$ and $Var[X], Var[Y]$ for the marginal distribution.
2. $E[X|y], E[Y|x]$ and $Var[X|y], Var[Y|x]$ for the conditional distribution.

Discrete	Continuous	
$E[X] = \sum_x x f_X(x)$	$= \int_{-\infty}^{+\infty} x f_X(x) dx$	$= \mu_X$
$Var[X] = \sum_x x^2 f_X(x) - \mu_X^2$	$= \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \mu_X^2$	$= \sigma_X^2$
$E[X|y] = \sum_x x f_{X|y}(x)$	$= \int_{-\infty}^{+\infty} x f_{X|y}(x) dx$	$= \mu_{X|y}$
$Var[X|y] = \sum_x x^2 f_{X|y}(x) - \mu_{X|y}^2$	$= \int_{-\infty}^{+\infty} x^2 f_X|y(x) dx - \mu_{X|y}^2$	$= \sigma_{X|y}^2$

ILLINOIS

Expectations and variances

We define two types of expectations and variances:

1. $E[X]$, $E[Y]$ and $Var[X]$, $Var[Y]$ for the marginal distribution.
2. $E[X|y]$, $E[Y|x]$ and $Var[X|y]$, $Var[Y|x]$ for the conditional distribution.

	Discrete	Continuous	
$E[X]$	$=\sum_x x f_X(x)$	$=\int_{-\infty}^{+\infty} x f_X(x)dx$	$=\mu_X$
$Var[X]$	$=\sum_x x^2 f_X(x) - \mu_X^2$	$=\int_{-\infty}^{+\infty} x^2 f_X(x)dx - \mu_X^2$	$=\sigma_X^2$
$E[X\|y]$	$=\sum_x x f_{X\|y}(x)$	$=\int_{-\infty}^{+\infty} x f_{X\|y}(x)dx$	$=\mu_{X\|y}$
$Var[X\|y]$	$=\sum_x x^2 f_{X\|y}(x) - \mu_{X\|y}^2$	$=\int_{-\infty}^{+\infty} x^2 f_X\|y(x)dx - \mu_{X\|y}^2$	$=\sigma_{X\|y}^2$

Expectation of a function

Recall that for random variable X and function $g(X)$, we have:

discrete : $$E[g(X)] = \sum_x g(x)p(x)$$

continuous : $$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

For two jointly distributed random variables X, Y and a function $g(X, Y)$, this becomes:

discrete : $$E[h(X,Y)] = \sum_x \sum_y h(x,y)f_{XY}(x,y)$$

continuous : $$E[h(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f_{XY}(x,y)dxdy$$

Expectation of a function

Recall that for random variable X and function $g(X)$, we have:

$$\text{discrete}: \quad E\left[g(X)\right] = \sum_x g(x)p(x)$$

$$\text{continuous}: \quad E\left[g(X)\right] = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

For two jointly distributed random variables X, Y and a function $g(X, Y)$, this becomes:

$$\text{discrete}: \quad E\left[h(X, Y)\right] = \sum_x \sum_y h(x, y)f_{XY}(x, y)$$

$$\text{continuous}: \quad E\left[h(X, Y)\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x, y)f_{XY}(x, y)dxdy$$

ILLINOIS

Example

Back to the chemical mixture volumes: X and Y are continuous random variables between 0 and 1 with $f_{XY}(x,y) = \frac{2}{5}(2x+3y)$. Recall that $f_Y(y) = \frac{6y+2}{5}$ and that $f_{X|y}(x) = \frac{4x+6y}{6y+2}$.

1. *What is the expectation of Y?*
2. *What is the expectation of X given that $Y = 0.6$?*
3. *What is the expectation of $3X + 7Y$?*

1. $\mu_Y = E[Y] = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^1 y \frac{6y+2}{5} dy = 0.6.$
2. $\mu_{X|y} = E[X|y] = \int_{-\infty}^{+\infty} x f_{X|y}(x) dx = \int_0^1 x \frac{4x+6y}{6y+2} dx.$

   Since we are told that $Y = 0.6$: $\mu_{X|y} = \int_0^1 x \frac{4x+3.6}{5.6} dx = 0.5595.$
3. $E[h(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y) f_{XY}(x,y) dxdy =$

   $\int_0^1 \int_0^1 (3x+7y)\frac{2}{5}(2x+3y)\, dxdy = 5.9.$

ILLINOIS

Example

Back to the chemical mixture volumes: X and Y are continuous random variables between 0 and 1 with $f_{XY}(x,y) = \frac{2}{5}(2x+3y)$. Recall that $f_Y(y) = \frac{6y+2}{5}$ and that $f_{X|y}(x) = \frac{4x+6y}{6y+2}$.

1. *What is the expectation of Y?*
2. *What is the expectation of X given that $Y = 0.6$?*
3. *What is the expectation of $3X + 7Y$?*

1. $\mu_Y = E[Y] = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^1 y \frac{6y+2}{5} dy = 0.6.$

2. $\mu_{X|y} = E[X|y] = \int_{-\infty}^{+\infty} x f_{X|y}(x) dx = \int_0^1 x \frac{4x+6y}{6y+2} dx.$

   Since we are told that $Y = 0.6$: $\mu_{X|y} = \int_0^1 x \frac{4x+3.6}{5.6} dx = 0.5595.$

3. $E[h(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) f_{XY}(x,y) dxdy =$

   $\int_0^1 \int_0^1 (3x+7y) \frac{2}{5}(2x+3y)\, dxdy = 5.9.$

ILLINOIS

Example

Back to the chemical mixture volumes: X and Y are continuous random variables between 0 and 1 with $f_{XY}(x,y) = \frac{2}{5}(2x+3y)$. Recall that $f_Y(y) = \frac{6y+2}{5}$ and that $f_{X|y}(x) = \frac{4x+6y}{6y+2}$.

1. *What is the expectation of Y?*
2. *What is the expectation of X given that $Y = 0.6$?*
3. *What is the expectation of $3X + 7Y$?*

1. $\mu_Y = E[Y] = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^1 y \frac{6y+2}{5} dy = 0.6.$
2. $\mu_{X|y} = E[X|y] = \int_{-\infty}^{+\infty} x f_{X|y}(x) dx = \int_0^1 x \frac{4x+6y}{6y+2} dx.$

   Since we are told that $Y = 0.6$: $\mu_{X|y} = \int_0^1 x \frac{4x+3.6}{5.6} dx = 0.5595.$
3. $E[h(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) f_{XY}(x,y) dx dy =$

   $\int_0^1 \int_0^1 (3x+7y) \frac{2}{5}(2x+3y) dx dy = 5.9.$

ILLINOIS

Example

Back to the chemical mixture volumes: X and Y are continuous random variables between 0 and 1 with $f_{XY}(x,y) = \frac{2}{5}(2x+3y)$. Recall that $f_Y(y) = \frac{6y+2}{5}$ and that $f_{X|y}(x) = \frac{4x+6y}{6y+2}$.

1. *What is the expectation of Y?*
2. *What is the expectation of X given that $Y = 0.6$?*
3. *What is the expectation of $3X + 7Y$?*

1. $\mu_Y = E[Y] = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^1 y \frac{6y+2}{5} dy = 0.6.$

2. $\mu_{X|y} = E[X|y] = \int_{-\infty}^{+\infty} x f_{X|y}(x) dx = \int_0^1 x \frac{4x+6y}{6y+2} dx.$

   Since we are told that $Y = 0.6$: $\mu_{X|y} = \int_0^1 x \frac{4x+3.6}{5.6} dx = 0.5595.$

3. $E[h(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) f_{XY}(x,y) dxdy =$

   $\int_0^1 \int_0^1 (3x+7y) \frac{2}{5}(2x+3y) dxdy = 5.9.$

ILLINOIS

Independence

Recall independence of events: events A and B are independent if

$$P(A|B) = P(A) \text{ or } P(A \cap B) = P(A) \cdot P(B).$$

Random variables X, Y are independent if any of the following hold:

1. $f_{XY}(x, y) = f_X(x) f_Y(y), \forall x, y$
2. $f_{X|y}(x) = f_X(x), \forall x, y$
3. $f_{Y|x}(y) = f_Y(y), \forall x, y$
4. $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B), \forall A, B$

The first one is typically the easiest to check.

ILLINOIS

Independence

Recall independence of events: events A and B are independent if

$$P(A|B) = P(A) \text{ or } P(A \cap B) = P(A) \cdot P(B).$$

Random variables X, Y are independent if any of the following hold:

1. $f_{XY}(x, y) = f_X(x) f_Y(y), \forall x, y$
2. $f_{X|y}(x) = f_X(x), \forall x, y$
3. $f_{Y|x}(y) = f_Y(y), \forall x, y$
4. $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B), \forall A, B$

The first one is typically the easiest to check.

Independence: examples

Example

Consider that for two continuous random variables X, Y, we have $f_{XY}(x, y) = x \cdot y$ for $0 \leq x \leq 1, 0 \leq y \leq 2$. Are X and Y independent?

Answer: Yes. First, find $f_X(x) = \int_0^2 xydy = 2x$ and $f_Y(y) = \int_0^1 xydx = \frac{1}{2}y$. Then, we check whether $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$, which is true.

Example

Let $f_{XY}(x, y) = \frac{1}{2}x \cdot y$, with $0 \leq x \leq y \leq 2$. Are X and Y independent?

Answer: Again, find $f_X(x) = \int_x^2 \frac{1}{2}xydy = x - \frac{x^3}{4}$ and $f_Y(y) = \int_0^y \frac{1}{2}xydx = \frac{y^3}{4}$.

When checking whether $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$, we get $\frac{1}{2}xy \neq \frac{xy^3 - x^3y^3}{4}$.

ILLINOIS

Independence: examples

Example

Consider that for two continuous random variables X, Y, we have $f_{XY}(x, y) = x \cdot y$ for $0 \leq x \leq 1, 0 \leq y \leq 2$. Are X and Y independent?

Answer: Yes. First, find $f_X(x) = \int_0^2 xy dy = 2x$ and $f_Y(y) = \int_0^1 xy dx = \frac{1}{2}y$.

Then, we check whether $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$, which is true.

Example

Let $f_{XY}(x, y) = \frac{1}{2}x \cdot y$, with $0 \leq x \leq y \leq 2$. Are X and Y independent?

Answer: Again, find $f_X(x) = \int_x^2 \frac{1}{2}xy dy = x - \frac{x^3}{4}$ and $f_Y(y) = \int_0^y \frac{1}{2}xy dx = \frac{y^3}{4}$.

When checking whether $f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$, we get $\frac{1}{2}xy \neq \frac{xy^3 - x^3y^3}{4}$.

Independence: examples

Example

Consider that for two continuous random variables X, Y, we have $f_{XY}(x,y) = x \cdot y$ for $0 \leq x \leq 1, 0 \leq y \leq 2$. Are X and Y independent?

Answer: Yes. First, find $f_X(x) = \int_0^2 xydy = 2x$ and $f_Y(y) = \int_0^1 xydx = \frac{1}{2}y$.

Then, we check whether $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$, which is true.

Example

Let $f_{XY}(x,y) = \frac{1}{2}x \cdot y$, with $0 \leq x \leq y \leq 2$. Are X and Y independent?

Answer: Again, find $f_X(x) = \int_x^2 \frac{1}{2}xydy = x - \frac{x^3}{4}$ and $f_Y(y) = \int_0^y \frac{1}{2}xydx = \frac{y^3}{4}$.

When checking whether $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$, we get $\frac{1}{2}xy \neq \frac{xy^3 - x^3y^3}{4}$.

Independence: examples

Example

Consider that for two continuous random variables X, Y, we have $f_{XY}(x,y) = x \cdot y$ for $0 \le x \le 1, 0 \le y \le 2$. Are X and Y independent?

Answer: Yes. First, find $f_X(x) = \int_0^2 xydy = 2x$ and $f_Y(y) = \int_0^1 xydx = \frac{1}{2}y$.

Then, we check whether $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$, which is true.

Example

Let $f_{XY}(x,y) = \frac{1}{2}x \cdot y$, with $0 \le x \le y \le 2$. Are X and Y independent?

Answer: Again, find $f_X(x) = \int_x^2 \frac{1}{2}xydy = x - \frac{x^3}{4}$ and $f_Y(y) = \int_0^y \frac{1}{2}xydx = \frac{y^3}{4}$.

When checking whether $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$, we get $\frac{1}{2}xy \neq \frac{xy^3 - x^3y^3}{4}$.

ILLINOIS

Independence: examples

Example

Consider that for two continuous random variables X, Y, we have $f_{XY}(x,y) = x \cdot y$ for $0 \leq x \leq 1, 0 \leq y \leq 2$. Are X and Y independent?

Answer: Yes. First, find $f_X(x) = \int_0^2 xydy = 2x$ and $f_Y(y) = \int_0^1 xydx = \frac{1}{2}y$.
Then, we check whether $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$, which is true.

Example

Let $f_{XY}(x,y) = \frac{1}{2}x \cdot y$, with $0 \leq x \leq y \leq 2$. Are X and Y independent?

Answer: Again, find $f_X(x) = \int_x^2 \frac{1}{2}xydy = x - \frac{x^3}{4}$ and $f_Y(y) = \int_0^y \frac{1}{2}xydx = \frac{y^3}{4}$.
When checking whether $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$, we get $\frac{1}{2}xy \neq \frac{xy^3 - x^3y^3}{4}$.

ILLINOIS

Covariance

- A measure of the *association* between two random variables.
- For two random variables X and Y, we define *covariance* as:

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

Recall that for a single random variable:

$$\sigma_X^2 = Var[X] = E\left[(X - E[X])^2\right] = E\left[X^2\right] - (E[X])^2.$$

Observations:

1. If $X \geq E[X]$ whenever $Y \geq E[Y]$ and if $X \leq E[X]$ whenever $Y \leq E[Y]$, then the covariance will be positive.
2. If $X \geq E[X]$ whenever $Y \leq E[Y]$ and if $X \leq E[X]$ whenever $Y \geq E[Y]$, then the covariance will be negative.

Two independent random variables X, Y will have $\sigma_{XY} = Cov[X, Y] = 0$. The inverse is not necessarily true.

ILLINOIS

Covariance

- A measure of the *association* between two random variables.
- For two random variables X and Y, we define *covariance* as:

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

Recall that for a single random variable:

$$\sigma_X^2 = Var[X] = E\left[(X - E[X])^2\right] = E\left[X^2\right] - (E[X])^2.$$

Observations:

1. If $X \geq E[X]$ whenever $Y \geq E[Y]$ and if $X \leq E[X]$ whenever $Y \leq E[Y]$, then the covariance will be positive.
2. If $X \geq E[X]$ whenever $Y \leq E[Y]$ and if $X \leq E[X]$ whenever $Y \geq E[Y]$, then the covariance will be negative.

Two independent random variables X, Y will have $\sigma_{XY} = Cov[X,Y] = 0$. The inverse is not necessarily true.

ILLINOIS

Covariance

- A measure of the *association* between two random variables.
- For two random variables X and Y, we define *covariance* as:

$$\sigma_{XY} = Cov\left[X, Y\right] = E\left[(X - E\left[X\right]) \cdot (Y - E\left[Y\right])\right] = E\left[XY\right] - E\left[X\right] \cdot E\left[Y\right].$$

Recall that for a single random variable:

$$\sigma_X^2 = Var\left[X\right] = E\left[(X - E\left[X\right])^2\right] = E\left[X^2\right] - (E\left[X\right])^2.$$

Observations:

1. If $X \geq E[X]$ whenever $Y \geq E[Y]$ and if $X \leq E[X]$ whenever $Y \leq E[Y]$, then the covariance will be positive.
2. If $X \geq E[X]$ whenever $Y \leq E[Y]$ and if $X \leq E[X]$ whenever $Y \geq E[Y]$, then the covariance will be negative.

Two independent random variables X, Y will have $\sigma_{XY} = Cov[X, Y] = 0$. The inverse is not necessarily true.

ILLINOIS

Correlation

- Covariance is not normalized.
- It would be nice to have a measure that directly relates its result to the magnitude of dependence.

Correlation is a measure of the linear relationship between two random variables X and Y.

- It is calculated by $\rho_{XY} = \frac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$.
- By definition, $-1 \leq \rho_{XY} \leq 1$.
  1. $\rho_{XY} = 0$: X and Y are not correlated.
     When X and Y are independent then $\sigma_{XY} = \rho_{XY} = 0$.
  2. $\rho_{XY} = 1$: X and Y are fully positively correlated.
  3. $\rho_{XY} = -1$: X and Y are fully negatively correlated.

ILLINOIS

Correlation

- Covariance is not normalized.
- It would be nice to have a measure that directly relates its result to the magnitude of dependence.

Correlation is a measure of the linear relationship between two random variables X and Y.

- It is calculated by $\rho_{XY} = \frac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$.
- By definition, $-1 \leq \rho_{XY} \leq 1$.
  1. $\rho_{XY} = 0$: X and Y are not correlated.
     When X and Y are independent then $\sigma_{XY} = \rho_{XY} = 0$.
  2. $\rho_{XY} = 1$: X and Y are fully positively correlated.
  3. $\rho_{XY} = -1$: X and Y are fully negatively correlated.

ILLINOIS

Correlation

- Covariance is not normalized.
- It would be nice to have a measure that directly relates its result to the magnitude of dependence.

Correlation is a measure of the linear relationship between two random variables X and Y.

- It is calculated by $\rho_{XY} = \frac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$.
- By definition, $-1 \leq \rho_{XY} \leq 1$.
  1. $\rho_{XY} = 0$: X and Y are not correlated.
     When X and Y are independent then $\sigma_{XY} = \rho_{XY} = 0$.
  2. $\rho_{XY} = 1$: X and Y are fully positively correlated.
  3. $\rho_{XY} = -1$: X and Y are fully negatively correlated.

Correlation

- Covariance is not normalized.
- It would be nice to have a measure that directly relates its result to the magnitude of dependence.

Correlation is a measure of the linear relationship between two random variables X and Y.

- It is calculated by $\rho_{XY} = \frac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$.
- By definition, $-1 \leq \rho_{XY} \leq 1$.
  1. $\rho_{XY} = 0$: X and Y are not correlated.
     When X and Y are independent then $\sigma_{XY} = \rho_{XY} = 0$.
  2. $\rho_{XY} = 1$: X and Y are fully positively correlated.
  3. $\rho_{XY} = -1$: X and Y are fully negatively correlated.

ILLINOIS

Covariance and correlation: an example

Example

Consider continuous random variables X, Y with $f_{XY}(x, y) = 10x^2y$, defined for $0 \le y \le x \le 1$.

a) Independent X&Y? *b) What is σ_{XY}?* *c) What is ρ_{XY}?*

Answer: We need a **lot of information**:

- $f_X(x) = \int_0^x 10x^2 y dy = 5x^4.$
- $f_Y(y) = \int_y^1 10x^2 y dx = y\frac{10-10y^3}{3} = \frac{10}{3}y\left(1-y^3\right).$
- $E[X] = \int_0^1 x \cdot 5x^4 dx = \frac{5}{6}.$
- $E[Y] = \int_0^1 y \cdot \frac{10}{3}y\left(1-y^3\right) dy = \frac{5}{9}.$
- $E[XY] = \int_0^1 \int_0^x xy 10x^2 y dy dx = \frac{10}{21}.$

Covariance and correlation: an example

Example

Consider continuous random variables X, Y *with* $f_{XY}(x,y) = 10x^2y$*, defined for* $0 \le y \le x \le 1$.

a) *Independent* $X\&Y$*?* b) *What is* σ_{XY}*?* c) *What is* ρ_{XY}*?*

Answer: We need a **lot of information**:

- $f_X(x) = \int_0^x 10x^2y\,dy = 5x^4.$
- $f_Y(y) = \int_y^1 10x^2y\,dx = y\frac{10-10y^3}{3} = \frac{10}{3}y\left(1-y^3\right).$
- $E[X] = \int_0^1 x \cdot 5x^4\,dx = \frac{5}{6}.$
- $E[Y] = \int_0^1 y \cdot \frac{10}{3}y\left(1-y^3\right)dy = \frac{5}{9}.$
- $E[XY] = \int_0^1 \int_0^x xy10x^2y\,dy\,dx = \frac{10}{21}.$

ILLINOIS

Covariance and correlation: an example

Example

Consider continuous random variables X, Y with $f_{XY}(x,y) = 10x^2y$, defined for $0 \leq y \leq x \leq 1$.

a) Independent X&Y? *b) What is σ_{XY}?* *c) What is ρ_{XY}?*

Answer: We need a **lot of information**:

- $f_X(x) = \int_0^x 10x^2 y dy = 5x^4$.
- $f_Y(y) = \int_y^1 10x^2 y dx = y\frac{10-10y^3}{3} = \frac{10}{3}y\left(1-y^3\right)$.
- $E[X] = \int_0^1 x \cdot 5x^4 dx = \frac{5}{6}$.
- $E[Y] = \int_0^1 y \cdot \frac{10}{3} y\left(1-y^3\right) dy = \frac{5}{9}$.
- $E[XY] = \int_0^1 \int_0^x xy 10x^2 y dy dx = \frac{10}{21}$.

ILLINOIS

Covariance and correlation: an example

Example

Consider continuous random variables X, Y *with* $f_{XY}(x,y) = 10x^2y$*, defined for* $0 \le y \le x \le 1$.

a) Independent $X \& Y$*?* *b) What is* σ_{XY}*?* *c) What is* ρ_{XY}*?*

Answer: We need a **lot of information**:

- $f_X(x) = \int_0^x 10x^2 y dy = 5x^4$.
- $f_Y(y) = \int_y^1 10x^2 y dx = y\frac{10-10y^3}{3} = \frac{10}{3}y\left(1-y^3\right)$.
- $E[X] = \int_0^1 x \cdot 5x^4 dx = \frac{5}{6}$.
- $E[Y] = \int_0^1 y \cdot \frac{10}{3}y\left(1-y^3\right)dy = \frac{5}{9}$.
- $E[XY] = \int_0^1 \int_0^x xy10x^2 ydydx = \frac{10}{21}$.

ILLINOIS

Covariance and correlation: an example

Example

Consider continuous random variables X, Y with $f_{XY}(x,y) = 10x^2y$, defined for $0 \leq y \leq x \leq 1$.

a) Independent X&Y? *b) What is σ_{XY}?* *c) What is ρ_{XY}?*

Answer: We need a **lot of information**:

- $f_X(x) = \int_0^x 10x^2y\,dy = 5x^4.$
- $f_Y(y) = \int_y^1 10x^2y\,dx = y\frac{10-10y^3}{3} = \frac{10}{3}y\left(1-y^3\right).$
- $E[X] = \int_0^1 x \cdot 5x^4 dx = \frac{5}{6}.$
- $E[Y] = \int_0^1 y \cdot \frac{10}{3}y\left(1-y^3\right) dy = \frac{5}{9}.$
- $E[XY] = \int_0^1 \int_0^x xy10x^2y\,dy\,dx = \frac{10}{21}.$

ILLINOIS

Hence, we have:

$$\sigma_{XY} = E[XY] - E[X] \cdot E[Y] = \frac{10}{21} - \frac{5}{6} \cdot \frac{5}{9} = 0.01323.$$

To calculate the correlation, we also need the variances. We have:

$$Var[X] = E\left[X^2\right] - (E[X])^2 = \int_0^1 x^2 \cdot 5x^4 dx - \left(\frac{5}{6}\right)^2 = \frac{5}{7} - \frac{25}{36} = 0.0198.$$

$$Var[Y] = E\left[Y^2\right] - (E[Y])^2 = \int_0^1 y^2 \cdot \frac{10}{3} y\left(1 - y^3\right) dy - \left(\frac{5}{9}\right)^2 = \frac{5}{14} - \frac{25}{81} = 0.0485.$$

Overall:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{Var[X]} \cdot \sqrt{Var[Y]}} = 0.4265.$$

Hence, we have:

$$\sigma_{XY} = E[XY] - E[X] \cdot E[Y] = \frac{10}{21} - \frac{5}{6} \cdot \frac{5}{9} = 0.01323.$$

To calculate the correlation, we also need the variances. We have:

$$Var[X] = E\left[X^2\right] - (E[X])^2 = \int_0^1 x^2 \cdot 5x^4 dx - \left(\frac{5}{6}\right)^2 =$$

$$= \frac{5}{7} - \frac{25}{36} = 0.0198.$$

$$Var[Y] = E\left[Y^2\right] - (E[Y])^2 = \int_0^1 y^2 \cdot \frac{10}{3} y \left(1 - y^3\right) dy - \left(\frac{5}{9}\right)^2 =$$

$$= \frac{5}{14} - \frac{25}{81} = 0.0485.$$

Overall:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{Var[X]} \cdot \sqrt{Var[Y]}} = 0.4265.$$

Hence, we have:

$$\sigma_{XY} = E[XY] - E[X] \cdot E[Y] = \frac{10}{21} - \frac{5}{6} \cdot \frac{5}{9} = 0.01323.$$

To calculate the correlation, we also need the variances. We have:

$$Var[X] = E\left[X^2\right] - (E[X])^2 = \int_0^1 x^2 \cdot 5x^4 dx - \left(\frac{5}{6}\right)^2 =$$
$$= \frac{5}{7} - \frac{25}{36} = 0.0198.$$
$$Var[Y] = E\left[Y^2\right] - (E[Y])^2 = \int_0^1 y^2 \cdot \frac{10}{3} y\left(1 - y^3\right) dy - \left(\frac{5}{9}\right)^2 =$$
$$= \frac{5}{14} - \frac{25}{81} = 0.0485.$$

Overall:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sqrt{Var[X]} \cdot \sqrt{Var[Y]}} = 0.4265.$$