

Measuring Dependency With Volume Tests

Chiara SABATTI

Volume tests in a regression context were introduced in 1939. The concept was revisited in 1985, this time in the case of contingency tables to refine the power of explanation of the χ^2 test. This article considers volume tests as a measure of dependency for tables with fixed margins. Building on earlier contributions, this article suggests the use of the volume test statistic as a measure of dependence to be applied, for example, in the evaluation of linkage disequilibrium between markers.

KEY WORDS: Chi-square tests; Contingency tables; Linkage disequilibrium; Measure of association.

1. INTRODUCTION

Consider the qualitative random variables X, Y with modalities A_1, \dots, A_r and B_1, \dots, B_c , characterized by the frequencies p_1, \dots, p_r , and q_1, \dots, q_c . We are interested in the problem of measuring the dependence that characterizes their bivariate distribution π_{ij} $i = 1, \dots, r$, and $j = 1, \dots, c$ using counts n_{ij} observed in a sample of size n , as follows:

$$\{n_{ij}\} = \begin{array}{cccccc} & B_1 & B_2 & \cdots & B_c & \\ A_1 & n_{11} & n_{12} & \cdots & n_{1c} & n_{1.} \\ A_2 & n_{21} & n_{22} & \cdots & n_{2c} & n_{2.} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ A_r & n_{r1} & n_{r2} & \cdots & n_{rc} & n_{r.} \\ & n_{.1} & n_{.2} & \cdots & n_{.c} & n \end{array} \quad (1)$$

This problem appears in many contexts. One that is particularly relevant today is the study of linkage disequilibrium in statistical genetics. In this setting, the realizations of the random variables X and Y represent the alleles present at a given pair of loci A and B on one chromosome. By genotyping a set of n chromosomes, we gather counts as in (1). Because of the mechanism that regulates the transmission of chromosomes from parents to progeny in humans, in some circumstances there is a direct correspondence between the amount of dependence between X and Y and the distance between A and B on the chromosome, which, generally speaking, is difficult to measure directly. Given a variety of loci, scientists are then interested in measuring the relative association between all the pairs of corresponding random variables, in order to infer the relative distance and position of the loci. The sample sizes available for these comparisons are often different and the number of categories in each of them quite variable. Additionally, the genetic context makes it of interest to study dependency conditionally on the marginal allele

frequencies, as they carry no information on relative distances. We are interested in discussing a measure of dependence that can be profitably used in these cases.

There are two related problems that are well known to statisticians: (I) measuring the association between two qualitative random variables when their distribution $\{\pi_{ij}\}$ is known and (II) testing for independence the contingency table $\{n_{ij}\}$. Both procedures are based on measuring the distance of the $\{\pi_{ij}\}$ (or $\{n_{ij}/n\}$) from the distribution under independence $\{p_i q_j\}$ (or $\{n_{i.} n_{.j}/n^2\}$) and comparing this value with an appropriate reference. A very popular choice is the Mahalanobis distance:

$$M(\{\pi_{ij}\}, \{p_i q_j\}) = \sum_{ij} \frac{(\pi_{ij} - p_i q_j)^2}{p_i q_j},$$

which, when evaluated on the counts $\{n_{ij}\}$ is equivalent to the χ^2 statistics divided by n .

When interested in measuring dependence, $M(\{\pi_{ij}\}, \{p_i q_j\})$ is standardized by the maximum possible value it could take on the space of all tables of the given structure, so that the resulting measure takes on value 0 in case of independence and 1 in case of perfect dependence. As we are particularly interested in the case of fixed marginal distributions, it is worth to recall that, with the exception of the 2×2 case, where the solution has been popularized under the name D' in statistical genetics (see Devlin and Risch 1995), there is no general closed form solution for the problem

$$\max_{\pi_{ij}: \pi_{i.} = p_i \ \& \ \pi_{.j} = q_j} M(\{\pi_{ij}\}, \{p_i q_j\}).$$

In a test of hypothesis, the measured distance between $\{n_{ij}/n\}$ and $\{n_{i.} n_{.j}/n^2\}$ is evaluated, instead, with reference to the chance of recording such a big difference if the observations were generated randomly under independence. In this context, sample size plays a crucial role in determining if the distance from independence can be considered significant.

The problem we are interested in shares aspects of both these approaches: as in (I) we are seeking a measure of dependence that allows comparisons across tables, but as in (II) we do not observe $\{\pi_{ij}\}$, but only $\{n_{ij}\}$, hence we are subject to the effects of random fluctuations that are particularly serious when n is small. Neither (a) simply translating the measures of dependence defined for π_{ij} on $\{n_{ij}/n\}$ nor (b) use the p value of a test of hypothesis as a measure of association offer satisfactory solutions to our problem.

To clarify the fallacies of (a), consider the following example. Suppose that X and Y are bivariate. Then $M(\{\pi_{ij}\}, \{p_i q_j\})$ can be interpreted as the square of the correlation coefficient R^2 , has maximum value 1, and attains it only when the marginal distributions of the random variables are the same. All of this

Chiara Sabatti is Assistant Professor of Statistics and Human Genetics at UCLA, Los Angeles, CA 90095-1554 (E-mail: csabatti@mednet.ucla.edu). This work was completed while she was a post-doctoral fellow in the laboratory of Neil Risch in the Genetics department of the Stanford School of Medicine, Stanford, CA 94305.

can be easily seen with the following parameterization $x = \pi_{11}$,

$$\pi = \begin{array}{c|cc|c} & 1 & 0 & \\ \hline 1 & x & p-x & p \\ \hline 0 & q-x & 1-p-q+x & 1-p \\ \hline & q & 1-q & 1 \end{array}$$

$$M(\{\pi_{ij}\}, \{p_i q_j\}) = \frac{(x-pq)^2}{pq(1-p)(1-q)} = R^2. \quad (2)$$

Consider now a case in which we record only four observations and $n_{1.} = n_{2.} = n_{.1} = n_{.2} = 2$. There are only three possible contingency tables leading to such outcome:

1	1
1	1

2	0
0	2

0	2
2	2

The last two tables, conditionally on the marginal counts, have probability 1/6 under independence and lead to a R^2 value of 1: one third of the time, then, an independent model would lead to a measured value corresponding to perfect dependence. This is clearly a spurious effect due to small sample size.

Let us now take, instead, the test of hypothesis viewpoint (b) and show how it also induces shortcomings. Consider the following two tables, each recorded with 100 observations:

.10	.10	.2
.3	0	.3
0	.5	.5
.4	.6	

.02	.08	.1
.38	.02	.4
0	.5	.5
.4	.6	

If we calculate the values of the χ^2 test for these tables and normalize them to obtain a “measure” of dependence, we get $\chi^2/n = .791$ for Table 1 and, $\chi^2/n = 0.854$, for Table 2. This leads to the impression that the second distribution is characterized by higher dependence than the first one. However, this conclusion is questionable. Indeed, let us look at the space of all tables $\{\pi_{ij}\}$ with marginals equal to the reported ones. We can parameterize this space in terms of $x = \pi_{11}$ and $y = \pi_{21}$. Figure 1 illustrates with shaded areas these two spaces. We can then

identify the positions of the observed tables in the spaces (filled circles) and their Mahalanobis distance from independence (the ellipses represent points of equal distance from the independence distribution identified with an empty circle). One can notice, then, that Table 1 has the highest possible distance from independence among the Tables with the specified marginals, while Table 2 does not. The different relative ordering that we obtain considering simply the χ^2 values is due to the fact that the metric changes in the two cases corresponding to the two independence distributions of reference. This example illustrates how p values obtained from a test of significance, which are directly related to χ^2 values, cannot be taken as satisfactory measures of dependence, even if the number of observations collected for each compared table is the same. This is especially true when we are interested in considering fixed marginal counts.

To address the difficulties outlined above, often one adopts the following approach: first conduct a test of independence, to screen for artifacts due to random fluctuations, and then use a plug-in measure of dependence on the tables for which the hypothesis of independence had to be rejected. Alternatively, we advocate the rediscovery of volume tests.

2. VOLUME TESTS

Hotelling (1939) was the first to use geometrical properties to test statistical hypothesis in an article on nonlinear regression. The applications of such “volume tests” in the context of contingency tables have been reviewed and analyzed providing a novel interpretation by Diaconis and Efron (1985). The terminology volume tests is due to them and we follow here their exposition. The authors suggested the use of volume tests to obtain p values that are a better indication of the degree of dependence than the ones obtained through a χ^2 test. One of the problems of χ^2 is, they recalled, that for large n it tends to almost always reject the null hypothesis. In general, they remark, that once the independence hypothesis has been rejected, one can obtain remarkably little information from the value of the χ^2 statistics. They illustrated this with the consideration of two datasets: a contingency table of hair and eye color and one of yearly in-

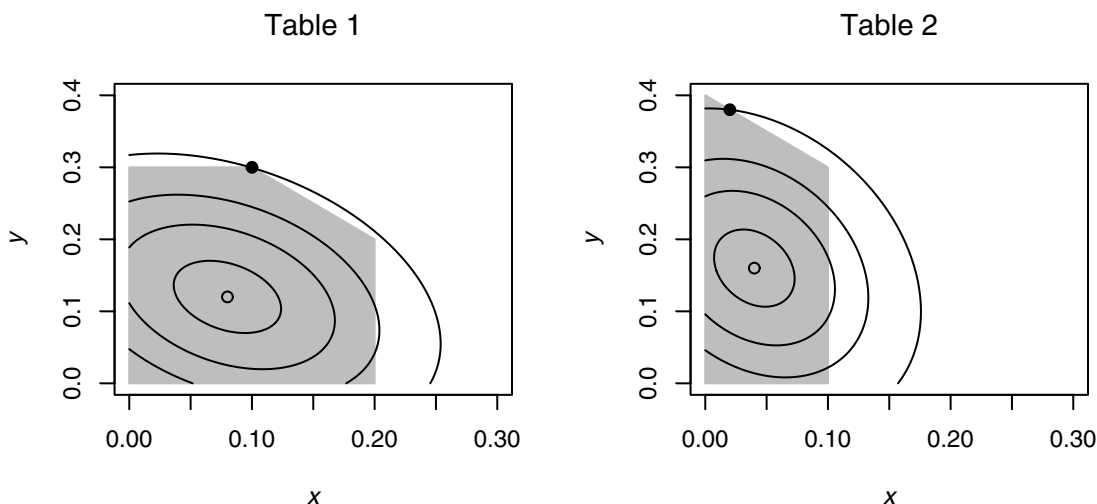


Figure 1. Measures of dependency vs tests of independence. The spaces of all possible tables with same marginals as Table 1 and Table 2 are illustrated by the shaded areas. The points corresponding to independence are marked by empty circles and the one corresponding to the actual tables by a filled circle. The ellipses are contour plot of the Mahalanobis distance from independence.

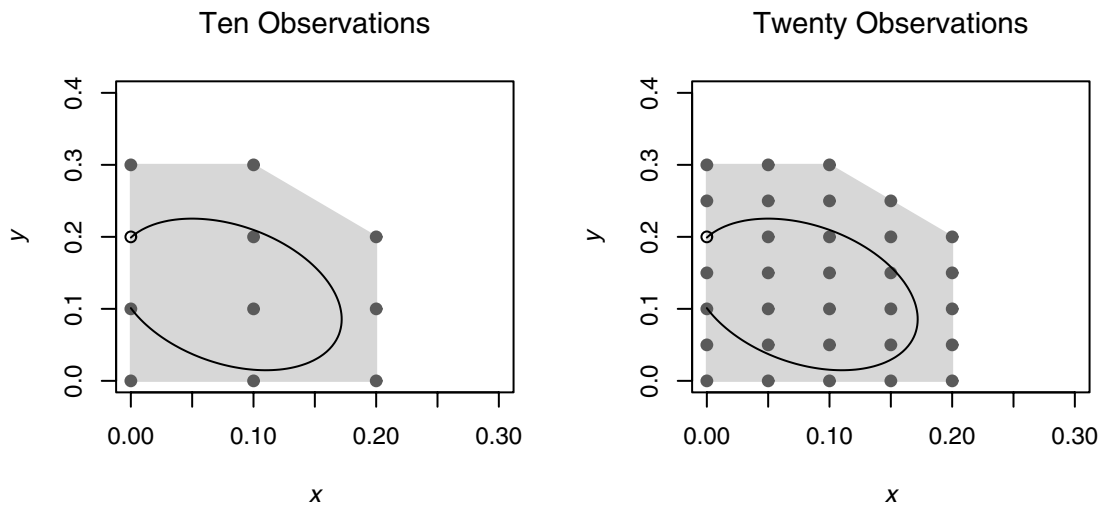


Figure 2. Evaluation of ChiVol. The space of all possible tables with the same marginals as Table 1 (see Figure 1) is here represented emphasizing the effect of sample size. The shaded area represents the space of $\{\pi_{ij}\}$; the filled circles, the tables $\{n_{ij}\}$ that can be actually observed for a given sample size. On the left, it is illustrated the case of 10 observations and on the right the case of 20 observations. The empty circles identify the tables actually observed and ellipses are loci of the tables with the same Mahalanobis distance from independence as the observed ones.

come and number of children. The first leads to a $\chi^2_9 = 138.29$ and the second one to $\chi^2_{12} = 568.59$. However, they showed how the second table is much closer to independence than the first one in a sense that can be made explicit with geometrical considerations. Intuitively, the idea is to look among all the possible contingency tables with the same number of rows and columns of the observed and evaluate which proportion of them leads to a χ^2 statistics with values smaller or equal than the recorded one. If this proportion is small, the table is close to independence, while, if this proportion is high, the table is “far away” from independence. In these tests, then, the hypothesis of independence, in which the marginal probabilities totally determine the interior probabilities of the contingency table, is contrasted with the somewhat antagonistic hypothesis that the interior probabilities are selected from a uniform distribution, that is, are highly undetermined.

To give a formal justification of this procedure, we restrict ourselves to the case of fixed margins, which is the one we are interested in considering in general. For clarity, we have to introduce the notion of a Bayesian hypothesis: instead of selecting a particular class of distributions $\{\pi_{ij}\}$, we specify a probability on the space of all possible distributions $\{\pi_{ij}\}$. In particular, let H_0 assume that $\{\pi_{ij}\}$ is chosen according to a uniform distribution among all tables with a given number of rows and columns. Then, if $\{n_{ij}\}$ is obtained with multinomial sampling from π , the distribution of $\{n_{ij}\}|n_{i.}, n_{.j}$ is uniform on the space of all contingency tables with the same margins. Let the alternative H_1 be the hypothesis of independence; when conditioning on the observed marginal counts, we have for $\{n_{ij}\}$ the Fisher–Yates distribution. Then, both H_0 and H_1 are simple hypothesis. If we use χ^2 as a test statistics, the p value of the test is

$$P(\chi^2 < \text{observed} | H_0) = \frac{\#\{\text{tables} : \chi^2(\{n_{ij}\}) \leq \chi^2 | n_{i.}, n_{.j}\}}{\#\text{total tables} | n_{i.}, n_{.j}}. \quad (3)$$

We are interested in considering a variation of (3), which we will call ChiVol:

$$\text{ChiVol}(\{n_{ij}\}) = \frac{\#\{\text{tables} : \chi^2(\{n_{ij}\}) < \chi^2 | n_{i.}, n_{.j}\}}{\#\text{total tables} | n_{i.}, n_{.j}}. \quad (4)$$

The difference from (3) is relevant only for small n . In the following we propose the use of ChiVol as a measure of dependency, in agreement with the spirit of Diaconis and Efron (1985). Before discussing its properties, however, let us consider in more detail how it is calculated. Suppose the observed table is

$$\begin{array}{|c|c|c|} \hline .0 & .2 & .2 \\ \hline .2 & .1 & .3 \\ \hline .2 & .3 & .5 \\ \hline .4 & .6 & \\ \hline \end{array}, \quad (5)$$

with $n = 10$. The observed Mahalanobis distance of Table (5) from independence is .222. Since we have a total of 10 observations, the frequency in each cell can be only a multiple of 0.1. It is easily seen, then, that there are a total of 11 tables that with 10 observations have the same marginal counts: they are illustrated as filled circles in Figure 2(a). The table in (5) is identified with an empty circle. By looking at the locus of points with the same χ^2 value as (5) (drawn ellipse), we notice that there are two tables with smaller χ^2 values. The ChiVol statistic would then be $2/11$. If the same frequencies as in (5) were observed on the base of 20 observations, the set of possible tables leading to the same marginal counts is larger and depicted in Figure 2(b): there are 32 such tables, 11 of which lead to $\chi^2/n \leq .222$, determining a ChiVol of $11/32$. For $n \rightarrow \infty$, the number of possible tables increases, so that the discrete points cover the entire area corresponding to the admissible values for $\{\pi_{ij}\}$ (shaded in Figure 2), so that ChiVol approximates the ratio of two volumes: the one of the space of tables with smaller Mahalanobis distance from independence than the observed one and the one of the space of admissible tables.

3. CHIVOL AS A MEASURE OF DEPENDENCE

Consider now the measure of dependency defined by (4). We can make the following remarks with reference to its properties.

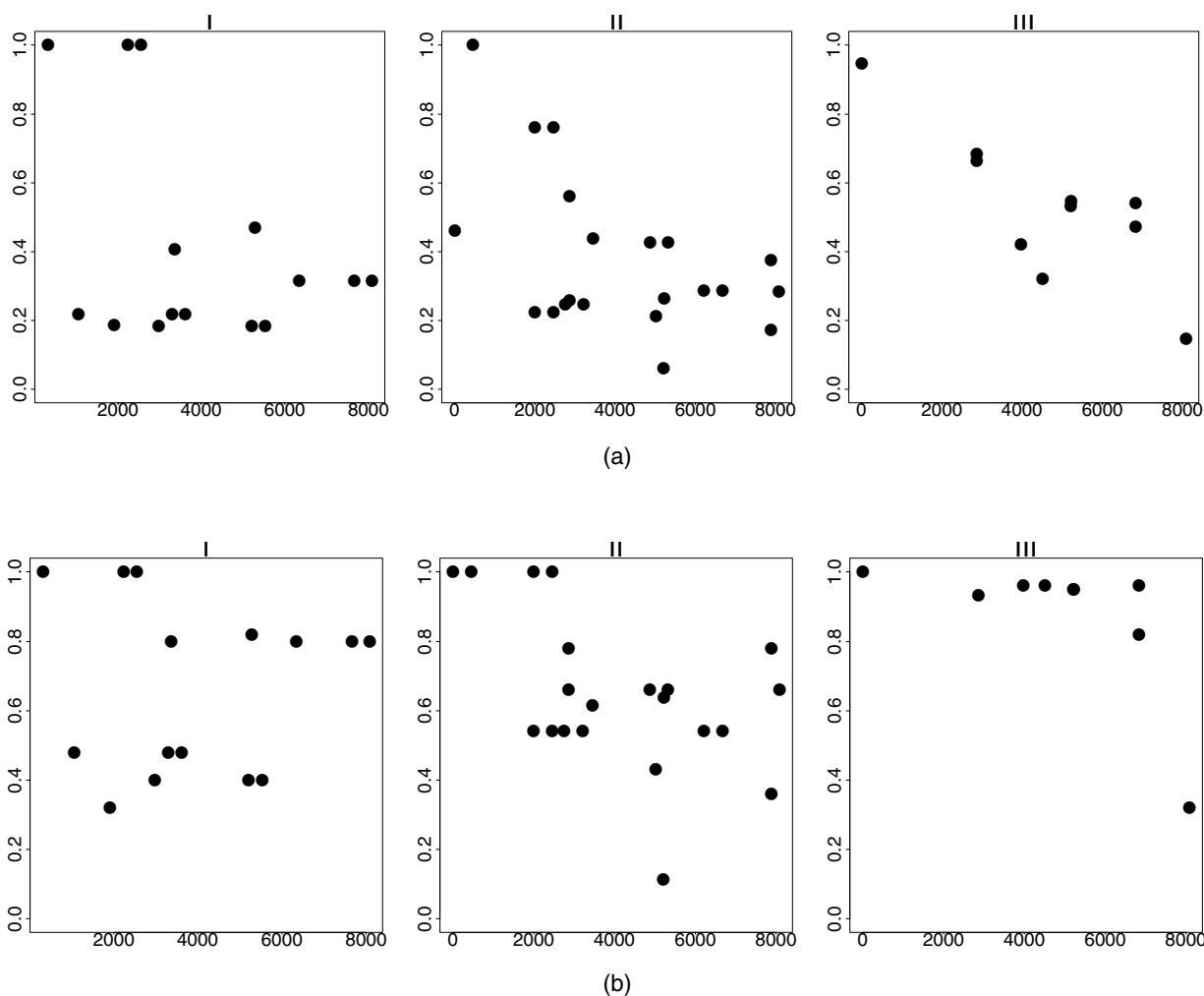


Figure 3. Analysis of mitochondria data using (a) R , and (b) $Chivol$. In each graph, the x-axis is the distance between markers (in number of base pairs) and the y-axis the measured dependence. Data in I was collected from Swedish and Finnish; in II, from Siberian; and in III, from Native American individuals.

(a) If the table observed coincides with the one expected under independence, $ChiVol$ takes on value 0. On the other hand, the value of $ChiVol$ approaches 1 (more closely as $n \rightarrow \infty$) as the distance between the observed table and the independent one increases. The value 1 would be attained irrespectively of the sample size n , if we adopted an alternative definition of $ChiVol$, substituting the numerator of (4) and its strict inequality with the one of (3). The advantage of the definition we have adopted is in its dependence on n in small sample sizes.

(b) As a measure of dependence, $ChiVol$ standardizes the χ^2 values with reference to the space of all possible tables with the same margins.

(c) $ChiVol$ depends on n in an advantageous way. On the one hand, when n is small, the limited number of possible tables influences the value of the ratio that defines $ChiVol$, appropriately discounting the observed dependence. On the other hand, when n is large, the measure becomes practically independent on n : as the number of possible tables $\{n_{ij}\}$ increases to cover the entire space of admissible $\{\pi_{ij}\}$, the value of $ChiVol$ approaches the ratio of two volumes.

One additional important remark pertains to the algorithmic evaluation of $ChiVol$. As Diaconis and Efron (1985) already

pointed out there is no general closed form solution to evaluate the ratio in (4). They proposed an analytical approximation that is more accurate for large n . When n is relatively small, instead, we propose using an approximation based on Markov chain Monte Carlo sampling. Indeed, using an appropriate Markov chain, it is possible to obtain a sample of tables with distribution as in H_0 . This sample can be used to estimate the value of the ratio in (4) simply counting the fraction of the tables in the sample that have χ^2 statistic smaller than the observed one. An algorithm to construct a Markov chain of this type was proposed by Diaconis and Sturmfels (1998). As initial state of the Markov chain, take the observed table; at each iteration, two rows and two columns are selected and the 2×2 table identified by their intersection is updated. This local move consists in replacing the identified 2×2 table with another one sampled uniformly among the tables with identical margins: because of these marginal constraints, this amounts to generate a uniform one-dimensional random variable. Diaconis and Sturmfels (1998) gave n^2 as a conservative upper-bound on the number of iterations necessary for stationarity. As a matter of fact, the use of MCMC methods enables one to use volume tests in contingency tables with much more

generality than was possible in the context originally studied by Diaconis and Efron (1985).

4. LINKAGE DISEQUILIBRIUM

In the introduction, we mentioned the need in statistical genetics for reliable measures of associations for comparisons across tables. Let us now illustrate one context where this is particularly true and give some background information. Human cells contain one chromosome of each type, one inherited from the mother and one from the father of the individual. During the process of gametes formation, these “homologous” chromosomes pair up forming one body that afterwards separates originating two novel chromosomes, patchwork of the original ones. Each one of these chromosomes will be transmitted to one gamete cell. The process with which the chromosomes “exchange” a portion of their DNA material is called recombination. Consider now two specific loci on a chromosome: if the alleles from each of them are coming from the same ancestral (paternal or maternal) chromosome we say that there is no recombination between the loci. If, instead, there has been an exchange of genetic material between the two loci, we say that there has been a recombination between them. The chance of observing a recombination between two loci increases with the distance between them. Recall now that for two loci A and B we considered the random variables X and Y , whose realizations are the alleles that a given chromosome has in correspondence of A and B . The presence or absence of recombination between A and B will translate in varying degrees of dependence between the random variables X and Y , so that the amount of dependence between X and Y , the frequency of recombination between A and B and the distance of these two loci are all positively related. Two markers whose allele distribution are not independent are said to be in linkage disequilibrium. Experimentally, it is easiest to measure the frequency of the alleles at the two loci than their distance or the amount of recombination between them. Hence, the observed dependency between the random variables X and Y (linkage disequilibrium) is used to infer information about the other quantities. Notice, that the marginal distribution of X and Y carry no information about the amount of recombination between the two loci. Hence, a satisfactory measure of dependency should condition on marginal frequencies.

Measures of linkage disequilibrium have been recently used to question a very well established hypothesis on the mechanism of transmission of mitochondria. Mitochondria are cell organelles that are responsible for respiratory activities and that contain some small amounts of chromosomal DNA, an exception with respect to the general fact that DNA is organized in chromosomes that reside in the cell nucleus. The current belief with regard to the mechanism of transmission of mitochondria is that they are inherited only from the mother and that their DNA experiences no recombination. Indeed, numerous studies that aim at dating species have been conducted using this hypothesis. Recently, the fact that the amount of disequilibrium between markers seemed to decrease as the distance between them increased has been used to suggest the presence of recombination in mitochondrial DNA, against this established hypothesis. Awadalla, Eyre-Walker, and Smith (1999) used R^2 on 2×2 tables as a measure of linkage disequilibrium and bases its claims on the

observation that the levels of linkage disequilibrium decrease as the distance between markers increases, as illustrated in Figure 3(a). As illustrated in (2), R^2 is equivalent to the χ^2/n statistics on 2×2 tables and it can be equal to 1 only when the marginal frequencies of X and Y are equal. This leads to the suspicion that the effect observed by Awadalla et al. (1999) may be at least partly due to the lack of consideration of marginal frequencies.

ChiVol can be profitably applied in this context: unlike R^2 it takes fully account of the marginal frequencies and unlike other measures of disequilibrium often used in genetics (Devlin and Risch 1995) is not defined only for 2×2 tables. We hence computed ChiVol for these three datasets.

As the number of observations n is relative large (≈ 200), we computed ChiVol as the ratio of relevant volumes, which is easy to calculate explicitly in this case. Indeed, recalling the parameterization described in (2), $\chi^2/n = (x - pq)^2/pq(1-p)(1-q)$, function of only one variable x , with $\max(0, p+q-1) \leq x \leq \min(p, q)$. The volume of all possible tables in this case corresponds simply to the length of the above segment: $V_1 = \min(p, q) - \max(0, p+q-1)$. We can then find the boundary values for x that lead to a χ^2 with values inferior to the observed one and take the ratio of the length V_2 of this segment and V_1 as ChiVol. The results are in Figure 3(b). The relation between dependence (disequilibrium) of alleles and distance of loci exhibited by R^2 is greatly diluted by ChiVol and its statistical significance disappears. This suggests (in agreement with other scientists—see the discussion by Kivisild et al. (2000)) that the evidence presented in favor of the hypothesis of recombination in mitochondria is rather weak.

5. CONCLUSION

The ideas of Hotelling (1939) and Diaconis and Efron (1985) should receive renewed attention. The availability of MCMC algorithms of known behavior makes it possible to estimate with reasonable precision and computation time the value of the ratio of volumes. Additionally, statistical genetics calls for measures of dependency conditional on the marginals of a contingency table and sensitive to the sample size, providing an ideal field of application for Hotelling’s original proposal.

[Received October 2000. Revised February 2002.]

REFERENCES

- Awadalla, P., Eyre-Walker, A., and Smith, J. (1999), “Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA,” *Science*, 286, 2524–2525.
- Devlin, B., and Risch, N. (1995), “A Comparison of Linkage Disequilibrium Measures for Fine-Scale Mapping,” *Genomics*, 29, 311–322.
- Diaconis, P., and Efron, B. (1985), “Testing for Independence in a Two-Way Table: New Interpretations of the Chi-Square Statistics,” *The Annals of Statistics*, 13, 845–874.
- Diaconis, P., and Sturmfels, B. (1998), “Algebraic Algorithms for Sampling from Conditional Distributions,” *The Annals of Statistics*, 26, 363–397.
- Fisher, R. A. (1958), *Statistical Methods for Research Workers* (13th ed.), New York, NY: Hafner Publishing Company Inc.
- Hotelling, H. (1939), “Tubes and Spheres in n -spaces, and a Class of Statistical Problems,” *American Journal of Mathematics*, 61, 440–460.
- Huttley, G., Smith, M., Carrington, M., and O’Brien, S. (1999), “A Scan for Linkage Disequilibrium Across the Human Genome,” *Genetics*, 152, 1711–1722.
- Kivisild, T. et al. (2000), “Questioning Evidence for Recombination in Human Mitochondrial DNA,” *Science*, 288, 1931a.
- Lewontin, R. C. (1964), “The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models,” *Genetics*, 49, 49–67.