Heredity (1983), **51**(3), 665–668

ESTIMATION OF RELATIVE DNA DENSITY IN HETEROCHROMATIN FROM C-BAND KARYOTYPES (CRITICAL REMARKS)

W. TIMISCHL* AND J. GREILHUBER†

** Institut für Algebra und Diskrete Mathematik, Technische Universität Wien, Argentinierstrasse 8, A-1040 Vienna*

† Institut für Botanik, Universität Wien, Rennweg 14, A-1030 Vienna

Received 15.iv.83

SUMMARY

The paper deals with problem of estimating the ratio of DNA densities in heterochromatin and euchromatin from C-band karyotypes of related species, when the observations are subject to errors of measurement. Using a set of data published for Lolium it is shown that under realistic assumptions about the error variances no significant difference of the densities in the two types of chromatin can be detected.

1. INTRODUCTION

In a recent investigation it was found by Thomas (1981) that in Lolium C-band heterochromatin has approximately seven times as much DNA per unit metaphase chromosome as euchromatin. This result was obtained from the data on 6 species listed in his table 1 by means of the simple regression model

$$M = c_h H + c_e E + f_m \tag{i}$$

with M being the DNA mass (in pg) per nucleus, H the length of C-band heterochromatin, E the length of euchromatin (both in μm) and f_m the error component of M (thus it is assumed that both H and E are measured without errors). The least-squares estimates of the parameters c_h and c_e were found to be $\hat{c}_h = 0{\cdot}388$ and $\hat{c}_e = 0{\cdot}0537$. These estimates were interpreted as "average concentrations" of DNA in the two types of chromatin and it was concluded from these figures that the ratio $c_e : c_h$ is as low as approximately 1:7. Unfortunately, no attempt was made to discuss the statistical significance of this ratio, either by the author or by other authors who seem to have accepted this result even if it is not supported by their own cytophotometric data (see Lukaszewski *et al.*, 1982). It is very doubtful if the value found for the ratio $c_e : c_h$ has any practical meaning at all because of the assumption that H and E can be observed without errors. We think that this assumption is not justified and it is our aim to show what happens if errors of measurement are taken into consideration. Two possible approaches for handling the problem are outlined in the next section; in model A a linear relationship between mathematical variables is assumed, in model B a structural relationship between random variables.

2. Models and discussion

Model A

It is first assumed that the variables $X_1 = H$, $X_2 = E$ and $X_3 = M$ are each subject to an error of measurement, *i.e.*,

(ii) $$X_i = X_i' + f_i \qquad (i = 1, 2, 3)$$

where the X_i' are the "true" variables which can not be observed because of some error random variables f_i. The X_i' are considered as non-random, *i.e.*, as mathematical variables, which are linearly related according to

(iii) $$X_3' = a_1 X_1' + a_2 X_2'.$$

Inserting (ii) into (iii) gives $X_3 = a_1 X_1 + a_2 X_2 + f$ (where $f = f_3 - a_1 f_1 - a_2 f_2$), which is different from (i) because X_1 as well as X_2 are correlated with f Thus this equation can not be handled as a simple regression problem unless both $\sigma_{f_1}^2$ and $\sigma_{f_2}^2$ are zero (see *e.g.*, Kendall and Stuart 1979, p. 401). What remains is to construct a confidence region for the parameters a_1 and a_2 which are the concentrations of the two types of chromatin. In addition to (ii) and (iii) the error variables f_i are supposed to be uncorrelated and each distributed as $N(0, \sigma_{f_i}^2)$. Let (X_{1j}, X_{2j}, X_{3j}) be the triple of variables corresponding to observation j $(j = 1, 2, \ldots, n)$. Then, according to Brown (1957, see also Brown *et al.*, 1958) the sum of squares of perpendiculars from the points $(X_{1j}/\sigma_{f_1}, X_{2j}/\sigma_{f_2}, X_{3j}/\sigma_{f_3})$ on to the hyperplane (iii), *i.e.*,

$$D = \sum_{j=1}^{n} (a_1 X_{1j} + a_2 X_{2j} - X_{3j})^2 / (\sigma_{f_1}^2 a_1^2 + \sigma_{f_2}^2 a_2^2 + \sigma_{f_3}^2)$$

has a χ_n^2-distribution, and thus with c satisfying $P(\chi_n^2 > c) = 1 - \gamma$ it follows that $D \leqq c$ is a $100(1-\gamma)$ per cent confidence region for (a_1, a_2). The equation $D = c$ defines a conic in the (a_1, a_2)-plane, which is—neglecting degenerate cases—a hyperbola if

$$\Delta = \left(\sum_j X_{1j}^2 - c\sigma_{f_1}^2\right)\left(\sum_j X_{2j}^2 - c\sigma_{f_2}^2\right) - \left(\sum_j X_{1j} X_{2j}\right)^2$$

is less than zero and an ellipse for $\Delta > 0$. Consequently the confidence region for (a_1, a_2) is bounded for $\Delta > 0$ and unbounded for $\Delta < 0$. An illustration of the latter case is given by figure 1, which is based on the data in Thomas' paper and on realistic assumptions about the error variances. Figure 1 also shows the straight lines $a_2/a_1 = 1$ as well as $a_2/a_1 = 1/7$ (the latter corresponds to the ratio of concentrations found by Thomas) and it is seen that there is no reason why one should rely more on $a_2/a_1 = 1/7$ than on *e.g.*, the ratio $a_2/a_1 = 1$ corresponding to equal concentrations.

For a detailed discussion we refer to the article by Brown (1958) and also to the comments in the book by Kendall and Stuart (1979).

Model B

We now assume that X_1' and X_2' are random variables from which the new variables $Y_1' = X_1'/X_3$ and $Y_2' = X_2'/X_3$ (being both of dimension length per mass) are derived. Instead of equation (iii) the relation

(iv) $$E(Y_2') = b_1 E(Y_1') + b_0$$

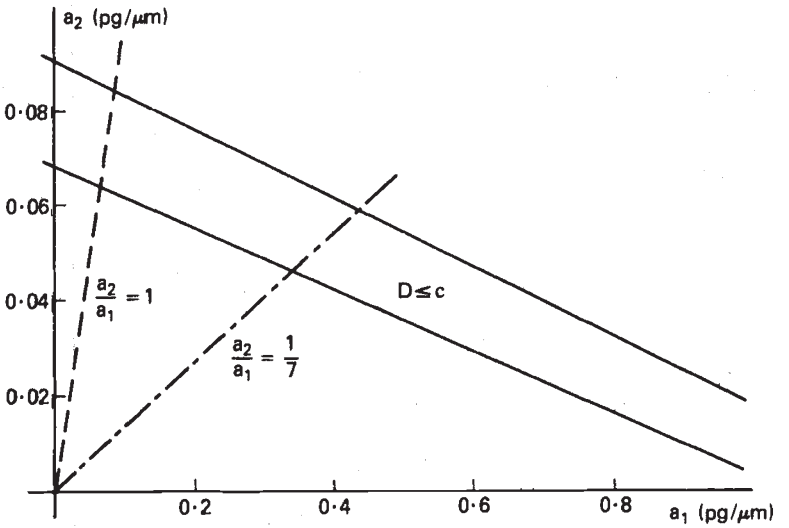

FIG. 1. 95 per cent-confidence region for the parameters a_1, a_2 in equation (iii) based on the data of Thomas (1981, table 1). The standard deviations σ_{f_1} and σ_{f_2} were taken to be 10 per cent of the sample means of X_1 and X_2; σ_{f_3} was assumed to be 5 per cent of the corresponding sample mean.

is postulated between the expectations of Y_2' and Y_1', where b_1 is just the negative ratio of DNA concentration of C-band heterochromatin and euchromatin and b_0 a constant. Model B is thus based on a structural relationship between two (unobservable) random variables Y_2' and Y_1'. What can be observed are the variables $Y_i = Y_i' + g_i$ $(i = 1, 2)$ where the error variables g_i are again assumed to be uncorrelated and distributed as $N(0, \sigma_{g_i}^2)$ and the $Y_i'^{s}$ as $N(\mu_{Y_i}, \sigma_{Y_i}^2)$. If the ratio $\lambda = \sigma_{g_2}^2/\sigma_{g_1}^2$ of error variances is known, the *ML*-estimate $\hat{b}_1^*$ of $b_1^* = b_1/\sqrt{\lambda}$ is found to be (see *e.g.*, Kendall and Stuart 1979)

$$\hat{b}_1^* = \frac{1}{2s_{Y_1Y_2^*}}[(s_{Y_2^*}^2 - s_{Y_1}^2) + \sqrt{(s_{Y_2^*}^2 - s_{Y_1}^2)^2 + 4s_{Y_1Y_2^*}^2}]$$

where $s_{Y_1}^2$, $s_{Y_2^*}^2$ are the sample variances of Y_1 and $Y_2^* = Y_2/\sqrt{\lambda}$ and $s_{Y_1Y_2^*}$ the sample covariance between Y_1 and Y_2^*. In order to test the hypothesis $b_1 = b_{10}$ or to construct a confidence interval for b_1 the test statistic

$$\text{(v)} \qquad T = \sqrt{(n-2)\sin^2 2(\beta_1^* - \beta_{10}^*)\frac{(s_{Y_2^*}^2 - s_{Y_1}^2)^2 + 4s_{Y_1Y_2^*}^2}{4(s_{Y_1}^2 s_{Y_2^*}^2 - s_{Y_1Y_2^*}^2)}}$$

may be used which follows the t-distribution with $(n-2)$ degrees of freedom (n is the number of observations (y_1, y_2) and $\beta_1^* = \arctan b_1^*$, $\beta_{10}^* = \arctan b_{10}^*$). Assuming the error variances to be proportional to the squares of the corresponding means μ_{Y_i}, we get from the data by Thomas $\lambda \approx 15$ and $\hat{b}_1 = -6{\cdot}18$. From (v) the 95 per cent-confidence interval $(-18{\cdot}3, 1{\cdot}5)$ may be derived for b_1 which is rather large. It includes the b_1-value corresponding to the ratio 1/7 of concentrations as well as the b_1-value corresponding to *e.g.*, equal concentrations.

3. References

BROWN, R. L. 1957. Bivariate structural relation. *Biometrica, 44*, 84–96.

BROWN, R. L. AND FEREDAY, F. 1958. Multivariate linear structural relations. *Biometrica, 45*, 136–153

KENDALL, M. AND STUART, A. 1979. The advanced theory of statistics. Vol. 2. Griffin, London.

LUKASZEWSKI, A. J., WOZNA, J., APOLINARSKA, B. AND NAGANOWSKA, B. 1982. The DNA content of the individual chromosome of rye. *Theor. Appl. Genet. 62*, 145–148.

THOMAS, H. M. 1981. The Giemsa C-band karyotypes of six *Lolium* species. *Heredity, 46*, 263–267.