THE MULTINOMIAL SOLID AND THE CHI TEST*

BY

B. H. CAMP

1. INTRODUCTION

Let p_i be the relative frequency of $x_i (i=1, 2, \cdots, m)$ in a frequency distribution. If the number of elements in the distribution be infinite and a sample of size n be drawn, the probability of obtaining exactly s_1 elements of size x_1, and at the same time s_2 of size x_2, etc., is known to be

$$(1) \qquad P = \frac{n!}{s_1! \cdots s_m!} p_1^{s_1} \cdots p_m^{s_m}, \qquad p_1 + p_2 + \cdots + p_m = 1,$$

where the s's are integers whose sum is n. P is a function of $m-1$ independent variables, and is called the multinomial solid. When $m=2$, it degenerates into the general term of the point binomial, $(p_1+p_2)^n$. Although much has been known about this solid, its usefulness has not been sufficiently recognized. In this paper I shall point out certain of its properties and two applications, one to a well known theorem in sampling, and one to the χ-test. The proof of the χ-test here given may be thought of as supplementing Pearson's original proof† in that it replaces an assumed approximation by a demonstrated approximation. It is not related to Pearson's second proof‡ which is a variant of Soper's.§

2. GENERAL PROPERTIES¶

(a) P may be written in the form $P=P_1 \cdots P_m=P_1 \cdots P_{m-1}$; where P_1 is a function of s_1 only, and is a point binomial; P_2 is a function of s_1, s_2 only, and is a point binomial if s_1 be fixed; etc.; P_i is a function of $s_1 \cdots s_i$ only, and is a point binomial if $s_1 \cdots s_{i-1}$ be fixed; $P_m=1$; thus:

---

* Presented to the Society, December 29, 1926; received by the editors in June, 1928.

† K. Pearson, *On deviations from the probable in random sampling*, Philosophical Magazine, (5), vol. 50 (1900), pp. 157–175.

‡ K. Pearson, *On the general theory of multiple contingency*, Biometrika, vol. 11 (1916), pp. 145–158.

§ Soper, *Frequency Arrays*, Cambridge University Press, 1922, p. 15, footnote.

This proof makes other approximations which do not conveniently lend themselves to measurement. I shall be particularly interested here in the degree of approximation attained in the application of the chi test to ordinary problems.

¶ The properties grouped in this section are not essentially new.

$$P = \left[\frac{n!}{s_1!(n-s_1)!}p_1^{s_1}(1-p_1)^{n-s_1}\right]\left[\frac{(n-s_1)!}{s_2!(n-s_1-s_2)!}\left(\frac{p_2}{1-p_1}\right)^{s_2}\cdot\left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-s_1-s_2}\right]$$

$$\left[\frac{(n-s_1-s_2)!}{s_3!(n-s_1-s_2-s_3)!}\left(\frac{p_3}{1-p_1-p_2}\right)^{s_3}\left(\frac{1-p_1-p_2-p_3}{1-p_1-p_2}\right)^{n-s_1-s_2-s_3}\right]\cdots$$

$$\cdot\left[\frac{s_m!}{s_m!(s_m-s_m)!}\left(\frac{p_m}{1-p_1-\cdots-p_{m-1}}\right)^{s_m}=1\right].$$

(b) The form (a) shows that equation (1) might have been derived by the following reasoning: Find the probability of exactly s_1 successes with p_1 as probability and n as the number of drawings; multiply by the probability of exactly s_2 successes with $p_2/(1-p_1)$ as probability and $(n-s_1)$ as the number of drawings; multiply, etc.

Definitions. (α) Any positive integrable function of $(m-1)$ independent variables may be considered an $(m-1)$-*way frequency solid* (or surface or curve), and may be represented in m-dimensional space.

(β) If, in an $(m-1)$-way solid, r variables be fixed, there is left a function of $(m-1-r)$ variables, therefore an $(m-1-r)$-way solid; this is to be called an $(m-1-r)$-*way array*.

Example. If $m=4$, the $(m-1)$-way solid is a 3-way correlation solid, say a function of x, y, and z. If one fixes x and y, thus taking $r=2$, the one-way arrays resulting are straight lines or columns in the z-direction. If, further, one added up the frequencies in each of these columns, one would arrive at a set of totals which might be thought of as spread over the xy-plane, thus comprising a two-way correlation table.

(γ) If one adds the frequencies in each of the $(m-1-r)$-way arrays, the result is a function of r variables; this is an r-*way total*.

(c) **Totals.** If, in P above, the variable s_1 be fixed, and if the frequencies of each of the $(m-2)$-way arrays thus determined be added, the one-way total finally determined is P_1. If both the variables s_1 and s_2 be fixed, and if the frequencies in each of the $(m-3)$-way arrays thus determined be added, the two-way total determined is P_1P_2. If $s_1, s_2, \cdots, s_{m-2}$ be fixed, and the totals of the one-way arrays be found, the $(m-2)$-way total that results is $P_1P_2\cdots P_{m-2}$.

These facts follow easily from the form (a). E.g., fixing s_1,

$$\sum_{s_2}\sum_{s_3}\cdots\sum_{s_{m-1}}P_1\cdots P_{m-1} = P_1\left(\sum_{s_2}P_2\right)\cdots\left(\sum_{s_{m-1}}P_{m-1}\right) = P_1.$$

(d) **Means of totals.** The mean of the one-way total of P in the s_1 direction is p_1n; the mean of the two-way total in s_1s_2-space is (p_1n, p_2n); the general mean of the whole solid is $(p_1n, p_2n, \cdots, p_{m-1}n)$.

These facts follow also from the form (a) and the well known theorems that the coördinates of the general mean of a correlation solid are the means of its one-way totals.

3. Regression surfaces

(a) The mean of the one-way array in the s_{m-1} direction made by fixing $s_1, s_2, \cdots, s_{m-2}$ is

$$\frac{p_{m-1}}{1-p_1-\cdots-p_{m-2}}(n-s_1-s_2-\cdots-s_{m-2}).$$

This may be thought of as a linear function of the $m-2$ variables, $s_1, s_2, \cdots, s_{m-2}$, and thus yields the equation of the regression surface, s_{m-1} on $(s_{m-2}, \cdots, s_1)$.

Use the form of §2(a). The one-way array is $P_1P_2\cdots P_{m-1}$, and its total is, by §2(c), $P_1P_2\cdots P_{m-2}$. So its mean is

$$\frac{1}{P_1\cdots P_{m-2}}\sum_{s_{m-1}} s_{m-1}P_1\cdots P_{m-1} = \sum_{s_{m-1}} s_{m-1}P_{m-1}.$$

The last named sum is the mean number of successes in a point binomial distribution in which $p_{m-1}/(1-p_1-\cdots-p_{m-2})$ is the probability of a success, and $(n-s_1-\cdots-s_{m-2})$ is the number of drawings. This is well known to have the value given above.

(b) Every partial regression is also a linear function. (For one needs only to fix certain of the variables of the regression surface in (a) to obtain the partial regression surface.) It follows from (a) and (b) that, for $m=4$, this solid is an example of the all linear case I have considered elsewhere.*

Application. Let $m=3$. The mean of the one-way array made by fixing s_1 is the following linear function of s_1: $p_2(n-s_1)/(1-p_1)$. This fact, now proved to be true, was chosen by Pearson as a reasonable assumption in his proof of a fundamental theorem in sampling,† viz.: that the coefficient of correlation between two frequencies with which two fixed values of a variable will occur is $-(p_1p_2/(q_1q_2))^{1/2}$.

---

* Camp, *Mutually consistent multiple regression surfaces*, Biometrika, vol. 17 (1925), p. 450.

† K. Pearson, *On the probable errors of frequency constants*, Biometrika, vol. 2 (1902-3), p. 274. It has been copied, as an unproved assumption, in several texts.

4. The Laplace approximation

It has been generally assumed, specifically by Pearson in deriving the χ-test, that the multinomial solid just considered could be approximated when n is large, by the normal solid. I wish to examine the measure of this approximation, and to do so I shall also derive the normal solid from the multinomial solid in a manner similar to that used by Laplace for the one-dimensional case.

In (1) of §1, put $s_i=np_i+k_i$, $i=1, 2, \cdots, m$. Then P becomes

$$(2) \qquad P_{k_1\cdots k_m} = \frac{n!}{(np_1+k_1)!\cdots(np_m+k_m)!}p_1^{np_1+k_1}\cdots p_m^{np_m+k_m},$$

where

$$(3) \quad p_1+\cdots+p_m=1\ ;\ k_1+\cdots+k_m=0, \text{ since } s_1+\cdots+s_m=n\ ; \text{ or}$$

$$(4) \qquad P=n!(r_1\cdots r_m), \quad \text{where} \quad r_i=p_i^{np_i+k_i}/(np_i+k_i)!.$$

Let

$$(5) \qquad \epsilon=1/(np), \quad \omega=k\epsilon=k/(np), \quad \alpha=k\omega=k^2/(np).$$

Use Stirling's approximation, and the symbol $\cong$ to mean "is approximately equal to," and drop the subscript from r temporarily:

$$\begin{aligned}\log r &\cong \log p^{np+k}-(np+k+\tfrac{1}{2})\log(np+k)+(np+k)-\log(2\pi)^{1/2}\\ &=\log(p^{np(1+\omega)}/(2\pi)^{1/2})-[np(1+\omega)+\tfrac{1}{2}][\log np+\omega-\omega^2/2-\omega^3/3-\cdots]\\ &\qquad+np(1+\omega)\\ &=\log(p^{np(1+\omega)}/(2\pi)^{1/2})-\log(np)^{np(1+\omega)+1/2}\\ &\qquad-[np(1+\omega)+\tfrac{1}{2}][\omega-\omega^2/2+\omega^3/3-\cdots]+np(1+\omega)\\ &=A+\log p^{np(1+\omega)}-\log(2n\pi p)^{1/2}(np)^{np(1+\omega)},\end{aligned}$$

where

$$A=np(1+\omega)(1-\omega+\omega^2/2-\omega^3/2+\cdots)+\tfrac{1}{2}(-\omega+\omega^2/2-\omega^3/3+\cdots).$$

So

$$(6) \qquad r\cong e^A[(2n\pi p)^{1/2}n^{np(1+\omega)}]^{-1},$$

and the relative error thus introduced into r is known* to be $1-e^{-t/12}$, where $0<t<1/(np(1+\omega))$. It is now necessary to impose certain restrictions on ϵ, ω, α, m. These will seem arbitrary to the reader, and indeed they are.

---

* The usual statement is $x!=(2\pi x)^{1/2}x^xe^{-x}e^{\theta/(12x)}$, where $0<\theta<1$. Hence, if y is the approximate value, $y=x!e^{-\theta/(12x)}$. The relative error in $x!$ is $(-y+x!)/x!=-y/x!+1=-e^{-\theta/(12x)}+1$, where $0<\theta/x<1/x$.

They have been chosen after considerable experimentation so as to give results as favorable as possible to the general purposes of the χ-test. Doubtless more useful conditions might be found. Certainly it would be easy to discover better conditions if one restricted his view to special cases. For example, the point binomial case ($m=2$) is a very special one. These conditions are not chosen so as to measure well the closeness of the approximation in this case. Let

$$(7)\qquad \text{(a) } np - |k| \geqq 3, \qquad \text{(b) } |k|/(np) = |\omega| < 0.2,$$
$$\text{(c) } k^2/(np) = \alpha < 6, \qquad \text{(d) } m \leqq 20.$$

It follows that $np(1+\omega)=np+k\geqq 3$, and so $t<1/(np(1+\omega))<1/3$; and therefore the maximum relative error in r, as given by (6), is

$$(8)\qquad \delta_1 = 1 - e^{-t/12} \leqq 0.09202/(np + k).$$

Now let $A=np(1-\omega^2/2)+B$, where

$$(9)\qquad B = \omega\alpha\left(\frac{1}{3\cdot 2} - \frac{\omega}{4\cdot 3} + \frac{\omega^2}{5\cdot 4} - \cdots\right) + \frac{\omega}{2}\left(-1 + \frac{\omega}{2} - \frac{\omega^2}{3} + \cdots\right)$$
$$= (\omega/6)(\alpha - 3)(1 - \omega/2) + \omega\alpha(\omega^2/20 - \omega^3/30 + \omega^4/42 - \cdots)$$
$$+ (\omega^3/2)(-1/3 + \omega/4 - \omega^2/5 + \cdots).$$

So

$$(10)\qquad e^A = e^{np(1-\omega^2/2)}(e^B \cong 1 + B),$$

and the error thus made in e^B is $(B^2/2)(1+B/3+B^2/12+\cdots)$.

We can find an upper limit to this by the use of (9) and (7) (b) and (c):

$$|B| < \left|\frac{\omega}{6}(\alpha - 3)\left(1 - \frac{\omega}{2}\right)\right| + \left|\omega^3\left\{\left(\frac{\alpha}{20} - \frac{1}{6}\right)\right.\right.$$
$$\left.\left. - \omega\left(\frac{\alpha}{30} - \frac{1}{8}\right) + \omega^2\left(\frac{\alpha}{42} - \frac{1}{10}\right) - \omega^3\left(\frac{\alpha}{56} - \frac{1}{12}\right) + \cdots\right\}\right|$$
$$< |(\omega/6)(\alpha - 3)(1 - \omega/2)|$$
$$+ |\omega^3|\{1/6 + |\omega/8| + \omega^2/10 + |\omega^3/12| + \cdots\}$$
$$< (1/30)(3)(1.1) + 0.0014 < 0.1114.$$

$$(11)\qquad \text{Let } a = \omega(\alpha - 3)/6. \quad \text{By (7)}, \quad |a| < 0.1.$$

$$(12)\qquad B^2(1 + |B|/3)/2 < \{a^2(1 - \omega/2)^2$$
$$+ 0.000{,}308 + 0.000{,}001{,}96\}(1 + 0.0372 + 0.00104 + 0.0001 + \cdots)/2$$
$$< \{a^2(1 - \omega/2)^2 + 0.000{,}155\}(0.520) \cong 0.52a^2(1 - \omega/2)^2,$$

which we shall take as the maximum absolute error made in e^B by the approximation of (10). We shall now increase this error slightly by cutting off from (10) a part of B, writing $B \cong a$. The additional absolute error thus made is, by (11) and (9), at most equal to

$$\begin{aligned}
&|\,\omega\alpha(-\omega/12+\omega^2/20-\omega^3/30+\omega^4/42-\cdots)\\
&\qquad\qquad +(\omega/2)(\omega/2-\omega^2/3+\omega^3/4-\omega^4/5+\omega^5/6-\cdots)\,|\\
&=\left|\frac{\omega^2}{4}\left(1-\frac{\alpha}{3}\right)+\frac{\omega^3}{2}\left(\frac{\alpha}{10}-\frac{1}{3}\right)+\frac{\omega^4}{2}\left(-\frac{\alpha}{15}+\frac{1}{4}\right)+\frac{\omega^5}{2}\left(\frac{\alpha}{21}-\frac{1}{5}\right)+\cdots\right|\\
&<\omega^2/4+|\,\omega\,|^3/6+\omega^4/8+|\,\omega\,|^5/10+\cdots.
\end{aligned}$$

From this inequality and (12) it follows that the maximum error made in e^B by writing

$$e^B \cong 1+a \tag{13}$$

is

$$0.52a^2(1-\omega/2)^2+\omega^2/4+|\,\omega\,|^3/6+\omega^4/8+\cdots. \tag{14}$$

We need, however, the maximum *relative* error in e^B. To get this we divide (14) by the minimum value of e^B, which is the series $1-0.1114+(0.1114)^2/2-(0.1114)^3/6+\cdots=0.8945$. Hence the maximum relative error in e^B as given by (13) is

$$\begin{aligned}
\delta_2 &= 0.582a^2(1-\omega/2)^2+0.559(\omega^2/2+|\,\omega\,|^3/3+\omega^4/4+\cdots)\\
&<0.582(\omega^2/4)(1+|\,\omega\,|+\omega^2/4)+0.559(\omega^2/2+|\,\omega\,|^3/3+\omega^4/4+\cdots)\\
&=0.425\omega^2+0.332|\,\omega\,|^3+0.176\omega^4+\cdots<0.02.
\end{aligned} \tag{15}$$

By (6), (10), and (13) we have now made altogether the following approximation to r:

$$r \cong \frac{(1+a)e^{np(1-\omega^2/2)}}{(2\pi pn)^{1/2}n^{np(1+\omega)}}. \tag{16}$$

By (8) and (15) the relative error in r thus made is at most

$$\begin{aligned}
\rho &= \delta_1+\delta_2+\delta_1\delta_2\\
&<0.094/(np-|\,k\,|)+0.425\omega^2+0.332|\,\omega\,|^3+0.176\omega^4+\cdots,
\end{aligned} \tag{17}$$

and we shall hereafter omit the terms not printed in this series. The following known theorem has just been used and will be used again shortly: Let δ_i be the relative error in x_i, $i=1,\cdots,m$. The relative error in the product $x_1\cdots x_m$ is

$$\sum\delta_i-\sum\sum\delta_i\delta_j+\sum\sum\sum\delta_i\delta_j\delta_k-\cdots, \qquad \text{where } i\neq j\neq k, \text{ etc.}$$

It now follows that the maximum absolute relative error in the product $r_1 \cdots r_m$ of equation (4) is

$$\sum_1^m \rho_i + m(m-1)(0.0416)^2/2 + m(m-1)(m-2)(0.0416)^3/6 + \cdots \cong \sum_1^m \rho_i + m(m-1)(0.000,866) + m(m-1)(m-2)(0.000,0120). \tag{18}$$

More terms must be added to (18) unless condition (7)(d), that $m \leqq 20$, is satisfied, but it usually is in practice, if the other conditions of (7) are. In equation (4) we now write $n! \cong n^n e^{-n}(2\pi n)^{1/2}$ with a maximum relative error of $1 - e^{-1/(12n)} \cong 1/(12n)$, so that, finally,

$$P_{k_1 \cdots k_m} \cong \phi(1 + a_1) \cdots (1 + a_m), \tag{19}$$

where

$$\phi = \phi_0 e^{-\chi^2/2}, \quad \phi_0 = (2\pi n)^{(1-m)/2}(p_1 \cdots p_m)^{-1/2}, \quad \chi^2 = \sum_1^m k_i^2/np_i;$$

and the maximum relative error in P is numerically less than

$$\gamma = \sum_1^m \rho_i + m(m-1)(0.000,866) + m(m-1)(m-2)(0.000,012) + 1/(12n). \tag{20}$$

Up to this point we have derived in (19) an approximation to the ordinate of the point binomial solid at the point $(k_1, \cdots, k_{m-1})$. When $m=2$ this reduces to Laplace's well known form for the point binomial. What we really seek however is an approximation to the sum of such ordinates. We shall accordingly find the sum of all $P_{k_1 \cdots k_m}$'s for all values of the k's numerically less than a prescribed fixed set. This prescribed set we shall now call $(k_1, \cdots, k_m)$, using $(k_1', \cdots, k_m')$ for the sets of smaller numbers. It is necessary to explain in more detail precisely what ordinates, and parts of ordinates, our sum is to include. The case of the point binomial shall serve as our model. Putting $m=2$ for a moment we have

$$(p_1 + p_2)^n = p_1^n + \cdots + {}_nC_t p_1^s p_2^t + \cdots + p_2^n = u_0 + \cdots + u_t + \cdots + u_n.$$

Let us think of a histogram in which the base of each rectangle is unity and the altitude one of the u's. If it happens that the mean t, $\bar{t} = p_2 n$, is the coördinate of the middle or one end of one of these bases, then when we add a positive number k so that we come to the middle of a base on the right of the mean, it will also happen that when we subtract k we shall arrive at the middle of a base on the left. Then if we add all the ordinates at these two

end points and at the intermediate points we shall have the same result as if we added the corresponding rectangles of the histogram, i.e., we need to add those rectangles which lie in the interval $(\bar{\imath}-k-1/2,\ \bar{\imath}+k+1/2)$. But, if the mean is not so situated, then when we add k so as to arrive at the middle of a base on the right it will happen that when we subtract k we shall not arrive at the middle of a base on the left. The sum desired in this case is defined to be the total area of those rectangles or parts of rectangles included in the same interval as before. Now, in our general case when $m \neq 2$ each base of unit length is replaced by a unit cube of $m-1$ dimensions, and the sum desired is the sum of volumes of parallelepipeds found by multiplying the volume of each of these cubes by the length of the ordinate P supposed erected at the middle point of the cube. In case p_1n_1, p_2n_2, etc., are integers, as happens in our most important application, then whole parallelepipeds only are included in our sum. If not, partial parallelepipeds may be included; the summation in any event shall extend over the intervals

$$(p_in - k - \tfrac{1}{2},\ p_in + k + \tfrac{1}{2}) \qquad (i = 1, \cdots, m-1).$$

We now note from (11) that, if k_i, $i \leqq m-1$, is replaced by $-k_i$, a_{k_i} becomes $-a_{k_i}$, but ϕ is unchanged. To obtain then the approximation corresponding to (19) for any of the points whose coördinates, referred to the general mean of the solid, are $(\pm k_1, \cdots, \pm k_{m-1})$, we have only to change the signs of certain of the a_{k_i}'s, $i \leqq m-1$, and then obtain a_{k_m} from the formula

$$a_{k_m} = \frac{k_m}{6np_m}\left(\frac{k_m^2}{np_m} - 3\right), \quad k_m = -(k_1 + \cdots + k_{m-1})\ ;$$

but we must be careful to insist here that both np_i and k_i shall be integers, for by (2) np_i+k_i was an integer and so now np_i-k_i would have to be. In case these numbers are not integers we shall still assume that the formal approximation holds for those ordinates which are at points having integral coördinates and lying in unit cubes contiguous to the one enclosing the point $(\pm k_1, \cdots, \pm k_{m-1})$; for if n is fairly large the relative errors involved at two such adjacent ordinates are almost exactly the same.

In forming our sum, we begin by adding the P's at all the 2^{m-1} vertices of the outer parallelepiped, i.e., at the points $(\pm k_1, \cdots, \pm k_{m-1})$, and denote this sum by $\overline{P}_{k_1 \cdots k_m}$. The coefficient of ϕ in (19) is

$$(21) \qquad 1 + \sum a_i + \sum a_ia_j + \sum a_ia_ja_m + \cdots + (a_1 \cdots a_m),\ i \neq j \neq k, \text{ etc.}$$

When we form our sum for all the vertices mentioned, certain of the terms that appear like those of (21) will cancel in pairs. To prove this, we note first that, since $k_m = -(k_1 + \cdots + k_{m-1})$, if we change all the signs of

$k_1, k_2, \cdots, k_{m-1}$, we shall change the sign only of k_m, and hence the sign only of a_{k_m}; but that, if we change some only of the signs of $k_1, k_2, \cdots, k_{m-1}$, we may change the numerical value of k_m and hence of a_{k_m}. It follows that by pairing the vertices properly the terms of the type $\sum a_i$ will vanish, also, if m is odd, terms of the type* $(a_1 \cdots a_m)$. To see that this last group vanishes, we note that, if $(k_1, \cdots, k_{m-1})$ is a specified vertex, these k's being positive or negative, there exists also a vertex $(-k_1, \cdots, -k_{m-1})$. For the first vertex our term involves $a_1 \cdots a_m$, and for the second it involves $(-a_1) \cdots (-a_m)$, which equals $-a_1 \cdots a_m$, if m is odd. By analogous reasoning it may be shown now that, by proper pairing, all those terms of the other types involved in (21) will cancel in pairs, except the first, and except perhaps those that contain a_m. Now in $\sum a_i a_j$ there are ${}_{m-1}C_1$ terms that involve a_m, in $\sum a_i a_j a_k$ there are ${}_{m-2}C_2$ such terms, etc. Each of these must be counted 2^{m-1} times. So, finally,

$$\overline{P}_{k_1 \cdots k_m} \leqq (2^{m-1})(1 + 0 + {}_{m-1}C_1 \max|a_m a_i| + {}_{m-2}C_2 \max|a_m a_i a_j| + \cdots)\phi$$

$$\leqq (2^{m-1})\left(1 + \frac{m-1}{100} + \frac{(m-1)(m-2)}{2 \cdot 1000} + \frac{(m-1)(m-2)(m-3)}{3 \cdot 2 \cdot 10{,}000} + \cdots\right)\phi$$

and so

$$\overline{P}_{k_1 \cdots k_m} \cong 2^{m-1}\phi, \tag{22}$$

with a maximum absolute error in $\overline{P}$ of

$$\eta = 2^{m-1}\phi\left(\gamma + \frac{m-1}{100} + \frac{(m-1)(m-2)}{2 \cdot 1000} + \frac{(m-1)(m-2)(m-3)}{3 \cdot 2 \cdot 10{,}000} + \cdots\right)$$

$$\leqq 2^{m-1}\phi\left(\gamma + \frac{m-1}{100} + \frac{(m-1)(m-2)(1.86)}{2000}\right),$$

and therefore

$$\eta \leqq 2^{m-1}\phi\{\gamma + (m-1)/100 + (m-1)(m-2)/1000\}. \tag{23}$$

What we have just done for the parallelepiped $k_1 \cdots k_m$ we may repeat for each of the smaller parallelepipeds $k_1' \cdots k_m'$, obtaining a sum in each case which we may denote by $\overline{P}_{k_1' \cdots k_m'}$. Then by adding all these $\overline{P}$'s together we shall obtain the sum originally sought, viz.:

$$Q = \sum \phi_0 e^{-\chi^2/2}. \tag{24}$$

This is Pearson's expression which he has evaluated by means of a definite integral and tabulated,† but his $(1-P)$ equals my Q, and his n' equals my m.

* This fact is important only when $m=3$.

† Tables, loc. cit., pp. 26 ff.

In order to obtain an estimate of the error in Q we might use the fact that the absolute error in $\overline{P}_{k_1'\cdots k_m'}$ is no greater, *relatively*, than in $\overline{P}_{k_1\cdots k_m}$, but this, although true, would lead us to an inequality too wide to be useful, and so it is better to make an assumption. We shall assume that the *absolute* error in $\overline{P}_{k_1'\cdots k_m'}$ is not greater than in $\overline{P}_{k_1\cdots k_m}$. This is a reasonable assumption, and we know that when the k's are nearly zero, the error also is nearly zero. However, it must be used with caution, for if the k's were allowed to be quite large the value of ϕ would be so exceedingly small that the error η would also be exceedingly small, no matter what the other factors of η might be, and we could not be sure that this absolute error would not be exceeded at points nearer the general mean of the solid; but, as the k's have been considerably restricted by (7), we are not in serious danger on account of that possibility. We shall keep in mind the fact that from now on we are only estimating our maximum error instead of computing it by using the phrase "estimated maximum error." The estimated maximum absolute error in Q as approximated by (23) is $E=\eta(|k_1|+\frac{1}{2})\cdots(|k_{m-1}|+\frac{1}{2})$; or, in a form suitable for computation:

$$
\begin{aligned}
&E = \theta\phi\tau,\ \theta = (|k_1| + \tfrac{1}{2})\cdots(|k_{m-1}| + \tfrac{1}{2})2^{m-1},\\
&\phi = e^{-\chi^2/2}/D,\ D^2 = (np_1)\cdots(np_{m-1})(p_m)(2\pi)^{m-1},\ \chi^2 = \sum_1^m k_i^2/(np_i),\\
&\tau = \sum_1^m \rho_i + (m-1)(0.000,866m + 0.01)\\
&\qquad + (m-1)(m-2)(0.000,012m + 0.001),\qquad (25)\\
&\sum_1^m \rho_i = 0.094\sum_1^m (np_i - |k_i|)^{-1} + 0.425\sum_1^m (k_i/(np_i))^2\\
&\qquad + 0.332\sum_1^m |k_i/(np_i)|^3 + 0.176\sum_1^m (k_i/(np_i))^4.
\end{aligned}
$$

Sometimes it is quite sufficient to know that always $\tau<1.55$. If this gives a sufficiently small value of E, the computation is greatly shortened. Finally, it is necessary to notice that the conditions (7) need not be satisfied exactly in order that the above expression for E may hold approximately, small changes in the conditions (except in (7)(d)) producing small changes in E; but we must not depart from them widely.

5. Numerical examples

The χ-test may now be used to establish rigorously statements which heretofore one could not be truly sure of. As an example we choose the first one used by Pearson in which he appeared to show a bias in dice.* I

* Philosophical Magazine, loc. cit., p. 167.

give below Pearson's figures and also a set of revised figures. My conditions (7) were not nearly satisfied in his last two groups, and so in my proposed revision I have lumped together his last three groups. This of course amounts to a restatement of the whole problem, but the new problem thus created will serve the general purpose Pearson had in view substantially as well as the old one. In this new problem all my conditions are satisfied except for the figures in parentheses, and in these cases they are nearly satisfied, and so, as observed at the close of §4, we may be sure that our E is approximately correct. The second column, marked f, consists of the "observed frequency of dice with five or six points when a cast of twelve dice was made 26,306 times." The column marked np (Pearson's m) consists of the corresponding theoretical frequencies. The first column, marked "dice", indicates the number of dice in a cast which showed 5 or 6 points. Pearson found that $1-Q$ was 0.000,016. This was the probability that a set of dice with no bias would give results as far removed from the theoretical as those observed. It is so small as to make one practically certain that there was bias. But, since in this case we are not sure that it is proper to compute $1-Q$ by the use of the normal function, as he did, I have modified the problem slightly so as to be sure. I find $1-Q=0.000,086$, and, although this is not so small as Pearson's value, it is still small enough to be convincing.

Pearson's Problem				Author's Revision				
Dice	f	np	k	f	k	$np-\|k\|$	ω	α
0	185	203	−18				.0887	1.596
1	1149	1217	−68				.0559	3.800
2	3265	3345	−80				.0239	1.913
3	5475	5576	−101				.0181	1.829
4	6114	6273	−159				.0253	4.030
5	5194	5018	176		No Change		.0351	6.173
6	3067	2927	140				.0478	(6.696)
7	1331	1254	77				.0614	4.728
8	403	392	11				.0281	0.309
9	105	87	18				(.2069)	3.724
10	14	13	1					
11	4	1	3	18	4	10	(.2857)	1.143
12	0	0	0					
$x^2=43.872$, $m=13$, $1-Q=0.000,016$.				$x^2=35.941$, $m=11$, $1-Q=0.000,086$. By (25) $E=0.000,004$; by the shortened form of (25), taking $r<1.55$, $E<0.000,017$.				

It is not always feasible to rephrase a problem in this fashion and still keep the general purpose of the investigation in view. Sometimes, however, another sort of revision is possible, which may as well be illustrated in a further, though in this case needless, refinement of the example above. In the revision of that example the last value of ω was 0.2857, instead of the 0.2 permitted by (7), and accordingly we had to think of E and therefore of $1-Q$ as possibly a little larger than as stated. Let us now decrease this ω to less than 0.2 by arbitrarily changing the observed frequency opposite it from 18 to 16 and compensating for this by adding arbitrarily 2 to the frequency 6144, where np is large and k is negative. The total effect is again to set up a new problem: Pearson's was the first, our revision a second, and now we have a third; but, since no k has been numerically increased in passing from the second to the third, the value of Q for the third is known to be smaller than the value of Q sought in the second, since we would now be adding up frequencies over a parallelepiped enclosed in the one used before. We should thus arrive rigorously at a lower limit for the second Q, i.e., at an upper limit for the second $1-Q$, and this might still be small enough to support the general conclusion that there was bias in the dice. We do here in fact obtain $\chi^2=34.983$, $1-Q=0.000,13$, and so we are sure that the value of $1-Q$ in the second problem was less than 0.000,13. E has very nearly the same value as before.

WESLEYAN UNIVERSITY,
MIDDLETOWN, CONN.