NOTE ON A PAPER BY C. W. COTTERMAN AND L. H. SNYDER

By H. B. Mann[1]

Ohio State University

C. W. Cotterman and L. H. Snyder [1] gave a method to test simple Mendelian inheritance in randomly collected data. From a population assumed to be at equilibrium a sample is taken. The number of homozygous recessives in the sample is known. We wish to estimate the number of heterozygous individuals in the sample.

Let α be the proportion of recessive genes among all genes in the population; π, ρ, τ the proportion in the population of homozygous recessives, heterozygous and homozygous dominant individuals respectively and p, r, t the sampling values of π, ρ, τ. Then

$$\pi = \alpha^2, \rho = 2\alpha(1 - \alpha), \tau = (1 - \alpha)^2, p + r + t = 1. \tag{1}$$

Cotterman and Snyder use as an estimate of r the quantity $2\sqrt{p}(1 - \sqrt{p})$. It is the purpose of this note to show that this estimate is for all practical purposes equivalent to the maximum likelihood estimate of r.

The joint distribution of p, r and t in samples of n is given by

$$P(p, r, t) = \frac{n!\pi^{np}\rho^{nr}\tau^{nt}}{(np)!(nr)!(nt)!} = \frac{n!\alpha^{2np}[2\alpha(1 - \alpha)]^{nr}(1 - \alpha)^{2nt}}{(np)!(nr)!(nt)!}, \tag{2}$$

where $P(p, r, t)$ is the probability of obtaining the values p, r, t in samples of n.

We wish to maximize $P(p, r, t)$ for fixed values of p with respect to α and r. Maximizing first with respect to α one easily obtains

$$2\alpha = 2p + r. \tag{3}$$

We can regard α as a continuous parameter and hence (3) must hold at any maximum of $P(p, r, t)$. For any maximum of $P(p, r, t)$ we must further have

$$\frac{n!\pi^{np}\rho^{nr}\tau^{nt}}{(np)!(nr)!(nt)!} \geqslant \frac{n!\pi^{np}\rho^{nr+1}\tau^{nt-1}}{(np)!(nr + 1)!(nt - 1)!}$$

and

$$\frac{n!\pi^{np}\rho^{nr}\tau^{nt}}{(np)!(nr)!(nt)!} \geqslant \frac{n!\pi^{np}\rho^{nr-1}\tau^{nt-1}}{(np)!(nr - 1)!(nt + 1)!}.$$

This leads to the inequalities

$$\frac{\tau}{nt} \geqslant \frac{\rho}{nr + 1}, \qquad \frac{\rho}{nr} \geqslant \frac{\tau}{nt + 1}. \tag{4}$$

Substituting $t = 1 - p - r$, $\tau = 1 - \pi - \rho$ one easily obtains from (4)

[1] Research under a grant of the research foundation of Ohio State University.

$$\frac{\rho n - \rho n p + \rho}{n(1 - \pi)} \geqslant r \geqslant \frac{\rho n - \rho n p - \tau}{n(1 - \pi)}. \tag{5}$$

The difference of the two bounds is $\frac{1}{n}$. Hence r must satisfy an equation

$$r = \frac{\rho n - \rho n p + \rho}{n(1 - \pi)} - \frac{\epsilon}{n}, \qquad 0 \leqslant \epsilon \leqslant 1.$$

Substituting the values for ρ, π and r from (1) and (3) we obtain

$$\alpha^2 - \frac{\alpha}{n}(1 - \epsilon/2) - p + \frac{\epsilon}{2n} = 0,$$

$$\alpha = \frac{2 - \epsilon}{4n} + \frac{1}{2}\sqrt{\frac{(2 - \epsilon)^2}{4n^2} + 4p - \frac{2\epsilon}{n}}.$$

Since $0 \leqslant \epsilon \leqslant 1$ we obtain from (3)

$$\frac{1}{n} + \sqrt{4p + \frac{1}{n^2}} - 2p \geqslant r \geqslant \frac{1}{2n} + \sqrt{4p + \frac{1}{4n^2} - \frac{2}{n}} - 2p. \tag{6}$$

From (6) we see that for all practical purposes we may use the estimate

$$r = 2\sqrt{p}(1 - \sqrt{p}).$$

REFERENCE

[1] C. W. Cotterman and L.H. Snyder, "Tests of simple Mendelian inheritance in randomly collected data of one and two generations," *Jour. Am. Stat. Assn.*, Vol. 34 (1939), pp. 511–523.