# List of other possible dissimilarities

To compare histograms $H = (h_1, h_2, \ldots, h_n)$ and $K = (k_1, k_2, \ldots, k_n)$

## Bin-by-bin dissimilarity measures

Only pairs of bins that have the same index are matched – dissimilarity is a combination of all the pairwise differences:

$L_1$ distance $d_{L_1}(H,K) = \sum_i |h_i - k_i|$

Kullback-Leibler divergence $d_{KL}(H,K) = \sum_i h_i \log \frac{h_i}{k_i}$

Jeffrey divergence

$$d_J(H,K) = \sum_i \left( h_i \log \frac{h_i}{m_i} + k_i \log \frac{k_i}{m_i} \right), \; m_i = \frac{h_i + k_i}{2}$$

Aitchison distance $d_A^2(H,K) = \sum_i \left( \log \frac{h_i}{g(H)} - \log \frac{k_i}{g(K)} \right)^2$

$g(H)$, $g(K)$ - the geometric mean.

# Cross-bin dissimilarity measures

Information on a ground distance are incorporated.

Qadratic-form distance $d_S^2(H,K) = (H-K)^T S (H-K)$

## similarity matrix

$$S = \left[1 - \frac{d_{ij}}{d_{\max}}\right], \; d_{ij} = |i-j|, \; d_{\max} = \max_{ij} d_{ij}$$

$$S = \begin{bmatrix} 1 & 1-\frac{1}{4} & 1-\frac{2}{4} & 1-\frac{3}{4} & 0 \\ 1-\frac{1}{4} & 1 & 1-\frac{1}{4} & 1-\frac{2}{4} & 1-\frac{3}{4} \\ 1-\frac{2}{4} & 1-\frac{1}{4} & 1 & 1-\frac{1}{4} & 1-\frac{2}{4} \\ 1-\frac{3}{4} & 1-\frac{2}{4} & 1-\frac{1}{4} & 1 & 1-\frac{1}{4} \\ 0 & 1-\frac{3}{4} & 1-\frac{2}{4} & 1-\frac{1}{4} & 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 4 & 3 & 2 & 1 & 0 \\ 3 & 4 & 3 & 2 & 1 \\ 2 & 3 & 4 & 3 & 2 \\ 1 & 2 & 3 & 4 & 3 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

# Match distance

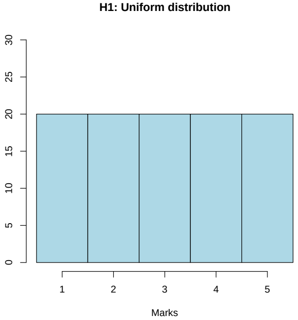
$$d_M(H,K) = \sum_i \left| \hat{h}_i - \hat{k}_i \right|$$

$\hat{h}_i = \sum_{j \leq i} h_j$ the cummulative histogram of $H$
$\hat{k}_i = \sum_{j \leq i} k_j$ the cummulative histogram of $K$

This distance represents a special case of the Earth Mover's Distance (EMD) for one-dimensional histograms with equal areas.

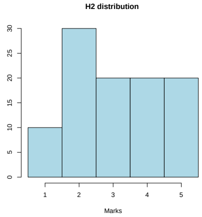
# Test data sets and perceptual dissimilarity

$$H1 = (20, 20, 20, 20, 20)$$

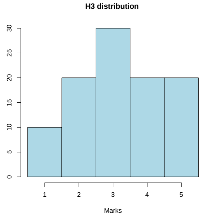


# Test data sets and perceptual dissimilarity

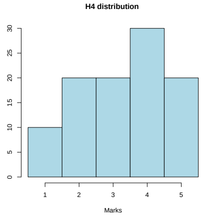
$H2 = (10, 30, 20, 20, 20)$



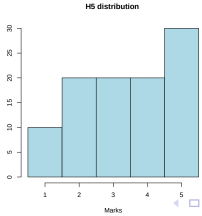

$H3 = (10, 20, 30, 20, 20)$




$H4 = (10, 20, 20, 30, 20)$




$H5 = (10, 20, 20, 20, 30)$



## Perceptual dissimilarities are **increasing**

$$d(H1, H2) < d(H1, H3) < d(H1, H4) < d(H1, H5) \quad (1)$$
$$d(H2, H3) < d(H2, H4) < d(H2, H5) \quad (2)$$
$$d(H3, H4) < d(H3, H5) \quad (3)$$

# Dissimilarity matrices

## $L_1$ **distance** is not increasing:

| $d_{L_1}$ | H2 | H3 | H4 | H5 |
|---|---|---|---|---|
| H1 | 20 | 20 | 20 | 20 |
| H2 | | 20 | 20 | 20 |
| H3 | | | 20 | 20 |
| H4 | | | | 20 |

# Dissimilarity matrices

**Kullback-Leibler divergence** is not increasing:

| $d_{KL}$ | H2 | H3 | H4 | H5 |
|---|---|---|---|---|
| H1 | 5.75 | 5.75 | 5.75 | 5.75 |
| H2 | | 4.05 | 4.05 | 4.05 |
| H3 | | | 4.05 | 4.05 |
| H4 | | | | 4.05 |

# Dissimilarity matrices

**Quadratic form distance** is increasing:

| $d_S^2$ | H2 | H3 | H4 | H5 |
|---|---|---|---|---|
| H1 | 50 | 100 | 150 | 200 |
| H2 | | 50 | 100 | 150 |
| H3 | | | 50 | 100 |
| H4 | | | | 50 |

**Match distance** is increasing:

| $d_M$ | H2 | H3 | H4 | H5 |
|---|---|---|---|---|
| H1 | 10 | 20 | 30 | 40 |
| H2 | | 10 | 20 | 30 |
| H3 | | | 20 | 30 |
| H4 | | | | 10 |

# Dissimilarity matrices

**Directed dissimilarity** is increasing:

| $D$ | H2 | H3 | H4 | H5 |
|---|---|---|---|---|
| H1 | 0.04 | 0.08 | 0.012 | 0.16 |
| H2 | | 0.05 | 0.09 | 0.13 |
| H3 | | | 0.05 | 0.09 |
| H4 | | | | 0.05 |

# Discussion and Suggestions

Quadratic form distance, Match distance and Directed dissimilarity are increasing in accordance with perceptual dissimilarities.

**Further work:**

- To select the most appropriate dissimilarity by applying other test sets of distributions;
- To perform comparison of achievements through grading (school vs. national grades, teachers vs. national grades, teacher vs. teacher in the same subject in a school, subject vs. subject . . . ) with selected dissimilarity and adequate classification method and
- to define school's, teacher's, student's or class *profile*.