Independent statistical observables for ultrametric disordered populations

B. G. Giraud
Service de Physique Théorique, Centre d'Etudes de Saclay, 91191 Gif sur Yvette, France
(Received 11 December 1998; revised manuscript received 20 August 1999)

It is not exceptional that a sample of N random data X_i, $i=1,\ldots,N$ contains ultrametric covariations, namely the matrix $\mathcal{C}$ with matrix elements $\langle X_iX_j\rangle-\langle X_i\rangle\langle X_j\rangle$ is ultrametric. We define independent (decorrelated) ''collective'' observables by diagonalizing this matrix. Symmetry properties of such eigenvectors are discussed. Often also, however, while the existence of an ultrametric tree is known, the degrees of parentage of the data are unknown, because a random perturbation confuses the labeling of the leaves of the tree. We sort out those observables which are more robust with respect to such labeling mistakes. [S1063-651X(99)08612-2]

PACS number(s): 87.10.+e, 87.23.−n

I. INTRODUCTION

Ultrametricity [1] is a useful concept in several fields of data analysis. For instance, in physics, overlaps of replicas for spin glasses most likely show, at low temperatures, the phenomenon of replica symmetry breaking [2]. In biology, taxonomic or genetic trees may sometimes show, at least approximately, ultrametric covariations [3–5]. More generally, every time one knows that random variables are not independent, deviations from the central limit theorem are very likely and may even become quite strong [6,5]. The minimal precaution to be implemented is then to analyze such likely deviations and, furthermore, to rearrange the degrees of freedom into independent observables. When only linear rearrangements are considered, a list of uncorrelated observables is obtained by the diagonalization of the covariation matrix $\mathcal{C}$. (As usual, this matrix is defined by its elements $C_{ij}=\langle X_iX_j\rangle-\langle X_i\rangle\langle X_j\rangle$, where $\langle\ \rangle$ denotes the probabilistic average with respect to the probability governing the variables X_i.)

This paper is concerned with statistics in the special case where any average property of each element of the sample under study is the same for each element. This is an important symmetry of many practical problems: $\langle X_i\rangle$ is a constant μ, independent of i. Moreover, we restrict our subject to those cases where the matrix $\mathcal{C}$ is ''binary ultrametric,'' because of the additional symmetries of such matrices. These symmetries will be reflected in special properties of the eigenvectors, naturally. There are other ultrametricities than ''binary'' ones, but the binary scheme is not a severe restriction for biological models, at least.

Such results, however, are often devalued by a lack of a precise knowledge of the parentage relationships between individuals. While straight (symmetric) statistical means of individual properties across the whole sample are insensitive to permutations between elements, measures of heterogeneity between and across subgroups of the sample (families, superfamilies, etc.) may loose significance when a proper labeling of the leaves of the ultrametric tree is missing. Subgroups become ill defined and measures of differences between subgroups loose their relevance. There is thus a need for ''collective observables,'' which minimize labeling errors.

This paper is organized as follows. Section II states the model we are studying and the relevant notations. Section III contains the diagonalization of $\mathcal{C}$ and lists several properties of its eigenvectors. Section IV investigates consequences of ''confusion of labels'' in the sample under study and states a few theorems for minimizing the consequences of this confusion. Section V slightly restricts the model to a case, inspired by biology, where the problem of robustness can be solved analytically. Section VI, finally, contains a discussion and a conclusion. An Appendix briefly investigates non binary ultrametricity.

II. MODEL AND NOTATIONS

In the following, we consider binary trees only, as illustrated by Fig. 1. Namely, the sample of degrees of freedom contains $N=2^G$ elements, with G the number of ''generations.'' For notational simplicity, Fig. 1 shows three generations only, $G=3$, and the degrees of freedom are labeled $s,t,\ldots,z$ instead of $X_1,X_2,\ldots,X_8$, respectively.

Degrees u and v, e.g., have parentage 1, because of their nearest common ancestor, o. In turn, e.g., degrees x and z have parentage 2 because of ancestor m. And so on. Ultrametricity is implemented if, whenever X_i and X_j have parentage ν, then C_{ij} depends on ν only, $C_{ij}=c_\nu$.

For notational convenience again, we now slightly change the definition of $\mathcal{C}$, by adding μ^2 to all its matrix elements. Namely, now, $C_{ij}=\langle X_iX_j\rangle$ in the following. This adds a con-

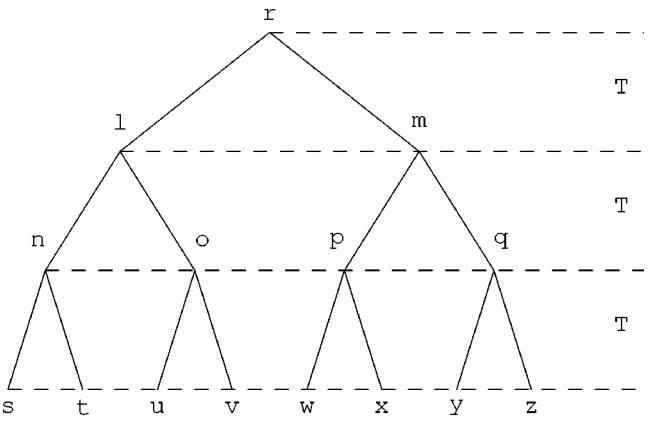

FIG. 1. Binary tree with 3 generations. For simplicity of the model, each generation has the same lifetime T.

stant μ^2 to all the numbers c_ν, without changing the definition of ultrametricity. We then assume that the degrees of freedom are normalized according to the condition, $c_0 \equiv \langle X_i^2 \rangle = 1$. This normalization identifies covariation with correlation. Then, for $\nu \geqslant 1$, and naturally for $\nu \leqslant G$, the real numbers c_ν are arbitrary within the trivial constraint, $|c_\nu| < 1$. In the case of Fig. 1, the matrix $\mathcal{C}$ then reads,

$$\mathcal{C}_3 = \begin{bmatrix} 1 & c_1 & c_2 & c_2 & c_3 & c_3 & c_3 & c_3 \\ c_1 & 1 & c_2 & c_2 & c_3 & c_3 & c_3 & c_3 \\ c_2 & c_2 & 1 & c_1 & c_3 & c_3 & c_3 & c_3 \\ c_2 & c_2 & c_1 & 1 & c_3 & c_3 & c_3 & c_3 \\ c_3 & c_3 & c_3 & c_3 & 1 & c_1 & c_2 & c_2 \\ c_3 & c_3 & c_3 & c_3 & c_1 & 1 & c_2 & c_2 \\ c_3 & c_3 & c_3 & c_3 & c_2 & c_2 & 1 & c_1 \\ c_3 & c_3 & c_3 & c_3 & c_2 & c_2 & c_1 & 1 \end{bmatrix}. \tag{1}$$

Obviously, $\mathcal{C}$ is invariant under a large subgroup of the permutation group of $N = 2^G$ elements. This subgroup has ''parities'' as factors, in the sense that such factors are permutations $\mathcal{P}$ whose square is the identity permutation, $\mathcal{P}^2 = \mathcal{I}$. These are (i) $N/2$ parities $\mathcal{P}_1^\pi$, $\pi = 1, \ldots, N/2$, that switch the two members of a ''$\nu = 1$ minifamily,'' e.g., $\mathcal{P}_1^1$ induces the exchange $X_1 \leftrightarrow X_2$, and $\mathcal{P}_1^{N/2}$ induces the exchange $X_{N-1} \leftrightarrow X_N$, (ii) $N/4$ parities $\mathcal{P}_2^\pi$, $\pi = 1, \ldots, N/4$ that switch the two $\nu = 1$ minifamilies of a ''$\nu \leqslant 2$ family,'' but without disturbing the internal order within each minifamily, e.g., $\mathcal{P}_2^1$ induces the exchange $(X_1 X_2) \leftrightarrow (X_3 X_4)$, and $\mathcal{P}_2^{N/4}$ induces the exchange $(X_{N-3} X_{N-2}) \leftrightarrow (X_{N-1} X_N)$, and so on, until (iii) the ''superparity'' $\mathcal{P}_G$, which exchanges, at parentage G, the two ''superfamilies'' with population numbers 2^{G-1} at inner parentage $\nu \leqslant G-1$. For Fig. 1 this reads, $(stuv) \equiv (X_1 \ldots X_{N/2}) \leftrightarrow (wxyz) \equiv (X_{N/2+1} \ldots X_N)$. The generalization to any value of G is trivial.

In order to diagonalize $\mathcal{C}$ the next Section, Sec. III, will take advantage of the fact that $\mathcal{P}_1^1, \mathcal{P}_2^1, \ldots, \mathcal{P}_G$ provide a complete set of labels, classifying all the eigenstates. As will be seen in the sequel, any other list of ''parities'' $\mathcal{P}_1^\pi, \mathcal{P}_2^{\pi'}, \ldots, \mathcal{P}_G$ gives the same labeling, hence we shall use the shorter notation $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_G$.

To avoid confusions in the wording used by this paper, we specify that three kinds of averages are considered: (i) probabilistic averages $\langle\ \rangle$, governed by the (correlated) probability distribution which drives the set of N degrees of freedom $\{X_i\}$, (ii) statistical means, namely weighted sums of the variables X_i, and (iii) averages over permutations of the labels i. As will be seen in the following, the name ''observable,'' which we shall use for a statistical mean, should be clear enough. The notation $\langle\ \rangle$ should also clarify what belongs in (i). Finally, we retain an explicit factor $(N!)^{-1}$ whenever permutation averaging occurs; this factor should make a convenient signature for (iii).

III. EIGENOBSERVABLES AND THEIR FOURIER INTERPRETATION

The following list of almost obvious results, without detailed proofs, makes a constructive derivation of the eigenvectors and eigenvalues of $\mathcal{C}$.

Case $G = 1$: $N = 2$, and only $\mathcal{P}_1$ exists. Then, $\mathcal{C}$ reads,

$$\mathcal{C}_1 = \begin{bmatrix} 1 & c_1 \\ c_1 & 1 \end{bmatrix}, \tag{2}$$

and the matrix of row (left) eigenvectors, completed by the corresponding eigenvalues of $\mathcal{C}$ and $\mathcal{P}_1$, respectively, reads,

$$\mathcal{E}_1 = \begin{bmatrix} +1 & +1 & 1+c_1 & + \\ +1 & -1 & 1-c_1 & - \end{bmatrix}. \tag{3}$$

Case $G = 2$: $N = 4$, and now $\mathcal{P}_2$ joins $\mathcal{P}_1$. Then

$$\mathcal{C}_2 = \begin{bmatrix} 1 & c_1 & c_2 & c_2 \\ c_1 & 1 & c_2 & c_2 \\ c_2 & c_2 & 1 & c_1 \\ c_2 & c_2 & c_1 & 1 \end{bmatrix}, \tag{4}$$

and the matrix of row eigenvectors, completed by the corresponding eigenvalues of $\mathcal{C}$, $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, reads,

$$\mathcal{E}_2 = \begin{bmatrix} +1 & +1 & +1 & +1 & 1+c_1+2c_2 & + & + \\ +1 & +1 & -1 & -1 & 1+c_1-2c_2 & + & - \\ +1 & -1 & +1 & -1 & 1-c_1 & - & + \\ +1 & -1 & -1 & +1 & 1-c_1 & - & - \end{bmatrix}. \tag{5}$$

This table of eigenvectors is obtained by a duplication of the previous table of eigenvectors and then, as indicated by $\mathcal{P}_2$, a symmetrization and an antisymmetrization. An equivalent table of eigenvectors is

$$\mathcal{E}_2' = \begin{bmatrix} +1 & +1 & +1 & +1 & 1+c_1+2c_2 & + & + \\ +1 & +1 & -1 & -1 & 1+c_1-2c_2 & + & - \\ +1 & -1 & 0 & 0 & 1-c_1 & - & \\ 0 & 0 & -1 & +1 & 1-c_1 & - & \end{bmatrix}, \tag{6}$$

where degenerate eigenvectors are simplified, at the cost of their $\mathcal{P}_2$ labels. This form, Eq. (6), better illustrates the construction of eigenvectors and eigenvalues ''at level G'' when those ''at level $G-1$'' are known. The key of the derivation, besides the labeling parities already mentioned, is the fact that a matrix with constant matrix elements is essentially null, except for its only nondegenerate eigenvector, of the form $(1,1,\ldots,1)$.

Case $G = 3$: $N = 8$, and now $\mathcal{P}_3$ comes in. For $\mathcal{C}$ we refer to Eq. (1). The matrix of row (left) eigenvectors, completed by the corresponding eigenvalues of $\mathcal{C}, \mathcal{P}_1, \mathcal{P}_2$ and $\mathcal{P}_3$, respectively, reads

$$\mathcal{E}_3=\begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & 1+c_1+2c_2+4c_3 & + & + & + \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & 1+c_1+2c_2-4c_3 & + & + & - \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 & 1+c_1-2c_2 & + & - & + \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & 1+c_1-2c_2 & + & - & - \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & 1-c_1 & - & + & + \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & 1-c_1 & - & + & - \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 & 1-c_1 & - & - & + \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & 1-c_1 & - & - & - \end{bmatrix}. \tag{7}$$

This table is obtained by a duplication of the previous table of eigenvectors and then, as indicated by $\mathcal{P}_3$, a symmetrization and an antisymmetrization. An equivalent table of row eigenvectors is

$$\mathcal{E}_3'=\begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & 1+c_1+2c_2+4c_3 & + & + & + \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & 1+c_1+2c_2-4c_3 & + & + & - \\ +1 & +1 & -1 & -1 & 0 & 0 & 0 & 0 & 1+c_1-2c_2 & + & - & \\ 0 & 0 & 0 & 0 & +1 & +1 & -1 & -1 & 1+c_1-2c_2 & + & - & \\ +1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1-c_1 & - & & \\ 0 & 0 & +1 & -1 & 0 & 0 & 0 & 0 & 1-c_1 & - & & \\ 0 & 0 & 0 & 0 & +1 & -1 & 0 & 0 & 1-c_1 & - & & \\ 0 & 0 & 0 & 0 & 0 & 0 & +1 & -1 & 1-c_1 & - & & \end{bmatrix}, \tag{8}$$

where degenerate eigenvectors are simplified, at the cost of their $\mathcal{P}_3$ labels. For those eigenvectors which actually relate to the $G=1$ substructure, the simplification also erases the $\mathcal{P}_2$ labels. This form, Eq. (8), again illustrates the construction of eigenvectors and eigenvalues at level G when those at level $G-1$ are known. The key of the derivation, besides the labeling parities, is again the fact that a matrix with constant matrix elements is essentially null, except for its only non degenerate eigenvector, the ''fully symmetric'' set of constant components $(1,1,\ldots,1)$.

General case: $N=2^G$. (i) The eigenvalue $E_1=c_0-c_1$ occurs $N/2$ times. The corresponding eigenvectors are easy to list if one uses the simpler form introduced by Eqs. (6) and (8). Namely the first of them is $(+1,-1,0,0,\ldots,0,0)$ and the other ones are deduced from this one by all the permutations $\mathcal{P}_2^{\pi}$ which exchange ''$\nu=1$ minifamilies.'' (ii) Then the eigenvalue $E_2=c_0+c_1-2c_2$ occurs $N/4$ times, with eigenvectors the pattern $(+1,+1,-1,-1,0,0,\ldots,0,0)$ and all the vectors induced by the exchanges $\mathcal{P}_3^{\pi'}$ of ''$\nu\leqslant 2$ families.'' (iii) In turn, the degree of degeneracy of the eigenvalue $E_3=c_0+c_1+2c_2-4c_3$ is 2^{G-3}. The corresponding eigenvectors derive from the pattern $(+1,+1,+1,+1,-1,-1,-1,-1,0,0,\ldots,0,0)$ by all the exchanges $\mathcal{P}_4^{\pi''}$ of ''$\nu\leqslant 3$ families.'' And so on, until one reaches (iv) the nondegenerate eigenvalue $E_G=c_0+\Sigma_{\nu=1}^{G-1}2^{\nu-1}c_\nu-2^{G-1}c_G$, with eigenvector a ''one node'' sequence of $N/2$ positive components followed by $N/2$ negative components, $(+1,+1,\ldots+1,-1,-1\ldots,-1)$. (v) Finally, the (fully symmetric) vector with equal components is again a nondegenerate eigenvector, with eigenvalue $E_0=c_0+\Sigma_{\nu=1}^{G}2^{\nu-1}c_\nu$.

For the sake of definiteness and future reference, we list here the projectors which, for $G=3$, define the various subspaces,

$$\mathcal{Q}^{E_0}=\frac{1}{8}\begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \end{bmatrix}, \tag{9}$$

$$\mathcal{Q}^{E_1}=\frac{1}{2}\begin{bmatrix} +1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & +1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & +1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & +1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & +1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & +1 \end{bmatrix}, \tag{10}$$

$$\mathcal{Q}^{E_2}=\frac{1}{4}\begin{bmatrix} +1 & +1 & -1 & -1 & 0 & 0 & 0 & 0 \\ +1 & +1 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & +1 & +1 & 0 & 0 & 0 & 0 \\ -1 & -1 & +1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & +1 & +1 & -1 & -1 \\ 0 & 0 & 0 & 0 & +1 & +1 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & +1 & +1 \\ 0 & 0 & 0 & 0 & -1 & -1 & +1 & +1 \end{bmatrix}, \tag{11}$$

$$\mathcal{Q}^{E_3}=\frac{1}{8}\begin{bmatrix} +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 \\ -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 \end{bmatrix}. \tag{12}$$

In view of their trivial ''growth'' block structures, the rules governing the construction of such projectors are obvious. Let the matrix elements of such operators be denoted by $\mathcal{Q}_{E_n ij}$. The quantity

$$\Lambda_n=\sum_{i,j=1}^{N} X_i \mathcal{Q}_{E_n ij} X_j, \tag{13}$$

is an observable which obviously tells how much a given set of X_i's belongs to the subspace with eigenvalue E_n. The sum rule,

$$\sum_{n=1}^{G} \Lambda_n = 1, \tag{14}$$

obtains identically.

It is useful at this stage to reinstate the symmetrizations and antisymmetrizations prescribed by the various parities $\mathcal{P}_\nu^\pi$ of the problem. This alternate representation of the eigenvectors replaces their zeroes by either $+1$ or -1 and recovers the signatures which generalize those shown by Eqs. (5) and (7).

Two properties of this representation are then obvious, (i) except for the fully symmetric eigenvector, all the other eigenvectors, obtained by such rearrangements in their respective degenerate eigenspace, show an equal number of positive and negative components, and can thus be transformed into one another by suitable permutations of leaf labels, and (ii) given a signature read from left to right, corresponding to the ordered list $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_G$, the first ''minus'' sign of the signature specifies the eigenvalue. The existence of a representation of eigenvectors ''without zeroes,'' see again, e.g., Eqs. (5) and (7), is of importance for the next section, Sec. IV. This representation, incidentally, is nothing but the well known set of Hadamard matrices [7]. Such matrices can be defined by two constraints: (i) be made of orthogonal vectors, (ii) have matrix elements ± 1 only.

It must be recalled here that the present Section, Sec. III, was dedicated to the diagonalization of $\mathcal{C}$, for a linear decorrelation of the degrees of freedom X_i. It is only fitting that systematically one of the eigenvectors, the fully symmetric pattern $(1,1,\ldots,1)$, defines as a suitable collective observable, after an obvious normalization, the statistical mean $\Sigma_i X_i/N$. More interesting, maybe, is the fact that the other eigenvectors define ''heterogeneity observables'' of the form $\Sigma_{i\in I} X_i - \Sigma_{j\in J} X_j$. Here I and J are complementary subsets of $N/2$ leaf labels, and these subsets depend upon the eigenvector, naturally. Assume that closer parentage induces greater covariation, namely, that $c_0 \geqslant c_1 \geqslant c_2 \geqslant \cdots \geqslant c_G$. This induces an eigenvalue hierarchy $E_1 \leqslant E_2 \leqslant \cdots \leqslant E_G$. In so far as strong eigenvalues of the covariation matrix might be favored, the last (but one) eigenvector thus seems to be a favorite, all the more so because its E_G is not degenerate. The uniqueness of this eigenvector might make the contrast,

$$\mathcal{O}_G=\sum_{i=1}^{N/2} X_i - \sum_{j=N/2+1}^{N} X_j, \tag{15}$$

between the superfamilies $\nu \leqslant G-1$ a preferred statistical measure of heterogeneity.

Let $u_{i\sigma}$, $i=1,\ldots,N$ be the components of a ''no zero'' heterogeneity eigenvector, identified by its signature σ. As stated above, the knowledge of σ defines also the eigenvalue E. The observable

$$\mathcal{O}_\sigma=\sum_{i=1}^{N} u_{i\sigma} X_i \tag{16}$$

is nothing but the scalar product of the observed pattern of X_i's with the eigenvector. This scalar product has a vanishing probabilistic average $\langle \mathcal{O}_\sigma \rangle = \mu \Sigma_{i=1}^{N} u_{i\sigma}$, since $\mu = \langle X_i \rangle$ does not depend on i and since the -1 components of the eigenvector exactly compensate its $+1$ components. In actual measurements, deviations from this prediction $\langle \mathcal{O}_\sigma \rangle = 0$ must, to be statistically significant, be compared with the square root of the variance,

$$\langle \mathcal{O}_\sigma^2 \rangle = \sum_{i=1}^{N}\sum_{j=1}^{N} u_{i\sigma} \langle X_i X_j \rangle u_{j\sigma} = NE. \tag{17}$$

Clearly, the choice of the highest eigenvalue(s), namely the non degenerate E_G if the abovementioned hierarchy occurs,

creates the most demanding significance threshold. In a spirit slightly analogous to that of singular value decompositions, where largest eigenvalues are preferred, this choice might justify a preference for the observable $\mathcal{O}_G$.

To summarize this section, the diagonalization of the correlation matrix defines a Fourier analysis well suited to the situation created by ultrametric correlations. The modes upon which field observations are expanded are defined by the eigenvectors of the correlation matrix, naturally, and the corresponding decorrelated observables, see Eq. (16) are, when multiplied by an obvious $N^{-1/2}$ normalization, the associated Fourier coefficients. From the patterns of ± 1 components that have been discussed in this section, it is clear that finer structures (high frequencies in the oscillation of the ± 1 components) relate to lower eigenvalues. An observer, however, may resent the slight ambiguity in the information carried by such coefficients, because most eigenvalues are degenerate, and thus the eigenvectors are not uniquely defined and only the eigensubspaces are defined without ambiguity. There is then no difficulty in lumping together the squares of the Fourier coefficients pertaining to each degenerate eigenspace, see Eq. (13). More explicitly, according to Eqs. (13) and (16), given all those observables $\mathcal{O}_\sigma$, which belong to a given degenerate subspace with eigenvalue E_n, this lumping reads,

$$\frac{1}{N}\sum_{\sigma|E(\sigma)=E_n}[\mathcal{O}_\sigma]^2=\Lambda_n. \tag{18}$$

IV. CONSEQUENCES OF LABELING CONFUSION

The previous section, Sec. III, achieved more than the derivation of independent observables $\mathcal{O}_\sigma$, defined by Eq. (16). It also found a hierarchy between them. Namely, a sequence of X_i's can be analyzed in terms of Fourier coefficients $N^{-1/2}\mathcal{O}_\sigma$, relating to contrasts between minifamilies, or families, ... or superfamilies. The zoology of eigenstates makes this statement transparent, according to the ''frequencies'' at which the ± 1 components of the eigenvectors oscillate, or, equivalently, according to which (degenerate) subspace(s) such eigenstates belong. Hence one may detect in the X_i data fine structures pertaining to the influence of minifamilies, ..., superfamilies.

How reliable is this detection based on the consideration of such Fourier coefficients? The decorrelation has minimized the rôle of statistical fluctuations, indeed, but another problem may arise, namely labeling confusion. In practical situations of sample analysis, the existence of a tree may be known to be likely, but the labeling of the tree leaves (the elements of the sample under study) is either badly known or contains at least some errors. Accordingly, there are mistakes in the parentages between individuals. There exist non linear observables for measuring the sample heterogeneity, such as $\Sigma_{i<j}(X_i-X_j)^2$, which are as insensitive to a permutation of labels as the linear mean $\Sigma X_i/N$. But, in the realm of linear rearrangements, once we disregard this symmetric mean observable, perfect robustness is not available. Hence the question: among the ''heterogeneity'' linear combinations defined by the eigenvectors of $\mathcal{C}$, is there a preferable choice when a random permutation perturbs the labeling?

First question, first criterion: Let $|E\rangle$ be an eigenstate of $\mathcal{C}$, with eigenvalue E. The ''symmetric'' $|E_0\rangle$ being excluded, a random permutation P of components (differing from the identity) actually converts $|E\rangle$ into a different pattern $P|E\rangle$. Is the new pattern still an eigenstate, defining a legitimate observable? A first criterion of the robustness of $|E\rangle$ is thus the fluctuation (mean square deviation),

$$\begin{aligned} f(P|E\rangle)&=\langle E|P^{-1}[\mathcal{C}-\langle E|P^{-1}\mathcal{C}P|E\rangle]^2P|E\rangle \\ &=\langle E|P^{-1}\mathcal{C}^2P|E\rangle-(\langle E|P^{-1}\mathcal{C}P|E\rangle)^2. \end{aligned} \tag{19}$$

Indeed, a necessary and sufficient condition for eigenstates is the cancelation of this fluctuation.

Actually, since P is unknown, a best choice should result from minimizing the average of f over all permutations,

$$F(|E\rangle)=(N!)^{-1}\sum_P[\langle E|P^{-1}\mathcal{C}^2P|E\rangle-(\langle E|P^{-1}\mathcal{C}P|E\rangle)^2]. \tag{20}$$

Consider the ''no zero'' representation of eigenvectors, see again, e.g., Eqs. (5) and (7). As stated earlier for this Hadamard representation, given two distinct ''heterogeneity'' eigenvectors $|E\rangle$ and $|\mu\rangle$, there is always a permutation Q relating them, $|\mu\rangle=Q|E\rangle$. Then, since the summations upon permutations P and permutations $P'=QP$ make the same summation, one finds two ''indifference theorems,''

$$\begin{aligned} (N!)^{-1}\sum_P\langle\mu|P^{-1}\mathcal{C}P|\mu\rangle&=(N!)^{-1}\sum_P\langle E|(PQ)^{-1}\mathcal{C}PQ|E\rangle \\ &=(N!)^{-1}\sum_{P'}\langle E|P'^{-1}\mathcal{C}P'|E\rangle, \end{aligned} \tag{21}$$

$$\begin{aligned} F(|\mu\rangle)=F(Q|E\rangle)&=(N!)^{-1}\sum_P[\langle E|(PQ)^{-1}\mathcal{C}^2PQ|E\rangle \\ &\quad-(\langle E|(PQ)^{-1}\mathcal{C}PQ|E\rangle)^2] \\ &=(N!)^{-1}\sum_{P'}[\langle E|P'^{-1}\mathcal{C}^2P'|E\rangle \\ &\quad-(\langle E|P'^{-1}\mathcal{C}P'|E\rangle)^2]=F(|E\rangle). \end{aligned} \tag{22}$$

These results, Eqs. (21) and (22), may be summarized by the statement that the permutation average value of $\mathcal{C}$ is the same for a perturbed $|E\rangle$ and for any other perturbed eigenstate $|\mu\rangle$, and that the same equality is true for the fluctuation of $\mathcal{C}$. In other words, once perturbed by an arbitrary, unknown permutation of its components $u_{i\sigma}$, no heterogeneity observable is preferable for robustness, whether one considers the expectation value of $\mathcal{C}$ or the corresponding fluctuation. Our first criterion, namely the approximate conversion of an eigenstate into another one, therefore fails to suggest an observable more robust than the others. This failure is not too dramatic, however, because, anyhow, the replacement of an eigenvector by another one might induce the replacement of an eigensubspace by another one, confusing the interpretation of information in terms of lower and higher ''frequencies.'' The question of eigensubspace robustness will be studied later in the present section.

Paradoxically, a non diagonal form of Eq. (21) states that, even perturbed, the eigenvectors are still eigenstates, in the sense that the operator,

$$\mathcal{A}=(N!)^{-1}\sum_P P^{-1}\mathcal{C}P, \tag{23}$$

has vanishing off diagonal matrix elements between them. Consider indeed again two heterogeneity eigenvectors $|\sigma\rangle$ and $|\tau\rangle$, where σ and τ denote their signatures and not just their eigenvalues, the latter being ambiguous because of degeneracies. Let R be one of the ''parities,'' which make these signature differ, for instance $R|\sigma\rangle=|\sigma\rangle$, while $R|\tau\rangle=-|\tau\rangle$. Then, because it makes no difference whether one sums upon P or upon PR, one obtains,

$$N!\langle\tau|\mathcal{A}|\sigma\rangle=\sum_P\langle\tau|P^{-1}\mathcal{C}P|\sigma\rangle=\sum_P\langle\tau|(PR)^{-1}\mathcal{C}PR|\sigma\rangle$$

$$=\sum_P\langle\tau|R^{-1}P^{-1}\mathcal{C}P|\sigma\rangle=-\sum_P\langle\tau|P^{-1}\mathcal{C}P|\sigma\rangle. \tag{24}$$

Hence, $\langle\tau|\mathcal{A}|\sigma\rangle=0$. A similar statement holds for $\mathcal{C}^2$, naturally. Hence, ''off diagonal fluctuations'' also vanish.

Second question, second criterion: Slightly modifying Eq. (20) we now take advantage of the eigenvalue of the candidate and ask: does the permutation, which perturbs the eigenvector, respect the eigenvalue? In other words, does the observable, while disturbed, belong to the same eigensubspace and essentially yields a similar information? One possible answer lies in a minimization of the averaged square norm, $(N!)^{-1}\Sigma_P|(\mathcal{C}-E)P|E\rangle|^2$. This amounts to minimize,

$$\mathcal{G}=\sum_P[\langle E|P^{-1}\mathcal{C}^2P|E\rangle-2E\langle E|P^{-1}\mathcal{C}P|E\rangle+E^2]. \tag{25}$$

Here eigenvectors are square renormalized to unity, namely the ''no zero'' representation becomes multiplied by $N^{-1/2}$.

Because we found that both the $\mathcal{C}^2$ and the $\mathcal{C}$ terms are indifferent to the choice of the *eigenvector* $|E\rangle$, the only way to minimize $\mathcal{G}$ is to choose that *eigenvalue* E which is nearest to the number $a=\langle E|\mathcal{A}|E\rangle$. In the eigensubspace specified by the optimal E under search, any (normalized) vector will then make an optimal observable.

To calculate a we take advantage of the huge degeneracy, of degree $N-1=2^G-1$, of $\mathcal{A}$ and evaluate a with respect to the simplest possible relevant vector of the corresponding subspace. This vector is, in the representation with zeroes, the normalized vector $2^{-1/2}(1,-1,0,0,\ldots,0)$, with two nonvanishing components only. Then a reads

$$a=\frac{1}{2N!}\sum_P[(P^{-1}\mathcal{C}P)_{11}-(P^{-1}\mathcal{C}P)_{12}-(P^{-1}\mathcal{C}P)_{21}+(P^{-1}\mathcal{C}P)_{22}]. \tag{26}$$

Here the numbers $(P^{-1}\mathcal{C}P)_{ij}$ are the matrix elements of $\mathcal{C}$ under the perturbations (permutations) P. Diagonal matrix elements remain unchanged, and equal to c_0, under these permutations. Under the same permutations, off diagonal matrix elements sample equally all the off diagonal elements of the symmetric, ultrametric, initial $\mathcal{C}$. The result, an average eigenvalue, reads

$$a=c_0-\frac{1}{2^G-1}\sum_{\nu=1}^{G}2^{\nu-1}c_\nu. \tag{27}$$

To summarize this section, we found that the most robust observables under labeling confusion are those whose eigenvalues are as close as possible to the number a defined by Eq. (27). The next section, Sec. V, shows a case where it is easy to locate a with respect to the eigenvalues.

V. ROBUSTNESS OF HETEROGENEITY MEASUREMENTS IN A GENETIC MODEL

Imagine a viral epidemy in a large geographical area, divided into two ''subcontinents'' A and B. Assume each subcontinent to be divided into two ''regions,'' the regions splitting in turn into ''plains'' versus ''mountains,'' then the mountains splitting into two ''valleys'' while the plains show two main ''rivers basins.'' And so on until ''villages'' and ''clans'' within villages, ''families'' within clans, etc. One suspects, conjecture (i), that the virus actually splits into two distinct classes of strains and, furthermore, that a strain might be more specific to A while the other might pertain more to B. One also suspects, conjecture (ii), that some social structures and mechanisms within ''villages'' may create additional contrasts and separate strains on a shorter scale. However, the concerned populations travel enough to make somewhat dubious any assessment such as ''this individual belongs to this subcontinent, this region, etc., this family.'' Hence, for the discovery of strains with a sufficient contrast, is the observable $\Sigma_{i\in A}X_i-\Sigma_{i\in B}X_i$ significant, or should one use a different observable $\Sigma_{i\in I}X_i-\Sigma_{i\in J}X_i$ where I and J are complementary subsets across A and B? Moreover, in the search for evidences of fine structure mechanisms under conjecture (ii), when a few ''Fourier coefficients'' turn out to be dominant, do they relate to a reasonably robust eigenspace?

For this question, we consider a genetic model used earlier [5] where the matrix elements of $\mathcal{C}$ are most simple, $c_\nu=p^\nu$. The parameter p is a positive number, slightly smaller than 1. This model was used for strains of DNA, or RNA, or proteins, with individuals duplicating after a constant lifetime T for each generation. During this lifetime, a property $X=\pm1$ of each individual drifts stochastically, in such a way that two individuals a and b starting identical, $X_a=X_b$, at one of the nodes of the tree, see Fig. 1, finish with a correlation $\langle X_aX_b\rangle=p$ just before duplicating. The normalization of such variables X reads, obviously, $c_0=\langle X^2\rangle=p^0=1$.

The list of eigenvalues then reads

$$E_1=1-p,E_2=1+p-2p^2,\ldots,$$

$$E_m=1+\frac{p[(2p)^{m-1}-1]}{2p-1}-p(2p)^{m-1},\ldots,$$

$$E_G=1+\frac{p[(2p)^{G-1}-1]}{2p-1}-p(2p)^{G-1}, \tag{28}$$

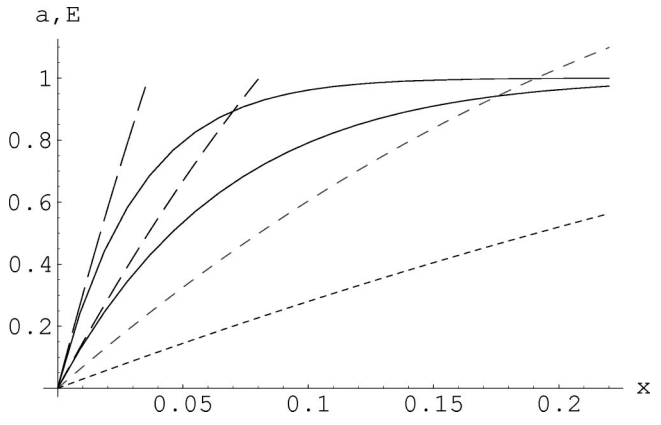

FIG. 2. Genetic model. Plots of the average eigenvalue a for $G=32$ (upper full line) and $G=16$ (lower full line), plots of eigenvalues E_5 (long dashed line), E_4 (dashed line), E_3 (short dashed line) and E_2 (dotted line), as functions of the mutation rate, $x=1-p$.

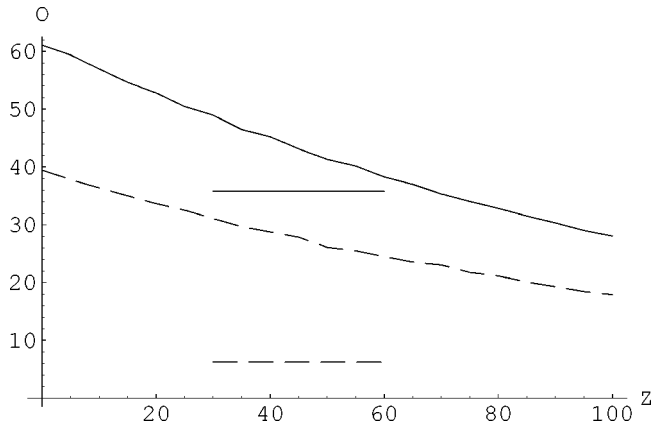

FIG. 3. For $G=8$, average values of $\mathcal{O}_8$ (full line curve) and $\mathcal{O}_3$ (dashed curve), as functions of the number Z of final transpositions of individuals in the population. Horizontal full line: confidence threshold (variance, see Eq. (17)) for $\mathcal{O}_8$. Dashed horizontal line: the same for $\mathcal{O}_3$.

while a reads

$$a=1-\frac{p[(2p)^G-1]}{(2^G-1)(2p-1)}. \tag{29}$$

Expansions of E_m and a in the vicinity of $p=1$ give, respectively,

$$E_m=(2^m-1)(1-p)+\mathcal{O}(1-p)^2, \quad \text{and}$$

$$a=\left[\frac{G}{1-2^{-G}}-1\right](1-p)+\mathcal{O}[(1-p)^2]. \tag{30}$$

As soon as 2^{-G} can be neglected with respect to 1, the best choice, according to the second criterion of Sec. V, corresponds to that integer m closest to $\log_2 G$. This is valid, however, for values of p close to 1 only. As shown by Fig. 2, which plots a for $G=32$ and $G=16$, with E_5, E_4, E_3, and E_2 as functions of $x=1-p$, the best choice depends on p. Indeed, when p is close to 1, E_5 is a good approximation of a if $G=32$. The same is true for E_4 if $G=16$. For $p\simeq 0.93$, and for $p\simeq 0.81$, however, better approximates of a are E_4 and E_3, respectively, if $G=32$. For $G=16$, in turn, E_3 coincides with a if $p\simeq 0.82$. For $p\simeq 0.5$, the best approximate of a seems to be E_2, independently from G, but such extreme values of p are not of a great practical significance. For all practical purposes, p is quite close to 1.

The following figure, Fig. 3, lists the results obtained with 10^4 independent runs of the model when $G=8$, namely, 256 individuals, and $p=1-2.25\times 10^{-2}$. An artificial mutation is enforced during the first ''generation,'' namely, if the notations of Fig. 1 are used, $X_l=1$ and $X_m=-1$. With such an initial condition, which differentiates A from B, conjecture (i) should be observable by means of $\mathcal{O}_8$ up to some extent at least, even after 7 levels of duplication and mutations, and after a reasonable amount Z of random transpositions of individuals at the end.

Except for this initial condition, flips ± 1 are random, with probability $\varepsilon=\frac{1}{2}(1-p)^{1/2}=0.075$. Generation $\#G-2=6$, however, is artificially modified, in order to superimpose data compatible with conjecture (ii). For this, with probability 25%, the pairs X_{2i-1},X_{2i} normally generated in this 6th generation are replaced by pairs $+1,-1$. Such pairs, two generations later, are meant to induce octuplets of the form $+1,+1,+1,+1,-1,-1,-1,-1$, easily detectable by an observable $\mathcal{O}_3$, unless they have been destroyed by mutations and/or population reshuffling.

Except for their identical first generations, the 10^4 runs differ by their random mutations along the tree, by the random enforcement of $+1,-1$ pairs at generation #6, and by the Z random transpositions of leaves of the tree. The code generating such runs is available upon request.

The full line curve shown in Fig. 3 plots, as a function of Z, the value, averaged over the 10^4 runs, taken by the observable $\mathcal{O}_8$. In a transparent notation, this observable may be listed as $[128+,128-]$, naturally. With $p=1-2.25\times 10^{-2}$, the corresponding eigenvalue and square root of the variance (SRV) are 5.0 and 36., respectively. Then the dashed line curve in Fig. 3 shows values taken by the simplest among observables $\mathcal{O}_3$, namely, in the same transparent notation again, $[4+,4-,4+,4-,\ldots,4+,4-]$ (32 sequences $++++----$). In order to appraise significance, the average value of $\mathcal{O}_3$ must be scaled against an SRV 6.2, coming from eigenvalue 0.15.

The robustness of $\mathcal{O}_3$ compared to that of $\mathcal{O}_8$ is transparent from Fig. 3. It will be noticed that the noise brought by the random transpositions does not prevent the kind of Fourier analysis, brought by observables $\mathcal{O}_\sigma$, from being efficient a long time, since a simultaneous detection of both modes #8 and #3 is still possible when $Z\simeq 50$, compared to a population of 256 here. But beyond $Z\simeq 50$, average values of $\mathcal{O}_8$, compared to $N^{1/2}E_8$, become too small, while $\mathcal{O}_3$ remains reliable a long way still.

Rather than using a specific eigenvector for testing robustness, one could also consider, for a similar detection of contrast (heterogeneity) modes, the projectors $\mathcal{Q}^{E_n}$ on eigensubspaces for each (degenerate) eigenvalue E_n, naturally. It is likely that such projectors would show the same robustness, because of their invariance groups. The present section, Sec. V, however, already gives sufficient evidence for the hierarchy of robustness created by the second criterion discussed at

the end of the previous section, Sec. IV. We retain the question of projector robustness for a further study.

VI. DISCUSSION AND CONCLUSION

There are two kinds of results in this paper. The first kind deals with disentangling correlated degrees of freedom, in order to define a set of ''collective'' observables, better amenable to statistical observations. The idea of correlation matrix diagonalization is not original, nor is new, at least for the representation with zeroes [8], the zoology of the eigenvectors and eigenvalues of ultrametric matrices. But the results, in terms of two groups of observables, the symmetric statistical mean on the one hand, and a large set of heterogeneity observables on the other hand, are more interesting. Indeed there emerges a natural hierarchy of such heterogeneity observables: depending on the eigensubspace retained for the observable, one measures heterogeneity within ''minifamilies,'' or ''families,'' etc. each scale being characterized by a maximum degree of parentage, ν. All told, this paper tells that only a few selected lists of $+$ and $-$ signs define proper measures $\Sigma_{i\in I}X_i-\Sigma_{j\in J}X_j$ of heterogeneity. This question is of some importance for biology, in particular, where speciation may occur if a population becomes heterogeneous enough to form contrasting clusters. While a strict ultrametricity of correlations is an obvious oversimplification [9], the hierarchy we found is likely to tolerate some deviations from this modelization hypothesis.

The second kind of results deals with consequences of mistakes in individual identities inside the population under statistical observation. Section III defined ''eigenmeasurements'' invariant under subgroups of permutations, those with a $+$ sign in the signatures. The robustness of such an observable thus depends on the size of the subgroup and the answer $\nu\simeq\log_2 G$ essentially identifies a compromise between the subgroup size and the fluctuation of the observable. If fluctuation is retained as the only criterion, the heterogeneity observable with maximum fluctuation seems preferable, as it creates the most demanding threshold of significance when statistical deviations are found.

There could be situations where the number of mistakes in the labeling, while large, makes a small proportion of all possible permutations. A modified procedure of averaging over permutations, with larger weights for permutations close to the identity, is then necessary. Alternately, one might consider what happens if only a few eigensubspaces, neighbors according to their labels ν, are mixed by the perturbations. This may mean replacing the minimization of $\mathcal{G}$, Eq. (25), by that of $(N!)^{-1}\Sigma_P|(\mathcal{C}-E)P|E\rangle|^k$, with $k>2$. We are investigating such questions.

All these considerations hold when only linear rearrangements of variables are considered. Probabilistic averages of such heterogeneity observables vanish for the models considered in this paper, hence, significance depends upon experimental deviations larger than the expected scales of fluctuations. Such scales, in turn, depend on an underlying model, namely the values of the correlations. It is unknown whether non linear observables define measurements which are less model dependent. But it is likely that such non linear observables will show spurious correlations between one another, and the question of disentangling them will rise again. This problem is also under consideration.

In practice, the methodology advocated by this paper boils down to the following recipe: (i) When there is a reasonable motivation for coding experimental observations in terms of stochastic variables, $X_i=\pm 1$, making the leaves of an ultrametric binary tree with G generations, decide on a set of such labels i according to reasonable assumptions about parentage relationships. (ii) Then take advantage of the fact that the eigenstates and eigenprojectors of the correlation matrix *do not depend* on its matrix elements c_ν. Obtain eigenstates and projectors from Sec. III. Hence, calculate the Fourier coefficients $N^{-1/2}\mathcal{O}_\sigma$, see Eq. (16). Or at least calculate the projector expectation values Λ_m, see Eq. (13). Because of the sum rules, Eqs. (14) and (18), the global heterogeneity of the sampled population is estimated by the number $1-\Lambda_0$. It may happen that the ''heterogeneity'' Fourier coefficients or Λ's *cluster* in two groups, namely ''small ones'' and ''larger ones.'' (iii) If so, the ''larger ones'' indicate specific heterogeneities inside substructures of the sampled population. Such heterogeneities can be trusted if the labeling is believed to be accurate. If, however, given G generations in the ultrametric graph, the labeling of its leaves suffers from serious uncertainties, one may consider integers m close to $\log_2 G$, and a qualitative amount of confidence can still be retained for those observables detecting heterogeneity between subpopulations at parentages of order m. (iv) A more detailed analysis, along the arguments detailed in Secs. IV and V, is then in order, at the cost of a refined model implying values for the correlations c_ν's and a comparison of the spectrum of the correlation matrix with the average eigenvalue a, see Eq. (27). For want of bounds on the number of label mistakes, only those observables derived from eigenvalues close to a are likely to be reliable.

ACKNOWLEDGMENT

The author is indebted to Alan Lapedes for a discussion of criterions for ''best'' observables.

APPENDIX: GENERALIZATION

The property that eigenvalues ''at level $G-1$'' remain valid ''at level G'' and that eigenvectors, completed by zeroes, extend from level to level, is not restricted to binary branching. Consider for instance the ternary ultrametric matrix,

$$\mathcal{T}_2=\begin{bmatrix} c_0 & c_1 & c_1 & c_2 & c_2 & c_2 & c_2 & c_2 & c_2 \\ c_1 & c_0 & c_1 & c_2 & c_2 & c_2 & c_2 & c_2 & c_2 \\ c_1 & c_1 & c_0 & c_1 & c_2 & c_2 & c_2 & c_2 & c_2 \\ c_2 & c_2 & c_1 & c_0 & c_1 & c_1 & c_2 & c_2 & c_2 \\ c_2 & c_2 & c_2 & c_1 & c_0 & c_1 & c_2 & c_2 & c_2 \\ c_2 & c_2 & c_2 & c_1 & c_1 & c_0 & c_2 & c_2 & c_2 \\ c_2 & c_2 & c_2 & c_2 & c_2 & c_2 & c_0 & c_1 & c_1 \\ c_2 & c_2 & c_2 & c_2 & c_2 & c_2 & c_1 & c_0 & c_1 \\ c_2 & c_2 & c_2 & c_2 & c_2 & c_2 & c_1 & c_1 & c_0 \end{bmatrix}, \tag{A1}$$

and its diagonal and off-diagonal substructures, the submatrices

$$\mathcal{T}_1=\begin{bmatrix} c_0 & c_1 & c_1 \\ c_1 & c_0 & c_1 \\ c_1 & c_1 & c_0 \end{bmatrix} \quad \text{and} \quad \mathcal{N}_1=\begin{bmatrix} c_2 & c_2 & c_2 \\ c_2 & c_2 & c_2 \\ c_2 & c_2 & c_2 \end{bmatrix}. \tag{A2}$$

It is trivial to observe that both $\mathcal{T}_1$ and $\mathcal{N}_1$ have the vector (1,1,1) as an eigenvector with eigenvalue c_0+2c_1, and, moreover, that $\mathcal{N}_1$ is null in the orthogonal subspace spanned by the other eigenvectors of $\mathcal{T}_1$. Let (a,b,c) be any such ''other'' eigenvector of $\mathcal{T}_1$. Returning to $\mathcal{T}_2$, one sees at once, by a trivial argument using block matrix algebra, that the three extensions $(a,b,c,0,0,0,0,0,0)$, $(0,0,0,a,b,c,0,0,0)$ and $(0,0,0,0,0,0,a,b,c)$ are eigenvectors of $\mathcal{T}_2$, retaining as eigenvalue the corresponding eigenvalue c_0-c_1 of $\mathcal{T}_1$.

This eigenvalue c_0-c_1 is twofold degenerate for $\mathcal{T}_1$, hence sixfold degenerate for $\mathcal{T}_2$. For $\mathcal{T}_1$ a possible basis of the corresponding subspace consists of the vectors $(+1,-1,0)$ and $(+1,0,-1)$, making a two-dimensional representation of the mixed representation for the permutations of 3 elements. A table of row eigenvectors and eigenvalues of $\mathcal{T}_2$ reads,

$$\mathcal{F}_2=\begin{bmatrix} +1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & c_0-c_1 \\ +1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & c_0-c_1 \\ 0 & 0 & 0 & +1 & -1 & 0 & 0 & 0 & 0 & c_0-c_1 \\ 0 & 0 & 0 & +1 & 0 & -1 & 0 & 0 & 0 & c_0-c_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & +1 & -1 & 0 & c_0-c_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & -1 & c_0-c_1 \\ +1 & +1 & +1 & -1 & -1 & -1 & 0 & 0 & 0 & c_0+2c_1-3c_2 \\ +1 & +1 & +1 & 0 & 0 & 0 & -1 & -1 & -1 & c_0+2c_1-3c_2 \\ +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & c_0+2c_1+6c_2 \end{bmatrix}. \tag{A3}$$

The sum of the vectors in rows 1, 3 and 5 of this matrix generates the vector $(+1,-1,0,+1,-1,0,+1,-1,0)$ which differs from the vector $(+1,+1,+1,-1,-1,-1,0,0,0)$ by just a permutation of components. The point is, the former vector belongs to the subspace with eigenvalue c_0-c_1, while the latter belongs to that with eigenvalue $c_0+2c_1-3c_2$. The ''indifference'' results of Sec. IV thus extends to this ternary case. It is easy to generalize such arguments to other degrees of ultrametricity.

[1] R. Rammal, G. Toulouse, and M. A. Virasoro, Rev. Mod. Phys. **58**, 765 (1985).

[2] G. Parisi, J. Phys. A **13**, L115 (1980).

[3] B. Derrida and L. Peliti, Bull. Math. Biol. **53**, 355 (1991).

[4] A. S. Lapedes, B. G. Giraud, L. C. Liu, and G. D. Stormo (unpublished).

[5] B. G. Giraud, A. S. Lapedes, and L. C. Liu, Phys. Rev. E **58**, 6312 (1998).

[6] B. Derrida, Non-self-averaging effects in sums of random variables, spin glasses, random maps and walks, *On three levels: micro, meso, and macroscopic approaches in physics*, edited by M. Fannes *et al.*, Leuwen, Belgium, July 19–23 (1993), pp. 125–137.

[7] M. L. Mehta, *Elements of Matrix Theory*, Sec. 7.5, Hindustan Publishing Corp.(I), Dehli-110007, 1977, pp. 97–99.

[8] A. T. Ogielski and D. L. Stein, Phys. Rev. Lett. **55**, 1634 (1985).

[9] R. Rammal, J. C. Angles d'Auriac, B. Doucot, J. Phys. (France) Lett. **46**, L-945 (1985).