

THE ESTIMATION OF GENE FREQUENCIES FROM FAMILY RECORDS

II. FACTORS WITH DOMINANCE

D. J. FINNEY

*Lecturer in the Design and Analysis of Scientific Experiment,
University of Oxford*

Received 2.iii.48.

I. INTRODUCTION

IN the previous paper (Finney, 1948), the theory of scoring records of related individuals, for the purpose of obtaining an estimate of gene frequency in the population from which the records were derived, was discussed. As Fisher (1940) has shown, application of the principle of maximum likelihood leads to a system of efficient scores, and so to an estimate of gene frequency which not only is unbiased but also has a minimal sampling variance. Formulæ for these scores have been derived and tabulated in the earlier paper for three types of record, two parents and their s children, one parent and one child, and two sibs with neither parent recorded, in respect of a genetic factor which shows no dominance. For sets of sibs with one or neither parent recorded, the complexity of the scores increases so rapidly as to make maximum likelihood scoring impracticable; a modification of a method due to Cotterman (1947), simple in use and of high efficiency, was therefore recommended for general use and tables of weights to be attached to the scores were given.

The purpose of the present paper is to set out a similar scoring system for use with a factor showing dominance. Fisher has already given results for maximum likelihood scores, and Cotterman has again pointed the way to a simplified system of high efficiency. Some recapitulation of these results is necessary for completeness here; the final system recommended is of higher efficiency than Cotterman's, but more easily applied to large families than Fisher's.

2. MAXIMUM LIKELIHOOD SCORING

A genetic factor involving two allelomorphous genes, T and t , of which T is dominant to t , gives rise to only two phenotypically distinguishable forms of individual. If the gene frequencies are denoted by μ , ν respectively, the frequency of recessive tt individuals in a population mating at random will be ν^2 . It is convenient to develop formulæ for the estimation of ν^2 , rather than of ν , since for the simplest case the procedure reduces to a count of recessives.

A sample of unrelated individuals consisting of a dominants and b recessives yields an estimate

$$\nu^2 = \frac{b}{a+b}, \quad \dots \quad (1)$$

with variance

$$V(\nu^2) = \frac{\nu^2(1-\nu^2)}{a+b} \quad \dots \quad (2)$$

For groups of related individuals, an estimate formed in this way will still be consistent and unbiased, but its variance will be greater than that given by equation (2). Furthermore, it will not in general be the most precise estimate obtainable from the data.

The theory of fully efficient scores, based on the principle of maximum likelihood, has been outlined by Fisher (1940) and Finney (1948). The probability that members of a family (as defined in the previous paper), having a specified pattern of relationship with one another, shall be of a particular observed set of phenotypes may be expressed as a function of ν , $P(\nu)$. A weighted maximum likelihood score for the family may then be calculated in terms of a provisional estimate of ν^2 as

$$x = W_L \nu^2 + \frac{\nu^2(1-\nu^2)}{P} \frac{dP}{d\nu^2}; \quad \dots \quad (3)$$

W_L , the weight of the score is given by

$$W_L = S \left\{ \frac{1}{P} \left(\frac{dP}{d\nu^2} \right)^2 \right\}, \quad \dots \quad (4)$$

the summation being over families of the same pattern with all possible sets of phenotypes. A revised estimate obtained from records of a number of families of any pattern is

$$\nu^2 = \frac{\Sigma x}{\Sigma W_L}, \quad \dots \quad (5)$$

with variance

$$V(\nu^2) = \frac{\nu^2(1-\nu^2)}{\Sigma W_L}, \quad \dots \quad (6)$$

Σ denoting summation over all records. The weight may be regarded as the number of unrelated individuals required to provide the same amount of information on ν^2 as is given by the family. Except for the occurrence of ν^2 instead of ν , these equations are the same as equations (3)-(6) of the previous paper, where they were derived for use with factors showing no dominance. Three examples follow.

(i) *Both parents recorded*

When both parents are recorded for a factor without dominance, children contribute no information (Cotterman, 1947; Finney, 1948). This is not so for a factor with dominance, since the phenotypes of

children may contribute information on whether a T parent is homozygous or heterozygous. For two parents with one child, the calculations leading to the maximum likelihood scores may be put in tabular form :—

Phenotypes		P	$\frac{dP}{d\nu^2}$	$\frac{1}{P} \frac{dP}{d\nu^2}$	$\frac{1}{P} \left(\frac{dP}{d\nu^2}\right)^2$
Parents	Child				
T, T	T	$\mu^2(1+2\nu)$	-3μ	$\frac{-3}{(1-\nu)(1+2\nu)}$	$\frac{9}{(1+2\nu)^2}$
T, T	tt	$\mu^2\nu^2$	$\mu(1-2\nu)$	$\frac{1-2\nu}{\nu^2(1-\nu)}$	$\frac{(1-2\nu)^2}{\nu^2}$
T, tt	T	$2\mu\nu^2$	$2-3\nu$	$\frac{2-3\nu}{2\nu^2(1-\nu)}$	$\frac{(2-3\nu)^2}{2\nu^2(1-\nu)}$
T, tt	tt	$2\mu\nu^3$	$\nu(3-4\nu)$	$\frac{3-4\nu}{2\nu^2(1-\nu)}$	$\frac{(3-4\nu)^2}{2\nu^2(1-\nu)}$
tt, tt	tt	ν^4	$2\nu^2$	$\frac{2}{\nu^2}$	4

Hence

$$S \left\{ \frac{1}{P} \left(\frac{dP}{d\nu^2} \right)^2 \right\} = \frac{6-\nu+\nu^2}{2\nu^2(1-\nu)(1+2\nu)},$$

so that, if $W_{a,L}(s)$ is the maximum likelihood weight for a family with a parents and s children recorded,

$$W_{2,L}(1) = \frac{(1+\nu)(6-\nu+\nu^2)}{2(1+2\nu)} \quad (7)$$

The weight is always greater than 2 (and rises to 3 when ν is very small) ; this is because the child provides evidence on the genotypes of the parents, so making the record of greater value than if the parents were scored as unrelated individuals and the child discarded.

From equation (3) and the working table above, the scores are derived as :—

$$\left. \begin{aligned} \text{Parent T, T ; child T : } x &= -\frac{\nu^3(1-\nu^2)}{2(1+2\nu)} \\ \text{Parents T, T ; child tt : } x &= \frac{(1-\nu^2)(2+2\nu-\nu^3)}{2(1+2\nu)} \\ \text{Parents T, tt ; child T : } x &= \frac{(1+\nu)(2+\nu-\nu^3+\nu^4)}{2(1+2\nu)} \\ \text{Parents T, tt ; child tt : } x &= \frac{(1+\nu)(3+2\nu-2\nu^2-\nu^3+\nu^4)}{2(1+2\nu)} \\ \text{Parents tt, tt ; child tt : } x &= \frac{(1+\nu)(4+4\nu-2\nu^2-\nu^3+\nu^4)}{2(1+2\nu)} \end{aligned} \right\} (8)$$

Table 1, analogous to Fisher's (1940) tables 1 and 3, gives numerical values for these scores and the weight, W , at different levels of ν^2 .

This table may be seldom required, however, as the modified scoring system described in section 3 is also fully efficient for families with one child and has the advantage of being applicable for larger values of s .

TABLE 1
Maximum likelihood score and weight for two parents and one child

ν^2 (provisional)	Scores for phenotypes as shown (parents on first line, child on second)					W_L
	T, T T	T, T tt	T, tt T	T, tt tt	tt, tt tt	
0.00	0.0000	1.0000	1.0000	1.5000	2.0000	3.0000
0.05	-0.0037	0.7995	0.9364	1.4114	2.0232	2.4631
0.10	-0.0087	0.7169	0.9251	1.3751	2.0332	2.3317
0.15	-0.0139	0.6506	0.9192	1.3442	2.0379	2.2525
0.20	-0.0189	0.5923	0.9159	1.3159	2.0395	2.1974
0.25	-0.0234	0.5391	0.9141	1.2891	2.0391	2.1562
0.30	-0.0274	0.4896	0.9134	1.2634	2.0373	2.1244
0.35	-0.0308	0.4430	0.9138	1.2388	2.0346	2.0990
0.40	-0.0335	0.3989	0.9152	1.2152	2.0314	2.0785
0.45	-0.0355	0.3570	0.9174	1.1924	2.0278	2.0618
0.50	-0.0366	0.3169	0.9205	1.1705	2.0240	2.0481
0.55	-0.0370	0.2786	0.9245	1.1495	2.0203	2.0369
0.60	-0.0365	0.2420	0.9293	1.1293	2.0166	2.0276
0.65	-0.0351	0.2069	0.9350	1.1100	2.0131	2.0202
0.70	-0.0329	0.1732	0.9416	1.0916	2.0099	2.0142
0.75	-0.0297	0.1410	0.9491	1.0741	2.0071	2.0094
0.80	-0.0257	0.1102	0.9574	1.0574	2.0046	2.0058
0.85	-0.0207	0.0807	0.9667	1.0417	2.0027	2.0031
0.90	-0.0147	0.0525	0.9769	1.0269	2.0012	2.0013
0.95	-0.0078	0.0256	0.9880	1.0130	2.0003	2.0003
1.00	0.0000	0.0000	1.0000	1.0000	2.0000	2.0000

This table is analogous to those given by Fisher (1940; tables 1 and 3) for one parent and one child and for a pair of sibs.

As usual, the complexity of the maximum likelihood scoring system is much greater for families of more than one child. The results for $s = 2$ have been obtained, but the scores have not been tabulated; the weight, which is required in the examination of the efficiency of the linear scores in section 3, is

$$W_{2,L}(2) = \frac{(1+\nu)(60-70\nu+42\nu^2-19\nu^3+5\nu^4)}{2(2-\nu)(4+8\nu-3\nu^2)}. \quad (9)$$

When the number of children is large, their segregations for T give a considerable amount of evidence on the genotypes of the parents. The limiting result is

$$W_{2,L}(s) \rightarrow \frac{(1+\nu)(6-3\nu+\nu^2)}{1+3\nu} \text{ as } s \rightarrow \infty, \quad (10)$$

a quantity which ranges from 6 when $\nu = 0$ to 2 when $\nu = 1$.

(ii) *One parent recorded*

For a record consisting of one parent and one child, Fisher (1940) used a similar analysis to show that the scores are:—

$$\left. \begin{aligned} \text{Parent T, child T} : x &= -\frac{\nu^3(1+\nu)(1-2\nu)}{4(1+\nu-\nu^2)} \\ \text{Parent T, child } tt \} : x &= \frac{(1+\nu)(4-2\nu-2\nu^2-\nu^3+2\nu^4)}{4(1+\nu-\nu^2)} \\ \text{Parent } tt, \text{ child T} \} \\ \text{Parent } tt, \text{ child } tt : x &= \frac{(1+\nu)(6-4\nu^2-\nu^3+2\nu^4)}{4(1+\nu-\nu^2)} \end{aligned} \right\} \quad (11)$$

with a weight for the pair of individuals

$$W_{1,L(1)} = \frac{(1+\nu)(8-7\nu+2\nu^2)}{4(1+\nu-\nu^2)} \quad (12)$$

Fisher shows values of these functions in his table 1. Again no attempt has yet been made to investigate the more complex expressions required for several children with one recorded parent, except to obtain the weights for $s = 2$ and for large s :

$$W_{1,L(2)} = \frac{(1+\nu)(264-568\nu+488\nu^2-219\nu^3+65\nu^4-20\nu^5+4\nu^6)}{4(2-\nu)(3-\nu)(4+4\nu-7\nu^2+\nu^3)} \quad (13)$$

and

$$W_{1,L(s)} \rightarrow \frac{(1+\nu)(5-6\nu+3\nu^2-\nu^3)}{1+2\nu-2\nu^2} \text{ as } s \rightarrow \infty \quad (14)$$

The last expression shows that for a small value of ν a large number of sibs with one recorded parent are almost as informative as 5 unrelated individuals, but that for large ν the value falls to 2.

(iii) *Neither parent recorded*

For two sibs with neither parent recorded, Fisher obtained the scores* :—

$$\left. \begin{aligned} \text{Children T T} : x &= W_{\nu^2} - \frac{\nu^2(1+\nu)(7-3\nu-2\nu^2)}{4+4\nu-3\nu^2-\nu^3} \\ \text{Children T } tt : x &= W_{\nu^2} + \frac{(1+\nu)(3-3\nu-2\nu^2)}{3+\nu} \\ \text{Children } tt, tt : x &= W_{\nu^2} + (1-\nu)(1+2\nu) \end{aligned} \right\} \quad (15)$$

where the weight per sib-pair is

$$W_{0,L(2)} = \frac{(1+\nu)(21-5\nu-8\nu^2+4\nu^3)}{(3+\nu)(4+4\nu-3\nu^2-\nu^3)} \quad (16)$$

These functions have been tabulated by Fisher in his table 3; once

* In Fisher's formula for the score when both children are T, the factor $(7+3\nu-2\nu^2)$ is a misprint for $(7-3\nu-2\nu^2)$; the numerical values in his table 3 are calculated from the correct formula.

again corresponding expressions for larger sibships have not been derived, except for the limiting value

$$W_{0,L}(s) \rightarrow \frac{(1+\nu)(4-3\nu+2\nu^2-\nu^3)}{1+2\nu-\nu^2} \text{ as } \nu \rightarrow \infty, \quad (17)$$

the value of which ranges from 4 when $\nu = 0$ to 2 when $\nu = 1$.

3. SIMPLIFIED SCORING SYSTEMS

The property of full efficiency possessed by a maximum likelihood estimation process is of little value to the investigator if the formulæ required for its use are not known to him. An occasional record of a group of relatives not covered by the scoring methods of section 2 might be dealt with by developing the maximum likelihood equations from first principles as required, but the application of section 2 is so limited in respect of family size that some alternative procedure is much to be desired. Cotterman (1947) showed that a score which was a simple count of genes was of high efficiency for the estimation of ν^2 . Just as in the previous paper (Finney, 1948), a sub-classification of types of record leads to a scoring system a little more efficient than Cotterman's which, with the aid of tables presented here, can be applied almost as easily as his very simple system. In this section, formulæ and tables will be given for sibships of s with two, one, or no parents recorded.

(i) Both parents recorded

Cotterman found that a score which is simply the total number of recessives amongst the $(s+2)$ individuals recorded carries the weight

$$W_{2,C}(s) = \frac{4(s+2)^2(1+\nu)}{(s^2+3s+8)+\nu(3s^2+17s+8)} \quad (18)$$

The weight for two unrelated individuals, scored by a simple count of recessives, is 2; when ν exceeds $(s+5)/(s+9)$, $W_{2,C}(s)$ is less than 2, and rejection of the children from the score will therefore improve the efficiency. By comparison of equation (18) for $s=1$ with the maximum likelihood weight, equation (7), the efficiency of Cotterman's score is seen to be

$$\text{Efficiency} = \left\{ \begin{array}{l} \frac{18(1+2\nu)}{(3+7\nu)(6-\nu+\nu^2)} \text{ for } \nu \leq 0.6 \\ \frac{4(1+2\nu)}{(1+\nu)(6-\nu+\nu^2)} \text{ for } \nu > 0.6 \end{array} \right\} \quad (19)$$

This efficiency approaches unity when ν is very small or very large, and never falls below 0.95. A similar comparison of equation (18) for $s=2$ with equation (9) gives the efficiency for families of two children; in general this is rather less than for one child, but it

never falls below 0.92. The large family efficiency, in the limit as $s \rightarrow \infty$, is obtained from equations (10) and (18) as

$$\text{Efficiency} = \frac{4}{6-3\nu+\nu^2}; \quad . \quad . \quad . \quad (20)$$

the situation here is very different from that in the absence of dominance, as Cotterman's score is not fully efficient except for $\nu = 1$, and may involve a sacrifice of as much as one-third of the information when ν is small. These three efficiency functions are shown graphically in fig. 1.

The efficiency may be increased by scoring all parents as unrelated individuals and, separately, scoring children according to a system specially appropriate to their parental phenotypes. The derivation of scores for the children requires detailed analysis of possible sibships. The probability that a parent phenotypically T shall be heterozygous, in the absence of any information on children or other relatives, is $2\nu/(1+\nu)$. Hence the probabilities of the various phenotype combinations among the children of a T × T mating are :—

Parents	No. of children		P
	T	tt	
TT × T	s	0	$1 - \frac{4\nu^2}{(1+\nu)^2}$
Tt × Tt	s-n	n	$\frac{4\nu^2}{(1+\nu)^2} \binom{s}{n} \frac{3^{s-n}}{4^s}$

If families of this type were scored with n, the total number of recessive children, the mean score would be

$$\begin{aligned} E(n) &= \frac{4\nu^2}{(1+\nu)^2} \sum_{n=1}^s n \binom{s}{n} \frac{3^{s-n}}{4^s} \\ &= \frac{s\nu^2}{(1+\nu)^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21) \end{aligned}$$

Summation of n^2 similarly leads to the variance

$$V(n) = \frac{s\nu^2}{4(1+\nu)^4} \{s(1-\nu)(1+3\nu) + 3(1+\nu)^2\} \quad . \quad (22)$$

Suppose now that ν_0^2 is a first approximation to the desired estimate, ν^2 , and that

$$\nu^2 = \nu_0^2 + \delta.$$

Define the score, y, as

$$y = -s\nu_0^2 + n(1+\nu_0)^2 \quad . \quad . \quad . \quad (23)$$

Then $E(y) = s(\nu_0^2 + \delta)$ (24)
to the first order in δ , and

$$V(y) = \frac{1}{4} s\nu^2(1+\nu)^2 \{s(1-\nu)(1+3\nu) + 3(1+\nu)^2\}.$$

Consequently the information on ν^2 provided by the score is

$$i(\nu^2) = \frac{4s}{\nu^2(1+\nu)^2\{s(1-\nu)(1+3\nu)+3(1+\nu)^2\}} \quad (25)$$

For a $T \times tt$ mating with s children, similar analysis shows that the mean number of recessive children is

$$E(n) = \frac{s\nu}{1+\nu}, \quad (26)$$

with variance

$$V(n) = \frac{s\nu}{2(1+\nu)^2}\{s(1-\nu)+(1+\nu)\} \quad (27)$$

Taking ν_0^2 as before, a score

$$y = -s\nu_0^2(1+2\nu_0) + 2n\nu_0(1+\nu_0)^2 \quad (28)$$

has expectation

$$E(y) = s(\nu_0^2 + \delta), \quad (29)$$

to the first order in δ , and variance

$$V(y) = 2s\nu^3(1+\nu)^2\{s(1-\nu)+(1+\nu)\}.$$

The information on ν^2 given by this score is therefore

$$i(\nu^2) = \frac{s}{2\nu^3(1+\nu)^2\{s(1-\nu)+(1+\nu)\}} \quad (30)$$

A mating of $tt \times tt$ can produce only tt children, and these clearly provide no additional information on gene frequencies. Fisher (1940) has pointed out that a scoring of all families of two recorded parents with an average weight has the apparently irrational effect of increasing the weight assigned to the record of a $tt \times tt$ mating for each additional child, in spite of the fact that all children must be tt . The objection is not serious, but it is perhaps preferable to score these children as zero with zero weight. For the remaining families, the mean information per family is :

$$i(\nu^2) = \frac{1-\nu^2}{1+\nu^2} \times \text{Information for } (T \times T) + \frac{2\nu^2}{1+\nu^2} \times \text{Information for } (T \times tt)$$

Hence the mean weight per individual scored in the sibships is

$$\left. \begin{aligned} \text{Parents } T, T, \text{ or } T, tt: w &= \frac{4(1-\nu)^2}{(1+\nu^2)\{s(1-\nu)(1+3\nu)+3(1+\nu)^2\}} \\ &+ \frac{\nu(1-\nu)}{(1+\nu)(1+\nu^2)\{s(1-\nu)+(1+\nu)\}} \end{aligned} \right\} \quad (31)$$

Parents tt, tt : $w = 0$

and the scores are

$$\left. \begin{aligned} \text{Parents } T, T: & y = -s\nu^3 + n(1+\nu)^3 \\ \text{Parents } T, tt: & y = -s\nu^2(1+2\nu) + 2n\nu(1+\nu)^2 \\ \text{Parents } tt, tt: & y = 0 \end{aligned} \right\} \quad (32)$$

where s is the total number of children, of whom n are recessive. The various functions required for scoring with any trial value of ν^2 are given in table 2, and the weight per sib, w , is tabulated against ν^2 and s in table 3.

TABLE 2

Functions required in forming most efficient linear scores for children of two recorded parents

ν^2	ν^4	$(1 + \nu)^2$	$\nu^2(1 + 2\nu)$	$2\nu(1 + \nu)^2$
0.00	0.0000	1.0000	0.0000	0.0000
0.05	0.0112	1.8320	0.0724	0.6696
0.10	0.0316	2.2803	0.1632	1.0957
0.15	0.0581	2.6700	0.2662	1.4908
0.20	0.0894	3.0311	0.3789	1.8733
0.25	0.1250	3.3750	0.5000	2.2500
0.30	0.1643	3.7075	0.6286	2.6241
0.35	0.2071	4.0319	0.7641	2.9973
0.40	0.2530	4.3504	0.9060	3.3709
0.45	0.3019	4.6643	1.0537	3.7454
0.50	0.3536	4.9749	1.2071	4.1213
0.55	0.4079	5.2828	1.3658	4.4990
0.60	0.4648	5.5886	1.5295	4.8787
0.65	0.5240	5.8927	1.6981	5.2605
0.70	0.5857	6.1956	1.8713	5.6446
0.75	0.6495	6.4976	2.0490	6.0311
0.80	0.7155	6.7988	2.2311	6.4199
0.85	0.7837	7.0995	2.4173	6.8112
0.90	0.8538	7.3999	2.6076	7.2050
0.95	0.9259	7.7000	2.8019	7.6012
1.00	1.0000	8.0000	3.0000	8.0000

Scores are calculated from s , the total number of sibs, and n , the number of recessive sibs, according to equations (32).

On an average, the total weight derived from parents and children in these families is

$$W_{2,M}(s) = 2 + sw(1 - \nu^4) = 2 + \frac{4s(1 + \nu)(1 - \nu)^3}{\{s(1 - \nu)(1 + 3\nu) + 3(1 + \nu)^2\}} + \frac{s\nu(1 - \nu)^2}{s(1 - \nu) + (1 + \nu)}. \quad (33)$$

When $s = 1$, this expression reduces to the maximum likelihood weight, equation (7), so that the scores, though different from those of table 1, are just as good and have the advantage of belonging to a system applicable to any value of s . When $s = 2$

$$W_{2,M}(2) = \frac{2(27 + 2\nu - 11\nu^2 + 9\nu^3 - 4\nu^4 + \nu^5)}{(3 - \nu)(5 + 10\nu - 3\nu^2)}, \quad (34)$$

and comparison with equation (9) gives the efficiency shown graphically in fig. 1, always a little higher than Cotterman's. As $s \rightarrow \infty$, the

TABLE 3

Weight per sib for most efficient linear scoring of children of two recorded parents

$\frac{s}{\nu^2}$	1	2	3	4	5	6	7	8
0.00	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444	0.4000	0.3636
0.05	0.4642	0.3727	0.3120	0.2684	0.2356	0.2100	0.1895	0.1726
0.10	0.3350	0.2719	0.2293	0.1984	0.1749	0.1564	0.1415	0.1292
0.15	0.2583	0.2120	0.1801	0.1567	0.1387	0.1245	0.1130	0.1034
0.20	0.2056	0.1706	0.1460	0.1278	0.1137	0.1024	0.0932	0.0855
0.25	0.1667	0.1398	0.1206	0.1062	0.0949	0.0858	0.0783	0.0720
0.30	0.1367	0.1159	0.1008	0.0893	0.0801	0.0727	0.0666	0.0614
0.35	0.1128	0.0968	0.0848	0.0756	0.0682	0.0621	0.0571	0.0528
0.40	0.0935	0.0811	0.0717	0.0643	0.0583	0.0533	0.0492	0.0456
0.45	0.0775	0.0680	0.0606	0.0547	0.0499	0.0459	0.0424	0.0395
0.50	0.0641	0.0569	0.0512	0.0465	0.0427	0.0394	0.0366	0.0342
0.55	0.0528	0.0474	0.0431	0.0394	0.0364	0.0338	0.0315	0.0296
0.60	0.0432	0.0392	0.0360	0.0332	0.0308	0.0288	0.0270	0.0254
0.65	0.0349	0.0321	0.0297	0.0276	0.0258	0.0243	0.0229	0.0217
0.70	0.0278	0.0258	0.0241	0.0226	0.0213	0.0202	0.0191	0.0182
0.75	0.0215	0.0202	0.0191	0.0181	0.0172	0.0164	0.0156	0.0150
0.80	0.0161	0.0153	0.0146	0.0140	0.0134	0.0128	0.0124	0.0119
0.85	0.0113	0.0109	0.0105	0.0101	0.0098	0.0095	0.0092	0.0089
0.90	0.0071	0.0069	0.0067	0.0066	0.0064	0.0063	0.0061	0.0060
0.95	0.0033	0.0033	0.0032	0.0032	0.0032	0.0031	0.0031	0.0030
1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\frac{s}{\nu^2}$	9	10	11	12	13	14	15	16
0.00	0.3333	0.3077	0.2857	0.2667	0.2500	0.2353	0.2222	0.2105
0.05	0.1585	0.1465	0.1362	0.1273	0.1195	0.1125	0.1064	0.1009
0.10	0.1189	0.1101	0.1025	0.0959	0.0901	0.0850	0.0804	0.0763
0.15	0.0953	0.0884	0.0825	0.0773	0.0727	0.0686	0.0650	0.0617
0.20	0.0790	0.0734	0.0685	0.0643	0.0606	0.0572	0.0543	0.0516
0.25	0.0667	0.0621	0.0581	0.0546	0.0514	0.0487	0.0462	0.0439
0.30	0.0570	0.0532	0.0498	0.0469	0.0443	0.0419	0.0398	0.0379
0.35	0.0491	0.0459	0.0431	0.0406	0.0384	0.0364	0.0347	0.0330
0.40	0.0425	0.0399	0.0375	0.0354	0.0335	0.0319	0.0303	0.0290
0.45	0.0370	0.0347	0.0327	0.0310	0.0294	0.0279	0.0266	0.0255
0.50	0.0321	0.0302	0.0286	0.0271	0.0258	0.0245	0.0234	0.0224
0.55	0.0278	0.0263	0.0249	0.0237	0.0226	0.0215	0.0206	0.0198
0.60	0.0240	0.0228	0.0216	0.0206	0.0197	0.0188	0.0181	0.0174
0.65	0.0205	0.0195	0.0186	0.0178	0.0171	0.0164	0.0157	0.0151
0.70	0.0173	0.0166	0.0159	0.0152	0.0146	0.0141	0.0136	0.0131
0.75	0.0143	0.0138	0.0132	0.0128	0.0123	0.0119	0.0115	0.0111
0.80	0.0115	0.0111	0.0107	0.0104	0.0100	0.0097	0.0095	0.0092
0.85	0.0087	0.0084	0.0082	0.0080	0.0078	0.0076	0.0074	0.0072
0.90	0.0059	0.0058	0.0056	0.0055	0.0054	0.0053	0.0052	0.0051
0.95	0.0030	0.0030	0.0029	0.0029	0.0029	0.0029	0.0028	0.0028
1.00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

These weights are to be used only when at least one parent is T; when both parents are #, the weight is zero.

weight approaches the same limit as that in equation (10); thus for large families the score tends to full efficiency, and at small values of ν is therefore much better than Cotterman's.

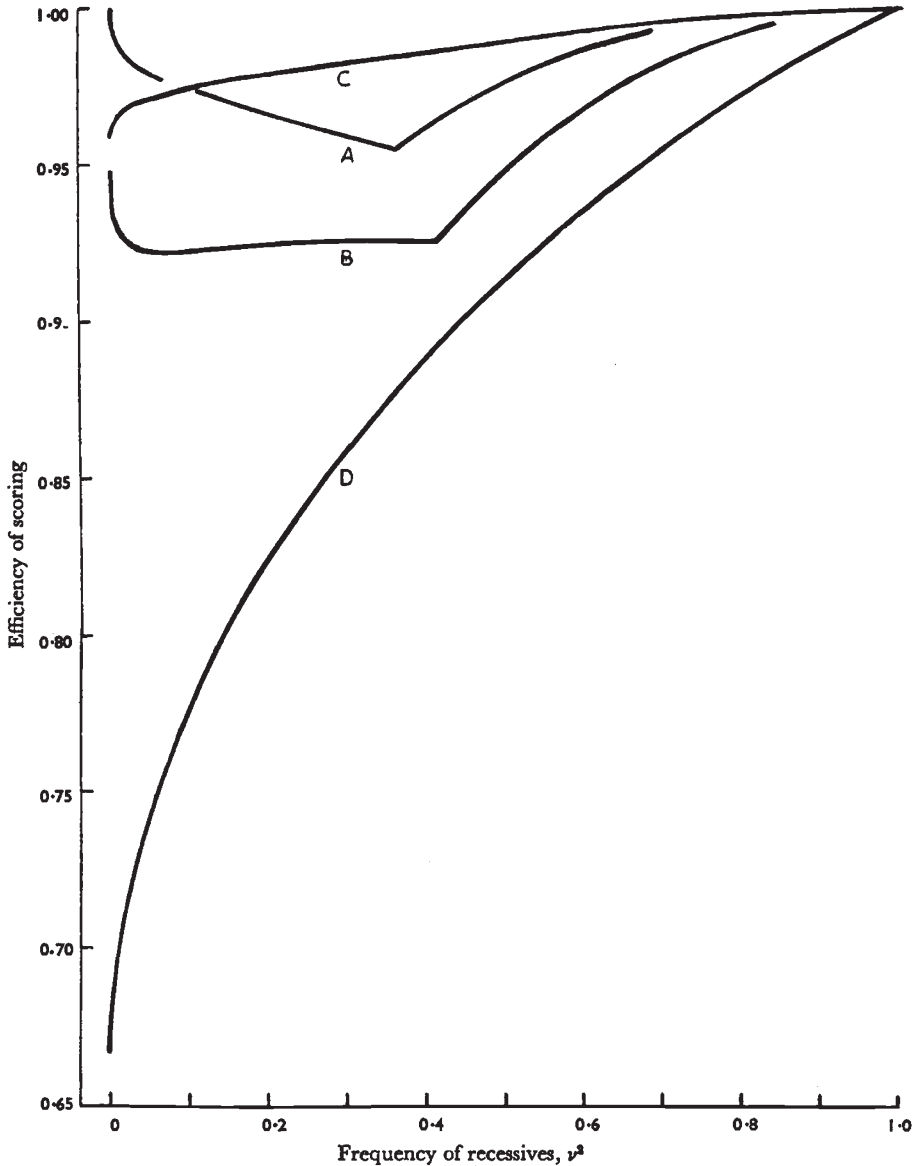


FIG. 1.—Efficiencies of Cotterman's scores and of most efficient linear scores, for families with two recorded parents, type (i).

- Curve A : Cotterman, $s = 1$;
- Curve B : Cotterman, $s = 2$;
- Curve C : Most efficient linear, $s = 2$;
- Curve D : Cotterman, $s \rightarrow \infty$.

(For $s = 1$ or $s \rightarrow \infty$, the most efficient linear score is fully efficient.)

(ii) *One parent recorded*

Cotterman's score for a family record of one parent and s children, the total number of recessives amongst the $(s+1)$ individuals, has weight

$$W_{1,c(s)} = \frac{4(s+1)^2(1+\nu)}{(s^2+3s+4) + \nu(3s^2+9s+4)} \quad (35)$$

This weight is always greater than 1 : scoring of the children is always better than discarding them and scoring only the parent. Cotterman points out that, for parent-child pairs, his weight

$$W_{1,c(1)} = \frac{2(1+\nu)}{1+2\nu}$$

shows an efficiency compared with that for maximum likelihood scoring, equation (12),

$$\text{Efficiency} = \frac{8(1+\nu-\nu^2)}{(1+2\nu)(8-7\nu+2\nu^2)}, \quad (36)$$

which expression reaches 1 for $\nu = 0$ or $\nu = \frac{1}{2}$, but falls to $8/9$ at $\nu = 1$. Insertion of $s = 2$ in equation (35) and comparison with equation (13) gives an efficiency ranging from $72/77$ at $\nu = 0$ to $6/7$ at $\nu = 1$, with rather higher intermediate values. For large s , comparison with equation (14) shows an efficiency of $4/5$ at $\nu = 0$ rising to 1 at $\nu = 1$. These efficiencies are illustrated in fig. 2.

Again the information can be increased by scoring the one parent as an unrelated individual and forming a y -score for the s sibs after separate examination of the possible progeny of a T and of a tt parent. If the recorded parent is T, the probability of which is $(1-\nu^2)$, the various possibilities for the other parent and the s children have the probabilities :—

Parents		No. of children		P
Recorded	Unrecorded	T	tt	
T	TT	s	0	$\frac{2\mu\nu(1-\nu)}{1+\nu}$
TT	Tt	s	0	
Tt	Tt	$s-n$	n	$\frac{4\mu\nu^2}{1+\nu} \binom{s}{n} \frac{3^{s-n}}{4^n}$
TT	tt	s	0	$\frac{\nu^2(1-\nu)}{1+\nu}$
Tt	tt	$s-n$	n	$\frac{2\nu^3}{1+\nu} \binom{s}{n} 2^{-s}$

From these probabilities are found the mean and variance of the number of recessive children

$$E(n) = \frac{s\nu^2}{1+\nu} \quad (37)$$

$$V(n) = \frac{s\nu^2}{4(1+\nu)^2} \{s(1-\nu)(1+3\nu) + (1+\nu)(3-\nu)\} \quad (38)$$

Similarly, for families having the recorded parent *tt*, the probability of which is ν^2 ,

$$E(n) = s\nu, \quad (39)$$

$$V(n) = \frac{1}{2}\mu\nu s(s+1). \quad (40)$$

Again taking ν_0^2 as a first approximation to ν^2 , the number of recessives in either type of family may be used to give an estimate of δ , a first

TABLE 4

Functions required in forming most efficient linear scores for children of one recorded parent

ν^2	$\frac{\nu^3}{2+\nu}$	$\frac{2(1+\nu)^2}{2+\nu}$	2ν
0.00	0.0000	1.0000	0.0000
0.05	0.0050	1.3467	0.4472
0.10	0.0137	1.4959	0.6325
0.15	0.0243	1.6124	0.7746
0.20	0.0365	1.7117	0.8944
0.25	0.0500	1.8000	1.0000
0.30	0.0645	1.8805	1.0954
0.35	0.0799	1.9549	1.1832
0.40	0.0961	2.0247	1.2649
0.45	0.1130	2.0905	1.3416
0.50	0.1306	2.1530	1.4142
0.55	0.1488	2.2127	1.4832
0.60	0.1675	2.2700	1.5492
0.65	0.1867	2.3252	1.6125
0.70	0.2065	2.3784	1.6733
0.75	0.2266	2.4299	1.7320
0.80	0.2472	2.4798	1.7889
0.85	0.2682	2.5284	1.8439
0.90	0.2896	2.5756	1.8974
0.95	0.3113	2.6217	1.9494
1.00	0.3333	2.6667	2.0000

Scores are calculated from *s*, the total number of sibs, and *n*, the number of recessive sibs, according to equations (41).

order adjustment to ν_0^2 . As before, the procedure may conveniently be put into the form of a score, *y*, which will lead directly to a revised estimate, ν^2 . The scores are :—

$$\left. \begin{aligned} \text{Recorded parent T : } y &= -\frac{s\nu^3}{2+\nu} + \frac{2n(1+\nu)^2}{2+\nu} \\ \text{Recorded parent tt : } y &= -s\nu^2 + 2n\nu, \end{aligned} \right\} \quad (41)$$

with a mean weight per sib scored, derived from equations (38), (40), of

$$w = \frac{(1-\nu)^2(2+\nu)^2}{s(1-\nu)(1+3\nu)+(1+\nu)(3-\nu)} + \frac{\nu(1+\nu)}{2(s+1)} \quad (42)$$

Table 4 (page 381) gives the functions required in scoring with a trial value of ν^2 , and w is tabulated in table 5.

The total weight per family, including the parental contribution, is

$$W_{1, M}(s) = 1 + sw \\ = 1 + \frac{s(1-\nu)^2(2+\nu)^2}{s(1-\nu)(1+3\nu)+(1+\nu)(3-\nu)} + \frac{s\nu(1+\nu)}{2(s+1)} \quad (43)$$

This expression is necessarily at least as great as $W_{1, C}(s)$ for any combination of s and ν . For parent-child pairs ($s = 1$), it reduces to $W_{1, L}(1)$, equation (12), showing the scoring then to be fully efficient. For $s = 2$, comparison with equation (13) shows the scoring to be more than 0.95 efficient (except when ν is very small), and therefore considerably more efficient than the simple count for large ν . As $s \rightarrow \infty$, the modified score does not tend to full efficiency, but comparison with equation (14) shows that

$$\text{Efficiency } (y) = \frac{(1+2\nu-2\nu^2)(10-3\nu+\nu^2)}{2(1+3\nu)(5-6\nu+3\nu^2-\nu^3)},$$

an expression which reaches unity at $\nu = 0, 0.5$ and 1 , and never falls below 0.994. In large families, the efficiency is thus almost complete, a great improvement on that of the simple count, especially for small ν . These efficiencies are also shown in fig. 2.

(iii) Neither parent recorded

For a record relating only to s sibs, no sub-classification can increase the information provided by Cotterman's score ; this is

$$y = n, \quad (44)$$

the number of recessives, and

$$E(y) = s\nu^2. \quad (45)$$

Cotterman has shown that the weight per sib is

$$w = \frac{4(1+\nu)}{s(1+3\nu)+(3+\nu)} \quad (46)$$

a function which he has tabulated against ν , and which is shown in table 6 tabulated against ν^2 . The total weight per sibship is

$$W_{0, C}(s) = \frac{4^s(1+\nu)}{s(1+3\nu)+(3+\nu)} \quad (47)$$

By comparison with equation (16), Cotterman has shown that when $s = 2$ the efficiency rises to unity at about $\nu^2 = 0.4$, but that for large or small ν it falls to about 0.9. Though no detailed examination has been made, there is reason to suspect that as s is increased the

TABLE 5

Weight per sib for most efficient linear scoring of children of one recorded parent

$\frac{s}{\nu^2}$	1	2	3	4	5	6	7	8
0.00	1.0000	0.8000	0.6667	0.5714	0.5000	0.4444	0.4000	0.3636
0.05	0.7033	0.5430	0.4431	0.3745	0.3244	0.2861	0.2560	0.2316
0.10	0.6197	0.4741	0.3851	0.3247	0.2807	0.2473	0.2211	0.1999
0.15	0.5666	0.4306	0.3487	0.2935	0.2536	0.2233	0.1995	0.1803
0.20	0.5286	0.3991	0.3224	0.2710	0.2339	0.2059	0.1839	0.1662
0.25	0.5000	0.3750	0.3021	0.2536	0.2188	0.1925	0.1719	0.1553
0.30	0.4780	0.3559	0.2858	0.2396	0.2065	0.1816	0.1621	0.1465
0.35	0.4610	0.3406	0.2726	0.2281	0.1964	0.1726	0.1541	0.1392
0.40	0.4480	0.3282	0.2617	0.2185	0.1879	0.1650	0.1472	0.1330
0.45	0.4385	0.3184	0.2527	0.2104	0.1807	0.1586	0.1414	0.1276
0.50	0.4320	0.3107	0.2453	0.2037	0.1747	0.1531	0.1364	0.1231
0.55	0.4282	0.3049	0.2395	0.1982	0.1696	0.1484	0.1321	0.1191
0.60	0.4269	0.3009	0.2349	0.1938	0.1654	0.1445	0.1285	0.1157
0.65	0.4280	0.2986	0.2317	0.1903	0.1620	0.1413	0.1254	0.1128
0.70	0.4314	0.2980	0.2298	0.1879	0.1594	0.1387	0.1229	0.1104
0.75	0.4370	0.2991	0.2292	0.1865	0.1577	0.1369	0.1210	0.1086
0.80	0.4449	0.3020	0.2299	0.1863	0.1569	0.1358	0.1198	0.1073
0.85	0.4551	0.3067	0.2322	0.1873	0.1572	0.1356	0.1193	0.1066
0.90	0.4676	0.3133	0.2361	0.1897	0.1587	0.1365	0.1198	0.1068
0.95	0.4826	0.3221	0.2419	0.1938	0.1617	0.1387	0.1215	0.1081
1.00	0.5000	0.3333	0.2500	0.2000	0.1667	0.1429	0.1250	0.1111
$\frac{s}{\nu^2}$	9	10	11	12	13	14	15	16
0.00	0.3333	0.3077	0.2857	0.2667	0.2500	0.2353	0.2222	0.2105
0.05	0.2114	0.1945	0.1801	0.1677	0.1569	0.1474	0.1390	0.1314
0.10	0.1824	0.1677	0.1552	0.1445	0.1351	0.1269	0.1197	0.1132
0.15	0.1645	0.1512	0.1400	0.1303	0.1218	0.1144	0.1079	0.1020
0.20	0.1516	0.1394	0.1290	0.1201	0.1123	0.1055	0.0994	0.0940
0.25	0.1417	0.1302	0.1205	0.1122	0.1049	0.0985	0.0929	0.0878
0.30	0.1336	0.1228	0.1137	0.1058	0.0989	0.0929	0.0876	0.0829
0.35	0.1269	0.1167	0.1080	0.1005	0.0940	0.0883	0.0832	0.0787
0.40	0.1212	0.1114	0.1031	0.0960	0.0897	0.0843	0.0795	0.0752
0.45	0.1164	0.1069	0.0989	0.0921	0.0861	0.0808	0.0762	0.0721
0.50	0.1121	0.1030	0.0953	0.0887	0.0829	0.0779	0.0734	0.0694
0.55	0.1085	0.0996	0.0921	0.0857	0.0801	0.0752	0.0709	0.0671
0.60	0.1053	0.0967	0.0894	0.0831	0.0777	0.0729	0.0687	0.0650
0.65	0.1026	0.0941	0.0870	0.0808	0.0755	0.0709	0.0668	0.0631
0.70	0.1003	0.0920	0.0849	0.0789	0.0737	0.0691	0.0651	0.0615
0.75	0.0985	0.0902	0.0832	0.0773	0.0721	0.0676	0.0636	0.0601
0.80	0.0972	0.0889	0.0819	0.0760	0.0708	0.0664	0.0625	0.0590
0.85	0.0964	0.0880	0.0810	0.0751	0.0699	0.0655	0.0616	0.0581
0.90	0.0964	0.0879	0.0807	0.0747	0.0695	0.0650	0.0611	0.0576
0.95	0.0974	0.0886	0.0813	0.0751	0.0698	0.0652	0.0612	0.0576
1.00	0.1000	0.0909	0.0833	0.0769	0.0714	0.0667	0.0625	0.0588

TABLE 6

Weight per sib for most efficient linear scoring of children of unrecorded parents

$\frac{s}{p^2}$	2	3	4	5	6	7	8	9
0.00	0.8000	0.6667	0.5714	0.5000	0.4444	0.4000	0.3636	0.3333
0.05	0.7455	0.5943	0.4940	0.4227	0.3694	0.3281	0.2950	0.2680
0.10	0.7299	0.5746	0.4738	0.4031	0.3508	0.3105	0.2785	0.2525
0.15	0.7196	0.5621	0.4611	0.3909	0.3392	0.2996	0.2683	0.2429
0.20	0.7120	0.5528	0.4518	0.3820	0.3308	0.2918	0.2610	0.2361
0.25	0.7059	0.5455	0.4444	0.3750	0.3243	0.2857	0.2553	0.2308
0.30	0.7008	0.5394	0.4384	0.3693	0.3190	0.2808	0.2507	0.2265
0.35	0.6965	0.5343	0.4334	0.3645	0.3145	0.2766	0.2469	0.2229
0.40	0.6927	0.5298	0.4290	0.3604	0.3107	0.2731	0.2435	0.2198
0.45	0.6893	0.5259	0.4251	0.3568	0.3073	0.2699	0.2407	0.2171
0.50	0.6863	0.5224	0.4217	0.3536	0.3044	0.2672	0.2381	0.2147
0.55	0.6836	0.5193	0.4186	0.3507	0.3017	0.2647	0.2358	0.2126
0.60	0.6811	0.5164	0.4158	0.3481	0.2993	0.2625	0.2338	0.2107
0.65	0.6788	0.5138	0.4133	0.3457	0.2971	0.2605	0.2319	0.2090
0.70	0.6767	0.5114	0.4110	0.3435	0.2951	0.2586	0.2302	0.2074
0.75	0.6747	0.5091	0.4088	0.3415	0.2932	0.2569	0.2286	0.2059
0.80	0.6729	0.5071	0.4068	0.3396	0.2915	0.2553	0.2271	0.2046
0.85	0.6712	0.5051	0.4049	0.3379	0.2899	0.2539	0.2258	0.2033
0.90	0.6696	0.5033	0.4032	0.3363	0.2884	0.2525	0.2245	0.2021
0.95	0.6681	0.5016	0.4015	0.3348	0.2870	0.2512	0.2233	0.2010
1.00	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500	0.2222	0.2000

$\frac{s}{p^2}$	10	11	12	13	14	15	16
0.00	0.3077	0.2857	0.2667	0.2500	0.2353	0.2222	0.2105
0.05	0.2456	0.2266	0.2103	0.1962	0.1839	0.1730	0.1634
0.10	0.2309	0.2127	0.1972	0.1838	0.1721	0.1618	0.1526
0.15	0.2219	0.2043	0.1892	0.1762	0.1649	0.1549	0.1461
0.20	0.2155	0.1982	0.1835	0.1708	0.1598	0.1501	0.1415
0.25	0.2105	0.1935	0.1791	0.1667	0.1558	0.1463	0.1379
0.30	0.2065	0.1898	0.1755	0.1633	0.1527	0.1433	0.1351
0.35	0.2031	0.1866	0.1726	0.1605	0.1500	0.1408	0.1327
0.40	0.2003	0.1839	0.1700	0.1581	0.1477	0.1387	0.1306
0.45	0.1978	0.1816	0.1678	0.1560	0.1458	0.1368	0.1288
0.50	0.1955	0.1795	0.1659	0.1542	0.1440	0.1351	0.1273
0.55	0.1936	0.1776	0.1641	0.1526	0.1425	0.1337	0.1259
0.60	0.1918	0.1760	0.1626	0.1511	0.1411	0.1324	0.1246
0.65	0.1902	0.1745	0.1612	0.1497	0.1398	0.1312	0.1235
0.70	0.1887	0.1731	0.1599	0.1485	0.1387	0.1301	0.1225
0.75	0.1873	0.1718	0.1587	0.1474	0.1376	0.1291	0.1215
0.80	0.1861	0.1706	0.1576	0.1464	0.1366	0.1281	0.1206
0.85	0.1849	0.1695	0.1565	0.1454	0.1357	0.1273	0.1198
0.90	0.1838	0.1685	0.1556	0.1445	0.1349	0.1265	0.1190
0.95	0.1828	0.1676	0.1547	0.1436	0.1341	0.1257	0.1183
1.00	0.1818	0.1667	0.1538	0.1429	0.1333	0.1250	0.1176

Unrelated individuals, including parents of recorded children as explained earlier, are scored as 0 if T, 1 if *tt*, with $w = 1, s = 1$ in the formula. It is quite legitimate to score certain sizes of families with maximum likelihood x -scores, where tables for these are available,

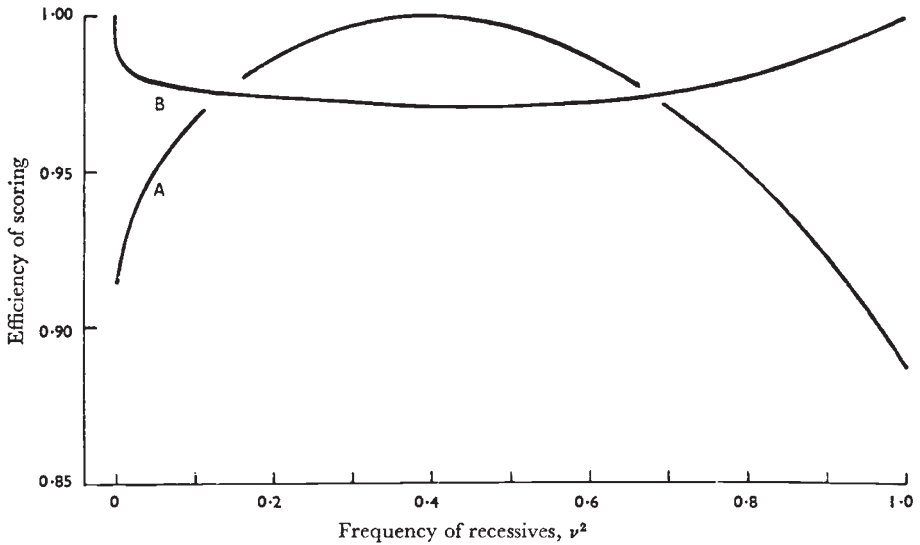


FIG. 3.—Efficiencies of Cotterman's or most efficient linear scores, for families with no recorded parent, type (iii).
 Curve A : $s = 2$;
 Curve B : $s \rightarrow \infty$.

in which case x replaces the corresponding wy in the numerator of equation (47), W_L replaces the corresponding ws in the denominator. If ν^2 differs much from ν_0^2 , the whole data should be re-scored using ν^2 as a provisional estimate. When the final estimate has been formed, its variance is

$$V(\nu^2) = \frac{\nu^2(1 - \nu^2)}{\sum ws} \dots \dots \dots (49)$$

An example follows in section 4.

4. NUMERICAL EXAMPLE

Boyd and Boyd (1941) record presence or absence of the ability to taste phenyl-thiocarbamide for 58 families of Armenians from Ghazir and Beyrouth (Syria). Of these families, 5 have both parents classified, 14 have one parent, and 39 consist of sibs only (in 3 instances, only one sib being classified). The factor t for taste-blindness is generally considered to be inherited as a simple recessive, and the records of the 58 families are consistent with this hypothesis. On the assumption that the population studied was mating at random

in respect of the taste factor, the frequency of the *t* gene may be estimated by the methods of section 3. The first stage of the calculations is shown in table 7.

TABLE 7

Estimation of frequency of t gene from records of 58 Armenian families

(a) Parents

Frequencies of		Sum of values of	
T	<i>tt</i>	W	<i>x</i>
20	4	24	4

(b) Children (scored for $\nu_0^2 = 0.25$)

<i>s</i>	Recorded parents	No. of families	Numbers of children		Sum of values of		
			T	<i>tt</i>	<i>w</i>	<i>y</i>	<i>s</i>
2	T, T	1	2	0	0.140	-0.25	2
	T, <i>tt</i>	1	2	0	0.140	-1.00	2
3	T, T	1	3	0	0.121	-0.38	3
	T, <i>tt</i>	1	2	1	0.121	0.75	3
4	T, T	1	4	0	0.106	-0.50	4
1	T	2	1	1	0.500	1.70	2
	<i>tt</i>	2	1	1	0.500	0.50	2
2	T	7	9	5	0.375	8.30	14
3	T	3	5	4	0.302	6.75	9
1	none	3	1	2	1.000	2.00	3
2	none	26	38	14	0.706	14.00	52
3	none	8	21	3	0.546	3.00	24
4	none	2	7	1	0.444	1.00	8

Part (a) of table 7 shows the phenotypes of all recorded parents for families of types (i) or (ii) (sections 2 and 3). In part (b), the phenotypes of the children are classified according to parental types and numbers of sibs; the three sibs with no recorded parents and *s* = 1, of course, might have been grouped with the parents as unrelated individuals. Of a total of 152 persons recorded, 36, or 24 per cent., are recessive non-tasters. Scoring with a trial value $\nu_0^2 = 0.25$, therefore seems appropriate. For the parents, the maximum likelihood *x*-score is simply the total number of recessives, and the weight to be attached to the score is the total number of persons. The most efficient linear scores in part (b) of the table are constructed as described in section 3. For example, the *y*-score for

the family of 3 sibs with parents T, *tt* is taken from the second of equations (32), which, using table 2 for $\nu^2 = 0.25$, gives

$$\begin{aligned} y &= -3 \times 0.500 + 1 \times 2.250 \\ &= 0.75. \end{aligned}$$

The weight per sib is read from table 3, in the column for $s = 3$ and the row for $\nu^2 = 0.25$, as $w = 0.121$. Similarly, the total score for the 7 families of 2 sibs with one T parent and the other parent unrecorded is obtained from the first of equations (41), with the aid of table 4, as

$$\begin{aligned} y &= -14 \times 0.050 + 5 \times 1.800 \\ &= 8.30, \end{aligned}$$

with a weight per sib $w = 0.375$ (from table 5). By summation over the whole of part (b) of table 7

$$\begin{aligned} \Sigma wy &= 20.034, \\ \Sigma ws &= 68.046. \end{aligned}$$

Therefore the revised estimate of the population frequency of recessives is

$$\begin{aligned} \nu^2 &= \frac{4 + 20.034}{24 + 68.046} \\ &= \frac{24.034}{92.046} \\ &= 0.2611, \end{aligned}$$

a result so close to the provisional ν_0^2 that obviously no re-scoring is needed. The variance of the estimate is

$$\begin{aligned} V(\nu^2) &= \frac{0.261 \times 0.739}{92.046} \\ &= 0.00210. \end{aligned}$$

The conclusion is that in the population studied 26.1 per cent. \pm 4.6 per cent. of persons are non-tasters, an estimate which corresponds to a frequency 0.511 for the *t* gene.

5. SUMMARY

Most efficient linear scores have been developed for the estimation of the frequency of recessives, in a population mating at random for a factor involving two allelomorphous genes which show simple dominance. The general theory of estimation by iterative scoring systems has been applied in order to obtain formulæ for these scores, and for families of sibs having two, one, or no recorded parents. The method is more efficient than Cotterman's (1947) estimation from a simple count of recessives, the gain being greater than in the absence of dominance; it is, of course, in general less efficient than

maximum likelihood estimation, but the latter is scarcely practicable as a routine method except for very small families.

With the aid of tables given in section 3, the method may be used almost as simply and expeditiously as Cotterman's. An example in section 4 shows its application to the estimation of the frequency of phenyl-thiocarbamide taste-blindness from a sample of 58 Armenian families; the estimate formed is that 26.1 per cent. \pm 4.6 per cent. of the population are non-tasters.

REFERENCES

- BOYD, W. C., AND BOYD, L. S. 1941.
Data for testing for genetic linkage on 500 pairs of sibs.
Ann. Eugen., Lond. 11, 1-9.
- COTTERMAN, C. W. 1947.
A weighting system for the estimation of gene frequencies from family records.
Cont. Lab. Vert. Biol., Univ. Michigan 33.
- FINNEY, D. J. 1948.
The estimation of gene frequencies from family records. I. Factors without dominance.
Heredity 2, 199-218.
- FISHER, R. A. 1940.
The estimation of the proportion of recessives from tests carried out on a family not wholly unrelated.
Ann. Eugen., Lond. 10, 160-170.