Case 3. Two non-intersecting neighbourhoods Δ_r and Δ_s in X *possess j common complete duplicates*. The mean value of their relation is then equal to $c\alpha + jE_0(\Delta_r, \Delta_s)$.

We have to separate Cases 1 and 3, for which we shall try to find the optimal radius of the determining neighbourhood (k). Note that, by increasing k we decrease the scattering with respect to α (variance of the relation L_0), which in turn increases the precision of the separation. However, for too large k, the duplicate completion degree lessens, thus leading to actually decreasing the factor. The value of k must not exceed the typical length of the elementary chronicle Z_i (see § 11). The optimal value is chosen from experience.

Remark. Since the system $\{\Delta_{k+1}, \Delta_{k+2}, \ldots, \Delta_{n-k-1}\}$ is that of 'current' neighbourhoods in the list X, then less pure neighbourhood duplicates than the given one are neighbouring to each other. To distinguish the most complete duplicates, we will only retain local maxima in the relation matrix $a_{rs} = L_0(\Delta_r, \Delta_s)$ and consider the relation only in the case where

$$L_0(\Delta_r, \Delta_s) \geqslant L_0(\Delta_r, \Delta_{s-1}) - \epsilon, \quad L_0(\Delta_r, \Delta_s) \geqslant L_0(\Delta_r, \Delta_{s+1}) - \epsilon$$

or else it is replaced by zero. This remark does not concern the construction of the frequency histograms (see below). The value of ϵ was chosen to be equal to the length of the interval to be divided in constructing the frequency histogram (see above).

Before turning our attention to the results, we describe certain statistical peculiarities 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 of this and the subsequent items are confirmed a posteriori, because they lead to a more precise picture of the distribution of essential relations with respect to the matrix. However, its general characteristics are stable under the threshold value oscillations, parameters k and p (that is lengths of the determining and connecting neighbourhoods), and also certain changes in the definition of the relation (see above).

14 The relation matrix

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 part. (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 there are identical names in them, namely

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

We denote by P *the list of the Roman popes' names*, and by N *the list of Roman emperors' names*. It turns out that the frequency histograms $L_0(\Delta_r, \Delta_s)$ with respect to the matrix for P and N, provided that $0(\Delta_r, \Delta_s)$ is fixed, indicate that the dependence of $L_0(\Delta_r, \Delta_s)$ is expressed explicitly, namely, as the number of common names increases, the relation increases too (in the statistical sense). It may seem that relation L_0 increases