The cumulants of the distributions of Fisher's "u_{11}" and "u_{31}" scores used in the detection and estimation of ~~partial sex~~ linkage in man.

By J. B. S. Haldane, F.R.S.

The detection of linkage in man, other than complete sex-linkage or linkage between sex-linked genes ~~in man~~, almost invariably involves statistical methods which have been gradually developed since the pioneer work of Bernstein (1931). Wherever one of the genes concerned is an autosomal or partially sex-linked recessive, Fisher's "u" score is the method of choice, and by far the most important is the u_{31} score. For example this is invariably used in the case of partial sex-linkage, and in the case of an autosomal recessive Finney (1940) used u_{31} in 43 out of 55 of ~~all~~ the families analysed for linkage with the ~~new~~ blood-group genes, and in 38 out of 40 ~~all but one~~ of those tested for linkage with M and N. As the data on partial sex linkage are at present more extensive and much more significant of linkage than those on autosomal linkage we shall be mainly concerned with them. It will be demonstrated ~~shown~~ that ~~a prec~~ a new step in the precision of the method shows that the evidence for partial sex linkage is somewhat less decisive than appeared, but still cogent. However in ~~some other~~ one case~~s~~ the significance of the evidence for an autosomal linkage becomes doubtful.

Fisher's (1935 a, 1935 b) original methods for the detection and estimation of linkage assumed, either that the probability of recording a family containing recessives was independent of the number of recessives in it, or that it was proportional to that number. In either case he used a score

$$u_{31} = (a-3b-c+3d)^2 - (a+9b+c+9d).$$

where (in the case of suspected partial sex-linkage) a sibship consists of a normal males, b affected males, c normal females, and d affected females. However in the data on which Haldane (1936) inferred the existence of partial sex-linkage, neither of ~~these~~ the above conditions was fulfilled. Fisher (1936) therefore considered the case where a sibship consisted of s_1 normals and s_2 abnormals. In what follows I shall use N to denote s_1 and n to denote s_2, since in my experience suffixes render algebra hard for the average biologist to follow.

Fisher did not consider any moments of the distribution of u beyond the second, and assumed that a positive value of S(u) from a group of sibships is to be regarded as significant according to the ratio which it bears to ~~its~~ ~~the square slan~~ its standard sampling error in the absence of linkage; in fact that the sampling distribution of S(u) may be taken as normal. We shall see that this is far from being the case for some of the groups of sibships on which the evidence for linkage is based. Of course when more data accumulate, the distribution will become more nearly normal, but at the same time the need for critical tests of significance will be lessened. It is unfortunately just where the data are on the borderline of significance that the present criteria are least satisfactory. Hence, just as in the case of the coefficient of correlation, a ~~device for approx~~ transformation which will approximately normalize the distribution of S(u) is ~~much to be desired~~ desirable.

To study the deviation from normality we must calculate one or more of the cumulants, beyond the second, of the distribution of u. Fisher gives an operational method of calculating them. However in the case of N normals and n ~~recessives~~ abnormals a more elementary method is available, although it is less generally applicable. For a and c are derived by simple sampling in a sample of N, b and d in a sample of n. Let $a - c = x$, $b - d = y$. Then $u_{31} = (x-3y)^2 - (N+9n)$.

The distribution of x and that of y for any linkage value are given by a binomial expansion. Hence their cumulants can be calculated. Further x and y are uncorrelated, provided the intensity of linkage does not vary from one family to another. Hence the cumulants of $x-3y$ can be written down, and those of $(x-3y)^2$ derived from them.

The distribution of u_{31} in the absence of linkage.

First consider the case of partial sex-linkage, where it is ~~known~~ assumed that every sibship, even if it consists of one sex only, is potentially segregating for ~~sex~~ the sexes with equal frequencies. The fact that the sex ratio is not unity introduces a slight error. If the true sex-ratio is ~~½(1+k)~~ $\frac{1}{2}(1+k)$ ♂ : $\frac{1}{2}(1-k)$ ♀, $E(u) = k^2\left[(N+3n)^2-(N+9n)\right]$ in the absence of linkage. Thus the estimate of ~~linkage~~ partial sex-linkage ~~for~~ an autosomal gene on a sufficiently vast sample would be $\chi = \frac{1}{2} - \dfrac{3k\,S\left[(N+3n)^2-(N+9n)\right]}{2S\left[(N+9n)^2-(N+81n)\right]}$, ~~that~~ that is to say the apparent recombination frequency would be less than ~~½~~ 50% by a value probably slightly greater than $\frac{k}{6}$, or 0·4%. This correction will be neglected in* The correction where a dominant test factor takes the place of sex will be considered later.

$x = 2a - N$, and the distribution of $\underline{a}$ is given by the symmetrical binomial $2^{-N}(1+\theta)^N$. The cumulant-generating function is $N \log \cosh t$. Odd cumulants vanish, and the even cumulants (Haldane 1940) are: —

$$k_2 = N,\ k_4 = -2N,\ k_6 = 2^4 N,\ k_8 = -2^4.17N,\ k_{10} = 2^8.31N,\ k_{12} = -2^9.691N, \text{~~etc~~}.$$

$$k_{2r} = (-1)^{r-1} r^{-1} B_r . 2^{2r-1}(2^{2r}-1)N.$$

* though it should be made to calculated cross-over values.

Similarly the cumulants of $3y$ are:—

$\kappa_2 = qn,\ \kappa_4 = -2.q^2n,\ \kappa_6 = 2^4.q^3n,\ \kappa_8 = -2^4.17.q^4n,\ \kappa_{10} = 2^8.31.q^5n,\ \kappa_{12} = -2^9.691.q^6n.$

So the cumulants of $x-3y$ are:

$\kappa_2 = N+qn,\ \kappa_4 = -2(N+q^2n),\ \kappa_6 = 2^4(N+q^3n),\ \kappa_8 = -2^4.17(N+q^4n),$

$\kappa_{10} = 2^8.31(N+q^5n),\ \kappa_{12} = -2^9.691(N+q^6n)$, etc.

Now the mean of u is zero. Its other cumulants are the same as those of $(x-3y)^2$. Haldane (1941) has given formulae[3] for the cumulants of the distribution of the square of a variate, in terms of the ~~dist~~ cumulants of the distribution of that variate. Thus if z is symmetrically distributed with cumulants κ_2, κ_4, etc, the fourth cumulant of the distribution of z^2 is $48\kappa_2^4 + 144\kappa_2^2\kappa_4 + 8(3\kappa_2\kappa_6 + 4\kappa_4^2) + \kappa_8$. Applying these formulae, we find for the cumulants of u_{31}:—

$\kappa_1 = 0.$

$\kappa_2 = 2\left[(N+qn)^2 - (N+q^2n)\right].$

$\kappa_3 = 8\left[(N+qn)^3 - 3(N+qn)(N+q^2n) + 2(N+q^3n)\right].$

$\kappa_4 = 16\left[3(N+qn)^4 - 18(N+qn)^2(N+q^2n) + 8\{3(N+qn)(N+q^3n) + (N+q^2n)^2\} - 17(N+q^4n)\right].$

~~$= 16\left[3(N+qn)^4 - 18(N+qn)^2(N+q^2n) + 32(N^2 + 66.qNn + q^4n^2) - 17(N+q^4n)\right].$~~

~~$\kappa_5 = 128\left[3(N+qn)^5 - 30(N+qn)^3(N+q^2n) + 20\{3(N+qn)(N+q^3n) + 2(N+q^2n)^2\}\right.$~~

~~$\left. - 5\{17(N+qn)(N+q^4n) + 10(N+q^2n)(N+q^3n)\} + 62(N+q^5n)\right]$~~

~~$= 128\left[3(N+qn)^5 - 30(N+qn)^3(N+q^2n) + 20(5N^2 + 282.qNn + 5.q^4n^2)\right.$~~

~~$-5($~~

1

$N + 9x = a$

$N + 9^2 x = a + b$

$N + 9^3 x = a + 10b$

$N + 9^4 x = a + 91b$

$N + 9^5 x = a + 820b$

$N + 9^6 x = a + 7381b$

$\frac{1}{72}(8561\frac{1}{4} - 4) = \frac{1}{8}.428$

$1 + 10$

$1 + 2 + 1$

91 / 729 / 820 / 6561 / 7381

$7 + x + x^2 + x^3 + x^4$

$$K_6 = \tfrac{1}{2}\Big[15a^6 - 225a^4(a+b) + 600a^2\{a(a+10b) + (a+b)^2\} - 15\{85a^2(a+91b) + 100a(a+b)(a+10b) + 11(a+b)^3\}$$
$$+ 4\{465a(a+820b) + 255(a+b)(a+91b) + 113(a+10b)^2\} - 1382(a+7381b)\Big]$$

$$= \tfrac{1}{2}\Big[15a^6 - 225a^4(a+b) + 600a^2(2a^2 + 12ab + b^2) - 15(196a^3 + 8868a^2b + 1033ab^2 + 11b^3)$$
$$+ 4(833a^2 + 407{,}020ab + 34{,}505b^2) - 1382(a + 7381b)\Big]$$

$$= \tfrac{1}{2}\Big[15a^6 - 225a^5 + 1200a^4 - 2940a^3 + 3{,}332a^2 - 1382a\Big] - \tfrac{1}{2}b\Big[225a^4 - 7200a^3 + {}^{30}15\times 4434a^2 - 4\times 407{,}020a + 1382\times 7381\Big)$$
$$+ \tfrac{1}{2}b^2(600a^2 - 15\times 1033a + 2\times 64{,}010) - \tfrac{165}{2}b^3$$

$$= \tfrac{1}{2}a(a-1)(15a^4 - 210a^3 + 990a^2 - 1950a + 1382) - \tfrac{1}{2}b\Big[225(a-8)^4 + 46{,}620(a^2 - 25a + 200)$$
$$+ 1480a - 44958\Big]$$
$$+ 10b^2(30a^2 - \cancel{258}\ 774a + 6401) - \tfrac{165}{2}b^3$$

$$= \tfrac{1}{2}a(a-1)\Big[15(a^4 - 14a^3 + 86a^2 - 130a + 92) + 2\Big] - \tfrac{1}{2}b\Big[225(a-8)^4 + 46{,}620(a^2 - 25a + 200) + 1480(a+25) + 458\Big]$$

$\frac{1380}{15} = \frac{460}{5}$

$255 \times 92 = 510 \times 46$

230 / 460 / 255

7381 / 1382 / 7381 / 22143 / 59048 / 14762 / 10200542 | 64 / 512 / 4 | 6

96 / 512 / 4 / 4608

930×410 / 3720 / 43 / 381300 / 1420 100 | 85 / 465

$1 - 32 + 384 - 2048 + 4096\frac{990}{15} = 3\overset{30}{\times}\frac{1032}{4}$ (above: 2×990)

$225 - 7200 + 86{,}400 - 460{,}800 + 921{,}600$ 258

$225 - 7200 + \cancel{66{,}510}\ 133{,}020 - 1{,}628{,}080 + 10{,}200{,}542$

$-225 + 7200 - 86{,}400 + 460{,}800 - 921{,}600$

$46{,}620 - 1{,}167{,}280 + 9{,}278{,}942$

$-46{,}620 + 1{,}165{,}500 - 9{,}324{,}000$

1780 45{,}058

$15(a^4 - 14a^3 + 86a^2 - 130a + 92) + 2$

$1480)\,44958\,(25$
3560
9358
8900
458

$2214085 + 7435$

2.3.4.5

$100 + 1100 + 1000$

$11 + 33 + 33 + 1$

$196 + 8868 + 1033$

$465 + 381{,}300$

$255 + 23{,}460 + 23{,}205$

$113 + 2{,}260 + 11{,}300$

833   407{,}020   34{,}505

$4)\,466{,}200$
116{,}550

$$k_5 = 128\Big[3(N+qn)^5 - 30(N+qn)^3(N+q^2n) + 20(N+qn)\{3(N+qn)(N+q^3n) + 2(N+q^2n)^2\}$$
$$-5\{17(N+qn)(N+q^4n) + 10(N+q^2n)(N+q^3n)\} + 62(N+q^5n)\Big].$$

$$k_6 = 256\Big[15(N+qn)^6 - 225(N+qn)^4(N+q^2n) + 600(N+qn)^2\{(N+qn)(N+q^3n) + (N+q^2n)^2\}$$
$$-15\{85(N+qn)^2(N+q^4n) + 100(N+qn)(N+q^2n)(N+q^3n) + 11(N+q^2n)^3\}$$
$$+4\{465(N+qn)(N+q^5n) + 255(N+q^2n)(N+q^4n) + 113(N+q^3n)^2\} - 1382(N+q^6n)\Big] \ldots (1)$$

These expressions are considerably simplified if we write $N+qn = A$, $722n = B$. We then have:–

$$k_2 = 4\underline{K}_2, \quad \underline{K}_2 = \tfrac{1}{2}A(A-1) - \tfrac{1}{2}B.$$
$$k_3 = 48\underline{K}_3, \quad \underline{K}_3 = \tfrac{1}{6}A(A-1)(A-2) - \tfrac{1}{6}B(3A-20).$$
$$k_4 = 32\underline{K}_4, \quad \underline{K}_4 = \tfrac{1}{2}A(A-1)[3(A-2)(A-3)-1] - \tfrac{1}{2}B[18(A-7)^2 - 4A + 665] + 4B^2.$$
$$k_5 = 768\underline{K}_5, \quad \underline{K}_5 = \tfrac{1}{6}A(A-1)(A-2)[3(A-3)(A-4)-5] - \frac{5B}{6}[6(A-8)^3 + 8(A-12)(A+45) + A + 104]$$
$$+ \frac{10}{3}B^2(2A-25)$$
$$k_6 = 512\underline{K}_6, \quad \underline{K}_6 = \tfrac{1}{2}A(A-1)[15(A-2)\{(A-3)(A-4)(A-5)-1\}+2] - \tfrac{1}{2}B[15(A-8)^4 + 46,560(A^2 - 25A + 2000)$$
$$-3,280A - 34,058] + \frac{5}{2}B^2[120(A-13)^2 + 23A + 17,324] - \frac{165}{2}B^3 \; \ldots \; (1a).$$

PTO

$\underline{K}_2$, $\underline{K}_3$, etc are integers, not always positive. Fisher (1936b) has tabulated $2\underline{K}_2$ in his Table XIV. I give tables of $\underline{K}_2$, $\underline{K}_3$, and $\underline{K}_4$ in Tables 1, 2, and 3.

The distribution of u_3, deviates considerably from normality even when N and n are not very small. Thus for $N = 6$, $n = 4$, $\gamma_1 = 2.00$, $\gamma_2 = 4.33$. Of course when a number of u values are summed, these values are reduced, but not always very greatly so, since a single family often contributes most of $S(k_3)$ and $S(k_4)$, and a large fraction of $S(k_2)$, as appears for example, in Table 4.

f.5 v

$$
\begin{aligned}
K_6 = {} & \tfrac{1}{2} A(A-1)\left[15(A-2)\left\{(A-3)\left[(A-4)(A-5)-5\right]-1\right\}+2\right] \\
& -\tfrac{1}{2} B\left[225(A-8)^4 + 46,620(A^2-25A+200)+1,480(A+25)+558\right] \\
& +\tfrac{5}{2} B^2\left[120(A-13)^2+21(A+348)+16\right]-\tfrac{165}{2} B^3 \quad \ldots\ldots (1a)
\end{aligned}
$$

f.6 5(a)

$\frac{1}{4}$ Table 1

$K_2 = \frac{1}{4} K_1$

N \ n	1	2	3	4	5	6	7	8	9	10
0	0	81	243	486	810	1215	1701	2268	2916	3645
1	9	99	270	522	855	1269	1764	2340	2997	3735
2	19	118	298	559	901	1324	1828	2413	3079	3826
3	30	138	327	597	948	1380	1893	2487	3162	3918
4	42	159	357	636	996	1437	1959	2562	3246	4011
5	55	181	388	676	1045	1495	2026	2638	3331	4105
6	69	204	420	717	1095	1554	2094	2715	3417	4200
7	84	228	453	759	1146	1614				
8	100	253	487	802	1198	1675				
9	117	279	522	846						
10	135	306	558	891						
11	154	334								
12	174	363								

Table ~~7~~ 2

$\frac{1}{48} k_3$

N \ n	1	2	3	4	5	6	7	8	9	10
0	0	0	729	2,916	7,290	14,580	25,515	40,824	61,236	87,480
1	0	81	972	3,402	8,100	15,795	27,216	43,092	64,152	91,125
2	9	180	1,242	3,924	8,955 ~~9,856~~	17,064 ~~17,037~~	28,980	45,432	67,149	94,860
3	28	298	1,540	4,483	9,856	18,388	30,808	47,845	70,228	98,686
4	58	436	1,837	5,080	10,804	19,768	32,701	50,332	73,390	102,604
5	100	595	2,224	5,716	11,800	21,205	34,660	52,894	76,636	106,615
6	155	776	2,612	6,392	12,845	23,700	36,686	55,532	79,967	110,720
7	224	980	3,032	7,109	13,940	24,254				
8	308	1,208	3,485	7,868	15,086	~~25,476~~ 25,868				
9	408	1,461	3,972	8,640						
10	525	1,740	4,494	9,516						
11	660	2,046 ~~1,848~~								
12	814	2,380								

f.8 5(c)

Table x^3

$\frac{1}{32} K_4$

N \ n	1	2	3	4	5	6	7
0	0	-6,561	-19,683	196,830	1,115,370	3,444,525	8,129,079
1	-81	-6,723	6,318	301,482	1,377,405	3,968,919	9,047,052
2	-163	-3,970	41,066	423,629	1,668,599	4,537,052	10,026,260
3	78	2346	85,533	564,567	1,990,572	5,150,868	11,068,971
4	1,002	12,909	140,727	725,628	2,344,980	5,812,347	12,177,489
5	3,005	28,439	207,692	908,180	2,733,515	6,523,505	13,354,154
6	6,519	49,692	287,508	1,113,627	3,157,905	7,286,394	14,601,342
7	12,012	77,460	381,291	1,343,409	3,619,914	8,103,102	
8	19,988	112,571	499,193	1,599,002	4,121,342	8,975,753	
9	30,987	155,889	615,402	1,881,918			
10	45,585	208,314	758,142	2,193,705			
11	64,394	279,782					
12	88062	344,265					

N \ n	8	9	10
0	16,350,012	29,524,500	49,305,915
1	17,819,028	31,728,267	52,454,385
2	19,369,691	34,037,009	55,734,074
3	21,004,593	36,453,642	59,148,222
4	22,726,362	38,981,118	62,700,105
5	24,537,662	41,622,425	66,393,035
6	26,441,193	44,380,587	70,230,360

When a dominant test factor, such as the gene for the H agglutino-gen, ~~is used in~~ or for the power of tasting phenylthiourea, is used in place of maleness, a correction must be made for the fact that ~~no~~ families including no member recessive for ~~the~~ its allelomorphic gene are excluded. In such families c = d = 0, so

$$u_{31} = (a-3b)^2 - (a+9b)$$
$$= (N-3n)^2 - (N+9n),$$

and the frequency of such families among those which ~~would segregate if they contained~~ have a heterozygous dominant parent is $2^{-s} = 2^{-(N+n)}$. In such families $K_1 = (N-3n)^2 - (N+9n)$, whilst the other cumulants are zero. Hence in summing the u_{31} scores for the remaining families, we must put in each case,

$$K_1 = \frac{N+9n-(N-3n)^2}{2^{N+n}-1},$$

whilst the other cumulants of equations (1) or (1a) are to be multiplied by $\frac{2^{N+n}}{2^{N+n}-1}$.

The u_{31} score is also applicable to families where one parent ~~was~~ is recessive for the rare gene, e.g. albinism, the other being heterozygous for it, whilst both are heterozygous for the test factor, e.g. that for an agglutinogen. This case demands analysis along quite different lines, and will not be considered further, except in the special case of Finney's Family 25, for which we develop equations (4) on p.

correction must be made for the fact that families including no member recessive for the allelomorphic gene are excluded. In such families $c = d = 0$, so $u_{31} = (a-3b)^2 - (a+9b)$
$$= (N-3n)^2 - (N+9n),$$
and the frequency of such families is $2^{-s} = 2^{-(N+n)}$. ~~So in this case the~~ cumulants of the distribution of ~~u_{31}~~ u are ~~as above in~~ equations (1), except ~~that~~ $k_1 = 2^{-(N+n)}\left[(N+9n)-(N-3n)^2\right]$. ~~x 6a~~

~~This is generally negative, but may be positive.~~

The distribution of u_{31} in the presence of linkage.

Let χ be the recombination frequency, and let $\zeta = (1-2\chi)^2$. Thus $\zeta = 0$ in the absence of linkage, and varies between 0 and 1 in presence of linkage. Its use leads to slightly simpler expressions than that of $\xi = \chi(1-\chi) = \frac{1}{4}(1-\zeta)$. We have:

$E(a) = \frac{1}{3}(2-\chi)N$, $E(c) = \frac{1}{3}(1+\chi)N$, $E(b) = \chi n$, $E(d) = (1-\chi)n$, or

$E(a) = \frac{1}{3}(1+\chi)N$, $E(c) = \frac{1}{3}(2-\chi)N$, $E(b) = (1-\chi)n$, $E(d) = \chi n$.

The cumulants of $y = b-d = 2b-n$ can readily be found. For the distribution of y is binomial, the probability that $b = r$ being the coefficient of θ^r in $(1-\chi+\chi\theta)^n$ or $\left[\chi+(1-\chi)\theta\right]^n$. Haldane (1940a) has given expressions for the first 12 cumulants of the binomial distribution in terms of c, the product of the frequencies, which is here $\chi(1-\chi)$, or $\frac{1}{4}(1-\zeta)$, and g, their difference, which is here $\pm(1-2\chi)$, or $\pm\zeta^{\frac{1}{2}}$. To obtain the appropriate values for y we must multiply the values of k_r given by Haldane by 2^r. The cumulants of y up to k_8 are:

$\kappa_1 = -gn = -\, y^{\frac{1}{2}} n$
$\kappa_2 = 4cn = (1-y)n$
$\kappa_3 = 8gcn = 2y^{\frac{1}{2}}(1-y)n$
$\kappa_4 = 16c(1-6c)n = -2(1-y)(1-3y)n$
$\kappa_5 = 32gc(1-12c)n = -8y^{\frac{1}{2}}(1-y)(2-3y)n$.
$\kappa_6 = 64c(1-30c+120c^2)n = 8(1-y)(2-15y+15y^2)n$
$\kappa_7 = 128gc(1-60c+360c^2)n = 16y^{\frac{1}{2}}(1-y)(\overset{17-60}{\cancel{32}-\cancel{45}}y+45y^2)n$.
$\kappa_8 = 256c(1-126c+1680c^2-5040c^3)n = -16(1-y)(17-231y+525y^2-315y^3)n$. . . . (2)

Further cumulants can easily be calculated if desired. ~~To find the cumulants of~~ x. The signs of the odd cumulants are arbitrary, but if that of one is changed that of all must be changed. To find the cumulants of x we must substitute $c = \frac{1}{9}(1+X)(2-X) = \frac{1}{4}\left(1-\frac{y}{9}\right)$, $g = \frac{1}{2}(1-2X) = \mp\frac{1}{3}y^{\frac{1}{2}}$. That is to say we must substitute N for n in the above expressions, $\frac{y}{9}$ for y, and also change the sign of $y^{\frac{1}{2}}$. So putting $y = q\eta$, the cumulants of x-3y are:–

$\kappa_1 = \cancel{q^{\frac{1}{2}}(N} - \eta^{\frac{1}{2}}(N+qn)$
$\kappa_2 = N+qn-\eta(N+q^2n)$
$\kappa_3 = 2\eta^{\frac{1}{2}}[N+q^2n-\eta(N+q^3n)]$
$\kappa_4 = -2[N+q^2n-4\eta(N+q^3n)+3\eta^2(N+q^4n)]$
$\kappa_5 = -8\eta^{\frac{1}{2}}[2(N+q^3n)-5\eta(N+q^4n)+3\eta^2(N+q^5n)]$
$\kappa_6 = 8[2(N+q^3n)-17\eta(N+q^4n)+30\eta^2(N+q^5n)-15\eta^3(N+q^6n)]$.

etc. The cumulants of u, after the first, are those of $(x-3y)^2$. They are obtained from the above by the expressions (1) given by Haldane (1942) and are, putting $N+qn = A$, $\eta^2 n = B$, :–

$$K_1 = \frac{\zeta}{9}\left[A(A-1) - B\right].$$

$$\tfrac{1}{2}K_2 = A(A-1) - B + 2\frac{\zeta}{9}\left[A(A-1)(A-2) - B(3A-20)\right] - \left(\frac{\zeta}{9}\right)^2\left[A(A-1)(2A-3) + B(2A^2-42A+1273) - B^2\right].$$

$$\tfrac{1}{8}K_3 = A(A-1)(A-2) - B(3A-20) + \frac{\zeta}{9}\left[A(A-1)\{3(A-2)(A-3)-1\} - B(18A^2-256A+1547) + 8B^2\right]$$
$$- \left(\frac{\zeta}{9}\right)^2\left[3A(A-1)(2A^2-11A+10) + B(6A^3-270A^2+3,793A-24,600) - 5B^2(3A-44)\right]$$
$$+ \left(\frac{\zeta}{9}\right)^3\left[A(A-1)(3A^2-19A+15) + B(6A^3-954A^2+15,710A-110,715) + B^2(3A^2-123A+1,339) - B^3\right] \ldots (3)$$

(last 2 lines to be checked)

$$\kappa_1 = \frac{y}{q}\left[(N+qn)^2-(N+q^2n)\right]$$

$$\tfrac{1}{2}\kappa_2 = (N+qn)^2-(N+q^2n) + \frac{2y}{q}\left[(N+qn)^3-3(N+qn)(N+q^2n)+2(N+q^3n)\right]$$
$$-\left(\frac{y}{q}\right)^2\left[2(N+qn)^2(N+q^2n)-4(N+qn)(N+q^3n)-(N+q^2n)^2+3(N+q^4n)\right]$$

$$\tfrac{1}{8}\kappa_3 = (N+qn)^3-3(N+qn)(N+q^2n)+2(N+q^3n)$$
$$+\frac{y}{q}\left[3(N+qn)^4-18(N+qn)^2(N+q^2n)+32(N^2+66.qNn+q^4n^2)-14(N+q^4n)\right]$$
$$-\left(\frac{y}{q}\right)^2\left[8(N+qn)^3(N+q^2n)-3(13N^3+759.qN^2n+1799.q^2Nn^2+13.q^5n^3)\right.$$
$$\left.+(61N^2+30{,}450.qNn+61.q^5n^2)-30(N+q^5n)\right]$$
$$+\left(\frac{y}{q}\right)^3\left[(5N^4+228.qN^3n+846.q^2N^2n^2+1028.q^3Nn^3+5.q^6n^4)\right.$$
$$-2(11N^3+3{,}849.qN^2n+1{,}289.q^3Nn^2+11.q^6n^3)$$
$$\left.+8(4N^2+1{,}869.q^2Nn+4.q^6n^2)-15(N+q^6n)\right]. \quad \text{--- (3)}$$

The expression for κ_4 is very heavy, being the sum of ~~20~~ 18 products of the cumulants of $x-3y$. It is unlikely to be used, so I have not given it. It will be noted that the coefficient of $\frac{y}{q}$ in κ_r is ~~2 times~~ half the leading term of κ_{r+1}. The general behaviour of κ_3 can be seen from the following considerations. When $y=0$ (no linkage) it is positive provided $N+n>2$. When $N=0$ and yn is large it approximates to* ~~$q^3n^2(\ldots)y(y-1)(5y-3)$~~. This vanishes when $y=\frac{3}{5}$, or $\chi=\cdot 1127$, and becomes negative for closer linkages, vanishing again when $y=1$, $\chi=0$. Thus for fairly close linkages the distribution may be pretty symmetrical, since most of the

* $18^3 n^4 y(1-y)(3-5y)$

value of ~~$S(\kappa_2)$~~ $S(n)$ comes from families with large n.

The distribution of χ for a given value of ~~$S(\kappa_2)$~~ $S(n)$ can be deduced in a rough way. The asymmetry noted above means that a surprisingly high value of ~~n~~ $S(n)$, (considering the known variance of $S(n)$, is compatible with a given value of y.) It follows that unexpectedly low values of y are compatible with a given value of $S(n)$, i.e. that the distribution of y is negatively skew. It further follows that the distribution of $1-2\chi$ is even more negatively skew. That is to say, when linkage is not very tight, χ may well exceed the value deduced from the data by several times its standard error, whereas a large error in the opposite direction is less probable than with a normal distribution.

When a dominant test factor is used in place of sex, ~~the cumulants are as above, except that~~

~~$$\kappa_1 = \frac{y}{q}\left[(N+qn)^2-(N+q^2n)\right] - 2^{-(N+n)}\left[(N-3n)^2-(N+qn)\right]$$~~

The case of a back-cross for the ~~recessi~~ gene causing abnormality, ~~not~~ with both parents heterozygous for the test factor, will not be considered here.

$$\kappa_1 = \frac{y}{q}\cdot\frac{2^{N+n}}{2^{N+n}-1}\left[(N+qn)^2-(N+q^2n)\right] - \frac{\left[(N-3n)^2-(N+qn)\right]}{2^{N+n}-1},$$

whilst the other cumulants of equation (3) are to be multiplied by $\frac{2^{N+n}}{2^{N+n}-1}$.

value of $S(k_3)$ comes from families with large n.

When a ~~test~~ dominant test factor is used in place of sex, the cumulants are as above, except that ~~k_1~~

$$k_1 = \frac{4}{9}\left[(N+9n)^2 - (N+9^2n)\right] - 2^{-(N+n)}\left[(N-3n)^2 - (N+9n)\right].$$

The distribution of u_3, when the method of ascertainment is known.

The cases originally considered by Fisher (1935 a, b) have not yet arisen in practice, and it therefore does not seem worth while to give the full expressions for the cumulants of u_3, for them. However an example will show how they may be calculated. Let us suppose that a group of families in which partial sex linkage is suspected have been recorded by the method of single ascertainment, that is to say ~~that~~ the probability of recording a family of s members was proportional to the number of recessives, n, in it. It is required to find the value of k_3 in the absence of linkage. The frequency of families containing n recessives among families of s ~~which~~ derived from two heterozygous parents is $\frac{3^N s!}{4^s N! n!}$. The frequency with which they are recorded is proportional to n times this quantity, and is therefore $P_n = \frac{3^{N-1}(s-1)!}{4^{s-1} N!(n-1)!}$.

Hence $\Sigma N P_n = \frac{3}{4}(s-1)$, $\Sigma N(N-1)P_n = \frac{9}{16}(s-1)(s-2)$, $\Sigma N(N-1)(N-2)P_n = \frac{27}{64}(s-1)(s-2)(s-3)$,
$\Sigma(n-1)P_n = \frac{1}{4}(s-1)$, $\Sigma(n-1)(n-2)P_n = \frac{1}{16}(s-1)(s-2)$, $\Sigma N(n-1)P_n = \frac{3}{16}(s-1)(s-2)$, etc.

We may write $\frac{1}{8}k_3$ as:

$$N(N-1)(N-2) + 27N(N-1)(n-1) + 243N(n-1)(n-2) + 729(n-1)(n-2)(n-3)$$
$$+ 27N(N-1) + 486N(n-1) + 2187(n-1)(n-2).$$

Summing over the different values of P_n, we find

$\kappa_3 = 216s(s-1)(s-2)(s+6)$. Similarly
$\kappa_4 = 144s(s-1)(24s^3+81s^2-964s+1016)$.

The Distribution of u_{11}.

In the

~~In the absence of linkage~~,

$u_{11} = (a-b-c+d)^2-(a+b+c+d)$

Now since the family $\underline{s}$ falls into two sections $a+d$ and $b+c$, whose ~~proba~~ expectations are fixed, the expressions for its moments are greatly simplified. In the absence of linkage $u_{11} = s(\chi_s^2-1)$, where χ_s^2 is the exact value of ~~the~~ Pearson's measure of divergence from expectation for a sample of $\underline{s}$ members and one degree of freedom. The cumulants of the distribution of χ_s^2 in this case have been given in equations (6) by Haldane (1938). Those of the distribution of $\underline{u}_{11}$ are:

$\kappa_1 = 0$

$\kappa_2 = 2s(s-1)$

$\kappa_3 = 8s(s-1)(s-2)$

$\kappa_4 = 16s(s-1)(3s^2-15s+14)$

$\kappa_5 = 128s(s-1)(s-2)(3s^2-21s+31)$

$\kappa_6 = 256s(s-1)(15s^4-210s^3+990s^2-1950s+1382)$ - - - - (4).

These are independent of the value of n, and hence of the method of ascertainment. They are also unaffected by dominance of the test factor. If there is linkage, the cumulants are to be derived as before from equations (2) of this paper, ~~They~~ substituting $\underline{s}$ for $\underline{n}$. They are: —

$$k_1 = \zeta s(s-1)$$

$$k_2 = 2s(s-1)(1-\zeta)\left[1+(2s-3)\zeta\right]$$

$$k_3 = 8s(s-1)(1-\zeta)\left[s-2+(3s^2-14s+15)\zeta-(5s^2-17s+15)\zeta^2\right]$$

$$k_4 = 16s(s-1)(1-\zeta)\left[3s^2-15s+17+(12s^3-105s^2+277s-231)\zeta \right.$$
$$\left. -(50s^3-327s^2+715s-525)\zeta^2+(42s^3-237s^2+465s-315)\zeta^3\right] \quad \text{- - - - (5)}$$

The distribution of u_{33} is probably best studied by Fisher's operational method. As it is not used in the study of partial sex-linkage, it will not be discussed here.

Application to data on partial sex-linkage.

Let us first consider the data concerning 28 sibships with normal parents, segregating for <u>epidermolysis bullosa</u>, summarized in Haldane's (1936) Table XIV, and Fisher's (1936 b) Tables XVI and XVII. The values of u and its cumulants are given in Table 4. The fourth column is obtained by halving the values of $\frac{1}{2}k_2$ in Fisher's Table XIV, the fifth and sixth from my Tables 2 and 3. If we write:

$S(u) = U = 434,$

$\frac{1}{4}S(k_2) = K_2 = 5{,}379,$

$\frac{1}{48}S(k_3) = K_3 = 37{,}222,$

$\frac{1}{32}S(k_4) = K_4 = 5{,}754{,}297,$

then in the absence of linkage, U is distributed with cumulants:

$\kappa_1 = 0,\ \kappa_2 = 4K_2 = 21{,}516,\ \kappa_3 = 48K_3 = 1{,}786{,}656,\ \kappa_4 = 32K_4 = 184{,}137{,}504$

Hence $\gamma_1 = \dfrac{6K_3}{K_2^{3/2}} = \cdot 566107$, $\gamma_2 = \dfrac{2K_4}{K_2^2} = \cdot 397758$. So the distribution of U is comparable with that of a χ^2 distribution with 25 degrees of freedom, which has $\gamma_1 = \cdot 5658$, $\gamma_2 = \cdot 48$, and is thus rather more platykurtic.

Several methods are available for the approximate normalization of moderately skew distributions by transformation of the variate. One type, in which the variate x is transformed into $(x+a)^b$, has been discussed by Haldane (1938, 1941a) and was intended for use on these <u>u</u> scores. In the case of the χ^2 distribution for more than about three degrees of freedom it is found that the cube root is almost normally distributed. This depends on the fact that the χ^2 distribution has $\kappa_3 = \dfrac{2\kappa_2^2}{\kappa_1}$, $\kappa_4 = \dfrac{6\kappa_2^3}{\kappa_1^2}$, whilst the distribution of the cube of a normal variate

Table ~~X~~ 4.

Twenty-eight families segregating for epidermolysis bullosa.

Number of families	N	n	S(n)	$\frac{1}{4}S(k_2)$	$\frac{1}{48}S(k_3)$	$\frac{1}{32}S(k_4)$
5	1	1	+18	45	0	-405
5	2	1	-10	95	45	-815
1	3	1	-8	30	28	78
1	4	1	-12	42	58	1,002
1	5	1	+2	55	100	3,005
2	6	1	+20	138	310	13,038
1	7	1	+84	84	224	12,012
1	12	1	-12	174	814	88062, ~~86,436~~
1	1	2	-18	99	81	-6,723
1	2	2	-4	118	180	-3,970
1	5	2	+26	181	595	28,439
1	6	2	-8	204	776	49,692
1	1	3	-12	270	972	6,318
1	3	3	+34	327	1,540	85,533
1	5	3	-32	388	2,224	207,692
1	3	4	-14	597	4,483	564,567
1	4	4	+24	636	5,080	725,628
2	3	5	+356	1,896	19,712	3,981,144 ~~1,770,572~~
28	73	38	+434	5,379	37,222	~~3,752,099~~

5,7~~42,641~~

54,297

has approximately $\kappa_3 = \frac{2\kappa_2^2}{\kappa_1}$, $\kappa_4 = \frac{56\kappa_2^2}{9\kappa_1} = 6.\dot{2}\frac{\kappa_2^2}{\kappa_1}$, if the coefficient of variation is small. Where this does not hold, we may use Haldane's (1938) transformation "B". We put:

$$\Upsilon = \left[\left(1+\frac{U}{g}\right)^{bg} + 2bd - 1\right]\left[1 - d\left(b - \frac{d}{K_2}\right)\right] \div 2bK_2^{\frac{1}{2}} \quad \ldots\ldots \quad (6),$$

where $b = \frac{32K_3^2 - K_2K_4}{8K_2^2K_3}$, $d = \frac{K_3}{3K_2}$, $g = \frac{8K_2^2K_3}{40K_3^2 - K_2K_4}$.

The terms ommitted from equation (6) only affect the fifth decimal place of Υ in the case here considered. Υ is an almost normally distributed variable with mean zero, and unit standard deviation.

On the data of Table 4, U is 2.959 times its standard error, giving $P = .0015$, were it normally distributed. $b = .001,553309$, $d = 2.306,625$, $g = 352.1412$, $bg = .5469841$. Hence $\Upsilon = 2.4460$, and $P = .00722$. Thus the value of U must still be regarded as significant, but its significance is decidedly lessened.

If preferred the method of Cornish and Fisher (1937) may be used, putting $a = 0$, $b = 0$, $c = \gamma_1$, $d = \gamma_2$, in the formulae of their p. 9. This gives $P = .0071$. This method is perhaps a little longer than that here given unless tables of Hermitian polynomials are available; and in theory the values of γ_3 and γ_4, if not of higher deviations from normality, should be used. The Υ transformation has the merit that, although only based on γ_1 and γ_2, it is known to normalize the χ^2 distribution very accurately, and u is clearly analogous to χ^2. It is also of interest as giving, at least approximately, the median value of U. If the sign

$q\left[(1-2bd)^{\frac{1}{bq}}-1\right]$

distribution of ~~t~~ Υ is symmetrical, which is nearly the case, this is the value of U which makes Υ vanish, i.e. ~~$kq\left[1-(1-2bd)^{\frac{1}{bq}}\right]$~~, or approximately ~~to the~~ 2d. In this case the median is $-$ ~~13.42~~ 4.60. That is to say although, in the absence of linkage, the ~~median~~ mean value of U is zero, it is as likely to be less than ~~−13.42 as greater.~~ −5 as greater.

The other most doubtful case of partial sex-linkage, based on data of this kind, is that of Oguchi's disease. Here $U = 294$, $K_2 = 3{,}543$, $K_3 = 17{,}381$, $K_4 = 2{,}604{,}511$. Thus U is 2.898 times its standard error, ~~but Υ~~ giving $P = .0019$. But $\Upsilon = 2.2480$, giving $P = .01136$. However in this case there is further information of two kinds. Back-crosses to affected females give $S(u_{11}) = 24$, and from equations (4), $S(k_2) = 160$, $S(k_3) = 2{,}400$, $S(k_4) = 39{,}040$. Since $\bar{u}_{11} = \frac{1}{2}k_2\zeta$, as compared with $\bar{u}_{31} = \frac{1}{18}k_2\zeta$, the u_{11} must be given 9 times the weight of u_{31}. That is to say we must consider $V = S(u_{31}) + 9\,S(u_{11})$. The variance of u_{11} must be multiplied by 81, its k_r by 9^r. When this is done we find:

$V = 510$, $K_2 = 5{,}813$, $K_3 = 53{,}831$, $K_4 = 10{,}608{,}931$, whence ~~$\Upsilon = 2.4245$~~, ~~$P =$~~ $\Upsilon = 2.6148$, $P = .004425$. Further the direct data from cousin marriage give $\frac{1}{2} - x$ equal to 1.313 times its standard error, with $P = .0946$. Combining these probabilities by Fisher's (1934) method, we find $\chi^2 = 15.557$ for 4 degrees of freedom. By Wilson and Hilferty's (1931) theorem, $P = .00382$. Thus the results are decidedly significant. However the probability of an explanation by chance is some ten times greater than appeared at first sight; and as the sex ratio is very aberrant, it is perhaps possible that the gene is not partially sex-linked. For the other cases of partial sex-linkage based on evidence of this type, the ~~prob~~ significances of the data are still greater.

Application to data on Friedreich's ataxia

Hogben and Pollack (1935) collected data on 12 families segregating for Friedreich's ataxia and for the blood-group genes. Using Bernstein's score they found no evidence of linkage. But Fisher (1936a) used the u_3, score, and found a high positive value. On the method which he then used, it was 1.530 times its standard error. Using the methods of this paper we find $S(u) = +146$. But ~~its~~ owing to the dominance of the test factor its expectation in the absence of linkage is -2.45. Hence $U = 148.45$. On the method then (used by Fisher ~~S(u)~~) ~~was~~ the ~~variance was~~ sampling variance was 9,108, ~~On his later~~ so $S(u)$ was 1.53 times its standard error. On the method) used here, ~~V~~ the variance is 4,500, so U is 2.22 times its standard error, and would therefore be regarded as significant were it not for the correction for skewness.

Fisher points out that the positive value of $S(u)$ is entirely due to one family. This family, from an A x O marriage, consisted of

2 AF, ~~0~~ 0 OF, 0 Af, 4 Of, where f is the gene for Friedreich's ataxia, and F its normal allelomorph.* Fisher ~~remark~~ writes of ~~this family~~ it "The u score attained by this family is, however, over four times its standard error, and, if it is not to be attributed to linkage, it must be ascribed to some cause, or causes, of disturbance capable of obscuring the evidence for the presence or absence of linkage. It provides, in fact, decisive evidence either of linkage, or of the heterogeneity of the twelve families reported".

Actually on Fisher's new method of ~~scoring~~ estimating the variance, the corrected value, $u = 157.031$, is 3.320 times its standard error. But even this would

*, and gave $u = 158$

be fairly decisive evidence, provided the distribution of <u>u</u> were normal. However in the case of a single family it is grossly abnormal. The actual probability is best found by elementary methods. We can ask what fraction of all families consisting of 2 normals and 4 abnormals, would give thus, the highest possible value of <u>u</u>, in the absence of linkage. The probability that ~~all~~ both the two normals should belong to group A is $\frac{1}{4}$. The probability that all four abnormals should belong to group O is $\frac{1}{16}$, giving a cumulative probability of $\frac{1}{64}$. The probability of observing a family of 0 AF, 2 OF, 4 Af, 0 Of, which would give the same <u>u</u> score, is equal. Hence the probability of obtaining this score by chance is $\frac{1}{32}$, or ·03125, as compared with ·0005 if <u>u</u> were normally distributed.* There is nothing surprising in finding one such family among 12.

It is highly probable that there are several ~~genes~~ different genes for ~~Friedreich's ataxia~~ recessive Friedreich's ataxia (Haldane 1940b) and quite possible that some of them are in different chromosomes. One, but not all, of these genes may well prove to be linked with the blood group genes. But Hogben and Pollack's data do not furnish decisive evidence either of linkage or of heterogeneity. Fisher's ~~arguments~~ further arguments, based on the analysis of variance, seem to be inapplicable to this case for the same reasons as the simpler arguments given above.

* If allowance is made for the fact that a family of 2 AF, 0 OF, 4 Af, 0 Of would not be used for linkage work, though it has a probability $\frac{1}{64}$, the probability of obtaining a maximal <u>u</u> is $\frac{2}{63}$, or ·0317.

Application to data on allergy.

Zieve, Wiener, and Fries (1936) recorded the segregation of allergy along with blood group and other genes. They used a relatively inefficient method of searching for linkage, and found none. However Finney (1940) has developed the use of the u scores with great ingenuity and in great detail, and applied it to this case. He concludes that the recessive gene h for allergy shows evidence of linkage with the blood group genes. The sum of his weighted u scores, $S(\lambda)$, is 1·70 times its standard error, giving $P = \cdot 040$. Clearly this is so near the border-line of significance that a detailed analysis becomes of interest.

Table ~~4~~ 5 is a summary of the 31 families ~~including both~~ members ~~recessive both for allergy and a recessive blood group gene~~ which were certainly segregating ~~for~~ for two gene pairs, to which u_{31} is applicable.* ~~It~~ The data are given more fully ~~in~~ Finney's pp. 186 and 187. Besides these, ~~no~~ two segregating families, 44 and 66, were scored by u_{33}, and a number of families which were only certainly segregating for one gene pair. These latter only gave 5 % of the information, and can be ommitted without serious injustice. The u_{33} families contributed 2·5 % of Finney's weighted u score, so their ommission is also not serious.

The families belonging to Finney's type 7 are from $Tt\ Rr \times tt\ Rr$, and the cumulants are given by the modified form of equations (1) on p. 6(a). Type 7* refers to the segregation of A and h among those progeny of A × AB which inherit B from their AB parent. Here the correction is as for type 7, except that

* Where a family ~~from~~ with A and B parents is segregating both for the A and B genes, it is scored twice.

since N and n only refer to the AB and B children, we use 2^s instead of 2^{N+n}. In family 52 this makes no difference to the result. Type 11 refers to the segregation of A and B from an AB parent. Since the parent is known to be heterozygous there is no correction, as in the case of partial sex linkage where the ~~father~~ is known to be heterozygous for sex.

Family 25, from Tt rr x Tt Rr contained abnormals (allergics) only. In this case $N=0$, $n=5$, and

$$u_{3,1} = (c-3d)^2-(c+9d) = (4d-n-1)^2-(3n+1).$$

Neglecting for the moment the fact that $\underline{d}$ may be zero, and the family thus excluded from the record, we can calculate the cumulants of $\underline{d}$ from equations (2), putting $c=\frac{3}{16}$, $g=\frac{1}{2}$. The cumulants of $4d-n-1$ are therefore:

$\kappa_1=-1$, $\kappa_2=3n$, $\kappa_3=6n$, $\kappa_4=-6n$, $\kappa_5=-120n$, $\kappa_6=-312n$, $\kappa_7=3{,}696n$, $\kappa_8=39{,}504n$.

The cumulants of u, after the first, are those of $(4d-n-1)^2$. They are:

$\kappa_1=0$.

$\kappa_2=18n(n-1)$.

$\kappa_3=72n(n-1)(3n-4)$.

$\kappa_4=144n(n-1)(27n^2-27n-43)$.

But when $d=0$, $\kappa_1=n(n-1)$ and the other cumulants vanish. This occurs with a frequency $\left(\frac{3}{4}\right)^n$. When allowance is made for this we find:

$$\kappa_1=\frac{3^n}{3^n-4^n}\,n(n-1).$$

$$\kappa_2=\frac{4^n}{4^n-3^n}\,18n(n-1).$$

$$\kappa_3=\frac{4^n}{4^n-3^n}\,72n\,\cancel{(3n-4)}(n-1)(3n-4).$$

$$\kappa_4=\frac{4^n}{4^n-3^n}\,144n(n-1)(27n^2-27n-43). \quad \text{-----} \quad (7)$$

Table ~~4~~ 5

Families segregating for allergy and bloodgroups.

Families	Type	N	n	$S(u)$	$S(k_1)$	$\frac{1}{4}S(k_2)$	$\frac{1}{48}S(k_3)$	$\frac{1}{32}S(k_4)$
4, 15, 36a, 36b	7	1	1	+24	+8.000	48.00	0	−432
14, 16, 22, 38a, 38b	7	2	1	+22	+7.143	108.57	51.43	−931.4
23a, 23b	7	3	1	+ ~~12~~ 16	+1.600	64.00	59.73	166.4
28, 42	7	4	1	−24	+0.774	86.71	119.74	2,068.6
46	7	7	1	−12	0	84.33	224.88	12,059.1
24, 57a, 57b	7	0	2	−18	−18.000	324.00	0	−26,244
30, 34, 44, 58a, 58b	7	1	2	−18	−4.286	565.71	462.86	−38,417.1
33, 59	7	3	2	+88	+0.777	284.90	615.23	~~2,444.4~~ 4843.4
18a, 18b	7	4	2	+28	+0.571	323.05	885.84	26,224.8
13	7	7	3	−18	+0.029	453.44	3,034.96	381,663.7
52a	7*	1	2	+18	−0.857	113.14	92.57	−7,683.4
52b	11	1	2	+30	0	99.00	81	−6,723
60b	11	2	1	−2	0	19.00	9	−163
25	6	0	3	+54	−4.378	46.70	~~25.95~~ 77.84	5,558
Total				+18~~4~~8	−8.630	2,620.55	5,~~663.14~~ 715.08	~~349,664.1~~ 351,993.1

On this basis we can calculate the last four columns of Table 5. $V = 196.63$, which is 1.922 times its standard error, giving $P = .027$. Thus on the data included in Table 4 the evidence for linkage is rather stronger than when the data giving the remaining 7.5% of Finney's information are added. Clearly no injustice is done to his case by leaving the latter out of consideration. $K_2 = 2,620.55$, $K_3 = 5,715.08$, $K_4 = 351,993.1$. Hence $\gamma_1 = .2556$, so the distribution is a good deal more symmetrical than those considered above. We find $b = 3.9102 \times 10^{-4}$, $d = .726957$, $g = 817.498$, $bg = .319658$, $\Upsilon = 1.7958$, whence $P = .0363$

Hence the correction only diminishes the significance of Finney's result to a slight extent. Probably if all families were included we should have P about .05. It must however be remembered that men have 23 pairs of autosomes, and that White (1940) has shown that two human genes may be in the same chromosome, and yet show no appreciable linkage. Hence the *a priori* probability of finding linkage between a given pair of genes is probably less than .05. Thus the *a priori* probability that one of the two pairs tested by Finney should show linkage is probably less than .1. Thus the data in question must be regarded as giving a strong indication of linkage, but not as indicating it with a high degree of probability. About twice as much information would be needed to make linkage highly probable. At present Burks' (193) data seem to give stronger evidence, but a full appreciation of their significance must await their complete publication, which is much to be desired.

Discussion

This paper is not intended to be polemical. Both Fisher and Finney have improved existing methods for the determination of linkage, including my own. ~~Never~~ And it is clear that the method of this paper is very far from final. The methods of detecting and measuring linkage in man ~~are~~ have developed very rapidly since Bernstein's pioneer work, and have not reached ~~finality~~ perfection. ~~Perhaps~~ Probably some better method than my own of treating these borderline cases will be found. Nevertheless it would seem that whenever U is less than three times its standard error, it is desirable to make some allowance for the skewness of its distribution. I have to thank Dr. N. Karn for reading the manuscript, and detecting several numerical errors.

Summary.

The distribution of Fisher's u scores used in testing for human linkage, is not normal, but has a decided positive skewness. Hence large values of u ~~or~~ may occur more frequently in the absence of linkage than would be supposed from their standard errors. The cumulants κ_2, κ_3 and κ_4 of u are tabulated in certain cases. It is shown that ~~when~~ when correction is made for skewness, the data for partial sex-linkage are still significant, though a good deal less so than had been thought. On the other hand the evidence for linkage of the blood group genes with Friedreich's ataxia is not significant. ~~And that for linkage of the blood group genes with allergy is barely so. On the other hand~~ But the significance of ~~Finney~~ the data on linkage of allergy and blood groups is only very slightly diminished.

$4c = N-x$, $8c = 2N-2x$ $\quad b = 3d =$

Cumulants of $(a-3b-3c+9d)$ are:

~~$\kappa_1 = 0$~~ If $N-4c = x$, $n-4d = y$, $\therefore u_{33} = (x-3y)^2 - (3N-2x+27n-18y)$

$$= (x-3y)^2 + 2(x+9y) - 3(N+9n)$$

To find variance of u_{33}, compared with $u'_{33} = (x-3y)^2 - 3(N+9n)$, in absence of linkage.

$v = (x-3y)^2$. $\quad v' = (x-3y)^2 + 2(x+9y)$, $\bar{v} = \bar{v}' = 3(N+9n)$

~~$v^2 = x^4$~~ $\quad = v + 2(x+9y)$

$x^2 - 6xy + 9y^2$
$3+9y$
$1 - 6 + 9$
$+9 - 54 + 81$

~~Second~~ if κ_n is nth cumulant of $x-3y$, then

$$V_v = 2\kappa_2^2 + \kappa_4 = 2\left[3(N+9^2n)\right]^2 - 120(N+9^4n)$$
$$= 18(N+9^2n)^2 - 120(N+9^4n) = 6\left[3(N+9^2n)^2 - 20(N+9^4n)\right]$$

$$v'^2 = v^2 + 4v(x+9y) + 4(x+9y)^2$$

$$V_{v'} = \overline{v'^2} - (\bar{v}')^2 = V_v + 4\,\overline{v(x+9y)} + 4\,\overline{(x+9y)^2}$$
$$= V_v + 4\left[\overline{x^3+3x^2y - 45xy^2 + 81y^3}\right] + 4\left(\overline{x^2+18xy+81y^2}\right)$$
$$= V_v + 4\left[\overline{x^3}+\overline{x^2} + 9^2(\overline{y^3}+\overline{y^2})\right]$$
$$= V_v + 4\left[-6N+3N+9^2(-6n+3n)\right] = V_v - 12(N+9^2n)$$

405
32
437
360
77

$\therefore v'$ is the less variable $\therefore u_{33}$ is better than u'_{33} $\quad$ DAMN.

825040
690

231
4
924
4
3696

$$v'^3 = v^3 + 6v^2(x+9y) + 12v(x+9y) + 8(x+9y)^3$$

$c = \frac{3}{16}$, $q = \frac{1}{2}$, $\kappa_4 = 128.\frac{3}{16}.\frac{1}{2}\left(1 - \frac{60.3}{16} + \frac{360.3^2}{16^2}\right) = 12\left(1 - \frac{360}{32} + \frac{45.9}{32}\right) = \frac{77.3}{8} = \frac{231}{8}$

$\kappa_8 = 16.3\left(1 - \frac{126.3}{16} + \frac{1680.3^2}{16^2} - \frac{5040.3^3}{16^3}\right) = 3\left(16 - 378 + \frac{1680.9}{16^2} - \frac{5040.27}{16^2}\right) = 3\left(16 - 378 + \frac{24(560-315)}{16}\right)$

$= \frac{3}{16}(24\times245 - 16\times362) = \frac{3\times823}{16} = \frac{2469}{16}$

2469
4
9876
4
39504 ✓

245
9
2205
3
6615
5792
823

362
4
1448
4
5792

$$P = a^{\frac{1}{3}(2-x)} c^{\frac{1}{3}(1+x)} b^{x} d^{1-x} + a^{\frac{1}{3}(1+x)} c^{\frac{1}{3}(2-x)} b^{1-x} d^{x} =$$

$$P = \left(\frac{2-x}{3}\right)^a \left(\frac{1+x}{3}\right)^c x^b (1-x)^d + \left(\frac{1+x}{3}\right)^a \left(\frac{2-x}{3}\right)^c (1-x)^b x^d$$

$$3^{a+c} P = (2-x)^a (1+x)^c x^b (1-x)^d + (1+x)^a (2-x)^c (1-x)^b x^d \qquad x = \tfrac{1}{2} + y$$

$$= \left(\tfrac{3}{2} - y\right)^a \left(\tfrac{3}{2} + y\right)^c \left(\tfrac{1}{2} + y\right)^b \left(\tfrac{1}{2} - y\right)^d + \left(\tfrac{3}{2} + y\right)^a \left(\tfrac{3}{2} - y\right)^c \left(\tfrac{1}{2} - y\right)^b \left(\tfrac{1}{2} + y\right)^d$$

$$2^{a+b+c+d} 3^{a+c} P = (3-2y)^a (3+2y)^c (1+2y)^b (1-2y)^d + (3+2y)^a (3-2y)^c (1-2y)^b (1+2y)^d$$

Let $a < c$, $b < d$

$$= (9-4y^2)^a (1-4y^2)^b \left[(3+2y)^{c-a} (1-2y)^{d-b} + (3-2y)^{c-a} (1+2y)^{d-b}\right]$$

$$= (9-4y^2)^a (1-4y^2)^b \left[\right.$$

$7-11+10$
$-2+11-10$
$2-15+21-10$
$6-34+63-30$

$$u = (a-3b-3c+9d)^2 - (a+9b+9c+81d) \qquad a+c=N,\ b+d=n$$

$$x = N-4c,\ y = n-4d \quad \therefore a-3c = x,\ b-3d = y$$

$4c = N-x$, $a+9c = 3N-2x$
$b+9d = n+2n-8y$

$k_4 = k_4 + 3k_2^2 =$

$$\therefore u = (x-3y)^2 - (3N-2x+27n-18y)$$

$$= (x-3y)^2 + 2(x+9y) - 3(N+9n)$$

$$\bar{x} = 0,\ \overline{x^2} = 3N,\ \overline{x^3} = -6N,\ \overline{x^4} = 27N^2 - 6N$$

$$v = (x-3y)^2,\ v' = (x-3y)^2 + 2(x+9y)$$

$$v^2 = x^4 - 4x^3(3y) + 6 x^2 \cdot 9y^2 - 4x(3y)^3 + 9^2 y^4$$

$$\overline{v^2} = 27N^2 - 6N + 6 \cdot 9 \cdot 3N \cdot 3n + 9^2 \cdot (27n^2 - 6n)$$

$$= 27(N^2 + 18Nn + 9^2 n^2) - 6(N + 9^2 n)$$

$$= 27(N+9n)^2 - 6(N+9^2 n)$$

$XYY,\ XY,\ X,\ Y,\ 1$
$11,\ 8,\ 5,\ 4,\ 0$

one

References.

Burks, B. [Proc. Nat Ac. Sci]

Fisher, R.A. (1934). Statistical methods for research workers Ch IV, para 21.1.
" " (1935a). "The detection of linkage with 'dominant' abnormalities" Ann. Eug. 6, 187-201
" " (1935b). "The detection of linkage with recessive abnormalities." Ann. Eug. 6, 339-351.
" " (1936a) "Heterogeneity of linkage data for Friedreich's ataxia and the spontaneous antigens". Ann. Eug. 7, 17-21.

Finney, D. J., (1940) "The Detection of Linkage". Ann. Eug. 10, 171-214.

Bernstein, F. (1931). "Zur Grundlegung der Chromosomentheorie der Vererbung beim Menschen mit besonderer Berücksichtigung der Blutgruppen". Z. ind Abst. u. Vererb. 57, 113-138.

Haldane, J. B. S. (1936). "A search for incomplete sex-linkage in man" Ann. Eug. 7, 28-57

Fisher, R. A. (1936b). "Tests of significance applied to Haldane's data on partial sex linkage". Ann. Eug. 7, 87-104.

Haldane J. B. S. (1938a). "The approximate normalization of a class of frequency distributions". Biometrika 29, 392-404.

PTO

J. B. S. Haldane (1938b) ~~"The first six moments of χ^2 for an n-fold table with n degrees~~

J. B. S. Haldane (1940a). "The cumulants and moments of the binomial distribution, and the cumulants of χ^2 for a $(n \times 2)$-fold table. Biometrika 31, 392-395.

J. B. S. Haldane (1941a), (1941b) (Biometrika)

Hogben, L. and Pollack R. (1935) "A contribution to the relation of the gene loci involved in the isoagglutinin reaction, taste blindness, Friedreich's ataxia, and major brachydactyly of man". J. Gen. 31, 353-361.

JBS. Haldane (1940b). [Paper on modifiers in J. Gen.] Rel. Imp. of — Prim. & Modifying genes in det. some human diseases. XLI. 149-157

Wilson, E. B. and Hilferty M. M. (1931). "The distribution of Chi-square". Proc Nat. Ac. Sci. 17, 684- .

White, T. (1940) [J. Gen.] Linkage & Crossing Over in the Human Sex Chromosomes. 40. 3. 403-437

Zieve, I, Wiener, A. S., and Fries J. H. (1936). "On the linkage relations of the genes for allergic disease, and the genes determining the blood groups, MN types and eye colour in man". Ann. Eug. 7, 163-178.

Cornish, E.A. and Fisher R. A. (1937) "Moments and cumulants in the specification of distributions" Revue de l'Institut International de Statistique, 1937, IV. 1-14.

f.31r

References.

Bernstein F.(1931). "Zur Grundlegung der Chromosomentheorie der Vererbungbeim Menschen mit besorderer Bueucksichtung der Blutgruppen.". Z. ind Abst. u. Vererb. 57, 113-138.

Burks, B. (1938) "Autosomal linkage in man- the recombination ratio between congenital tooth deficiency and hair colour. J.Nat.Acad.Sci. 24.7.276-282.

Fisher, R.A. (1934). Statistical methods for research workers. Ch. IV, para 21.1.

.. .. (1935a) "The detection of linkage with dominant abnormalities." Ann. Eug. 6, 187-201.

.. .. (1935b) "The detection of linkage with recessive abnormalities." Ann. Eug. 6, 339-551.

.. .. (1936a) "Heterogeneity of linkage data for Friedreich's ataxia and the spontaneous antigens." Ann. Eug. 7, 17-21.

.. .. (1936b) "Tests of significance applied to Haldane's data on partial sex linkage." Ann. Eug. 7,87-104.

Finney D.J. (1940) "The Detection of Linkage." Ann. Eug.10,171-214

Haldane J.B.S.(1936) "A search for incomplete sex-linkage in man." Ann. Eug. 7, 28-57.

.. .. (1938a) "The approximate normalization of a class of frequency distributions." Biometrika 29, 392-404.

.. .. (1940a) "The cumulants and moments of the binomial distribution, and the cumulants of for a (n 2) fold table. Biometrika 31, 392-395.

.. .. (1940b) "The Relative importance of principal and modifying genes in determining some human characters.J.Gen.XLI.149.

.. .. (1941a), ~~(1941b)~~ "The Cumulants of the Distribution of the Square of a Variate." Biometrika. XXXII 199-200.

Hogben,L and Pollack R. (1935) "A contribution to the relation of the gene loci involved in the isoagglutinin reaction, taste blindness, Friedreich's ataxia, and major brachydactyly of man". J.Gen.31.353-361.

Wilson, E.B. and Hefferty M.M. (1931) "The distribution of Chi-Square Proc. Nat. Ac.Sci. 17,684-

White, T. (1940) "Linkage and Crossing over in the Human Sex Chromosomes. 40. 403-437.

Zieve,I, Wiener,A.S., and Fries J.H. (1936) "On the linkgge relations of the genes for allergic disease, and

f.31 v

the genes determining the blood groups, MN types and eye colour in man." Ann. Eug. 7, 163-178.

Cornish, E.A. and Fisher R.A. (1937) "Moments and cumulants in the specification of distributions.". Revue de l' Institut International de Statistique, 1937, IV.1-14.

f. 32r

$\frac{K_3'}{8}$ = ~~8K_1^2(K_1K_3+3K_2~~ ~~K~~ $K_1^2(K_1K_3+3K_2^2)+\frac{1}{2}(3K_1^2K_4+12K_1K_2K_3+2K_2^3)+\frac{1}{4}(3K_1K_5+6K_2K_4+5K_3^2)+\frac{1}{8}K_6$

$$= K_1^3K_3+3K_1^2K_2^2+\frac{3}{2}K_1^2K_4+6K_1K_2K_3+K_2^3+\frac{3}{4}K_1K_5+\frac{3}{2}K_2K_4+\frac{5}{4}K_3^2+\frac{1}{8}K_6$$

$$= -2\eta^2a^3[a+b-\eta(a+10b)]+3\eta a^2[a-\eta(a+b)]^2-3\eta a^2[a+b-4\eta(a+10b)+3\eta^2(a+91b)]$$
$$-12\eta a[a-\eta(a+b)][a+b-\eta(a+10b)]+[a-\eta(a+b)]^3+6\eta a[2(a+10b)-5\eta(a+91b)+3\eta^2(a+820b)]$$
$$-3[a-\eta(a+b)][a+b-4\eta(a+10b)+3\eta^2(a+91b)]+5\eta[a+b-\eta(a+10b)]^2+\frac{1}{8}K_6$$

$$= -2\eta^2a^3[a+b-\eta(a+10b)]+3\eta a^2[a^2-2\eta a(a+b)+\eta^2(a+b)^2]-3\eta a^2[a+b-4\eta(a+10b)+3\eta^2(a+91b)]$$
$$-12\eta a[a(a+b)-\eta(a+b)^2-\eta a(a+10b)+\eta^2(a+b)(a+10b)]+a^3-3\eta a^2(a+b)+3\eta^2a(a+b)^2-\eta^3(a+b)^3$$
$$+6\eta a[2(a+10b)-5\eta(a+91b)+3\eta^2(a+820b)]-3[a(a+b)-\eta(a+b)^2-4\eta a(a+10b)+4\eta^2(a+b)(a+10b)$$
$$+3\eta^2a(a+91b)-3\eta^2(a+b)(a+91b)]+5\eta[(a+b)^2-2\eta(a+b)(a+10b)+\eta^2(a+10b)^2]$$
$$+2(a+10b)-14\eta(a+91b)+30\eta^2(a+820b)-15\eta^3(a+4381b)$$
$$= a(a-1)(a-2)-b(3a-20)+\eta[3a^4-18a^2(a+b)+24a(a+10b)+8(a+b)^2-14(a+91b)]$$
$$+\eta^2[-8a^3(a+b)+24a^2(a+10b)+15a(a+b)^2-39a(a+91b)-22(a+b)(a+10b)+30(a+820b)]$$
$$+\eta^3[2a^3(a+10b)+3a^2(a+b)^2-9a^2(a+91b)-12a(a+b)(a+10b)-(a+b)^3+18a(a+820b)+9(a+b)(a+91b)$$
$$+5(a+10b)^2-15(a+4381b)]$$
$$= a(a-1)(a-2)-b(3a-20)+\eta[a(a-1)(3a^2-15a+19)-b(18a^2-256a+1542)+8b^2]$$
$$-\eta^2[a(a-1)(8a^2-31a+30)+b(8a^3-240a^2+3491a-24{,}600)-5b^2(3a-44)]$$
$$+\eta^3[a(a-1)(5a^2-19a+15)+b(26a^3-954a^2+16{,}688a-110{,}915)+b^2(3a^2-123a+1319)-b^3]$$

$$= a(a-1)(a-2)-b(3a-20)+\eta\left[a(a-1)\{3(a-2)(a-3)-1\}-b\{18[(a-9)^2+34]-(4a+1)\}+8b^2\right]$$
$-\eta^2\left[a(a-1)(a-2)(8a-15)+b\{8(a-11)^3\right.$ ~~$-6(a-34)^2-(a+5638)\}$~~ $\left.-6(a-18)(a-130)-(a+912)\}-5b^2(3a-44)\right]$
$$+\eta^3\left[a(a-1)(5a^2-19a+15)+b\{26(a-12)^3-9(a-12)(a-494)+2a-2635\}+b^2\{3(a-20)(a-21)+59\}+b^3\right]$$

ct sn cp cv / ++ + cv ♀ (E) × olp fs ♂ ~~[? wrong]~~

2 regular daughters, one cv ♀ (alleged)

cv ♀ × olp fs ♂

1 cv olp fs ♂, 1 + olp fs ♂, 1 + + + ♂

1 + ♀, 1 + + fs ♀.

? was she $\frac{+}{cv}$ wrongly scored as cv?

$8 | 5040$, 630

$1680.9 - 315.27$
$= 27(560 - 315)$

315, 245, 2205, 6615, 256, 8871, 8048, 823, 2469, $\frac{348}{1572}$

.216, .216, .0432, 216, 1296, .046656

.04656, 4.056, .18624, 232, 249, .188247

$3.2 \times 6 = 1.2$, $28 \times .4 = 11.2$, $4.48 + 1.92 + .36 = 6.76$

0 ! cont of 8. If $p = .4$, prob is $.6^8 + 8 \times .6^7 \times .4 = .6^7(.6 + 8 \times .4)$
$= 3.8 \times 0.6^7$ | 2 out of 9. $P = .6^9 + 8 \times .4 \times .6^8 + 28 \times .4^2 \times .6^7 = .6^7(.36 + 1.92 + 4.48)$
$= 6.76 \times .6^7 = 4.056 \times .216^2 = .189$

u_{33} aND, cNR, bnD, dnR $\qquad$ $u = N - c$

$$E\, u_{33} = [a - 3c - (3b - 9d)]^2 - (a + 9b + 9c + 81d)$$
$$= [N - 4c - 3(N - 4d)]^2 - [N + 8c + 9(n + 8d)]$$

$-180 + 164$, $201 \times 16 = 804$, 434, 360

Cumulants of c (for $N=1$) are:

$K_1 = \frac{N}{4}\ \frac{1}{4}$

$K_2 = \frac{3}{16}$

$K_3 = \frac{3}{32}$

$K_4 = \frac{3}{16}\left(1 - \frac{9}{8}\right) = -3.2^{-7}$

$K_5 = \frac{1}{2} \cdot \frac{3}{16}\left(1 - \frac{9}{4}\right) = \frac{-15}{2^7}$

$K_6 = \frac{3}{16}\left(1 - \frac{30.3}{16} + \frac{120.3^2}{16^2}\right) = \frac{3}{16}\left(1 - \frac{180}{32} + \frac{135}{32}\right) = \frac{-39}{2^9}$

$K_7 = \frac{1}{2} \cdot \frac{3}{16}\left(1 - \frac{60.3}{16} + \frac{360.9}{16^2}\right) = \frac{3}{32}\left(1 - \frac{360}{32} + \frac{405}{32}\right) = \frac{3.67}{2^{10}}$

$K_8 = \frac{3}{16}\left(1 - \frac{126.3}{16} + \frac{1680.9}{16^2} - \frac{5040.27}{16^3}\right) = \frac{3}{16}\left(1 - \frac{378}{16} + \frac{245.27}{16^2}\right) = \frac{2469}{2^{12}}$

~~$c = \frac{3}{16}$, $q = \frac{1}{4}$~~ Cumulants of $-N + 4c$ ($= -x$) are:

$K_1 = 0$

$K_2 = 3N$

$K_3 = 6N$

$K_4 = -6N$

$K_5 = -120N = -15.2^3 N$

$K_6 = ~~-48~~ - 312N = -39.2^4 N$

$K_7 = 3216N = 201.2^4 N$

$K_8 = 2469.2^4 N$.

Cumulants of x are:

$K_2 = -3N$, $K_3 = -2.3N$, $K_4 = 2.3.N$, $K_5 = 2^3.3.5N$, $K_6 = 2^4.3.13N$, $K_7 = -2^4.3.67N$, $K_8 = -2^4.3^2.823N$

$c+d=n$ $\quad c+d=n,\ c=n-d \quad n-d$ f33v

$u_{31}=(c-3d)^2 \cdot (c+9d)$

$=(n-4d)^2-$

[*From* Biometrika, *Vol. XXXII. Part II. October*, 1941.]

(v) The Cumulants of the Distribution of the Square of a Variate

By J. B. S. HALDANE, F.R.S.

The following problem has arisen in several biometric investigations. The cumulants of the distribution of x are known, and it is desired to find the cumulants of the distribution of x^2. As this problem is likely to arise in future, it seems desirable to give the appropriate transformations for the first few cumulants.

Let $\kappa_1, \kappa_2, \kappa_3, \ldots$ be the cumulants of x.

Let $\mu_1', \mu_2', \mu_3', \ldots$ be the moments of x^2 about zero.

Let $\mu_2, \mu_3, \mu_4, \ldots$ be the moments of x^2 about its mean.

Let $\kappa_1', \kappa_2', \kappa_3', \ldots$ be the cumulants of x^2.

Then μ_r' is the $2r$th moment of x. These have been given in terms of the cumulants up to the 10th, i.e. μ_5', in the general case by Kendall (1940), and up to the 12th, i.e. μ_6', by Haldane (1938) when $\kappa_1 = 0$. We consider the general case first. We have such expressions as

$$\mu_2' = \kappa_1^4 + 6\kappa_1^2\kappa_2 + 4\kappa_1\kappa_3 + 3\kappa_2^2 + \kappa_4.$$

From these we calculate the moments μ_r, and hence the cumulants. The results are:

$$\left.\begin{aligned}
\kappa_1' &= \kappa_1^2 + \kappa_2,\\
\kappa_2' &= 4\kappa_1^2\kappa_2 + 2(2\kappa_1\kappa_3 + \kappa_2^2) + \kappa_4,\\
\kappa_3' &= 8\kappa_1^2(\kappa_1\kappa_3 + 3\kappa_2^2) + 4(3\kappa_1^2\kappa_4 + 12\kappa_1\kappa_2\kappa_3 + 2\kappa_2^3) + 2(3\kappa_1\kappa_5 + 6\kappa_2\kappa_4 + 5\kappa_3^2) + \kappa_6,\\
\kappa_4' &= 16\kappa_1^2(\kappa_1^2\kappa_4 + 12\kappa_1\kappa_2\kappa_3 + 12\kappa_2^3)\\
&\quad + 16(2\kappa_1^3\kappa_5 + 18\kappa_1^2\kappa_2\kappa_4 + 12\kappa_1^2\kappa_3^2 + 36\kappa_1\kappa_2^2\kappa_3 + 3\kappa_2^4)\\
&\quad + 8(3\kappa_1^2\kappa_6 + 18\kappa_1\kappa_2\kappa_5 + 32\kappa_1\kappa_3\kappa_4 + 18\kappa_2^2\kappa_4 + 30\kappa_2\kappa_3^2)\\
&\quad + 8(\kappa_1\kappa_7 + 3\kappa_2\kappa_6 + 7\kappa_3\kappa_5 + 4\kappa_4^2) + \kappa_8.
\end{aligned}\right\} \quad (1)$$

After this the expressions become very heavy. When $\kappa_1 = 0$, i.e. x has its mean zero, most of the terms vanish, and we have

$$\left.\begin{aligned}
\kappa_1' &= \kappa_2,\\
\kappa_2' &= 2\kappa_2^2 + \kappa_4,\\
\kappa_3' &= 8\kappa_2^3 + 2(6\kappa_2\kappa_4 + 5\kappa_3^2) + 6,\\
\kappa_4' &= 48\kappa_2^4 + 48\kappa_2(3\kappa_2\kappa_4 + 5\kappa_3^2) + 8(3\kappa_2\kappa_6 + 7\kappa_3\kappa_5 + 4\kappa_4^2) + \kappa_8,\\
\kappa_5' &= 384\kappa_2^5 + 960\kappa_2^2(\kappa_2\kappa_4 + 5\kappa_3^2) + 80(16\kappa_2\kappa_4^2 + 28\kappa_2\kappa_3\kappa_5 + 6\kappa_2^2\kappa_6 + 25\kappa_3^2\kappa_4)\\
&\quad + 2(20\kappa_2\kappa_8 + 60\kappa_3\kappa_7 + 100\kappa_4\kappa_6 + 63\kappa_5^2) + \kappa_{10},\\
\kappa_6' &= 3840\kappa_2^6 + 9600\kappa_2^3(3\kappa_2\kappa_4 + 10\kappa_3^2) + 4800(2\kappa_2^3\kappa_6 + 14\kappa_2^2\kappa_3\kappa_5 + 8\kappa_2^2\kappa_4^2 + 25\kappa_2\kappa_3^2\kappa_4 + 3\kappa_3^4)\\
&\quad + 40(30\kappa_2^2\kappa_8 + 180\kappa_2\kappa_3\kappa_7 + 300\kappa_2\kappa_4\kappa_6 + 226\kappa_3^2\kappa_6 + 189\kappa_2\kappa_5^2 + 672\kappa_3\kappa_4\kappa_5 + 132\kappa_4^3)\\
&\quad + 4(15\kappa_2\kappa_{10} + 55\kappa_3\kappa_9 + 120\kappa_4\kappa_8 + 198\kappa_5\kappa_7 + 113\kappa_6^2) + \kappa_{12}.
\end{aligned}\right\} \quad (2)$$

Finally, if x be symmetrically distributed, so that all its odd cumulants vanish,

$$\left.\begin{aligned}
\kappa_1' &= \kappa_2,\\
\kappa_2' &= 2\kappa_2^2+\kappa_4,\\
\kappa_3' &= 8\kappa_2^3+12\kappa_2\kappa_4+\kappa_6,\\
\kappa_4' &= 48\kappa_2^4+144\kappa_2^2\kappa_4+8(3\kappa_2\kappa_6+4\kappa_4^2)+\kappa_8,\\
\kappa_5' &= 384\kappa_2^5+1920\kappa_2^3\kappa_4+160\kappa_2(3\kappa_2\kappa_6+8\kappa_4^2)+40(\kappa_2\kappa_8+5\kappa_4\kappa_6)+\kappa_{10},\\
\kappa_6' &= 3840\kappa_2^6+28800\kappa_2^4\kappa_4+9600\kappa_2^2(\kappa_2\kappa_6+4\kappa_4^2)+240(5\kappa_2^2\kappa_8+50\kappa_2\kappa_4\kappa_6+22\kappa_4^3)\\
&\quad+4(15\kappa_2\kappa_{10}+120\kappa_4\kappa_8+113\kappa_6^2)+\kappa_{12}.
\end{aligned}\right\} \quad (3)$$

I have bracketed together terms which are products of the same number of κ_r's. If x is a linear function of observed numbers in a sample of n, every κ_n is proportional to x, so the terms in brackets will all be multiples of the same power of n.

REFERENCES

HALDANE, J. B. S. (1938). 'The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small.' *Biometrika*, **29**, 389–91.

KENDALL, M. G. (1940). 'The derivation of multivariate sampling formulae from univariate formulae by symbolic operation.' *Ann. Eugen., Lond.*, **10**, 392–402.

[*From* Biometrika, *Vol. XXXII. Part II. October,* 1941.]

(v) The Cumulants of the Distribution of the Square of a Variate

BY J. B. S. HALDANE, F.R.S.

The following problem has arisen in several biometric investigations. The cumulants of the distribution of x are known, and it is desired to find the cumulants of the distribution of x^2. As this problem is likely to arise in future, it seems desirable to give the appropriate transformations for the first few cumulants.

Let $\kappa_1, \kappa_2, \kappa_3, \ldots$ be the cumulants of x.

Let $\mu'_1, \mu'_2, \mu'_3, \ldots$ be the moments of x^2 about zero.

Let $\mu_2, \mu_3, \mu_4, \ldots$ be the moments of x^2 about its mean.

Let $\kappa'_1, \kappa'_2, \kappa'_3, \ldots$ be the cumulants of x^2.

Then μ'_r is the $2r$th moment of x. These have been given in terms of the cumulants up to the 10th, i.e. μ'_5, in the general case by Kendall (1940), and up to the 12th, i.e. μ'_6, by Haldane (1938) when $\kappa_1 = 0$. We consider the general case first. We have such expressions as

$$\mu'_2 = \kappa_1^4 + 6\kappa_1^2\kappa_2 + 4\kappa_1\kappa_3 + 3\kappa_2^2 + \kappa_4.$$

From these we calculate the moments μ_r, and hence the cumulants. The results are:

$$\left.\begin{aligned}
\kappa'_1 &= \kappa_1^2 + \kappa_2,\\
\kappa'_2 &= 4\kappa_1^2\kappa_2 + 2(2\kappa_1\kappa_3 + \kappa_2^2) + \kappa_4,\\
\kappa'_3 &= 8\kappa_1^2(\kappa_1\kappa_3 + 3\kappa_2^2) + 4(3\kappa_1^2\kappa_4 + 12\kappa_1\kappa_2\kappa_3 + 2\kappa_2^3) + 2(3\kappa_1\kappa_5 + 6\kappa_2\kappa_4 + 5\kappa_3^2) + \kappa_6,\\
\kappa'_4 &= 16\kappa_1^2(\kappa_1^2\kappa_4 + 12\kappa_1\kappa_2\kappa_3 + 12\kappa_2^3)\\
&\quad + 16(2\kappa_1^3\kappa_5 + 18\kappa_1^2\kappa_2\kappa_4 + 12\kappa_1^2\kappa_3^2 + 36\kappa_1\kappa_2^2\kappa_3 + 3\kappa_2^4)\\
&\quad + 8(3\kappa_1^2\kappa_6 + 18\kappa_1\kappa_2\kappa_5 + 32\kappa_1\kappa_3\kappa_4 + 18\kappa_2^2\kappa_4 + 30\kappa_2\kappa_3^2)\\
&\quad + 8(\kappa_1\kappa_7 + 3\kappa_2\kappa_6 + 7\kappa_3\kappa_5 + 4\kappa_4^2) + \kappa_8.
\end{aligned}\right\} \quad (1)$$

After this the expressions become very heavy. When $\kappa_1 = 0$, i.e. x has its mean zero, most of the terms vanish, and we have

$$\left.\begin{aligned}
\kappa'_1 &= \kappa_2,\\
\kappa'_2 &= 2\kappa_2^2 + \kappa_4,\\
\kappa'_3 &= 8\kappa_2^3 + 2(6\kappa_2\kappa_4 + 5\kappa_3^2) + 6,\\
\kappa'_4 &= 48\kappa_2^4 + 48\kappa_2(3\kappa_2\kappa_4 + 5\kappa_3^2) + 8(3\kappa_2\kappa_6 + 7\kappa_3\kappa_5 + 4\kappa_4^2) + \kappa_8,\\
\kappa'_5 &= 384\kappa_2^5 + 960\kappa_2^3(\kappa_2\kappa_4 + 5\kappa_3^2) + 80(16\kappa_2\kappa_4^2 + 28\kappa_2\kappa_3\kappa_5 + 6\kappa_2^2\kappa_6 + 25\kappa_3^2\kappa_4)\\
&\quad + 2(20\kappa_2\kappa_8 + 60\kappa_3\kappa_7 + 100\kappa_4\kappa_6 + 63\kappa_5^2) + \kappa_{10},\\
\kappa'_6 &= 3840\kappa_2^6 + 9600\kappa_2^3(3\kappa_2\kappa_4 + 10\kappa_3^2) + 4800(2\kappa_2^2\kappa_6 + 14\kappa_2^2\kappa_3\kappa_5 + 8\kappa_2^2\kappa_4^2 + 25\kappa_2\kappa_3^2\kappa_4 + 3\kappa_3^4)\\
&\quad + 40(30\kappa_2^2\kappa_8 + 180\kappa_2\kappa_3\kappa_7 + 300\kappa_2\kappa_4\kappa_6 + 226\kappa_3^2\kappa_6 + 189\kappa_2\kappa_5^2 + 672\kappa_3\kappa_4\kappa_5 + 132\kappa_4^3)\\
&\quad + 4(15\kappa_2\kappa_{10} + 55\kappa_3\kappa_9 + 120\kappa_4\kappa_8 + 198\kappa_5\kappa_7 + 113\kappa_6^2) + \kappa_{12}.
\end{aligned}\right\} \quad (2)$$

Finally, if x be symmetrically distributed, so that all its odd cumulants vanish,

$$\left.\begin{aligned}
\kappa_1' &= \kappa_2,\\
\kappa_2' &= 2\kappa_2^2+\kappa_4,\\
\kappa_3' &= 8\kappa_2^3+12\kappa_2\kappa_4+\kappa_6,\\
\kappa_4' &= 48\kappa_2^4+144\kappa_2^2\kappa_4+8(3\kappa_2\kappa_6+4\kappa_4^2)+\kappa_8,\\
\kappa_5' &= 384\kappa_2^5+1920\kappa_2^3\kappa_4+160\kappa_2(3\kappa_2\kappa_6+8\kappa_4^2)+40(\kappa_2\kappa_8+5\kappa_4\kappa_6)+\kappa_{10},\\
\kappa_6' &= 3840\kappa_2^6+28800\kappa_2^4\kappa_4+9600\kappa_2^2(\kappa_2\kappa_6+4\kappa_4^2)+240(5\kappa_2^2\kappa_8+50\kappa_2\kappa_4\kappa_6+22\kappa_4^3)\\
&\quad +4(15\kappa_2\kappa_{10}+120\kappa_4\kappa_8+113\kappa_6^2)+\kappa_{12}.
\end{aligned}\right\} \quad (3)$$

I have bracketed together terms which are products of the same number of κ_r's. If x is a linear function of observed numbers in a sample of n, every κ_r is proportional to n, so the terms in brackets will all be multiples of the same power of n.

REFERENCES

HALDANE, J. B. S. (1938). 'The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small.' *Biometrika*, **29**, 389–91.

KENDALL, M. G. (1940). 'The derivation of multivariate sampling formulae from univariate formulae by symbolic operation.' *Ann. Eugen., Lond.*, **10**, 392–402.

f36r

REPRINTED FROM
ANNALS OF EUGENICS, Vol. 11, Part 2, pp. 179–181, 1941
 PRINTED IN GREAT BRITAIN

THE FITTING OF BINOMIAL DISTRIBUTIONS

By J. B. S. HALDANE, F.R.S.

A number of cases arise where on different occasions an event has occurred $0, 1, 2, 3, \ldots, r, \ldots$ times. Sometimes a Poisson distribution gives a good fit, the probability of the event occurring r times being $P_r = e^{-m}\frac{m^r}{r!}$. In other cases a good fit may be obtained to a binomial distribution where P_r is the coefficient of t^r in $(1-p+pt)^k$. Here p and k are both positive or both negative. Where they are negative it is convenient to write $p' = -p$, $k' = -k$. Hence

$$P_r = \frac{k(k-1)\ldots(k-r+1)}{r!}p^r(1-p)^{k-r},$$

or

$$\frac{k'(k'+1)\ldots(k'+r-1)}{r!}p'^r(1+p')^{-k'-r}.$$

Such distributions have been discussed, with numerical examples, by Whitaker (1914), and Greenwood & Yule (1920) have paid special attention to the negative binomial distribution where p and k are negative.

In the past these distributions have, I think, always been fitted by the first two moments. For if m and v are the observed mean and variance, their expectations are $E(m) = kp$, $E(v) = kp(1-p)$. Whence we obtain consistent estimates, $\hat{p} = \frac{m-v}{m}$, $\hat{k} = \frac{m^2}{m-v}$. However, this method of estimation does not appear to be fully efficient. Fitting by maximum likelihood is so. Jeffreys (1939) states (pp. 260, 374) that tables of digamma functions are required for such fitting. It is the object of this note to show that the fitting may be done by elementary methods.

Let $q = 1-p$. Let n_r be the observed frequency of r, R the maximum value of r. Let $N = \sum_{r=0}^{R} n_r$, the total number of observations, and $m = \frac{1}{N}\sum_{r=0}^{R} rn_r$, the mean value of r. Then the logarithm of the likelihood is

$$L = \sum_{r=0}^{R} n_r \log P_r = \sum_{r=0}^{R} n_r\left[r\log p + (k-r)\log q + \sum_{s=0}^{r-1}\log(k-s) - \log r!\right].$$

$$\frac{\partial L}{\partial p} = \frac{1}{pq}\Sigma n_r(r-kp) = 0,$$

whence

$$kp = m. \tag{1}$$

$$\frac{\partial L}{\partial k} = \Sigma n_r\left(\log q + \sum_{s=0}^{r-1}\frac{1}{k-s}\right) = 0.$$

$$\therefore \quad N\log q + \sum_{r=0}^{R}\frac{1}{k-r}\sum_{s=r+1}^{R} n_s = 0,$$

or

$$N[\log k - \log(k-m)] = \frac{n_1+n_2+\ldots+n_R}{k} + \frac{n_2+n_3+\ldots+n_R}{k-1} + \ldots + \frac{n_R}{k-R+1}. \tag{2}$$

When k and p are negative, this becomes

$$N[\log(k'+m)-\log k'] = \frac{n_1+n_2+\ldots+n_R}{k'}+\frac{n_2+n_3+\ldots+n_R}{k'+1}+\ldots+\frac{n_R}{k'+R-1}. \quad (2\cdot1)$$

These equations can be stated in terms of digamma functions, but this is quite unnecessary, and they can be solved without great difficulty by trial and interpolation, rejecting the infinite root.

We further have

$$\frac{-\partial^2 L}{\partial p^2} = \frac{kN}{pq}, \quad \frac{-\partial^2 L}{\partial p\,\partial k} = \frac{N}{q},$$

$$\frac{-\partial^2 L}{\partial k^2} = \sum_{r=0}^{R}(k-r)^{-2}\sum_{s=r+1}^{R} n_s$$

$$= \frac{n_1+n_2+\ldots+n_R}{k^2}+\frac{n_2+n_3+\ldots+n_R}{(k-1)^2}+\ldots+\frac{n_R}{(k-R+1)^2}$$

$$= \frac{n_1+n_2+\ldots+n_R}{k'^2}+\frac{n_2+n_3+\ldots+n_R}{(k'+1)^2}+\ldots+\frac{n_R}{(k'+R-1)^2}.$$

Hence the amounts of information concerning p and k are

$$I_p = -\frac{\partial^2 L}{\partial p^2}+\left(\frac{\partial^2 L}{\partial p\,\partial k}\right)^2\Big/\frac{\partial^2 L}{\partial k^2}$$

$$= \frac{N}{q}\left[\frac{k}{p}-\frac{N}{q\sum_{r=0}^{R}(k-r)^{-2}\sum_{s=r+1}^{R} n_s}\right], \quad (3)$$

$$I_k = \sum_{r=0}^{R}(k-r)^{-2}\sum_{s=r+1}^{R} n_s-\frac{pN}{kq}. \quad (4)$$

Whitaker's values, in my terminology, are

$$I_p = \frac{kn}{1+(2k-3)\,p}, \quad I_k = \frac{q^2N}{2k(k-1)\,p^2}.$$

The numerical calculation of k to more than four significant figures is rather tedious; however, this does not matter in view of its large standard error. But as the numbers to be subtracted in the calculations of I_p and I_k are very nearly equal, p, q and $\frac{-\partial^2 L}{k^2}$ should be calculated to seven or eight significant figures.

As an example, we take Whitaker's data for the numbers of days out of 1096 on which r deaths of women over 80 were reported in *The Times* of 1910–12. They are: $n_0 = 162$, $n_1 = 267$, $n_2 = 271$, $n_3 = 185$, $n_4 = 111$, $n_5 = 61$, $n_6 = 27$, $n_7 = 8$, $n_8 = 3$, $n_9 = 1$. $k' = 10{\cdot}0$ gives the R.H.S. of equation (2·1) as 213·78, the L.H.S. as 213·97. $k' = 9{\cdot}9$ gives the R.H.S. as 216·029, the L.H.S. as 216·028. Hence $k' = 9{\cdot}900$. From equations (1), (3) and (4) we have

$$p = -0{\cdot}21787 \pm 0{\cdot}05292, \quad k = -9{\cdot}900 \pm 2{\cdot}492.$$

By the method of moments, Whitaker found

$$p = -0{\cdot}20770 \pm 0{\cdot}04862, \quad k = -10{\cdot}440 \pm 2{\cdot}702.$$

Thus the two results only differ by about one-fifth of the standard error, and in this case, at least, the method of moments is quite satisfactory. However with a smaller total, it would be less reliable. In any case it is useful to take $(m-v)/m^2$ as a first approximation to k in solving equation (2) or (2·1). It is also convenient to multiply both sides of this equation by k. If this is done, one side increases with k, whilst the other diminishes, and interpolation becomes easier. I have to thank Dr Jeffreys for a correction.

Summary

A binomial law can readily be fitted to observed data by the method of maximum likelihood.

REFERENCES

M. Greenwood & G. U. Yule (1920). 'An enquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents.' *J.R. Stat. Soc.* **83**, 255–79.

H. Jeffreys (1939). *Theory of Probability*. Oxford.

L. Whitaker (1914). 'On the Poisson Law of small numbers.' *Biometrika*, **10**, 36–71.

That is to say we must substitute N for n in the above expressions, $\frac{\zeta}{9}$ for ζ, and also change the sign of $\zeta^{\frac{1}{2}}$. So putting $\zeta = 9\eta$ the cumulants of x-3y are:-

$$k_1 = -\eta^{\frac{1}{2}}(N+9n)$$
$$k_2 = N+9n-\eta(N+9^2n)$$
$$k_3 = 2\eta^{\frac{1}{2}}\left[N+9^2n-\eta(N+9^3n)\right]$$
$$k_4 = -2\left[N+9^2n-4\eta(N+9^3n)+3\eta^2(N+9^4n)\right]$$
$$k_5 = -8\eta^{\frac{1}{2}}\left[2(N+9^3n)-5\eta(N+9^4n)+3\eta^2(N+9^5n)\right]$$

? $$k_6 = 8\left[2(N+9^3n)-17\eta(N+9^4n)+30\eta^2(N+9^5n) - 15\eta^3(N+9^6n)\right].$$

etc. The cumulants of u, after the first, are those of $(x-3y)^2$. They are obtained from the above by the expressions (1) given by Haldane (1941a) and are, putting N+9n=A, 72n=B, :-

$$k_1 = \frac{\zeta}{9}\left[A(A-1)-B\right]$$
$$\tfrac{1}{2}k_2 = A(A-1)-B+2\frac{\zeta}{9}\left[A(A-1)(A-2)-B(3A-20)\right] - \left(\frac{\zeta}{9}\right)^2\left[A(A-1)(2A-3)+B(2A^2-42A+273)-B^2\right].$$
$$\tfrac{1}{8}k_3 = A(A-1)(A-2)-B(3A-20)+\frac{\zeta}{9}\left[A(A-1)\{3(A-2)(A-3)-1\}-B(18A^2-256A+1547)+8B^2\right]$$
$$-\left(\frac{\zeta}{9}\right)^2\left[3A(A-1)(2A^2-11A+10)+B(6A^3-270A^2+3{,}793A-24{,}600)-5B^2(3A-44)\right]$$
$$+\left(\frac{\zeta}{9}\right)^3\left[A(A-1)(3A^2-19A+15)+B(6A^3-954A^2+15{,}710A-110{,}715) + B^2(3A^2-123A+1{,}339)-B^3\right] \quad \text{---(3)}$$

p11a

The expression for k_4 is very heavy, being the sum of 18 products of the cumulants of x-3y. It is unlikely to be used, so I have not given it. It will be noted that the coefficient of $\frac{\zeta}{9}$ in k_r is half the leading term of k_{r+1}. The general behaviour of k_3 can be seen from the following considerations. When $\zeta=0$ (no linkage) it is positive provided $N+n > 2$. When N=0 and ζn is large

$$\frac{1}{8}k_3 = A(A-1)(A-2) - B(3A-20) + \frac{\gamma}{9}\Big[A(A-1)\{3(A-2)(A-3)-1\} - B\{18[(A-4)^2+34] - (4A+1)\} + 8B^2\Big]$$

$$-\left(\frac{\gamma}{9}\right)^2\Big[A(A-1)(A-2)(8A-15) + B\{8(A-11)^3 - 6(A-18)(A-130) - (A+912)\} - 5B^2(3A-44)\Big]$$

$$+\left(\frac{\gamma}{9}\right)^3\Big[A(A-1)(5A^2-17A+15) + B\{26(A-12)^3 - 9(A-12)(A-494) + 2A - 2635\} + B^2\{3(A-20)(A-21)+59\}$$
$$+B^3\Big] \quad \cdots\cdots (3)$$

the cumulants of $\underline{u}_{31}$:-

$$k_1 = 0$$

$$k_2 = 2\left[(N+9n)^2 - (N+9^2 n)\right]$$

$$k_3 = 8\left[(N+9n)^3 - 3(N+9n)(N+9^2 n) + 2(N+9^3 n)\right]$$

$$k_4 = 16\left[3(N+9n)^4 - 18(N+9n^2) + 8\{3(N+9n)(N+9^3 n) + (N+9^2 n)^2\} - 17(N+9^4 n)\right]$$

$$k_5 = 128\left[3(N+9n)^5 - 30(N+9n)^3(N+9^2 n) + 20(N+9n)\{3(N+9n)(N+9^3 n) + 2(N+9^2 n)^2\} - 5\{17(N+9n)(N+9^4 n) + 10(N+9^2 n)(N+9^3 n)\} + 62(N+9^5 n)\right]$$

$$k_6 = 256\left[15(N+9n)^6 - 225(N+9n)^4(N+9^2 n) + 600(N+9n)^2\{(N+9n)(N+9^3 n) + (N+9^2 n)^2\} - 15\{85(N+9n)^2(N+9^4 n) + 100(N+9n)(N+9^2 n)(N+9^3 n) + 11(N+9^2 n)^3\} + 4\{465(N+9n)(N+9^5 n) + 255(N+9^2 n)(N+9^4 n) + 113(N+9^3 n)^2\} - 1382(N+9^6 n)\right] \text{ ------ } (1)$$

These expressions are considerably simplified if we write

$N+9n=A$, $72^n=B$. We then have:-

$$k_2 = 4K_2. \qquad K_2 = \tfrac{1}{2}A(A-1) - \tfrac{1}{2}B.$$

$$k_3 = 48K_3. \qquad K_3 = \tfrac{1}{6}(A(A-1)(A-2) - \tfrac{1}{6}B(3A-20)$$

$$k_4 = 32K_4. \qquad K_4 = \tfrac{1}{2}A(A-1)\left[3(A-2)(A-3)-1\right] - \tfrac{1}{2}B\left[18(A-7)^2 - 4A + 665\right] + 4B^2.$$

$$k_5 = 768K_5. \qquad K_5 = \tfrac{1}{6}A(A-1)(A-2)\left[3(A-3)(A-4)-5\right] - \frac{5B}{6}\left[6.(A-8)^3 9(A-12)(A+75) + A + 104\right] + \frac{10}{3}B^2(2A-25)$$

$$k_6 = 512K_6. \qquad K_6 = \tfrac{1}{2}A(A-1)\left[15(A-2)\{(A-3)[(A-4)(A-5)-5]-1\}+2\right] - \tfrac{1}{2}B\left[225(A-8)^4 + 46{,}620(A^2-25A+200) + 1{,}780(A+25) + 558\right] + \frac{5}{2}B^2\left[120(A-13)^2 + 21(A+348) + 16\right] - \frac{165}{2}B^3 \text{ -------- } (1a)$$

K_2, K_3, etc are integers, not always positive. Fisher (1936b) has tabulated $2K_2$ in his Table XIV. I give tables of K_1, K_2, and K_4 in Tables 1,2, and 3.

The distribution of u_{31}, deviates considerably from

One parent O other $\frac{n}{O}$, $n > 1$

Family	$\frac{A}{B}$ N	O N	$\frac{A}{B}$ P	O P	N	n	u
2	1	2	0	1	3	1	−8
94	1	1	0	1	2	1	−2
14	2	2	0	1	4	1	−4
18	1	3	3	0	4	3	+40
29	0	1	0	1	1	2	−6
31	0	2	1	0	2	1	+14
32	2	1	0	1	3	1	+4
34	1	0	1	1	1	2	−18
37	0	1	1	1	1	2	−18
41	1	2	0	1	3	1	−18
W	2	2	0	2	4	2	+14
10							+34

One parent O, other het

	x N	y N	x P	y P		
3	1	4	0	3	5	3
4	1	0	1	1	1	2
18	1	3	3	0	4	3
23	0	0	1	1	0	2
36	1	4	1	1		
44	1	0	0	1		
A, O } 48	0	2	1	3		
B O } 46	2	0	2	2		
A, O Bird	~~1~~	~~1~~	~~1~~	~~0~~		
A_2O } "	1	1	2	0		
Lockyer A	1	0	0	2		
" B	0	1	1	1		

$(n+1) - \frac{1}{2}(\varepsilon+1)$

$(1-4+12)^2-(5+36)$

$= 81-41$ $\quad (1-3-9)^2-(4+27)$ $\quad (1-4)^2-(5+18)$

$9-14$ $\quad 121-31$ $\quad 4-23$

$1-3+$ $\quad (1-3)^2-10=-6$ $\quad (4+3)^2-(4+9)$ $=49-13$

$(5+4)^2-(5+27)=146-32$

$(2+3)^2-11$

34 56 384 106

A B had A, B and A, children ∴ BO

Other parent A, A, or A, O, or A, O, not A'B

A×B gave A, B, O

∴ A O × B O

A

3673 2083 5456

226 18 244

3673 14892 2083 16,775

$(1+3)^2-10$ $\quad (2+3)^2-11$

$(1+3)^2-10$ $\quad 1-$

$1-1-$

$(1-2+3)^2-3-9=4-12$

$6^2-(4+18)=36-22$

$(1+9)^2-1781$

6|36|43(60.605

36

120 | 4300

1206 | 7236

1212 | 6400

121.21

$\alpha_{33}^2 = 324(5-1)(5+4)\left(\frac{4^8-3^{8-2}}{4^5 3^5}\right)$

$=324 \times 4 \times 5 \quad \frac{16-1}{16-9}$

$= 2160 \times \frac{15}{7} = \frac{1080 \times 30}{7} = \frac{32400}{7}$

7)32400 4

7)14580 2083 ~ 45^2

1|164|45(129.5

22 | 64 44

249 | 2345 2241

2585 | 13400

$(1-2-0+3)^2-12=4-12$ $\quad (1+3)^2-10$ $\quad 2^2-(2+30)$

$9-2-9$ $\quad (-2-3+4)^2-(2+30)=16-31$ $\quad 106$

$(1+6)^2-(17+18)$ $\quad 384$ $\quad 36$ $\quad 420$ $\quad 94$ A and B

1^2- $\quad$ A B X $\quad 9-(4+27)=9-31$

$(4+6)^2-(4+18)=100-22$

$(3-2-6+4)^2=50$

$=16-50$

84 -22 16 -34 100 .56

$6^2-(2+18)$ $\quad (9-3-6)^2-(1+24)$

$(2-9+18)^2-$ $\quad 3-$

243 39 282

1-3 $\quad (3-12+24)^2-(3+36+243)$

225- $\quad 324-282=42$

$(1-4+9)^2-(5+24)=36-34$

2√3285

18|1140(114.6

121 1040

224 | 896 14400

228

13140 2083

12|15223(12.3.4

144

243| 823 729

940

F41

F.42r

$N+9n = a$
$N+9^2n = a+b$
$N+9^3n = a+10b$
$N+9^4n = a+91b$
$N+9^5n = a+820b$
$N+9^6n = a+7381b$

91
729
820
6561
7381
765 / 85
7733

85 + 7735
100 + 1100 + 1000
11 + 33 + 33 + 11

430 × 410
410
372
93
381300

460
51
230
460
255
23205

44340 / 3
133020

7381
1382
7381
22143
59048
7462
10,200,542

465 + 381,300
255 + 23,460 + 23,205
113 + ~~3,390 + 33,900 + 113,000~~
2,260 11,300
833 + 407,020

98 × 30

10330
51650
154950

$R_6 = \frac{1}{2}\left[15a^6 - 225a^4(a+b) + 600a^2\{a(a+10b)+(a+b)^2\} - 15\{85a^2(a+91b) + 100a(a+b)(a+10b)\right.$
$\left. + 11(a+b)^3\} + 4\{465a(a+820b) + 255(a+b)(a+91b) + 113(a+10b)^2\} - 1382(a+7381b)\right]$
$= \frac{1}{2}\left[15a^6 - 225a^4(a+b) + 600a^2(a^2+10ab+a^2+2ab+b^2) - 15\{85a^2(a+91b) + 100a(a^2+11ab+10b^2)\right.$
$+ 11(a^3+3a^2b+3ab^2+b^3)\} + 4\{465a(a+820b) + 255(a^2+92ab+91b^2) + 113(a^2 + 20ab + 100b^2)\} - \ldots$
$= \frac{1}{2}\left[15a^6 - 225a^4(a+b) + 600a^2(2a^2+12ab+b^2) - 15(196a^3 + 8868a^2b + 1033ab^2 + 11b^3)\right.$
$\left. + 4(833a^2 + 407,020ab + 34,505b^2) + 1382(a+7381b)\right]$
$= \frac{1}{2}a(15a^5 - 225a^4 + 1200a^3 - 2940a^2 + 3332a - 1382) - \frac{b}{2}(225a^4 - 7200a^3 + 133,020a^2$
$- 1,628,080a + 10,100,542) + \frac{1}{2}b^2(600a^2 - 15,495a + 138,020) - \frac{165}{2}b^3$
$= \frac{1}{2}a(a-1)(15a^4 - 210a^3 + 990a^2 - 1950a + 1382) - \frac{1}{2}b\left[225(a-8)^4 + 46,620a^2 - 1,167,280a + 9,278,942\right]$
$+ \frac{5b^2}{2}(120a^2 - 3,099a + 27,604) - \frac{165}{2}b^3$
$= \frac{1}{2}a(a-1)\left[15(a^4 - 14a^3 + 66a^2 - 130a + 92) + 2\right] - \frac{1}{2}b\left[225(a-8)^4 + 46,620(a^2 - 25a + 200)\right.$
$\left. + 1780a - 445,058\right] + \frac{5}{2}b^2\left[120(a-13)^2 - 21a + 7,324\right] - \frac{165}{2}b^3$
$= \frac{1}{2}a(a-1)\left[15\{(a-2)(a-3)(a-4)(a-5) - 5(a-2)(a-3) - (a-2)\} + 2\right] - \frac{1}{2}b\left[225(a-8)^4 + 46,620(a^2-25a+200)\right.$
$\left. + 1780(a-25) + 558\right] + \frac{5}{2}b^2\left[120(a-13)^2 + 21(a+348) + 10\right] - \frac{165}{2}b^3$

$a(a-1) + \frac{15}{2}\{a(a-1)(a-2)[(a-3)(a-4)(a-5) - 5a + 14]$
$15(a-2)\{(a-3)(a-4)(a-5) - 5(a-3)\} - 1\} + 2$

$225(a-8)^4 = 225a^4 - 225.8.4a^3 + 225.6.8^2a^2 - 225.4.8^3a + 225.8^4)$
$= 225a^4 - 7,200a^3 + 86,400a^2 - 460,800a + 921,600$

8
400
512
9
225

133,020 − 1,628,080 + 10,200,542
− 86,400 + 460,800 − 921,600
46,620 − 1,167,280 + 9,278,942
−46,620 + 1,165,500 − 9,324,000
1,780 − 445,058

4)4662,000
1165,500

120 − 3099 + 27,604
−120 + 3120 − 20,280
1 − 5 + 6   + 21 + 7,324

$a^2 - 26a + 169$
338
2028

1−5
1−4
1−9+20
1−3
1−9+20
−3+27−60
1−12+47−60
−2+24−94+...
1−14+71−154+1...

1−14+66−130+92
−1+14−71+154−120
−5+24−28
45 −25 +30

~~−5a^2~~ −24a+28)
−(5a^2 − 24a + 28)

15 × (120 − 30 + 2) + 2 = 15 × 92 + 2 = 3014 + 2 = 1382

1780) 45058 (25 −1 +2
3560
9458
8900
558

1780) 445,058 (81
14240
2668
1780
888

21) 7324 (348
63
102
84
184
168
16

5(a−2)(a−3) + (a−2)
= (a−2)(5a − 15 + 1)

f42v

$a^3 - 3a(a+b) + 2(a+10b) = a^3 - 3a + 2 - b(3a - 20)$

$$\eta\Big[3a^4 - 3a^2(a+b) + 12a(a+10b) + 3(a+b)^2 + 12a(a+\ldots$$
$$-12a^2(a+b) + 12a(a+10b) + 5(a+b)^2 - 17(a+91b)\Big]$$
$$-3a^2(a+b)$$

$$\eta^2\Big[-2a^3(a+b) + 12a^2(a+10b) + 12a(a+b)^2 - 30a(a+91b) - 12(a+b)(a+10b)$$
$$-6a^3(a+b) + 12a^2(a+10b) + 3a(a+b)^2 - 9a(a+91b) - 10(a+b)(a+10b)$$
$$+30(a+820b)$$

$$\eta = \frac{y}{4}$$

$$k_1 = -\eta^{\frac{1}{2}} a$$

$$k_2 = a - \eta(a+b)$$

$$k_3 = 2\eta^{\frac{1}{2}}\left[a+b-\eta(a+10b)\right]$$

$$k_4 = -2\left[a+b-4\eta(a+10b)+3\eta^2(a+91b)\right]$$

$$k_5 = -8\eta^{\frac{1}{2}}\left[2(a+10b)-5\eta(a+91b)+3\eta^2(a+820b)\right]$$

$$k_6 = 8\left[2(a+10b)-14\eta(a+91b)+30\eta^2(a+820b)-15\eta^3(a+7381b)\right]$$

$$k_3' = 8k_1^2(k_1k_3+3k_2^2)+4(3k_1^2k_4+12k_1k_2k_3+2k_2^3)+2(3k_1k_5+6k_2k_4+5k_3^2)+k_6$$

$$= 8\eta a^2\left[-2\eta a\{a+b-\eta(a+10b)\}+3\{a-\eta(a+b)\}^2\right]+4\left[\sout{3\eta a^2} -6\eta a^2\{a+b-4\eta(a+10b)+3\eta^2(a+91b)\}\right.$$

$$\left.-24\eta a\{a-\eta(a+b)\}\{a+b-\eta(a+10b)\}+\sout{8\eta\{a+b-\eta(a+10b)\}^2}\right]\,8\{a-\eta(a+b)\}^3$$

$$+2\left[24\eta a\{2(a+10b)-5\eta(a+91b)+3\eta^2(a+820b)\}-12\{a-\eta(a+b)\}\{a+b-4\eta(a+10b)+3\eta^2(a+91b)\}\right.$$

$$\left.+20\eta\{a+b-\eta(a+10b)\}^2\right]+16(a+10b)-136\eta(a+91b)+240\eta^2(a+820b)-120\eta^3(a+7381b)$$

$$=8\eta a^2\left[3a^2-2\eta a(4a+3b)+\eta^2\{3(a+b)^2+2a(a+10b)\}\right]-24\eta a^2\left[a+b-4\eta(a+10b)+3\eta^2(a+91b)\right]$$

$$-96\eta a\left[a(a+b)-\eta\{(a+b)^2+a(a+10b)\}+\eta^2(a+b)(a+10b)\right]+\sout{32\eta\left[(a+b)^2-2\eta(a+b)(a+10b)+\eta^2(a+10b)^2\right]}\ *$$

$$+48\eta a\left[2(a+10b)-5\eta(a+91b)+3\eta^2(a+820b)\right]-24\left[a(a+b)-\eta\{(a+b)^2+4a(a+10b)\}+\eta^2\{4(a+b)(a+10b)+3a(a+91b)\}\right.$$

$$\left.-3\eta^3(a+b)(a+91b)\right]+40\eta\left[(a+b)^2-2\eta(a+b)(a+10b)+\eta^2(a+10b)^2\right]+16(a+10b)-136\eta(a+91b)$$

$$+240\eta^2(a+820b)-120\eta^3(a+7381b).$$

$$\frac{k_3'}{8} = a(a-1)(a-2)-b(3a-20)+\eta\left[3a^4-18a^2(a+b)+12a(a+10b)+3(a+b)^2+5(a+b)-14(a+91b)\right]$$

$$+\eta^2\left[-2a^3(4a+3b)+12a^2(a+10b)+12a(a+b)^2+12a^2(a+10b)+3a(a+b)^2-30a(a+91b)+3(a+b)^2+12a(a+10b)\right.$$

$$-12(a+b)(a+10b)-9a(a+91b)$$

$$\left.-10(a+b)(a+10b)+30(a+820b)\right]+\eta^3\left[3a^2(a+b)^2+6a^3(a+10b)-9a^2(a+91b)-12a(a+b)(a+10b)-(a+b)^3\right.$$

$$\left.+18a(a+820b)+9(a+b)(a+91b)+5(a+10b)^2-15(a+7381b)\right]$$

$$= a(a-1)(a-2)-b(3a-20)+\eta\left[\right.$$

$$*\ +\frac{8}{8}\left[a^3-3\eta a^2(a+b)^2+3\eta^2 a^2(a+b)^2-(a+b)^3\right]$$

$-4+3\eta-6a+20$

$-a^3(a+b)+12a^2(a+10b)$

$-2a^3(a+b)+12a^2(a+10b)+12a(a+b)^2$

f43v

$2a^4 + 20a^3b$
$3a^4 + 6a^3b + 3a^2b^2$
$-9a^3 - 819a^2b$
$-12a^3 - 132a^2b - 120ab^2$
$-a^3 - 3a^2b - 3ab^2 - b^3$
$+18a^2 + 14{,}760ab$
$+9a^2 + 828ab + 819b^2$
$+5a^2 + 100ab + 500b^2$
$-15a = 110{,}415b$

$5a^4 - 22a^3 + 32a^2 - 15a = a(a-1)(5a^2-14a+15)$

$b(26a^3 - 454a^2 + 16{,}688a - 110{,}415)$
$+b^2(3a^2 - 123a + 1319)$

1640
$a(1+92+91)$
$5(1+20+100)$
3381
$7)22143$
110415

$(a-34)^2 = a^2 - 74a + 1369$
$(a-74)^2 = a^2 - 148a + 5476$
$= (a-74)^2 - 3151$

$(a-74)^2$ $a-74$
$(a-130)(a-18)$
$= a^2 - 148a + 2340$

$8(1 - 33 + 363 - 1331)$
$1 - 30 + 300 - 1000$

$18a^2 - 256a + 1544$
$= 18(a^2 - 14a + 86) - 4a - 1$
$= 18\{(a-7)^2 + 37\} - (4a+1)$

86, 688, 1548

$6)13952$
2325
$32 - 62 + 3$

$8a^2 - 31a + 30 = 8a^2 - 32a + 32 + a - 2$
$= 8(a^2 - 4a + 4) + a - 2$
$= 8(a-2)^2 + a + 2$

$1-2)$ $8 - 31 + 30(8-15$
$8 - 16$
$-15 + 30$

$8a^3 - 270a^2 + 3741a - 24{,}600$
$-8 + 264 - 2904 + 10648$
$-6a^2 + 887a$
$+6a^2 - 888a + 5638$
13040
$-a - 912$

$8 - 270 + 3741 - 24{,}600$

$6)887$
148

$a^2 - 3a + 3$
$5a^2 - 15a + 10$

$18)1544(8$
144
104

$b(8a^3 - 270a^2 + 3741a - 24{,}600) - b^2(15a - 220)$

$8a^4 - 39a^3 + 61a^2 - 30a = a(a-1)(8a^2 - 31a + 30)$

$4a^4 - 27a^3 + 61a^2 - 30a$

$8a^3(a+b) - [12a^2(a+10b) + 15a(a+b)^2] + 39a(a+91b) + 22(a+b)(a+10b) - 30(a+820b)$

$8a^4 + 8a^3b$
$-24a^3 - 24a^2b$
$-15a^3 - 30a^2b - 15ab^2$
$+39a^2 + 3549ab$
$+22a^2 + 242ab + 220b^2$
$-30a - 24{,}600$

1369
5476

256, 18
$128 : 14$
$8)270$
$33 + 6$

Eggs in a raft, w q.

w q

75 75	74	~~65~~ 74	64	~~34~~ 62	64	62
52 44	94	64	87	74	87	74
66 71	116	78	77	2,136	77	59
70 79	109	47	102	1,73	102	78
99 121	68	38	101		101	~~5,~~ 273
103 123	65	5,309	81		81	138
107 116	75		56		56	
96 103	58		83		83	
102 106	84		72		72	
84 93	97		9,723		9,728	
10 981	51					
	11,904					

292
147
392
137
192
1210

Total ~~51 rafts, 4054~~ 4124 eggs, ~~81·02~~ 80·92 eggs per raft
= 81 − 4/51

10	981
11	904
5	309
9	723
2	136
9	728
4	273
50	4054

10	981
11	904
5	309
9	728
4	273
39	3195
1	73
40	3268
2	137
3	3131

Total ~~34 or 40 rafts, 3268 eggs mean 81·7~~
38 " , 3131 eggs, mean 82·4
= 82 15/19

51)4127(80·92
408
47·0
459
1·10

38)3131(82·394
304
91
76
150
114
360
342
180

53)4572(86·27
424
332
318
140
106
340

w

9	728
2	138
1	73
24	2104
7	548
10	981
53	4572

4245 72
1311 43·00
87·92

Mean 87·92
86·3

Eggs per raft + ♀s

f45

Rafts	Eggs
13	1222
4	297
10	808
18	1547
2	147
9	702
2	137
~~2~~	~~137~~
58	4860

73 249 3511
85 253 1964
64 304 1547
82 238
96 423
102 492
1964

58 rafts, 4860 eggs, mean 83.86 eggs

Grand total 96 rafts, 7991 eggs, mean 83.24 eggs

24)2430(83.8
232
112
87
250
232
180

4860
3131
98)7991(81.
784
151

96)7991(83.24
768
311
288
230
192
380

33)2801(84.9
264
161
132
290

79)6824(85.~~7~~ 6.8.
632
504
474
300
237
630

404
395
9

All families ~~66 rafts, 5602 eggs, mean 84.9~~
79 rafts, 6824 eggs, mean ~~86.3~~ 86.38

9	702
2	137
4	297
2	147
1	86
39	3511
9	722
66	5602
13	1222
79	6824

79	6824
53	4576
132	11396

11)11346
4)1036
3)259
86.3

f 46r

[*From* Biometrika, *Vol. XXXI. Parts III and IV. March*, 1940.]

(iii) The cumulants and moments of the binomial distribution, and the cumulants of χ^2 for a $(n\times 2)$-fold table

By J. B. S. HALDANE, F.R.S.

From the Department of Biometry, University College, London

The first four cumulants of the distribution of χ^2 for a $(n\times 2)$-fold table when samples are finite, have already been given (Haldane, 1937). These and higher cumulants and moments can be calculated by a simpler method. Consider a sample of s, the probability of a success being p, and $p+q=1$. Pearson (1919) pointed out that for moments of $(p+q)^s$ about its mean, the generating function is $(qe^{pt}+pe^{-qt})^s$, and Romanovsky (1923) gave a recurrence formula for the moments. That for the cumulants is much simpler.

Let $$U = qe^{pt}+pe^{-qt}.$$

Then the cumulant-generating function

$$K(t) \equiv \sum_{r=2}^{\infty} \frac{\kappa_r t^r}{r!} = s \log U.$$

To find the cumulants for $s=1$, we note that

$$\frac{\partial}{\partial q} K(t) = \frac{e^{pt}-e^{-qt}}{U} - t,$$

$$\frac{\partial}{\partial t} K(t) = \frac{pq(e^{pt}-e^{-qt})}{U}$$

So
$$\frac{\partial}{\partial t}K(t) = pq\frac{\partial}{\partial q}K(t)+pqt,$$
or
$$\sum_{r=2}^{\infty}\frac{\kappa_r t^{r-1}}{(r-1)!} \equiv pq\sum_{r=2}^{\infty}\frac{t^r}{r!}\frac{d\kappa_r}{dq}+pqt.$$

Equating the coefficients of $t^r/r!$, we find
$$\kappa_2 = pq,$$
and if $r>2$
$$\kappa_{r+1} = pq\frac{d\kappa_r}{dq}. \qquad \text{......(1)}$$

Let $pq = c$, $p-q = g$; then
$$\kappa_{r+1} = c\frac{d\kappa_r}{dq}, \quad \frac{dc}{dq} = g, \quad \frac{dg}{dq} = -2, \quad g^2 = 1-4c.$$

Hence if $\kappa_{2r} = f(c)$,
$$\left.\begin{aligned}\kappa_{2r+1} &= gcf'(c),\\ \kappa_{2r+2} &= c(1-6c)f'(c)+c^2(1-4c)f''(c).\end{aligned}\right\} \qquad \text{......(2)}$$

From these equations we can very rapidly calculate successive values of κ_r, since $\kappa_2 = c$, and find:
$$\begin{aligned}
\kappa_1 &= 0,\\
\kappa_2 &= c,\\
\kappa_3 &= cg,\\
\kappa_4 &= c-6c^2,\\
\kappa_5 &= g(c-12c^2),\\
\kappa_6 &= c-30c^2+120c^3,\\
\kappa_7 &= g(c-60c^2+360c^3),\\
\kappa_8 &= c-126c^2+1{,}680c^3-5{,}040c^4,\\
\kappa_9 &= g(c-252c^2+5{,}040c^3-20{,}160c^4),\\
\kappa_{10} &= c-510c^2+17{,}640c^3-151{,}200c^4+362{,}880c^5,\\
\kappa_{11} &= g(c-1{,}020c^2+52{,}920c^3-604{,}800c^4+1{,}814{,}400c^5),\\
\kappa_{12} &= c-2{,}046c^2+168{,}960c^3-3{,}160{,}080c^4+19{,}958{,}400c^5-39{,}916{,}800c^6. \qquad \text{......(3)}
\end{aligned}$$

If each of the above cumulants be multiplied by s, the moments about the mean can now be calculated from the expressions given by Fisher (1928) and Haldane (1938). If $p = \frac{1}{2}$ we have
$$K(t) = s\log\cosh\tfrac{1}{2}t,$$
so
$$\kappa_2 = \frac{s}{4}, \quad \kappa_4 = -\frac{s}{8}, \quad \kappa_6 = \frac{s}{4}, \quad \kappa_8 = -\frac{17s}{16}, \quad \kappa_{10} = \frac{31s}{4}, \quad \kappa_{12} = -\frac{691s}{8},$$
while if q is very small we have for the cumulant-generating function of a Poisson series
$$K(t) = sc(e^t-1-t).$$

The coefficient of c^2 is $-[s^r+(-1)^r-3]$. So when q is small, but its square is not neglected, the first order correction to the Poisson cumulant-generating function is
$$K(t) = sq(e^t-1-t)+sq^2(e^{2t}-e^t).$$

The numerical coefficient of the highest power of c in κ_r is $(r-1)!$ when r is even, and $\frac{1}{2}(r-1)!$ when r is odd.

Consider a sample of s, in which a successes are recorded. Then
$$\chi^2 = \frac{(a-sp)^2}{spq}.$$

But $a - sp$ is the departure from the mean of the binomial distribution $(p+q)^s$. Hence the rth moment of the distribution of χ^2 (for one degree of freedom) about zero, is

$$\nu_r' = \frac{\mu_{2r}}{s^r c^r},$$

where μ_{2r} is the $2r$th moment of $(p+q)^s$.

But if μ_r' and κ_r' be the rth moment about the mean, and the rth cumulant, of the χ^2 distribution, then

$$\mu_2 = \nu_2' - \nu_1^2, \text{ etc.,} \quad \kappa_2' = \mu_2', \text{ etc.}$$

Making the necessary substitutions, we find, for the cumulants of χ^2 in terms of those of the binomial distribution:

$$\kappa_1' = (sc)^{-1}\kappa_2,$$

$$\kappa_2' = (sc)^{-2}(2\kappa_2^2 + \kappa_4),$$

$$\kappa_3' = (sc)^{-3}[8\kappa_2^3 + 2(5\kappa_3^2 + 6\kappa_2\kappa_4) + \kappa_6],$$

$$\kappa_4' = (sc)^{-4}[48\kappa_2^4 + 48(5\kappa_2\kappa_3^2 + 3\kappa_2^2\kappa_4) + 8(4\kappa_4^2 + 7\kappa_3\kappa_5 + 3\kappa_2\kappa_6) + \kappa_8],$$

$$\kappa_5' = (sc)^{-5}[384\kappa_2^5 + 960(5\kappa_2^2\kappa_3^2 + 2\kappa_2^3\kappa_4) + 80(25\kappa_3^2\kappa_4 + 16\kappa_2\kappa_4^2 + 28\kappa_2\kappa_3\kappa_5 + 6\kappa_2^2\kappa_6) + 2(63\kappa_5^2 + 100\kappa_4\kappa_6 + 60\kappa_3\kappa_7 + 20\kappa_2\kappa_8) + \kappa_{10}],$$

$$\kappa_6' = (sc)^{-6}[3{,}840\kappa_2^6 + 9{,}600(10\kappa_2^3\kappa_3^2 + 3\kappa_2^4\kappa_4) + 4{,}800(3\kappa_3^4 + 25\kappa_2\kappa_3^2\kappa_4 + 8\kappa_2^2\kappa_4^2 + 14\kappa_2^2\kappa_3\kappa_5 + 2\kappa_2^3\kappa_6) + 40(132\kappa_4^3 + 672\kappa_3\kappa_4\kappa_5 + 189\kappa_2\kappa_5^2 + 226\kappa_3^2\kappa_6 + 300\kappa_2\kappa_4\kappa_6 + 180\kappa_2\kappa_3\kappa_7 + 30\kappa_2^2\kappa_8) + 4(113\kappa_6^2 + 198\kappa_5\kappa_7 + 120\kappa_4\kappa_8 + 55\kappa_3\kappa_9 + 15\kappa_2\kappa_{10}) + \kappa_{12}]. \quad \text{......(4).}$$

We now substitute the values of κ_r given in equations (3) multiplied by s, putting

$$k = (pq)^{-1} = c^{-1}.$$

We therefore have, for the cumulants of χ^2 with one degree of freedom:

$$\kappa_1 = 1,$$

$$\kappa_2 = 2 + (k-6)s^{-1},$$

$$\kappa_3 = 8 + 2(11k - 56)s^{-1} + (k^2 - 30k + 120)s^{-2},$$

$$\kappa_4 = 48 + 96(4k - 19)s^{-1} + 16(7k^2 - 125k + 420)s^{-2} + (k^3 - 125k^2 + 1{,}680k - 5{,}040)s^{-3},$$

$$\kappa_5 = 384 + 960(7k - 32)s^{-1} + 400(15k^2 - 214k + 648)s^{-2} + 6(81k^3 - 3{,}908k^2 + 38{,}420k - 98{,}496)s^{-3} + (k^4 - 510k^3 + 17{,}640k^2 - 151{,}200k + 362{,}880)s^{-4},$$

$$\kappa_6 = 3{,}840 + 9{,}600(13k - 58)s^{-1} + 9{,}600(26k^2 - 327k + 924)s^{-2} + 40(1{,}729k^3 - 56{,}236k^2 + 459{,}024k - 1{,}065{,}792)s^{-3} + 4(501k^4 - 59{,}398k^3 + 1{,}289{,}244k^2 - 8{,}824{,}320k + 18{,}555{,}840)s^{-4} + (k^5 - 2{,}046k^4 + 168{,}960k^3 - 3{,}160{,}080k^2 + 19{,}958{,}400k - 39{,}916{,}800)s^{-5}. \quad \text{......(5)}$$

When $p = \frac{1}{2}$, $k = 4$, and we have, for n degrees of freedom:

$$\kappa_1 = n,$$

$$\kappa_2 = 2ns^{-1}(s-1),$$

$$\kappa_3 = 8ns^{-2}(s-1)(s-2),$$

$$\kappa_4 = 16ns^{-3}(s-1)(3s^2 - 15s + 17),$$

$$\kappa_5 = 128ns^{-4}(s-1)(s-2)(3s^2 - 21s + 31).$$

$$\kappa_6 = 256ns^{-5}(s-1)(15s^4 - 210s^3 + 990s^2 - 1{,}950s + 1{,}382). \quad \text{......(6)}$$

If there are n samples, with different values of s, we have, for the cumulants of χ^2, where $h = \dfrac{1}{2pq}$, and $R_i = \Sigma s^{-i}$,

$$\begin{aligned}
\kappa_1 &= n,\\
\kappa_2 &= 2n[1+(h-3)\,R_1],\\
\kappa_3 &= 4n[2+(11h-28)\,R_1+(h^2-15h+30)\,R_2],\\
\kappa_4 &= 8n[6+12(8h-19)\,R_1+4(14h^2-125h+210)\,R_2+(h^3-63h^2+420h-630)\,R_3],\\
\kappa_5 &= 16n[24+120(7h-16)\,R_1+100(15h^2-107h+162)\,R_2+3(81h^3\\
&\qquad -1{,}954h^2+9{,}560h-12{,}312)\,R_3+(h^4-255h^3+4{,}410h^2-18{,}900h+22{,}680)\,R_4],\\
\kappa_6 &= 32n[120+600(13h-29)\,R_1+600(52h^2-327h+462)\,R_2+10(1{,}729h^3\\
&\qquad -28{,}118h^2+114{,}756h-133{,}228)\,R_3+2(501h^4-29{,}699h^3+322{,}311h^2\\
&\qquad -1{,}103{,}040h+1{,}159{,}740)\,R_4+(h^5-1{,}023h^4+42{,}240h^3-395{,}010h^2\\
&\qquad +1{,}247{,}400h-1{,}247{,}400)\,R_5]. \qquad \ldots\ldots(7)
\end{aligned}$$

When $p = q = \frac{1}{2}$, we have:

$$\begin{aligned}
\kappa_1 &= n,\\
\kappa_2 &= 2(n-R_1),\\
\kappa_3 &= 8(n-3R_1+4R_2),\\
\kappa_4 &= 16(3n-18R_1+32R_2-17R_3),\\
\kappa_5 &= 128(3n-30R_1+100R_2-135R_3+62R_4),\\
\kappa_6 &= 256(15n-225R_1+1{,}200R_2-2{,}940R_3+3{,}332R_4-1{,}382R_5). \qquad \ldots\ldots(8)
\end{aligned}$$

The first four of equations (5, 6, 7, 8) have already been given in a slightly different form by Haldane (1937). The limiting forms of equations (5) and (7) when s tends to infinity and k to zero, while $ks = g$, have been given by Haldane (1938). However, the expression for κ_6 there given is incorrect. The coefficient of R_1 in the expression for κ_6 should be 124,800.

The extension of equations (7) would be rather tedious. However, those of equations (6) and (8) would not be very difficult. The coefficient of $x^{2r}/2r!$ in the expansion of $\log\cosh t$ is the value of $(d/dx)^{2r-1}(1-\tanh^2 x)$ when $x = 0$, and can easily be calculated, since this differential coefficient is a polynomial in $\tanh x$. The equations for moments in terms of cumulants can easily be extended when all odd cumulants vanish. In this case a useful check can be obtained from the fact that when $s = 2$ the cumulant-generating function of χ^2 is $t + \log\cosh t$.

Summary

Expressions are obtained for the first twelve cumulants of the binomial distribution, and a simple recurrence formula for further cumulants. The first six cumulants of χ^2 for a $(n \times 2)$-fold table when expectations are small, are deduced.

REFERENCES

FISHER, R. A. (1928). "Moments and product moments of sampling distributions." *Proc. Lond. Math. Soc.* **30**, 200–38.

HALDANE, J. B. S. (1937). "The exact value of the moments of the distribution of χ^2, used as a test of goodness of fit, when expectations are small." *Biometrika*, **29**, 133–43.

—— (1938). "The first six moments of χ^2 for an n-fold table with n degrees of freedom when some expectations are small." *Biometrika*, **29**, 389–91.

PEARSON, K. (1919). "Peccavimus" (Footnote, p. 270). *Biometrika*, **12**, 259–81.

ROMANOVSKY, V. (1923). "Note on the moments of a binomial $(p+q)^2$ about its mean." *Biometrika*, **15**, 410–12.

f. 48

[published 1946]