Before turning our attention to the results, we describe certain statistical particularities discovered in the above example, and consider a qualitative method for determining the thresholds for separating cases (1) and (3). Note that all the qualitative arguments and the subsequent items are confirmed a posteriori, because they lead to a more precise picture of the distribution of the essential relations with respect to the matrix. However, general characteristics of relation matrix are stable under the threshold value oscillations, parameters k and p (i.e., lengths of the determining and connecting neighbourhoods), and also certain changes in the definition of the relation (see Item 13).

13. Frequency histograms for the appearance of relations. The choice of thresholds. Below, in constructing certain frequency histograms for the appearance of relations in a matrix, we shall have to break the interval where the relation is measured into equal disjoint segments. We will simply assume that the value of the relation is replaced by its integral parts (on account of the choice of the factor c in (5), we can reduce the general case to the above).

We now study how the relation between two neighbourhoods in X and the number of common names are connected. By definition, the number of common names (taken with multiplicities) of neighbourhoods Δ_r and Δ_s is the number of pairs from $\Delta_r \times \Delta_s$ such that they contain indentical names, viz.,

$$O(\Delta_r(k), \Delta_s(k)) = \sum_{i=r-k}^{r+k} \sum_{j=s-k}^{s+k} \delta_{a_i,a_j}; \qquad \delta_{a_i,a_j} = \begin{cases} 1, & u(a_i) = u(a_j), \\ 0 & \text{otherwise.} \end{cases}$$

We denote by Π the list of the names of Roman popes, and by N the list of the names of Roman emperors.

It turns out that, provided that $O(\Delta_r, \Delta_s)$ is fixed, the frequency histograms $L_0(\Delta_r, \Delta_s)$ with respect to the matrix for Π and N indicate that the dependence of $L_0(\Delta_r, \Delta_s)$ and $O(\Delta_r, \Delta_s)$ is expressed in explicit terms, viz., as the number of common names increases, the relation increases, too (in the statistical sense). It may seem that the relation L_0 increases directly on account of the common names, since the mechanisms leading to such an increase do exist. However, this is not so. For a demonstration, we introduce two additional relation measures. Consider a neighbourhood pair $\Delta_r(k)$ and $\Delta_s(k)$, then

$\Delta_r \supset \Delta'_r = \{$ set of entries of Δ_r with different names$\}$,

$\Delta_s \supset \Delta'_s = \{$ set of entries of Δ_s with different names$\}$,

$\Delta'_r \supset \Delta''_{r,s} = \{$ set of entries of Δ'_r whose names do not coincide with those from $\Delta'_s\}$.

Thus, the neighbourhoods Δ'_r and Δ'_s contain one representative of each name; besides, Δ'_s and $\Delta''_{r,s}$ contain no common names. Denote the length (number of terms) of a neighbourhood by $|\cdot|$. By definition, we put

$$L_1(\Delta_r, \Delta_s) = \frac{c}{|\Delta'_r| \times |\Delta'_s|} \sum_{a\in\Delta'_r,\ b\in\Delta'_s} l(a,b), \tag{6}$$

$$L_2(\Delta_r, \Delta_s) = \frac{c}{|\Delta''_{r,s}| \times |\Delta'_s|} \sum_{a\in\Delta''_{r,s};\ b\in\Delta'_s} l(a,b) \tag{7}$$