ESTIMATION OF LINKAGE IN PRESENCE OF INCOMPLETE PENETRATION FOR BOTH THE FACTORS

RANAJIT CHAKRABORTY
Indian Statistical Institute, Calcutta

1. INTRODUCTION AND SUMMARY. Measurement of penetrance is done by taking the proportion of individuals carrying the gene who are phenotypically different from the individuals of the population carrying the allele.

Sanchez-monge (1952) considered the problem of estimation of linkage between two genes, one of which has incomplete penetrance. Here the case when both the genes have got incomplete penetrance is considered.

Tests of significance of the relevant hypotheses using the estimates of linkage and penetrance co-efficients are also given for backcross as well as intercross data. A modified test of significance of the estimate of penetrance is also given using the Normal deviate.

2. BACKCROSS DATA. Consider the typical backcross mating GgL*l*× gg*ll*. Let 'p' denote the recombination fraction. We consider the case of linkage in coupling phase only. The different phenotypes of the offsprings of such crosses are GL, G*l*, gL and g*l*. The expected frequencies of these phenotypes are given in Table 1 under $u = 1$, $v = 1$; the totality of individuals being n.

Let $1-u$ = Fraction of 'Gg'—genotypes classified phenotypically as 'gg'.

and $1-v$ = Fraction of 'L*l*'—genotype classified phenotypically as '*ll*'.

The expected frequencies for the cases $u \neq 1$, $v = 1$ and $u \neq 1$, $v \neq 1$ are also presented in Table 1.

Table 1

Phenotypes	*Expected Frequencies* $\times 2/n$		
	$u = 1, v = 1$	$u \neq 1, v = 1$	$u \neq 1, v \neq 1$
GL	$1-p$	$(1-p)u$	Z
G*l*	p	pu	$u-Z$
gL	p	$p+(1-p)(1-u)$	$v-Z$
g*l*	$1-p$	$(1-p)+p(1-u)$	$2-u-v+Z$

where $Z = (1-p)uv$.

Let a, b, c and d denote the observed frequencies. Equating the observed and expected frequencies we obtain estimates of u, v, Z and hence of p given by

$$\left.\begin{array}{l} \hat{u} = \frac{2}{n}(a+b)\ ;\ \hat{v} = \frac{2}{n}(a+c)\ ;\ \hat{Z} = \frac{2a}{n} \\ \text{and } \hat{p} = 1-\frac{Z}{uv} = 1-\frac{an}{2(a+b)(a+c)} \end{array}\right\} \quad \ldots (2.1)$$

Writing $Y' = \frac{1}{n}(a,\ b,\ c)$; $X' = (u,\ v,\ Z)$ we have the relationship

$Y = A\ X$ where the matrix A is suitably defined.

The dispersion matrix of the estimates, D_X, is obtained from

$$D_X = A^{-1} D_Y\ (A^{-1})' \quad \ldots (2.2)$$

given by

$$D_X = \frac{1}{n}\begin{bmatrix} u(2-u) & 2Z-uv & Z(2-u) \\ 2Z-uv & v(2-v) & Z(2-v) \\ Z(2-u) & Z(2-v) & Z(2-Z) \end{bmatrix} \quad \ldots (2.3)$$

Applying the formula (Rao, 1965) an estimate of the variance of $\hat{p}$ is obtained as

$$\hat{V}_{\hat{p}} = (1-\hat{p})^2\left(\frac{1}{k}-\frac{1}{n}\right) \quad \ldots (2.4)$$

where

$$k = \frac{a(a+b)^2\,(a+c)^2}{(bc-a^2)^2+ab(a+c)^2+ac(a+b)^2}$$

It can be seen that $\hat{u}$, $\hat{v}$ and $\hat{p}$ are also the maximum likelihood estimates.

Let us now study the significance of the departure of u and v from unity admitting that there is linkage.

For $u = 1$, $v = 1$, the estimate of p is

$$\hat{p} = \frac{b+c}{n}$$

Hence under the hypothesis $u = 1 = v$, the linkage is $(b+c)/n$ and the expected frequencies are $\frac{1}{2}(a+d)$, $\frac{1}{2}(b+c)$, $\frac{1}{2}(b+c)$ and $\frac{1}{2}(a+d)$.

Comparing these with a, b, c and d we get

$$\chi^2 = \frac{(a-d)^2}{a+d}+\frac{(b-c)^2}{b+c} \quad \ldots (2.5)$$

with 2 d.f. This is the same χ^2 obtained by Sanchez-monge (1952) for testing the departure from $u = 1$.

Conversely, to detect the existence of linkage, admitting incomplete penetrance at both the factors we have for $p = \frac{1}{2}$, the estimates of u and v same as before.

The χ^2 in this case is given by

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)} \qquad \ldots (2.6)$$

with 1 d.f. This is the same χ^2 used for testing the independence of the two factors.

It can be noted here that the χ^2-test with 2 d.f. used by Sanchez-monge (1952) is not valid for testing the departure of u from unity. An appropriate test-statistic for the hypothesis $u = 1$ is $(u-1)/\sqrt{\hat{V}_{\hat{u}}}$ which is asymptotically normal (0, 1). The explicit form of the normal deviate is

$$t = (ab-cd)(a+c)(b+d)\sqrt{\frac{2}{n\{ac(b+d)^4+bd(a+c)^4\}}} \qquad \ldots (2.7)$$

3. Intercross Data. In case of a typical intercross of the type GgLl×GgLl with complete penetrance of both genes the frequencies of the four phenotypes are presented in Table 2 under $u = 1$, $v = 1$; p and p' are the recombination fractions for the gametes of the two sexes, giving $P = pp'$ for crossings in repulsion and $P = (1-p)(1-p')$ for crossings in coupling. But in case of incomplete penetrance with fractions u and v for 'GG', 'Gg' to be classified as 'gg' and 'LL', 'Ll' to be classified as 'll' respectively the frequencies will be those shown in Table 2 under $u \neq 1$, $v \neq 1$. Again for incomplete penetrance of G gene alone (i.e. for $u \neq 1$, $v = 1$) the frequencies are shown in Table 2 which are the same as obtained by Sanchez-monge (1952).

Table 2

Phenotypes	*Expected Frequencies* $\times\frac{4}{n}$		
	$u = 1, v = 1$	$u \neq 1, v = 1$	$u \neq 1, v \neq 1$
GL	$2+P$	$(2+P)u$	xv
Gl	$1-P$	$(1-P)u$	$y+x(1-v)$
gL	$1-P$	$1-P+(2+P)(1-u)$	$(3-x)v$
gl	P	$P+(1-P)(1-u)$	$1-y+(3-x)(1-v)$

where $x = (2+P)u$ and $y = (1-P)u$.

A further substitution of $Z = xv$ gives the expected frequencies as Z, $3u-Z$, $3v-Z$ and $4-3u-3v+Z$ respectively.

Equating observed and expected frequencies the estimates are obtained as

$$\left.\begin{aligned} \hat{u} &= \frac{4}{3n}(a+b) \; ; \; \hat{v} = \frac{4}{3n}(a+c) \; ; \; \hat{Z} = \frac{4a}{n} \\ \text{and hence} \quad \hat{P} &= \frac{\hat{Z}}{\hat{u}\,\hat{v}} - 2 \\ &= \frac{9an}{4(a+b)(a+c)} - 2 \end{aligned}\right\} \qquad \ldots (3.1)$$

Using the same set up and suitably changing the A-matrix, from (2.2) we can have the dispersion-matrix of the estimates $(\hat{u}, \hat{v}, \hat{Z})$ as

$$D_X = \frac{1}{9n}\begin{bmatrix} 3u(4-3u) & 4Z-9uv & 3Z(4-3u) \\ 4Z-9uv & 3Z(4-3v) & 3Z(4-3v) \\ 3Z(4-3u) & 3Z(4-3v) & 9Z(4-Z) \end{bmatrix} \qquad \ldots (3.2)$$

Variance of $\hat{P}$ can be estimated from variance of $\hat{Z}$ as

$$\hat{V}_{\hat{P}} = (2+\hat{P})^2 \left(\frac{1}{k} - \frac{1}{n}\right) \qquad \ldots (3.3)$$

Now as done for backcross data we can test the significance of the deviation of u and v from unity by estimating P assuming $u = 1 = v$ and then obtaining χ^2 (with 2 d.f.) which can be obtained by comparing the expected frequencies with a, b, c and d.

Conversely, to detect the existence of linkage assuming incomplete penetrance for both the genes, we test for $P = \frac{1}{4}\left(p = \frac{1}{2} = p'\right)$ $\hat{u}$ and $\hat{v}$ are again given by (3.1). The expected class frequencies are same as it was for χ^2 in (2.6).

For incomplete penetrance at one factor alone (i.e. for $v = 1$) Sanchez-monge's testing procedure [for testing the departure of u from unity] needs modification. Here the normal deviate is given by

$$t = \frac{\dfrac{b}{3(b+d)} - \dfrac{c}{a+c}}{\sqrt{\dfrac{n}{4}\left\{\dfrac{3ac}{(a+c)^3} + \dfrac{bd}{9(b+d)^3}\right\}}} \qquad \ldots (3.4)$$

4. ACKNOWLEDGEMENTS. I express my deep gratitude and regards to Prof. C. R. Rao whose valuable guidance has enabled me to produce this paper in this short time. I am also grateful to Dr. Y. R. K. Sarma for helpful discussions.

REFERENCES

Rao, C. R. (1965) : *Linear Statistical Inference and its applications*". John Wiley and Sons, New York

Sanchez-monge, E (1952) : *Heredity*, **6**, 121.