Biometrika Trust

ON THE LIKELIHOOD THAT ONE UNKNOWN PROBABILITY EXCEEDS ANOTHER IN VIEW OF THE EVIDENCE OF TWO SAMPLES.

By WILLIAM R. THOMPSON. From the Department of Pathology, Yale University.

Section 1.

In elaborating the relations of the present communication interest was not centred upon the interpretation of particular data, but grew out of a general interest in problems of research planning. From this point of view there can be no objection to the use of data, however meagre, as a guide to action required before more can be collected; although serious objection can otherwise be raised to argument based upon a small number of observations. Indeed, the fact that such objection can never be eliminated entirely—no matter how great the number of observations— suggested the possible value of seeking other modes of operation than that of taking a large number of observations before analysis or any attempt to direct our course. This problem is more general than that treated in *Section* 2, and is directly concerned with any case where probability criteria may be established by means of which we judge whether one mode of operation is *better* than another in some given sense or not.

Thus, if, in this sense, P is the probability estimate that one *treatment* of a certain class of individuals is *better* than a second, as judged by data at present available, then we might take some monotone increasing function of P, say $f_{(P)}$, to fix the fraction of such individuals to be treated in the *first manner;* until more evidence may be utilised, where $0 \leqslant f_{(P)} \leqslant 1$; the remaining fraction of such individuals $(1 - f_{(P)})$ to be treated in the *second manner*; or we may establish a probability of treatment by the two methods of $f_{(P)}$ and $1 - f_{(P)}$, respectively. If such a discipline were adopted, even though it were not the best possible, it seems apparent that a considerable saving of individuals otherwise sacrificed to the inferior treatment might be effected. This would be important in cases where either the rate of accumulation of data is slow or the individuals treated are valuable, or both.

If we arbitrarily decide to eliminate the second treatment in favour of the first at this time, then the expectation of sacrifice to the inferior treatment would be $(1 - P)$ for *all* subsequently treated individuals; whereas, if, for example, we take $f_{(P)} = P$, the expectation of such sacrifice would be *temporarily*

$$P (1 - P) + (1 - P) P = 2PQ,$$

where $Q = 1 - P$. Obviously, $2PQ \leqslant \frac{1}{2}$ and, if $P \neq \frac{1}{2}$, then $2PQ < \frac{1}{2}$; whence a saving is made in contrast to the so-called *alternate case method*. In the *long run*, if a real preference exists between the two *treatments*, the expected saving by continued application of this method of apportionment rather than by making immediate final decision is sensibly $1 - P$ of individuals subsequently treated.

Obviously, if we are to operate in this manner, we need methods of evaluation of P for *small* as well as *large* numbers of observations. In the latter case many approximate methods are available in all fields although bounds to approximation have not been considered generally.

In *Section* 2 a sampling problem is treated, which is equivalent to a special case, where we are to judge between two rival treatments upon the basis of the probability of occurrence of a given critical event following such treatment. These probabilities are assumed unknown, but denoted by $\tilde{p}_1$ and $\tilde{p}_2$; and it is assumed that, independently for each of these, *a priori* $\tilde{p}_i$ is equally likely to lie in either of any two equal intervals in its possible range, (0, 1). Our available experience consists solely of the data:

Of n_1 individuals treated by the first method, r_1 experienced the critical event and s_1 did not; and of n_2 treated by the second, r_2 and s_2 were the corresponding numbers with respect to the critical event.

In any given case it must be decided whether these requirements are met or not, and whether we may apply the well-known Principle of Bayes to convert the problem to the form of *Section* 2. Statistical criteria are often employed, however, in situations in which certain deviations from the conditions required in their development can be tolerated, when a better procedure is not available.

Section 2.

Consider the case of two infinite populations for which the unknown probabilities of occurrence of a given critical event are $\tilde{p}_1$ and $\tilde{p}_2$, and the probability of obtaining a sample containing r critical occurrences and s failures in $n = r + s$ trials is $\binom{n}{r} \cdot \tilde{p}_i^r (1 - \tilde{p}_i)^s$, where $i = 1$, 2, respectively. Furthermore, assume that one sample has been drawn at random from each population, the respective values of r and s being denoted by r_i and s_i (where $i = 1$, 2) and $n_i = r_i + s_i$; and that *independently* for $i = 1$ or 2 the probability that $\tilde{p}_i$ lies in the interval $(p, p + dp)$ is $P^{(i)}_{p, p+dp}$, where

$$(1) \qquad P^{(i)}_{p, p+dp} = \frac{\int_{p}^{p+dp} \binom{n}{r} \cdot p^r \cdot q^s \cdot dp}{\int_{0}^{1} \binom{n}{r} \cdot p^r \cdot q^s \cdot dp},$$

where $q = 1 - p$, $r = r_i$, $s = s_i$, and $n = r + s$. Then

$$(2) \qquad P^{(i)}_{p, 1} = \frac{(n+1)!}{r! \, . \, s!} \int_{p}^{1} p^r \cdot q^s \cdot dp = \sum_{a=0}^{r} \binom{n+1}{a} \cdot p^a \cdot q^{n+1-a};$$

the last expression having been indicated by K. Pearson* in this relation. In the notation employed by him and by Müller† we may write

$$(3) \qquad 1 - P^{(i)}_{p,1} = P^{(i)}_{0,p} = I_p(u,v) = \frac{B_p(u,v)}{B_1(u,v)} = 1 - I_q(v,u),$$

where $u = r_i + 1$ and $v = s_i + 1$. The object of the present communication is to give a reduced‡ rational algebraic evaluation of the probability ($P_{\tilde{p}_2 > \tilde{p}_1}$) that for the postulated systems $\tilde{p}_2$ exceed $\tilde{p}_1$, and to indicate certain relations between its value (later designated by $\psi_{(r_1, s_1, r_2, s_2)}$) and the sum of the first $r_2 + 1$ terms of a *hypergeometric series* which has appeared in the work of K. Pearson§‖ as well as in the *Incomplete B-* and *I-functions**† of (3).

Obviously, we may write

$$(4) \qquad P_{\tilde{p}_2 > \tilde{p}_1} = \frac{(n_1 + 1)!}{r_1! \, s_1!} \int_0^1 p_1^{r_1} \cdot q_1^{s_1} \cdot \frac{(n_2 + 1)!}{r_2! \, s_2!} \int_{p_1}^1 p_2^{r_2} \cdot q_2^{s_2} \cdot dp_2 \cdot dp_1 \qquad \text{(i)}$$

$$= \frac{(n_1 + 1)!}{r_1! \, s_1!} \int_0^1 \sum_{a=0}^{r_2} \binom{n_2 + 1}{\alpha} \cdot p^{r_1 + a} \cdot q^{s_1 + n_2 + 1 - a} \cdot dp \qquad \text{(ii)}$$

$$= \frac{(n_1 + 1)!}{r_1! \, s_1!} \cdot \sum_{a=0}^{r_2} \binom{n_2 + 1}{\alpha} \cdot \frac{(r_1 + \alpha)! \, (s_1 + n_2 + 1 - \alpha)!}{(n_1 + n_2 + 2)!} \qquad \text{(iii)}$$

$$= \frac{(n_1 + 1)! \, (n_2 + 1)!}{(n_1 + n_2 + 2)!} \cdot \sum_{a=0}^{r_2} \frac{(r_1 + \alpha)! \, (s_1 + n_2 + 1 - \alpha)!}{r_1! \, \alpha! \, . \, s_1! \, (n_2 + 1 - \alpha)!} \qquad \text{(iv)}$$

$$= \frac{(n_1 + 1)! \, (n_2 + 1)!}{(n_1 + n_2 + 2)!} \cdot \sum_{a=0}^{r_2} \frac{(r_1 + r_2 - \alpha)! \, (s_1 + s_2 + 1 + \alpha)!}{r_1! \, (r_2 - \alpha)! \, s_1! \, (s_2 + 1 + \alpha)!}, \qquad \text{(v)}$$

whence we have

$$(5) \qquad P_{\tilde{p}_2 > \tilde{p}_1} = \frac{\displaystyle\sum_{a=0}^{r_2} \binom{r_1 + r_2 - \alpha}{r_1} \cdot \binom{s_1 + s_2 + 1 + \alpha}{s_1}}{\dbinom{n_1 + n_2 + 2}{n_1 + 1}},$$

where, of course, $n_i = r_i + s_i$.

Now, it is obvious that $P_{\tilde{p}_2 > \tilde{p}_1} = P_{\tilde{p}_1 < \tilde{p}_2} = P_{\tilde{q}_1 > \tilde{q}_2} = P_{\tilde{q}_2 < \tilde{q}_1}$, where $\tilde{q}_i = 1 - \tilde{p}_i$, and thus in similar manner we have

$$(6) \qquad P_{\tilde{p}_2 > \tilde{p}_1} = \frac{\displaystyle\sum_{a=0}^{s_1} \binom{s_1 + s_2 - \alpha}{s_2} \cdot \binom{r_1 + r_2 + 1 + \alpha}{r_2}}{\dbinom{n_1 + n_2 + 2}{n_1 + 1}}.$$

Furthermore, $P_{\tilde{p}_2 > \tilde{p}_1} = 1 - P_{\tilde{p}_1 > \tilde{p}_2}$, as the probability that $\tilde{p}_1$ is exactly equal to $\tilde{p}_2$ is zero by hypothesis. Hence we have two other similar sums which may be used with this difference relation to evaluate the probability under consideration.

---

* Pearson, Karl: *Biometrika*, Vol. XVI (1924), pp. 202—203.

† Müller, J. H.: *Biometrika*, Vol. XXII (1930—31), pp. 284—297.

‡ The earliest work directed to this end is discussed by Todhunter. Cf. *A History of the Mathematical Theory of Probabilty*. Cambridge and London (1865), pp. 419—420.

§ Pearson, Karl: *Philosophical Magazine*, Series 6, Vol. 13 (1907), pp. 365—378.

‖ Pearson, Karl: *Biometrika*, Vol. XX^A (1928), pp. 149—174.

In actual evaluation we may make use of the well-known *pyramid form* of tabulation of the binomial coefficients,

$$
\begin{array}{ccccccccccc}
& & & & & 1 & & & & & \\
& & & & 1 & & 1 & & & & \\
& & & 1 & & 2 & & 1 & & & \\
& & 1 & & 3 & & 3 & & 1 & & \\
& 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
1 & 6 & & 15 & & 20 & & 15 & & 6 & 1
\end{array}
$$

(7)

which is readily constructed by the property that each entry except those on the boundaries (which are always unity) is the sum of the two nearest entries of the row next above*. The corresponding factors of the successive terms in the sum to be evaluated may be found in order on given diagonals, ascending in the case of the first and descending in the case of the second factor. The bounding diagonals of units may be deleted in practice (as may all entries to the right of the middle column). Then the first factor of the first term of the sum in (5) is the r_1-th entry in the row whose first entry (at the left) is $r_1 + r_2$, and the corresponding factors of successive terms lie successively above on the diagonal through the first and parallel to the left-hand boundary. Similarly, the other factors are found, but proceeding in the opposite direction on the appropriate diagonal.

Now, for any positive number, x, let $Q_{(x)} = x^{x+\frac{1}{2}}$. Then, by Stirling's formula, we have

(8)
$$
m! = Q_{(m)} \cdot \sqrt{2\pi} \cdot e^{-m} \cdot e^{\frac{\theta}{12m}},
$$

where $0 \leqslant \theta \leqslant 1$; whence (4)(v) gives

(9)
$$
P_{\tilde{p}_2 > \tilde{p}_1} = \frac{Q_{(n_1+1)} \cdot e^{w}}{Q_{(n_1+n_2+2)} \cdot Q_{(r_1)} \cdot Q_{(s_1)}} \cdot \sum_{a=0}^{r_2} \binom{n_2+1}{r_2-a} Q_{(r_1+r_2-a)} \cdot Q_{(s_1+s_2+1+a)},
$$

where
$$
|12w| < \frac{1}{r_1} + \frac{1}{s_1} + \frac{1}{n_1+1};
$$

which may be used to advantage in approximation of $P_{\tilde{p}_2 > \tilde{p}_1}$ when r_1 and s_1 are *large* and r_2 *small*.

Now we may define $\psi_{(r_1, s_1, r_2, s_2)}$ for any four rational integers (not negative) as identical with the right member of (5), where $n_i = r_i + s_i > 0$. Then we have shown several equivalent expressions of the same function and that

(10)
$$
\psi_{(r_1, s_1, r_2, s_2)} = \psi_{(s_2, r_2, s_1, r_1)} = 1 - \psi_{(s_1, r_1, s_2, r_2)}.
$$

* Glaisher, J. W. L.: "A Table of Binomial-Theorem Coefficients," *Messenger of Mathematics*, Vol. 47 (1917), pp. 97—107.

From the conditions stated it may be expected that if we set $p = \dfrac{r_1}{n_1}$ and $q = 1 - p$, then (provided $0 < p < 1$),

(11)
$$\lim_{n_1 \to \infty} \frac{\psi_{(r_1, s_1, r_2, s_2)}}{I_q (s_2 + 1, r_2 + 1)} = 1.$$

That this is true may be verified if we exclude the cases, $p = 0, 1$. Further bounds to approximation of this limit by the ratio $(\bar{R})$ of these functions for given values of n_1 may be found as follows:

By (4)(iv) we may write

(12)
$$\psi_{(r_1, s_1, r_2, s_2)} = \sum_{\alpha=0}^{r_2} \binom{n_2 + 1}{\alpha} \cdot \frac{\dfrac{(r_1 + \alpha)!}{r_1!} \cdot \dfrac{(s_1 + n_2 + 1 - \alpha)!}{s_1!}}{\dfrac{(n_1 + n_2 + 2)!}{(n_1 + 1)!}},$$

and by (2) and (3), introducing the appropriate values of $r = r_2$ and $s = s_2$, we have

(13)
$$I_q (s_2 + 1, r_2 + 1) = \sum_{\alpha=0}^{r_2} \binom{n_2 + 1}{\alpha} \cdot p^\alpha \cdot q^{n_2 + 1 - \alpha},$$

where $p = \dfrac{r_1}{n_1}$ and $q = \dfrac{s_1}{n_1}$. Obviously, therefore, as all terms of both sums in (12) and (13) are positive, if we exclude the special cases where p or $q = 0$, $\bar{R}$ is bounded by the greatest and least values attainable for the ratio of a term in the sum of (12) to the corresponding term in (13). Thus we may define

(14)
$$\omega_1 = \text{Min.} \left[\left(\frac{r_1 + 1}{r_1}\right)^\alpha \left(\frac{s_1 + 1}{s_1}\right)^{n_2 + 1 - \alpha} \left(\frac{n_1}{n_1 + n_2 + 2}\right)^{n_2 + 1} \right]$$

and
$$\omega_2 = \text{Max.} \left[\left(\frac{r_1 + \alpha}{r_1}\right)^\alpha \left(\frac{s_1 + n_2 + 1 - \alpha}{s_1}\right)^{n_2 + 1 - \alpha} \left(\frac{n_1}{n_1 + 2}\right)^{n_2 + 1} \right]$$

for $0 \leqslant \alpha \leqslant r_2$; and, obviously, then

(15)
$$\omega_1 < \bar{R} < \omega_2, \text{ and } \lim_{n_1 \to \infty} [\bar{R}] = 1.$$

In the excluded cases it is also readily verified that

(16)
$$\lim_{n_1 \to \infty} \left[\psi_{(r_1, s_1, r_2, s_2)} \right] = I_q (s_2 + 1, r_2 + 1).$$

The relation of this function to the sum of a given number of consecutive terms of a *hypergeometric series* is particularly interesting in view of the occurrence of such series in the investigations of K. Pearson [*][†]. By (4)(iv) we may write:

(17)
$$\psi_{(r_1, s_1, r_2, s_2)} = \frac{(n_1 + 1)!(n_2 + 1)!}{(n_1 + n_2 + 2)! \, r_1! \, s_1!} \cdot \sum_{\alpha=0}^{r_2} \frac{(r_1 + \alpha)!(s_1 + n_2 + 1 - \alpha)!}{\alpha! \, (n_2 + 1 - \alpha)!},$$

[*] Pearson, Karl: *Philosophical Magazine*, Series 6, Vol. 13 (1907), pp. 365—378.

[†] Pearson, Karl: *Biometrika*, Vol. xx^A (1928), pp. 149—174.

which is the sum of the first $r_2 + 1$ terms of a *hypergeometric series* multiplied by a constant. Similarly, we may write

$$(18) \quad \psi_{(s_1, r_1, s_2, r_2)} = \frac{(n_1 + 1)!\,(n_2 + 1)!}{(n_1 + n_2 + 2)!\,r_1!\,s_1!} \cdot \sum_{\alpha=0}^{s_2} \frac{(s_1 + \alpha)!\,(r_1 + n_2 + 1 - \alpha)!}{\alpha!\,(n_2 + 1 - \alpha)!}$$

$$= \frac{(n_1 + 1)!\,(n_2 + 1)!}{(n_1 + n_2 + 2)!\,r_1!\,s_1!} \cdot \sum_{\alpha=0}^{s_2} \frac{(r_1 + r_2 + 1 + \alpha)!\,(s_1 + s_2 - \alpha)!}{(r_2 + 1 + \alpha)!\,(s_2 - \alpha)!}$$

$$= \frac{(n_1 + 1)!\,(n_2 + 1)!}{(n_1 + n_2 + 2)!\,r_1!\,s_1!} \cdot \sum_{\alpha=r_2+1}^{n_2+1} \left[\frac{(r_1 + \alpha)!\,(s_1 + n_2 + 1 - \alpha)!}{\alpha!\,(n_2 + 1 - \alpha)!} \right],$$

obviously (as $n_2 = r_2 + s_2$); and by previous demonstration

$$(19) \qquad\qquad \psi_{(r_1, s_1, r_2, s_2)} + \psi_{(s_1, r_1, s_2, r_2)} = 1,$$

whence

$$(20) \qquad \frac{(n_1 + n_2 + 2)!\,r_1!\,s_1!}{(n_1 + 1)!\,(n_2 + 1)!} = \sum_{\alpha=0}^{n_2+1} \frac{(r_1 + \alpha)!\,(s_1 + n_2 + 1 - \alpha)!}{\alpha!\,(n_2 + 1 - \alpha)!}.$$

The last relation, demonstrated above by independent proof, has been established previously by Pearson[*][†] (with different notation). Thus we may regard $\psi_{(r_1, s_1, r_2, s_2)}$ as defined by the identity

$$(21) \qquad \psi_{(r_1, s_1, r_2, s_2)} \equiv \frac{\displaystyle\sum_{\alpha=0}^{r_2} \frac{(r_1 + \alpha)!\,(s_1 + n_2 + 1 - \alpha)!}{\alpha!\,(n_2 + 1 - \alpha)!}}{\displaystyle\sum_{\alpha=0}^{n_2+1} \frac{(r_1 + \alpha)!\,(s_1 + n_2 + 1 - \alpha)!}{\alpha!\,(n_2 + 1 - \alpha)!}},$$

extending the domain of definition to include the value, $r_2 = n_2 + 1$; but retaining the restrictions, $n_i = r_i + s_i \geqslant 0$, and that -1 be the least value of r_1, s_1, r_2, and s_2 (only one of which shall be admitted to be negative). Then, by this extension, we have

$$(22) \qquad \psi_{(r_1, s_1, 0, 0)} \equiv \frac{s_1 + 1}{r_1 + s_1 + 2}, \text{ and } \psi_{(r_1, s_1, n_2+1, -1)} \equiv 1,$$

which lie outside the domain of the initial discussion, and we extend to the new domain the relation of (10) *formally*; i.e.

$$(23) \qquad \psi_{(r, s, r', s')} \equiv \psi_{(s', r', s, r)} \equiv 1 - \psi_{(s, r, s', r')}.$$

K. Pearson[†] has considered the problem of likelihood of various values of R and S, the number of *marked* members and *unmarked* members, respectively, in a finite universe of aggregate number, $N = R + S$; assuming N fixed and all values of $R, S \geqslant 0$ equally likely *a priori* and that our sole experience from which judgment is to be made is that a *random sample* has been drawn containing exactly r marked and s unmarked members (R and S being used here in place of Pearson's p and q to avoid confusion). Then by (iii) and (iv) of the article[†] just mentioned, we have a means of evaluating the probability, $\bar{P}_R$, that the universe contains no more than R marked members by the relation,

$$(24) \qquad \bar{P}_R \equiv \psi_{(r, s, R-r, N-R-s-1)},$$

[*] Pearson, Karl: *Philosophical Magazine*, Series 6, Vol. 13 (1907), pp. 365—378.
[†] Pearson, Karl: *Biometrika*, Vol. xx^A (1928), pp. 149—174.

which may be verified readily. Similarly, in the case of the problem considered earlier by Pearson*—having drawn one random sample from a certain infinite population, the sample containing exactly r' *marked* and s' *unmarked* members, we are required to find the probability (under the given conditions) that if we draw another random sample of n'' individuals from the same population it will contain no more than r'' marked members—the required value is given for $r'' \leqslant n''$ by

$$\psi_{(r', s', r'', n''-r''-1)}.$$

In the tabulation of values of $\psi_{(r, s, r', s')}$ for ascending values of the arguments the work may be greatly simplified by certain relations in addition to those given in (22) and (23) in much the same manner as the binomial coefficients may be tabulated† by mere summation of two values already given. To this end let us examine two functions defined by

$$(25) \qquad N_{(r, s, r', s')} \equiv \sum_{\alpha=0}^{r'} \binom{r + r' - \alpha}{r} \binom{s + s' + 1 + \alpha}{s}$$

and $$D_{(r, s, r', s')} \equiv D_{(n, n')} \equiv \binom{n + n' + 2}{n + 1} \quad \text{for } n = r + s, \text{ and } n' = r' + s'.$$

Obviously, by the original definition of $\psi_{(r, s, r', s')}$, extended in (21), (22) and (23), then

$$(26) \qquad \psi_{(r, s, r', s')} \equiv \frac{N_{(r, s, r', s')}}{D_{(r, s, r', s')}},$$

where we extend the definition of (25) for $r, s, r', s' \geqslant 0$ by

$$(27) \qquad N_{(r, s, -1, s')} \equiv 0, \quad \text{and} \quad N_{(r, s, r', -1)} \equiv D_{(r+s, r'-1)},$$

and by (26) and (23) we have

$$(28) \qquad N_{(r, s, r', s')} \equiv N_{(s', r', s, r)} \equiv D_{(r+s, r'+s')} - N_{(s, r, s', r')},$$

as it is obvious that $D_{(n, n')} \equiv D_{(n', n)}$. Furthermore, by the well-known relation,

$$(29) \qquad \binom{a}{b} \equiv \binom{a-1}{b-1} + \binom{a-1}{b},$$

$$D_{(n, n')} \equiv D_{(n, n'-1)} + D_{(n-1, n')},$$

we have for $s > 0$ in (25)

$$(30) \qquad N_{(r, s, r', s')} \equiv \sum_{\alpha=0}^{r'} \binom{r + r' - \alpha}{r} \left[\binom{s + s' + \alpha}{s - 1} + \binom{s + s' + \alpha}{s} \right],$$

whence we have in any case under its definition, obviously, by (20), (27) and (28), the identities

$$(31) \qquad N_{(r, s, r', s')} \equiv N_{(r, s-1, r', s')} + N_{(r, s, r', s'-1)}$$
$$\equiv N_{(r-1, s, r', s')} + N_{(r, s, r'-1, s')}.$$

By (25) and (29), obviously, the same relation holds for the D-function; and we may write

$$(32) \qquad \psi_{(r, s, r', s')} \equiv \frac{N_{(r, s-1, r', s')} + N_{(r, s, r', s'-1)}}{D_{(r, s-1, r', s')} + D_{(r, s, r', s'-1)}}.$$

* Pearson, Karl: *Philosophical Magazine*, Series 6, Vol. 13 (1907), pp. 365—378.

† Glaisher, J. W. L.: "A Table of Binomial-Theorem Coefficients," *Messenger of Mathematics*, Vol. 47 (1917), pp. 97—107.

By means of these relations it is evident that it suffices to tabulate the N and the D-functions by additions of corresponding pairs of values already listed or readily obtained by the relations of (28) if we proceed from the lowest values of the arguments upward; and we need list values only for the cases where $r + s \geqslant r' + s'$ and $r \geqslant s$ where all four variables may be restricted to positive values, as by (25) we have

(33) $$N_{(r, s, 0, s')} \equiv D_{(s-1, s')}.$$

By (25) we may write

(34) $$N_{(r, s, r', s')} \equiv \sum_{a=0}^{r'} D_{(r-1, r'-1-a)} \cdot D_{(s-1, s'+a)},$$

which is of value if $D_{(n, n')}$ is tabulated for $n \geqslant n'$ through ascending values of n and n'. This may be done rapidly by means of (29) and the relation $D_{(n, n')} \equiv D_{(n', n)}$ and is more convenient than the pyramid form of the corresponding binomial coefficients. In the tabulation we may restrict n and n' to the positive integers, employing the relation $D_{(n, 0)} \equiv n + 2$. A short table of the N and D functions is appended as an illustration, the required probability being the ratio of these corresponding values.

Section 3.

If a system of operation such as suggested in *Section* 1 were adopted extensively for the case considered in *Section* 2, reference to values of $\psi_{(r, s, r', s')}$ for small values of the arguments should be required frequently; and, accordingly, a simple method of formation of a table of these would be valuable. The method given in *Section* 2 seems to serve this purpose; and, in conjunction with the relations of (28), many values need not be listed. The short table is given merely in illustration. It really deals completely with all cases of $n, n' \leqslant 5$ although certain cases are not listed where the values are readily obtainable from those given and (28). The several general evaluations at the head of the table would permit deletion of many more, e.g. any instance where one of the four arguments is zero; but they have been retained for illustrative purposes.

The function, $D_{(n, n')} \equiv \binom{n + n' + 2}{n + 1}$, is readily tabulated in a convenient form for increasing values of $n \geqslant n' \geqslant 1$, as has been mentioned under (34) above by adding to $n + 2$ successively the values of $D_{(n-1, n')}$, already listed, and taking a sub-total after each addition, and finally doubling the last sub-total. These sub-totals and the final double are the required successive values of $D_{(n, n')}$. The value of such a table extends far beyond that of the immediate problem; and, by means of it and relation (34), $\psi_{(r, s, r', s')}$ may be calculated rapidly or approximated with any required precision for considerably higher values of the arguments than it may be convenient to have tabulated $N_{(r, s, r', s')}$. In accord with some prescribed tolerance and limited extent, a table of approximate values of the D-function could be made with greater ease, but apparently not readily extensible within the same relative tolerance without revision. Just how far these tables should extend would depend upon demands for their use. All questions as to approximate methods should be decided

Short Table of $N_{(r,s,r',s')}$ and $D_{(n,n')}$ ($n = r + s$, and $n' = r' + s'$)

r	s	r'	s'	N	D
a	b	c	d	$\bar{N}$	$\bar{D}$
b	a	d	c	$\bar{D}-\bar{N}$	$\bar{D}$
d	c	b	a	$\bar{N}$	$\bar{D}$
r	s	-1	s'	0	
r	0	0	0	1	r+2
r	0	0	s'	1	
r	0	1	s'	r+2	
r	s	0	0	s+1	n+2
r	s	0	s'	$D_{(s-1,\,s')}$	
0	0	-1	0	0	1
0	0	0	0	1	2
1	0	0	0	1	3
0	1	0	0	2	
1	0	0	1	1	6
1	0	1	0	3	
2	0	0	1	1	10
2	0	1	0	4	
1	1	0	1	3	
2	0	0	2	1	20
2	0	1	1	4	
3	0	0	1	1	15
3	0	1	0	5	
2	1	0	1	3	
2	1	1	0	9	
3	0	0	2	1	35
3	0	1	1	5	
3	0	2	0	15	
2	1	0	2	4	
2	1	1	1	13	
2	1	2	0	25	
3	0	0	3	1	70
3	0	1	2	5	
3	0	2	1	15	
2	1	1	2	17	
4	0	0	1	1	21
4	0	1	0	6	
3	1	0	1	3	
3	1	1	0	11	
2	2	0	1	6	
4	0	0	2	1	56
4	0	1	1	6	
4	0	2	0	21	
3	1	0	2	4	
3	1	1	1	16	
3	1	2	0	36	
2	2	0	2	10	
2	2	1	1	28	
4	0	0	3	1	126
4	0	1	2	6	
4	0	2	1	21	
4	0	3	0	56	
3	1	0	3	5	
3	1	1	2	21	
3	1	2	1	51	
3	1	3	0	91	
2	2	0	3	15	
2	2	1	2	45	

r	s	r'	s'	N	D
4	0	0	4	1	252
4	0	1	3	6	
4	0	2	2	21	
4	0	3	1	56	
3	1	1	3	26	
3	1	2	2	66	
5	0	0	1	1	28
5	0	1	0	7	
4	1	0	1	3	
4	1	1	0	13	
3	2	0	1	6	
3	2	1	0	18	
5	0	0	2	1	84
5	0	1	1	7	
5	0	2	0	28	
4	1	0	2	4	
4	1	1	1	19	
4	1	2	0	49	
3	2	0	2	10	
3	2	1	1	34	
3	2	2	0	64	
5	0	0	3	1	210
5	0	1	2	7	
5	0	2	1	28	
5	0	3	0	84	
4	1	0	3	5	
4	1	1	2	25	
4	1	2	1	70	
4	1	3	0	140	
3	2	0	3	15	
3	2	1	2	55	
3	2	2	1	115	
3	2	3	0	175	
5	0	0	4	1	462
5	0	1	3	7	
5	0	2	2	28	
5	0	3	1	84	
5	0	4	0	210	
4	1	0	4	6	
4	1	1	3	31	
4	1	2	2	91	
4	1	3	1	196	
4	1	4	0	336	
3	2	0	4	21	
3	2	1	3	81	
3	2	2	2	181	
3	2	3	1	301	
3	2	4	0	406	
5	0	0	5	1	924
5	0	1	4	7	
5	0	2	3	28	
5	0	3	2	84	
5	0	4	1	210	
4	1	1	4	37	
4	1	2	3	112	
4	1	3	2	252	
3	2	2	3	262	

by several statisticians in consultation at a time when a definite programme for the use of these methods is formed.

In (24) and the paragraph in which it stands is given the relation between the hypergeometrical series studied by Pearson*† and $\psi_{(r,s,r',s')}$ by means of which it is obvious that any approximation methods valid for estimation of $\bar{P}_R$ in (24) are equally valid for the estimation of the corresponding ψ-function, several of which have been suggested by K. Pearson. The I-function of Pearson is related to the ψ-function by (11) to (15) also; and another approximation of $\psi_{(r_1,s_1,r_2,s_2)}$ is given in (9) with indicated domain of validity.

A further treatment of the ψ-function and the method of apportionment will be provided in a later paper.

* Pearson, Karl: *Philosophical Magazine*, Series 6, Vol. 13 (1907), pp. 365—378.

† Pearson, Karl: *Biometrika*, Vol. xx^A (1928), pp. 149—174.