Information Theory and Statistics
Lecture 1: Entropy and information

Łukasz Dębowski
ldebowsk@ipipan.waw.pl

Ph. D. Programme 2013/2014

Project co-financed by the European Union
within the framework of the European Social Fund

Claude Shannon (1916–2001)

Entropy of a random variable on a probability space is the fundamental concept of information theory developed by Claude Shannon in 1948.

Project co-financed by the European Union
within the framework of the European Social Fund

Probability as a random variable

Definition

Let $\mathbf{X}$ and $\mathbf{Y}$ be discrete variables and $\mathbf{A}$ be an event on a probability space $(\Omega, \mathcal{J}, \mathbf{P})$. We define $\mathbf{P}(\mathbf{X})$ as a discrete random variable such that

$$\mathbf{P}(\mathbf{X})(\omega) = \mathbf{P}(\mathbf{X} = \mathrm{x}) \iff \mathbf{X}(\omega) = \mathrm{x}.$$

Analogously we define $\mathbf{P}(\mathbf{X}|\mathbf{Y})$ and $\mathbf{P}(\mathbf{X}|\mathbf{A})$ as

$$\mathbf{P}(\mathbf{X}|\mathbf{Y})(\omega) = \mathbf{P}(\mathbf{X} = \mathrm{x}|\mathbf{Y} = \mathrm{y}) \iff \mathbf{X}(\omega) = \mathrm{x} \text{ and } \mathbf{Y}(\omega) = \mathrm{y},$$
$$\mathbf{P}(\mathbf{X}|\mathbf{A})(\omega) = \mathbf{P}(\mathbf{X} = \mathrm{x}|\mathbf{A}) \iff \mathbf{X}(\omega) = \mathrm{x},$$

where the conditional probability is $\mathbf{P}(\mathbf{B}|\mathbf{A}) = \mathbf{P}(\mathbf{B} \cap \mathbf{A})/\mathbf{P}(\mathbf{A})$ for $\mathbf{P}(\mathbf{A}) > 0$.

Project co-financed by the European Union
within the framework of the European Social Fund

Independence

We write $\mathbf{P(Y) = P(X_1, X_2, \ldots, X_n)}$ for $\mathbf{Y = (X_1, X_2, \ldots, X_n)}$.

Definition (independence)

We say that random variables $\mathbf{X_1, X_2, \ldots, X_n}$ are independent if

$$\mathbf{P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i).}$$

Analogously, we say that random variables $\mathbf{X_1, X_2, X_3, \ldots}$ are independent if $\mathbf{X_1, X_2, \ldots, X_n}$ are independent for any $\mathbf{n}$.

Example

Let $\mathbf{\Omega = [0, 1]}$ be the unit section and let $\mathbf{P}$ be the Lebesgue measure. Define real random variable $\mathbf{Y(\omega) = \omega}$. If we consider its binary expansion $\mathbf{Y = \sum_{i=1}^{\infty} 2^{-i} Z_i}$, where $\mathbf{Z_i : \Omega \to \{0, 1\}}$, then $\mathbf{P(Z_1, Z_2, \ldots, Z_n) = 2^{-n} = \prod_{i=1}^{n} P(Z_i)}$.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Expectation

Definition (expectation)

We define the expectation of a real random variable **X** as

$$\mathbb{E}\,X := \int X dP.$$

For discrete random variables we obtain

$$\mathbb{E}\,X = \sum_{x: P(X=x) > 0} P(X = x) \cdot x.$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Additivity of expectation

One of fundamental properties of the expectation is its additivity.

Theorem

Let $\mathbf{X}, \mathbf{Y} \geq \mathbf{0}$*. We have*

$$\mathbb{E}\,(\mathbf{X} + \mathbf{Y}) = \mathbb{E}\,\mathbf{X} + \mathbb{E}\,\mathbf{Y}.$$

Remark: The restriction $\mathbf{X}, \mathbf{Y} \geq \mathbf{0}$ is made because, e.g., for $\mathbb{E}\,\mathbf{X} = \infty$ and $\mathbb{E}\,\mathbf{Y} = -\infty$ the sum is undefined.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Entropy

Some interpretation of entropy $\mathbf{H(X)}$ is the average uncertainty carried by a random variable $\mathbf{X}$. We expect that uncertainty adds for probabilistically independent sources. Formally, for $\mathbf{P(X,Y) = P(X)P(Y)}$, we postulate $\mathbf{H(X,Y) = H(X) + H(Y)}$. Because $\mathbf{\log(xy) = \log x + \log y}$, the following definition comes as a very natural idea.

Definition (entropy)

The *entropy* of a discrete variable $\mathbf{X}$ is defined as

$$\mathbf{H(X) :=} \mathbb{E}\left[\mathbf{-\log P(X)}\right]. \tag{1}$$

Traditionally, it is assumed that $\mathbf{\log}$ is the logarithm to the base $\mathbf{2}$.

Because $\mathbf{\log P(X) \le 0}$, we put the minus sign in the definition (1) so that entropy be positive.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Entropy (continued)

Equivalently, we have

$$\mathbf{H}(\mathbf{X}) = -\sum_{\mathbf{x}:\mathbf{P}(\mathbf{X}=\mathbf{x})>0} \mathbf{P}(\mathbf{X}=\mathbf{x}) \log \mathbf{P}(\mathbf{X}=\mathbf{x}),$$

We can verify that for $\mathbf{P}(\mathbf{X},\mathbf{Y}) = \mathbf{P}(\mathbf{X})\mathbf{P}(\mathbf{Y})$,

$$\begin{aligned} \mathbf{H}(\mathbf{X},\mathbf{Y}) &= \mathbb{E}\left[-\log \mathbf{P}(\mathbf{X},\mathbf{Y})\right] = \mathbb{E}\left[-\log \mathbf{P}(\mathbf{X}) - \log \mathbf{P}(\mathbf{X})\right] \\ &= \mathbb{E}\left[-\log \mathbf{P}(\mathbf{X})\right] + \mathbb{E}\left[-\log \mathbf{P}(\mathbf{X})\right] = \mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y}). \end{aligned}$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Entropy for a binary variable

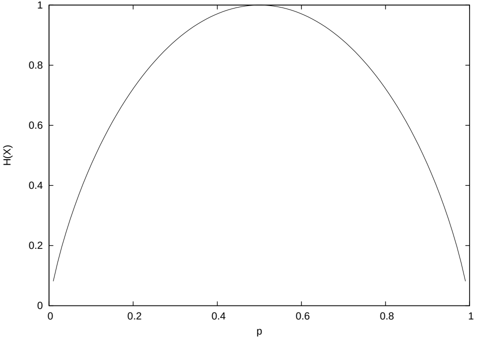

Figure: Entropy $\mathbf{H(X) = -p \log p - (1-p)\log(1-p)}$ for $\mathbf{P(X=0) = p}$ and $\mathbf{P(X=1) = 1-p}$.

Project co-financed by the European Union
within the framework of the European Social Fund

The range of entropy in general

Because function $\mathbf{f(p) = -p \log p}$ is strictly positive for $\mathbf{p \in (0, 1)}$ and equals $\mathbf{0}$ for $\mathbf{p = 1}$, it can be easily seen that:

Theorem

$\mathbf{H(X) \geq 0}$, *whereas* $\mathbf{H(X) = 0}$ *if and only if* $\mathbf{X}$ *assumes only a single value.*

This fact agrees with the idea that constants carry no uncertainty. On the other hand, assume that $\mathbf{X}$ takes values $\mathbf{x \in \{1, 2, ..., n\}}$ with equal probabilities $\mathbf{P(X = x) = 1/n}$. Then

$$\mathbf{H(X) = -\sum_{x=1}^{n} \frac{1}{n} \log \frac{1}{n} = \sum_{x=1}^{n} \frac{1}{n} \log n = \log n.}$$

As we will see, $\mathbf{\log n}$ is the maximal value of $\mathbf{H(X)}$ if $\mathbf{X}$ assumes values in $\mathbf{\{1, 2, ..., n\}}$. That fact agrees with the intuition that the highest uncertainty occurs for uniformly distributed variables.

Project co-financed by the European Union
within the framework of the European Social Fund

Kullback-Leibler divergence

A discrete probability distribution is a function $\mathbf{p} : \mathbb{X} \to [0, 1]$ on a countable set $\mathbb{X}$ such that $\mathbf{p(x)} \geq 0$ and $\sum_x \mathbf{p(x)} = 1$.

Definition (entropy revisited)

The entropy of a discrete probability distribution is denoted as

$$\mathbf{H(p)} := -\sum_{x:p(x)>0} \mathbf{p(x)} \log \mathbf{p(x)}.$$

Definition (KL divergence)

Kullback-Leibler divergence, or *relative entropy* of probability distributions **p** and **q** is defined as

$$\mathbf{D(p||q)} := \sum_{x:p(x)>0} \mathbf{p(x)} \log \frac{\mathbf{p(x)}}{\mathbf{q(x)}}.$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Convex and concave functions

Definition (convex and concave functions)

A real function $f : \mathbb{R} \to \mathbb{R}$ is *convex* if

$$p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2)$$

for $p_i \geq 0$ and $p_1 + p_2 = 1$. Moreover, f is called *strictly convex* if

$$p_1 f(x_1) + p_2 f(x_2) > f(p_1 x_1 + p_2 x_2)$$

for $p_i > 0$ and $p_1 + p_2 = 1$. We say that function f is *concave* if $-f$ is *convex*, whereas f is *strictly concave* if $-f$ is *stricly convex*.

Example

If function f has a positive second derivative then it is strictly convex. Hence functions $h(x) = -\log x$ and $g(x) = x^2$ are strictly convex.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Jensen inequality

Theorem (Jensen inequality)

If $\mathbf{f}$ *is a convex function and* $\mathbf{p}$ *is a discrete probability distribution over real values then*

$$\sum_{x:p(x)>0} p(x)f(x) \geq f\left(\sum_{x:p(x)>0} p(x)\cdot x\right).$$

Moreover, if $\mathbf{f}$ *is strictly convex then*

$$\sum_{x:p(x)>0} p(x)f(x) = f\left(\sum_{x:p(x)>0} p(x)\cdot x\right)$$

holds if and only if distribution $\mathbf{p}$ *is concentrated on a single value.*

The proof proceeds by induction on the number of values of $\mathbf{p}$.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

KL divergence is nonegative

Theorem

We have

$$D(p||q) \geq 0,$$

where the equality holds if and only if $p = q$.

Proof

By the Jensen inequality for $f(y) = -\log y$, we have

$$D(p||q) = -\sum_{x:p(x)>0} p(x) \log \frac{q(x)}{p(x)} \geq -\log \left(\sum_{x:p(x)>0} p(x) \frac{q(x)}{p(x)} \right)$$
$$= -\log \left(\sum_{x:p(x)>0} q(x) \right) \geq -\log 1 = 0.$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

The maximum of entropy

Theorem

Let X *assume values in* $\{1, 2, \ldots, n\}$. *We have* $H(X) \leq \log n$, *whereas* $H(X) = \log n$ *if and only if* $P(X = x) = 1/n$.

Remark: If the range of variable X is infinite then entropy $H(X)$ may be infinite.

Proof

Let $p(x) = P(X = x)$ and $q(x) = 1/n$. Then

$$0 \leq D(p||q) = \sum_{x:p(x)>0} p(x) \log \frac{p(x)}{1/n} = \log n - H(X),$$

where the equality occurs if and only if $p = q$.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Conditional entropy

The next important question is what is the behavior of entropy under conditioning. The intuition is that given additional information, the uncertainty should decrease. So should entropy. There are two distinct ways of defining conditional entropy.

Definition (conditional entropy)

Conditional entropy of a discrete variable $\mathbf{X}$ given event $\mathbf{A}$ is

$$\mathbf{H(X|A) := H(p)} \text{ for } \mathbf{p(x) = P(X = x|A)}.$$

Conditional entropy of $\mathbf{X}$ given a discrete variable $\mathbf{Y}$ is defined as

$$\mathbf{H(X|Y) := \sum_{y:P(Y=y)>0} P(Y = y)H(X|Y = y).}$$

Both $\mathbf{H(X|A)}$ and $\mathbf{H(X|Y)}$ are nonnegative.

Project co-financed by the European Union
within the framework of the European Social Fund

Minimum of conditional entropy

Theorem

$H(X|Y) = 0$ *holds if and only if* $X = f(Y)$ *for a certain function* f *except for a set of probability* 0.

Proof

Observe that $H(X|Y) = 0$ if and only if $H(X|Y = y) = 0$ for all y such that $P(Y = y) > 0$. This holds if and only if given $(Y = y)$ with $P(Y = y) > 0$, variable X is concentrated on a single value. Denoting this value as $f(y)$, we obtain $X = f(Y)$, except for the union of those sets $(Y = y)$ which have probability 0.

Project co-financed by the European Union
within the framework of the European Social Fund

Another formula for conditional entropy

We have

$$H(X|Y) = \mathbb{E}\left[-\log P(X|Y)\right].$$

Proof

$$\begin{aligned}
H(X|Y) &= \sum_{y} P(Y=y) H(X|Y=y) \\
&= -\sum_{x,y} P(Y=y) P(X=x|Y=y) \log P(X=x|Y=y) \\
&= -\sum_{x,y} P(X=x, Y=y) \log P(X=x|Y=y) \\
&= \mathbb{E}\left[-\log P(X|Y)\right].
\end{aligned}$$

Project co-financed by the European Union
within the framework of the European Social Fund

Conditional entropy and entropy

Because $\mathbf{P(Y)P(X|Y) = P(X,Y)}$, by the previous result,

$$\mathbf{H(Y) + H(X|Y) = H(X,Y).}$$

Hence

$$\mathbf{H(X,Y) \geq H(Y).}$$

Project co-financed by the European Union
within the framework of the European Social Fund

Inequality $H(X|A) \leq H(X)$ need not hold

Example

Let $P(X = 0|A) = P(X = 1|A) = 1/2$, whereas $P(X = 0|A^c) = 1$ and $P(X = 1|A^c) = 0$. Assuming $P(A) = 1/2$, we have $P(X = 0) = (1/2) \cdot (1/2) + (1/2) = 3/4$ and $P(X = 0) = (1/2) \cdot (1/2) = 1/4$ so

$$H(X) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = \log 4 - \frac{3}{4}\log 3 = 0.811\ldots.$$

On the other hand, we have $H(X|A) = \log 2 = 1$.

Despite that fact, $H(X|Y) \leq H(X)$ holds in general. Thus entropy decreases given additional information on average.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Mutual information

To show that $\mathbf{H(X)}$ is greater than $\mathbf{H(X|Y)}$, it is convenient to introduce another important concept.

Definition (mutual information)

Mutual information between discrete variables $\mathbf{X}$ and $\mathbf{Y}$ is

$$\mathbf{I(X;Y)} := \mathbb{E}\left[\log \frac{\mathbf{P(X,Y)}}{\mathbf{P(X)P(Y)}}\right].$$

We have $\mathbf{I(X;X) = H(X)}$. Thus entropy is sometimes called *self-information*.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Mutual information is nonnegative

Theorem

We have

$$\mathbf{I(X;Y) \geq 0},$$

where the equality holds if and only if $\mathbf{X}$ *and* $\mathbf{Y}$ *are independent.*

Proof

Let $\mathbf{p(x,y) = P(X = x, Y = y)}$ and $\mathbf{q(x,y) = P(X = x)P(Y = y)}$. Then we have

$$\mathbf{I(X;Y) = \sum_{(x,y):p(x,y)>0} p(x,y) \log \frac{p(x,y)}{q(x,y)} = D(p||q) \geq 0}$$

with the equality exactly for $\mathbf{p = q}$.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI
PhD STUDIES
UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Mutual information and entropy

By the definition of mutual information,

$$\mathbf{H}(\mathbf{X},\mathbf{Y}) + \mathbf{I}(\mathbf{X};\mathbf{Y}) = \mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y}),$$
$$\mathbf{H}(\mathbf{X}|\mathbf{Y}) + \mathbf{I}(\mathbf{X};\mathbf{Y}) = \mathbf{H}(\mathbf{X}).$$

Hence

$$\mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y}) \geq \mathbf{H}(\mathbf{X},\mathbf{Y}),$$
$$\mathbf{H}(\mathbf{X}) \geq \mathbf{H}(\mathbf{X}|\mathbf{Y}),\ \mathbf{I}(\mathbf{X};\mathbf{Y}).$$

Moreover, we have $\mathbf{H}(\mathbf{X}|\mathbf{Y}) = \mathbf{H}(\mathbf{Y})$ if $\mathbf{X}$ and $\mathbf{Y}$ are independent, which also agrees with intuition.

Project co-financed by the European Union
within the framework of the European Social Fund

Conditional mutual information

In a similar fashion to conditional entropy, we define:

Definition (conditional mutual information)

Conditional mutual information between discrete variables $\mathbf{X}$ and $\mathbf{Y}$ given event $\mathbf{A}$ is

$$\mathbf{I(X;Y|A) := D(p||q)} \text{ for } \mathbf{p(x,y) = P(X = x, Y = y|A)}$$
$$\text{and } \mathbf{q(x,y) = P(X = x|A)P(Y = y|A)}.$$

Conditional mutual information between discrete variables $\mathbf{X}$ and $\mathbf{Y}$ given variable $\mathbf{Z}$ is defined as

$$\mathbf{I(X;Y|Z) := \sum_{z:P(Z=z)>0} P(Z = z) I(X;Y|Z = z).}$$

Both $\mathbf{I(X;Y|A)}$ and $\mathbf{I(X;Y|Z)}$ are nonnegative.

Project co-financed by the European Union
within the framework of the European Social Fund

Another formula for CMI

As in the case of conditional entropy, this proposition is true:

Theorem

We have

$$\mathbf{I(X;Y|Z)} := \mathbb{E}\left[\log \frac{\mathbf{P(X,Y|Z)}}{\mathbf{P(X|Z)P(Y|Z)}}\right].$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Conditional independence

Definition (conditional independence)

Variables $X_1, X_2, \ldots, X_n$ are *conditionally independent* given Z if

$$P(X_1, X_2, \ldots, X_n|Z) = \prod_{i=1}^{n} P(X_i|Z).$$

Variables $X_1, X_2, X_3, \ldots$ are conditionally independent given Z if $X_1, X_2, \ldots, X_n$ are conditionally independent given Z for any n.

Theorem

We have

$$I(X; Y|Z) \geq 0,$$

with equality iff X *and* Y *are conditionally independent given* Z.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Conditional independence (examples)

Example

Let $\mathbf{Y} = \mathbf{f}(\mathbf{Z})$ be a function of variable $\mathbf{Z}$, whereas $\mathbf{X}$ be an arbitrary variable. Variables $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$. Indeed,

$$\begin{aligned} P(X = x, Y = y|Z = z) &= P(X = x|Z = z)1_{\{y=f(z)\}} \\ &= P(X = x|Z = z)P(Y = y|Z = z). \end{aligned}$$

Example

Let variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be independent. Variables $\mathbf{U} = \mathbf{X} + \mathbf{Z}$ and $\mathbf{W} = \mathbf{Y} + \mathbf{Z}$ are conditionally independent given $\mathbf{Z}$. Indeed,

$$\begin{aligned} &P(U = u, W = w|Z = z) = P(X = u - z, Y = w - z) \\ &= P(X = u - z)P(Y = w - z) \\ &= P(U = u|Z = z)P(W = w|Z = z). \end{aligned}$$

Project co-financed by the European Union
within the framework of the European Social Fund

Markov chains

Definition (Markov chain)

A stochastic process $(X_i)_{i=-\infty}^{\infty}$ is called a *Markov chain* if

$$P(X_i|X_{i-1}, X_{i-2}, ..., X_{i-n}) = P(X_i|X_{i-1})$$

holds for any n.

Example

For a Markov chain $(X_i)_{i=-\infty}^{\infty}$, variables X_i and X_k are conditionally independent given X_j if $i \leq j \leq k$. Indeed, after some algebra we obtain $P(X_k|X_i, X_j) = P(X_k|X_j)$, and hence

$$P(X_i, X_k|X_j) = P(X_i|X_j)P(X_k|X_i, X_j) = P(X_i|X_j)P(X_k|X_j).$$

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY

Conditional MI and MI

Theorem

We have

$$I(X;Y|Z) + I(X;Z) = I(X;Y,Z).$$

Remark: Hence, variables X and (Y,Z) are independent iff X and Z are independent and X and Y are independent given Z.

Proof

$$\begin{aligned} & I(X;Y|Z) + I(X;Z) \\ &= \mathbb{E}\left[\log \frac{P(X,Y,Z)P(Z)}{P(X,Z)P(Y,Z)}\right] + \mathbb{E}\left[\log \frac{P(X,Z)}{P(X)P(Z)}\right] \\ &= \mathbb{E}\left[\log \frac{P(X,Y,Z)}{P(X)P(Y,Z)}\right] = I(X;Y,Z). \end{aligned}$$

Project co-financed by the European Union
within the framework of the European Social Fund

CMI and entropy

Theorem

If entropies $\mathbf{H(X)}$, $\mathbf{H(Y)}$, *and* $\mathbf{H(Z)}$ *are finite, we have*

$$\begin{aligned} \mathbf{H(X|Y)} &= \mathbf{H(X,Y)} - \mathbf{H(Y)}, \\ \mathbf{I(X;Y)} &= \mathbf{H(X)} - \mathbf{H(X|Y)} = \mathbf{H(X)} + \mathbf{H(Y)} - \mathbf{H(X,Y)}, \\ \mathbf{I(X;Y|Z)} &= \mathbf{H(X|Z)} + \mathbf{H(Y|Z)} - \mathbf{H(X,Y|Z)} \\ &= \mathbf{H(X,Z)} + \mathbf{H(Y,Z)} - \mathbf{H(X,Y,Z)} - \mathbf{H(Z)}, \end{aligned}$$

where all terms are finite and nonnegative.

The proof is left as an easy exercise.

Project co-financed by the European Union
within the framework of the European Social Fund

KAPITAŁ LUDZKI NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA EUROPEJSKI FUNDUSZ SPOŁECZNY