Information Theory and Statistics
Lecture 2: Source coding

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

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Injections and codes

Definition (injection)

Function $\mathbf{f}$ is called an *injection* if $\mathbf{x} \neq \mathbf{y}$ implies $\mathbf{f(x)} \neq \mathbf{f(y)}$.

In coding theory we consider injections that map elements of a countable set $\mathbb{X}$ into strings over a countable set $\mathbb{Y}$. The set of these strings is denoted as $\mathbb{Y}^+ = \bigcup_{n=1}^{\infty} \mathbb{Y}^n$. Sometimes we also consider set $\mathbb{Y}^* = \{\lambda\} \cup \mathbb{Y}^+$ where λ is the empty string. Sets $\mathbb{X}$ and $\mathbb{Y}$ are called alphabets.

Definition (code)

Any injection $\mathbf{B} : \mathbb{X} \to \mathbb{Y}^*$ will be called a *code*.

We will consider mostly binary codes, i.e., codes for which $\mathbb{Y} = \{0, 1\}^*$. On the other hand, the alphabet $\mathbb{X}$ may consist of letters, digits or other symbols.

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

Example of a code

Example

An example of a code:

symbol $\mathbf{x}$:	*code word* $\mathbf{B(x)}$:
a	*0*
b	*1*
c	*10*
d	*11*

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

Uniquely decodable codes

The original purpose of coding is to transmit some representations of strings written with symbols from an alphabet $\mathbb{X}$ through a communication channel which passes only strings written with symbols from a smaller alphabet $\mathbb{Y}$. Thus the idea of a particularly good coding is that we should be able to reconstruct coded symbols from the concatenation of their codes. Formally speaking, the following property is desired.

Definition (uniquely decodable code)

Code $\mathbf{B} : \mathbb{X} \to \mathbb{Y}^*$ is called *uniquely decodable* if the code extension

$$\mathbf{B}^* : \mathbb{X}^* \ni (x_1, \dots, x_n) \mapsto \mathbf{B}(x_1) \dots \mathbf{B}(x_n) \in \mathbb{Y}^*$$

is also an injection.

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

Examples of codes

Example (a code which is not uniquely decodable)

symbol $\mathbf{x}$:	code word $\mathbf{B(x)}$:
a	0
b	1
c	10
d	11

Example (a uniquely decodable code)

symbol $\mathbf{x}$:	code word $\mathbf{B(x)}$:
a	$0c$
b	$1c$
c	$10c$
d	$11c$

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

Comma-separated codes

Definition (comma-separated code)

Let $\mathbf{c} \notin \mathbb{Y}$. Code $\mathbf{B} : \mathbb{X} \to (\mathbb{Y} \cup \{\mathbf{c}\})^*$ is called *comma-separated* if for each $\mathbf{x} \in \mathbb{X}$ there exists a string $\mathbf{w} \in \mathbb{Y}^*$ such that $\mathbf{B(x) = wc}$. Symbol $\mathbf{c}$ is called the comma.

Theorem

Each comma-separated code is uniquely decodable.

Proof

For a comma-separated code $\mathbf{B}$, let us decompose $\mathbf{B(x)} = \phi\mathbf{(x)c}$. We first observe that $\mathbf{B(x_1)...B(x_n) = B(y_1)...B(y_m)}$ holds only if $\mathbf{n = m}$ (the same number of $\mathbf{c}$'s on both sides of equality) and $\phi\mathbf{(x_i)} = \phi\mathbf{(y_i)}$ for $\mathbf{i = 1, ..., n}$. Next, we observe that function ϕ is a code. Hence string $\mathbf{B(x_1)...B(x_n)}$ may be only the image of $\mathbf{(x_1, ..., x_n)}$ under the mapping $\mathbf{B}^*$. This means that code $\mathbf{B}$ is uniquely decodable.

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

Fixed-length codes

Definition (fixed-length code)

Let $\mathbf{n}$ be a fixed natural number. Code $\mathbf{B} : \mathbb{X} \to \mathbb{Y}^{\mathbf{n}}$ is called a *fixed-length code*.

Example

An example of a fixed-length code:

symbol $\mathbf{x}$:	*code word* $\mathbf{B(x)}$:
a	*00*
b	*01*
c	*10*
d	*11*

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

Fixed-length codes (continued)

Theorem

Each fixed-length code is uniquely decodable.

Proof

Consider a fixed-length code $\mathbf{B}$. We observe that $\mathbf{B(x_1)\ldots B(x_n) = B(y_1)\ldots B(y_m)}$ holds only if $\mathbf{n = m}$ (the same length of strings on both sides of equality) and $\mathbf{B(x_i) = B(y_i)}$ for $\mathbf{i = 1, \ldots, n}$. Because $\mathbf{B}$ is an injection, string $\mathbf{B(x_1)\ldots B(x_n)}$ may be only the image of $\mathbf{(x_1, \ldots, x_n)}$ under the mapping $\mathbf{B^*}$. Hence, code $\mathbf{B}$ is uniquely decodable.

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

Expected code length

Let $|\mathbf{w}|$ denote the length of a string $\mathbf{w} \in \mathbb{Y}^*$, measured in the number in symbols. For a random variable $\mathbf{X} : \boldsymbol{\Omega} \to \mathbb{X}$, we will be interested in the expected code length

$$\mathbf{E}\,|\mathbf{B}(\mathbf{X})| = \sum_{\mathbf{x}\in\mathbb{X}} \mathbf{P}(\mathbf{X}=\mathbf{x})\,|\mathbf{B}(\mathbf{x})|\,.$$

Example

Consider the following distribution and a code:

symbol $\mathbf{x}$*:*	$\mathbf{P}(\mathbf{X}=\mathbf{x})$*:*	*code word* $\mathbf{B}(\mathbf{x})$*:*
a	$\mathbf{1/2}$	*0C*
b	$\mathbf{1/6}$	*1C*
c	$\mathbf{1/6}$	*10C*
d	$\mathbf{1/6}$	*11C*

We have $\mathbf{E}\,|\mathbf{B}(\mathbf{X})| = 2\cdot\frac{1}{2} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 3\cdot\frac{1}{6} = 2\frac{1}{3}$.

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

What is the shortest code?

- We are interested in codes that minimize the expected code length for a given probability distribution.
- In this regard, both comma-separated codes and fixed-length codes have advantages and drawbacks.
- If certain symbols appear more often than others then comma-separated codes allow to code them as shorter strings and thus to spare space.
- On the other hand, if all symbols are equiprobable then a fixed-length code without a comma occupies less space than the same code with a comma.

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

Kraft inequality

Theorem (Kraft inequality)

For any uniquely decodable code $\mathbf{B} : \mathbb{X} \to \{0,1\}^*$ *we have*

$$\sum_{x \in \mathbb{X}} 2^{-|B(x)|} \leq 1.$$

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

Kraft inequality (proof)

Proof

Consider an arbitrary L. Let $a(m, n, L)$ denote the number of sequences $(x_1, \ldots, x_n)$ such that $|B(x_i)| \leq L$ and the length of $B^*(x_1, \ldots, x_n)$ equals m. We have

$$\left(\sum_{x:|B(x)| \leq L} 2^{-|B(x)|} \right)^n = \sum_{m=1}^{nL} a(m, n, L) \cdot 2^{-m}.$$

Because the code is uniquely decodable, we have $a(m, n, L) \leq 2^m$. Therefore

$$\sum_{x:|B(x)| \leq L} 2^{-|B(x)|} \leq (nL)^{1/n} \xrightarrow{n \to \infty} 1.$$

Letting $L \to \infty$, we obtain the Kraft inequality.

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

Source coding inequality

Theorem (source coding inequality)

For any uniquely decodable code $\mathbf{B} : \mathbb{X} \to \{0,1\}^*$, *the expected length of the code satisfies inequality*

$$\mathbf{E}\ |\mathbf{B}(\mathbf{X})| \geq \mathbf{H}(\mathbf{X}),$$

where $\mathbf{H}(\mathbf{X})$ *is the entropy of* $\mathbf{X}$.

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

Source coding inequality (proof)

Proof

Introduce probability distributions $p(x) = P(X = x)$ and

$$r(x) = \frac{2^{-|B(x)|}}{\sum_{y \in \mathbb{X}} 2^{-|B(y)|}}.$$

We have

$$\begin{aligned} E\,|B(X)| - H(X) &= \sum_{x:p(x)>0} p(x) \log \frac{p(x)}{r(x)} - \log \left(\sum_{x \in \mathbb{X}} 2^{-|B(x)|} \right) \\ &= D(p||r) - \log \left(\sum_{x \in \mathbb{X}} 2^{-|B(x)|} \right). \end{aligned}$$

That difference is nonnegative by nonnegativity of Kullback-Leibler divergence and Kraft inequality.

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

Prefix-free and suffix-free codes

Definition (prefix-free code)

A code $\mathbf{B}$ is called *prefix-free* if no code word $\mathbf{B(x)}$ is a prefix of another code word $\mathbf{B(y)}$, i.e., it is not true that $\mathbf{B(y)} = \mathbf{B(x)u}$ for $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{u} \in \mathbb{Y}^*$.

Definition (suffix-free code)

A code $\mathbf{B}$ is called *suffix-free* if no code word $\mathbf{B(x)}$ is a suffix of another code word $\mathbf{B(y)}$, i.e., it is not true that $\mathbf{B(y)} = \mathbf{uB(x)}$ for $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{u} \in \mathbb{Y}^*$.

Example (a code which is prefix-free but not suffix-free)

symbol $\mathbf{x}$:	*code word* $\mathbf{B(x)}$:
a	*10*
b	*0*
c	*11*

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

Prefix-free and suffix-free codes (continued)

Theorem

Any prefix-free or suffix-free code is uniquely decodable.

Proof

Without loss of generality we shall restrict ourselves to prefix-free codes. The proof for suffix-free codes is mirror-like. Let $\mathbf{B}$ be a prefix-free code and assume that $\mathbf{B(x_1)...B(x_n) = B(y_1)...B(y_m)}$. By the prefix-free property the initial segments $\mathbf{B(x_1)}$ and $\mathbf{B(y_1)}$ must match exactly and $\mathbf{x_1 = y_1}$. The analogous argument applied by induction yields $\mathbf{x_i = y_i}$ for $\mathbf{i = 2, .., n}$ and $\mathbf{n = m}$. Thus code $\mathbf{B}$ is uniquely decodable.

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

A theorem converse to Kraft inequality

Theorem

If function $\mathbf{l} : \mathbb{X} \to \mathbb{N}$ *satisfies inequality*

$$\sum_{\mathbf{x} \in \mathbb{X}} \mathbf{2}^{-\mathbf{l}(\mathbf{x})} \leq \mathbf{1}$$

then we may construct a prefix-free code $\mathbf{B} : \mathbb{X} \to \{0,1\}^*$ *such that* $|\mathbf{B}(\mathbf{x})| = \mathbf{l}(\mathbf{x})$.

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

Complete codes

Definition (complete code)

A code $\mathbf{B} : \mathbb{X} \to \{0,1\}^*$ is called *complete* if

$$\sum_{x \in \mathbb{X}} 2^{-|B(x)|} = 1.$$

Example

A code which is prefix-free, suffix-free, and complete.

symbol $\mathbf{x}$*:*	*code word* $\mathbf{B(x)}$*:*
a	*00*
b	*01*
c	*10*
d	*11*

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

Complete codes (continued)

Example

Another code which is prefix-free, suffix-free, and complete.

symbol $\mathbf{x}$:	*code word* $\mathbf{B(x)}$:
a	*01*
b	*000*
c	*100*
d	*110*
e	*111*
f	*0010*
g	*0011*
h	*1010*
i	*1011*

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

Shannon-Fano code

Definition (Shannon-Fano code)

A prefix-free code $\mathbf{B} : \mathbb{X} \to \{0,1\}^*$ is called a *Shannon-Fano code* if

$$|\mathbf{B}(\mathbf{x})| = \lceil -\log \mathbf{P}(\mathbf{X} = \mathbf{x}) \rceil \,.$$

Theorem

Shannon-Fano codes exist for any distribution and satisfy

$$\mathbf{H}(\mathbf{X}) \leq \mathbf{E}\,|\mathbf{B}(\mathbf{X})| \leq \mathbf{H}(\mathbf{X}) + 1.$$

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

Shannon-Fano code (continued)

Proof

We have

$$\sum_{x \in \mathbb{X}} 2^{-\lceil -\log P(X=x) \rceil} \leq \sum_{x \in \mathbb{X}} 2^{\log P(X=x)} \leq 1.$$

Hence Shannon-Fano codes exist. The other claim follows by

$$-\log P(X=x) \leq |B(x)| \leq -\log P(X=x) + 1.$$

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

Drawbacks of the Shannon-Fano code

Shannon-Fano code is not necessarily the shortest possible code.

Example

Consider the following distribution and codes:

symbol $\mathbf{x}$:	$\mathbf{P(X = x)}$:	*code word* $\mathbf{B(x)}$:	*code word* $\mathbf{C(x)}$
a	$\mathbf{1 - 2^{-5}}$	*0*	*0*
b	$\mathbf{2^{-6}}$	*100000*	*10*
c	$\mathbf{2^{-6}}$	*100001*	*11*

Code **B** is a Shannon-Fano code, whereas code **C** is another code. We have $\mathbf{H(X) = 0.231...}$, $\mathbf{E\,|B(X)| = 1.15625}$, and $\mathbf{E\,|C(X)| = 1.03125}$. For no symbol code **C** is worse than code **B**, whereas for less probable symbols code **C** is much better.

A code that minimizes the expected code length is known under the name of Huffman code.

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

Trees and paths

Definition (binary tree)

A *binary tree* is a directed acyclic connected graph where each node has at most two children nodes and at most one parent node. The node which has no parents is called the *root node*. The nodes which have no children are called *leaf nodes*. We assume that links to the left children are labeled with $\mathbf{0}$'s whereas links to the right children are labeled with $\mathbf{1}$'s. Moreover, some nodes may be labeled with some symbols as well.

Definition (path)

We say that a binary tree contains a *path* $\mathbf{w} \in \{\mathbf{0}, \mathbf{1}\}^*$ if there is a sequence of links starting from the root node and labeled with the consecutive symbols of $\mathbf{w}$. We say that the path is ended with symbol $\mathbf{a} \in \mathbb{X}$ if the last link of the sequence ends in a node labeled with symbol $\mathbf{a}$.

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

Code trees

Definition (code tree)

The *code tree* for a code $\mathbf{B} : \mathbb{X} \to \{0,1\}^*$ is a labeled binary tree which contains a path $\mathbf{w}$ if and only if $\mathbf{B(a) = w}$ for some $\mathbf{a} \in \mathbb{X}$, and exactly in that case we require that path $\mathbf{w}$ is ended with symbol $\mathbf{a}$.

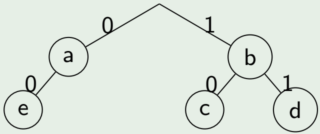
Project co-financed by the European Union
within the framework of the European Social Fund

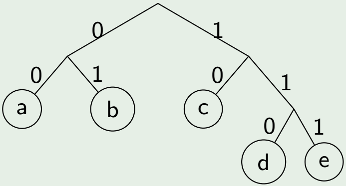
Examples of code trees

Example

symbol **x**:	*code word* **B(x)**:	*code word* **C(x)**:
a	*0*	*00*
b	*1*	*01*
c	*10*	*10*
d	*11*	*110*
e	*00*	*111*

The code trees for these codes are:





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

Weighted code trees

Definition (weighted code tree)

The *weighted code tree* for a prefix code $\mathbf{B} : \mathbb{X} \to \{0,1\}^*$ and a probability distribution $\mathbf{p} : \mathbb{X} \to [0,1]$ is the code tree for code $\mathbf{B}$ where the nodes are enhanced with the following weights: (1) for a leaf node with symbol $\mathbf{a}$, we add weight $\mathbf{p(a)}$, (2) to other (internal) nodes, we ascribe weights equal to the sum of weights of their children.

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

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

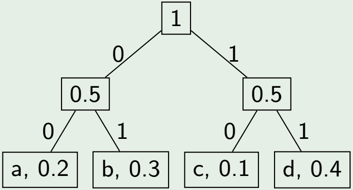
PhD STUDIES

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Example of a weighted code tree

Example

symbol **x**:	**p(x)**:	*code word* **C(x)**:
a	**0.2**	*00*
b	**0.3**	*01*
c	**0.1**	*10*
d	**0.4**	*11*



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

Huffman code

Definition (Huffman code)

The *Huffman code* for a probability distribution $\mathbf{p} : \mathbb{X} \to [\mathbf{0}, \mathbf{1}]$ is a code whose weighted code tree is constructed by the following algorithm:

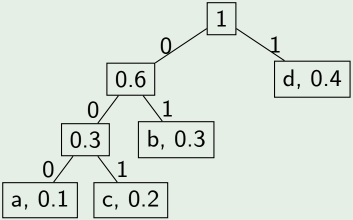
1. Create a leaf node for each symbol and add them to a list.
2. While there is more than one node in the list:
   1. Remove two nodes of the lowest weight from the list.
   2. Create a new internal node with these two nodes as children and with weight equal to the sum of the two nodes' weights.
   3. Add the new node to the list.
3. The remaining node is the root node and the tree is complete.

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

Example of Huffman code

Example

symbol **x**:	**p(x)**:	*Huffman code* **B(x)**:
a	**0.2**	*000*
b	**0.3**	*01*
c	**0.1**	*001*
d	**0.4**	*1*



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

KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Optimality of Huffman code

A code $\mathbf{B}$ will be called optimal for a given distribution $\mathbf{p(x) = P(X = x)}$ if $\mathbf{E}\ |\mathbf{B(X)}|$ achieves the minimum.

Theorem

For any probability distribution, the Huffman code is optimal.

To prove the theorem, we will use the this fact:

Lemma

Consider the two symbols $\mathbf{x}$ *and* $\mathbf{y}$ *with the smallest probabilities. Then there is an optimal code tree* $\mathbf{C}$ *such that these two symbols are sibling leaves in the lowest level of* $\mathbf{C}$*'s code tree.*

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

Proof of the lemma

Proof

Every internal node in a code tree for an optimal code must have two children. Then let $\mathbf{B}$ be an optimal code and let symbols $\mathbf{a}$ and $\mathbf{b}$ be two siblings at the maximal depth of $\mathbf{B}$'s code tree. Assume without loss of generality that $\mathbf{p(x)} \leq \mathbf{p(y)}$ and $\mathbf{p(a)} \leq \mathbf{p(b)}$. We have $\mathbf{p(x)} \leq \mathbf{p(a)}$, $\mathbf{p(y)} \leq \mathbf{p(b)}$, $|\mathbf{B(a)}| \geq |\mathbf{B(x)}|$, and $|\mathbf{B(b)}| \geq |\mathbf{B(y)}|$. Now let $\mathbf{C}$'s code tree differ from the $\mathbf{B}$'s code tree by switching $\mathbf{a} \leftrightarrow \mathbf{x}$ and $\mathbf{b} \leftrightarrow \mathbf{y}$. Then we obtain

$$\begin{aligned} &\mathbf{E}\ |\mathbf{C(X)}| - \mathbf{E}\ |\mathbf{B(X)}| \\ &= (\mathbf{p(a)} - \mathbf{p(x)})(|\mathbf{B(x)}| - |\mathbf{B(a)}|) \\ &\quad + (\mathbf{p(b)} - \mathbf{p(y)})(|\mathbf{B(y)}| - |\mathbf{B(b)}|) \leq \mathbf{0}. \end{aligned}$$

Hence code $\mathbf{C}$ is also optimal.

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

Proof of Huffman code's optimality

Now we will proceed by induction on the number of symbols in the alphabet $\mathbb{X}$. If $\mathbb{X}$ contains only two symbols, then Huffman code is optimal. In the second step, we assume that Huffman code is optimal for $n-1$ symbols and we prove its optimality for n symbols. Let C be an optimal code for n symbols. Without loss of generality we may assume that symbols x and y having the smallest probabilities occupy two sibling leaves in the lowest level of C's code tree. Then from the weighted code tree of C we construct a code C' for $n-1$ symbols by removing nodes with symbols x and y and ascribing a symbol z to its parent node. Hence we have

$$E\left|C'(X')\right| = E\left|C(X)\right| - p(x) - p(y),$$

where variable $X' = z$ if $X \in \{x, y\}$ and $X' = X$ otherwise.

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

Proof of Huffman code's optimality (continued)

On the other hand, let $\mathbf{B}'$ be the Huffman code for $\mathbf{X}'$ and let $\mathbf{B}$ be the code constructed from $\mathbf{B}'$ by adding leaves with symbols $\mathbf{x}$ and $\mathbf{y}$ to the node with symbol $\mathbf{z}$. By construction, code $\mathbf{B}$ is the Huffman code for $\mathbf{X}$. We have

$$\mathbf{E}\left|\mathbf{B}'(\mathbf{X}')\right| = \mathbf{E}\left|\mathbf{B}(\mathbf{X})\right| - \mathbf{p}(\mathbf{x}) - \mathbf{p}(\mathbf{y}).$$

Because $\mathbf{E}\left|\mathbf{B}'(\mathbf{X}')\right| \le \mathbf{E}\left|\mathbf{C}'(\mathbf{X}')\right|$ by optimality of Huffman code $\mathbf{B}'$, we obtain $\mathbf{E}\left|\mathbf{B}(\mathbf{X})\right| \le \mathbf{E}\left|\mathbf{C}(\mathbf{X})\right|$. Hence Huffman code $\mathbf{B}$ is also optimal.

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