ON THE NORMAL APPROXIMATION TO THE HYPERGEOMETRIC DISTRIBUTION[1]

By W. L. Nicholson

University of Illinois

1. Summary. In this paper a new normal approximation to a sum of hypergeometric terms is derived, which is a direct generalization of Feller's normal approximation to the binomial distribution [2]. For intervals that are asymmetric with respect to the mean, or when the distribution is skewed, the new approximation is a marked improvement over the classical procedure.

The hypergeometric distribution is discussed in Section 2, along with the classical norming and the resulting approximation. Feller's remarkable normal approximation for the related binomial distribution is given in Section 3 with an indication of how it can be extended to cover the hypergeometric case. The result of such an extension is presented in Theorem 2 of Section 4. This theorem gives upper and lower bounds on the hypergeometric sum and hence provides a useful estimate of the relative error. Preliminary results to proving Theorem 2 are exhibited in Section 5. The proof follows in Section 6.

2. Introduction. Let π be a finite population of N elements, D of which possess a specified characteristic S. In a random sample of size $n (n \leqq N)$, drawn without replacement from π, the probability G_k that exactly k of the n elements possess S is given by the hypergeometric function. Defining $H_{\lambda,\nu}$ as the probability that k satisfies the inequality $\lambda \leqq k \leqq v$, we have, symbolically,

$$(1) \qquad G_k = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}, \qquad H_{\lambda,\nu} = \sum_{k=\lambda}^{\nu} G_k .$$

The mean μ and the variance σ_h^2 of the distribution (1) are given by

$$(2) \qquad \mu = n\frac{D}{N} \quad \text{and} \quad \sigma_h^2 = \frac{n(N-n)}{N-1}\frac{D}{N}\frac{N-D}{N}.$$

If N, D, n, and k increase without bound in such a manner that

$$(3) \qquad \frac{D}{N} \to \text{limit}, \qquad \frac{n}{N} \to \text{limit}, \qquad \text{and } z_k = (k-\mu)\sigma_h^{-1} \to z, \text{ say},$$

then (see [1], page 146),

$$(4) \qquad G_k \sim (2\pi)^{-1/2}\sigma_h^{-1}e^{-z^2/2},$$

Received March 28, 1955; revised September 6, 1955.

[1]

where the symbol "$\sim$" means that the ratio of the two sides tends to one as the arguments increase. As a consequence of (4), approximations to G_k and $H_{\lambda,\nu}$ of (1) for large N, D, n, and k are

$$(5) \quad \eta_k = (2\pi)^{-1/2}\sigma_h^{-1}e^{-z_k^2/2}, \quad \text{and} \quad \pi_{\lambda,\nu} = \Phi\left(z_\nu + \frac{1}{2\sigma_h}\right) - \Phi\left(z_\lambda - \frac{1}{2\sigma_h}\right),$$

respectively, where

$$(6) \qquad \Phi(z) = (2\pi)^{-1/2}\int_{-\infty}^{z} e^{-t^2/2}\,dt.$$

Since (4) is an asymptotic result, we naturally are interested in the magnitude of the error involved when finite values of N, D, n, and k are used. The maximum error in the approximation (5) to (1) is $O(\sigma_h^{-1})$. For most values of z_k^2 (excluding only those values of z_k^2 that are close to zero), the contribution of the corresponding G_k terms to the sum $H_{\lambda,\nu}$ is negligible in comparison to σ_h^{-1}. Now, z_k^2 will be large if $|k - \mu|$ is large or if DN^{-1} is close to zero or one. Hence, for the cases of primary interest—an evaluation of the tail of the distribution (1) and an evaluation of (1) when a small percentage of the elements of π possess or do not possess S, as the case may be—the approximation (5) leaves much to be desired.

The above two instances will tend to invalidate the fit of any normal approximation to (1), since they constitute cases of extreme deviation from normality. For an approximation to (1) to be useful, it should be accompanied by a concrete bound on the error involved, preferably, a bound on the relative error that would not be affected seriously by the above extreme cases. Such a bound, as a function of N, D, n, λ, and ν, would explicitly tell in any given situation whether N, D, and n were sufficiently large to give the desired accuracy.

Approximations of the type in (5) that are functions of linear limits possess error terms which for the above extreme cases are at least $O(\sigma_h^{-1})$ over a uniformly bounded interval, an interval which does not increase with σ_h. Outside of this interval, the error is even larger. In an attempt to improve on the approximation (5), we consider the case where the limits are quadratic polynomials. The impetus for such an approach is due to the remarkable result of Feller [2] for the related problem of normal approximations to the binomial distribution. Since our development depends heavily on that of Feller, we include his result in detail.

3. Feller's result. For fixed n and $0 < p < 1$, $q = 1 - p$, Feller's designation of the binomial distribution is

$$(7) \qquad T_k = \binom{n}{k} p^k q^{n-k}, \qquad P_{\lambda,\nu} = \sum_{k=\lambda}^{\nu} T_k.$$

Set

$$(8) \qquad x_{1,k} = \{k + \tfrac{1}{2} - (n+1)p\}\sigma_1^{-1} \quad \text{and} \quad \sigma_1^2 = (n+1)pq.$$

Replacing the orthodox norming, $(k - np)(npq)^{-1/2}$, by $x_{1,k}$, Feller derives an exponential expansion for T_k which lacks the troublesome square root factor present in the classical expansion about $(k - np)(npq)^{-1/2}$. (For a discussion of the classical procedure see [1], Chap. 7, Sec. 2.) Using the new expansion (see Theorem 3 of this paper), he obtains upper and lower bounds for $P_{\lambda,\nu}$ as normal integrals with quadratic limits in $x_{1,\lambda}$ and $x_{1,\nu+1}$, respectively. The unique feature of the approximation is that the gap between the two bounds remain $O(\sigma_1^{-1})$ throughout an interval which increases with σ_1. Moreover, a useful upper bound on the relative error is provided.

Let $6a_1 = p - q$. Feller's normal approximation to the binomial distribution is contained in the following

THEOREM 1. *Suppose that*

$$\sigma_1 > 3 \tag{9}$$

and

$$\lambda \geqq (n+1)p, \qquad \nu + \tfrac{1}{2} \leqq (n+1)p + 2\sigma_1^2/3. \tag{10}$$

Then,

$$P_{\lambda,\nu} \leqq e^{5(1-pq)/36\sigma_1^2}\{\Phi(\eta_{\nu+1}) - \Phi(\eta_\lambda)\} \tag{11}$$

if

$$\eta_k = \frac{k-(n+1)p}{\sigma_1} + \frac{a_1}{\sigma_1}\left\{\frac{k-(n+1)p}{\sigma_1}\right\}^2 + \frac{2a_1}{\sigma_1} - \frac{1}{2\sigma_1^2}, \tag{12}$$

whereas the inequality in (11) *is reversed if*

$$\eta_k = \frac{k-(n+1)p}{\sigma_1} + \frac{a_1}{\sigma_1}\left\{\frac{k-(n+1)p}{\sigma_1}\right\}^2 + \frac{2a_1}{\sigma_1} + \frac{M_1}{6\sigma_1} + \frac{1}{7\sigma_1}, \tag{13}$$

where

$$M_1 = \frac{x_{1,\nu}^3}{\sigma_1} = \{\nu + \tfrac{1}{2} - (n+1)p\}^3\sigma_1^{-4}. \tag{14}$$

It should be stressed that this approximation holds for all combinations of n and p for which $(n+1)pq > 9$. λ must only be larger than the central value, and ν smaller than a monotone increasing function of σ_1, which for $\sigma_1 = 3$ is more than two standard units above the central value. An analogous result holds for (λ, ν) intervals to the left of the central value. The gap between the bounds is $O(\sigma_1^{-1})$, as long as $x_{1,\nu}^3 = O(\sigma_1)$, which covers most cases of interest.

Returning to the hypergeometric problem, we note that if N and D are large relative to n, sampling without replacement is closely approximated by sampling with replacement. In this case, (1) differs little from the binomial distribution (7), with $p = DN^{-1}$. This suggests that Feller's result could be generalized to the hypergeometric distribution if the ratio $G_kT_k^{-1}$ (with $p = DN^{-1}$) could be written in a suitable manner as an exponential expansion of the type (31). In

Section 5 it is shown that by slightly altering the definition of p to $p = (D + 1)(N + 2)^{-1}$, the corresponding ratio $G_k T_k^{-1}$ does have such an expansion. Multiplication by the Feller expansion for T_k gives an expansion of G_k, which admits almost identical treatment as that used by Feller to approximate T_k.

4. Normal approximation to hypergeometric. In order to simplify the notation, we introduce several auxiliary functions. Let

$$p = \frac{D+1}{N+2}, \quad q = 1 - p; \quad s = \frac{n+1}{N+2}, \quad t = 1 - s. \tag{15}$$

Thus, for large values of D, N, and n, p and s are approximately the proportion of elements in π that possess S and the sampling fraction, respectively. Set

$$a = \tfrac{1}{6}(p - q)(t - s). \tag{16}$$

For each value of k, define

$$x_k = [k + \tfrac{1}{2} - (n+1)p]\sigma^{-1}, \qquad \sigma^2 = (n+1)pqt. \tag{17}$$

The normal approximation for the hypergeometric distribution that is derived in this paper can now be stated in a form similar to that for the binomial distribution. The only changes are those due to the finite population.

THEOREM 2. *Suppose that*

$$\sigma > 3 \tag{18}$$

and

$$\lambda \geqq (n+1)p, \quad \nu + \tfrac{1}{2} \leqq (n+1)p + \tfrac{2}{3}\sigma^2, \quad n - \nu \geqq 4, \quad D - \nu \geqq 4. \tag{19}$$

Then,

$$H_{\lambda,\nu} \leqq \left(\frac{N+1}{N+2}\right) e^R \{\Phi(\eta_{\nu+1}) - \Phi(\eta_\lambda)\}, \tag{20}$$

where

$$R = \frac{5(1 - pq)(1 - st)}{36\sigma^2} + \frac{2}{3(N+2)} \tag{21}$$

and

$$\eta_k = \frac{k - (n+1)p}{\sigma} + \frac{a}{\sigma}\left\{\frac{k - (n+1)p}{\sigma}\right\}^2 + \frac{2a}{\sigma} - \frac{1}{2\sigma^2}, \tag{22}$$

whereas the inequality in (20) *is reversed if*

$$\eta_k = \frac{k - (n+1)p}{\sigma} + \frac{a}{\sigma}\left\{\frac{k - (n+1)p}{\sigma}\right\}^2 + \frac{2a}{\sigma} + \frac{M}{6\sigma} + \frac{1}{7\sigma}, \tag{23}$$

where

$$M = \frac{x_\nu^3}{\sigma} = \{\nu + \tfrac{1}{2} - (n+1)p\}^3 \sigma^{-4}. \tag{24}$$

The remarks immediately following Theorem 1 pertain equally to Theorem 2 if $(n + 1)pq$, σ_1, and $x_{1,\nu}$ are replaced by $(n + 1)pqt$, σ, and x_ν, respectively.

As a point estimate for $H_{\lambda,\nu}$, we suggest the right side of (20), with η_k defined by (22) plus $(2\sigma^2)^{-1}$. Designating this estimate, the lower bound and the upper bound by $\hat{H}$, L, and U, respectively, we obtain the following upper bound on the relative absolute error,

$$\frac{1}{L} \max [\hat{H} - L, U - \hat{H}]. \tag{25}$$

As an example of the increased accuracy afforded by the estimation procedure of Theorem 2 over that of (5), we consider the case $N = 5000$, $D = 500$, $n = 500$, $\lambda = 51$, and $\nu = 56$. Then, $\sigma^2 > 40$, which certainly satisfies (18). The correct value is $H_{51,56} \approx 0.30847$. Theorem 2 gives the bounding interval as (0.30426, 0.31050) with the point estimate $\hat{H} \approx 0.30770$. By using (25), the upper bound on the relative error is 0.92 per cent (calculation shows it to be 0.25 per cent). The classical procedure (5) estimates 0.31513 with a relative error of 2.16 per cent, about nine times as large as that for $\hat{H}$.

We can not expect the discrepancy to always favor our new procedure to such a marked degree. The symmetric cases when p and s are close to one-half (i.e., when a is close to zero) serve to illustrate this. Here, the limits (22) and (23) of Theorem 2 are almost linear functions of x. We can expect the two estimation schemes to give essentially the same result, and there is no guarantee that the estimate of Theorem 2 will be better. As an example of the symmetric case, we consider a case of perfect symmetry, $a = 0$. Let $N = 400$, $D = n = 200$, $\lambda = 101$, and $\nu = 105$. $\sigma^2 > 25$, which satisfies (18). The true value is $H_{101,105} \approx 0.32452$. Theorem 2 gives the bounding interval (0.31832, 0.32822) with 0.32476 as the point estimate. The bound on the relative error is 2.02 per cent, while the actual relative error is 0.07 per cent. The orthodox estimate (5) is 0.32426 with a relative error of 0.08 per cent. While the two estimates do not differ significantly, we still have the added attraction of the bounding interval provided by the new procedure.

As a rule of thumb, we suggest the use of Theorem 2 when the distribution (1) is skewed (i.e., when a is not close to zero). If only a point estimate is wanted for $H_{\lambda,\nu}$, the symmetric case can probably be handled just as effectively with the classical procedure (5).

5. Hypergeometric expansion. The following two lemmas and Theorem 3 are due to Feller [2]. We state them here for the sake of completeness (for proofs, see [2]). In the process of approximating $H_{\lambda,\nu}$, sums must be replaced by integrals of the normal type. Lemma 1 expresses the normal integral in a form that will be useful in this connection.

LEMMA 1. *For* $0 < h < 1$ *and* $|xh| < 1.4$,

$$\int_{x-h/2}^{x+h/2} e^{-u^2/2}\, du = h \exp \{-x^2/2 + (x^2 - 1)h^2/24 + \omega h^4\}, \tag{26}$$

with

$$-x^4/880 \leqq \omega \leqq 1/264. \tag{27}$$

We have slightly relaxed Feller's condition of $|xh| < 1$. As originally stated in [2], the lemma is not sufficient for our purpose. More care in handling Feller's inequalities shows that (16) of [2] is valid for $0 \leqq \alpha \leqq 0.7$. The remainder of the proof is identical to that of [2]. A modified form of Stirling's formula is provided by Lemma 2. This will be useful in expanding G_k as an exponential series.

LEMMA 2. *For* $n \geqq 4$,

$$n! = (2\pi)^{1/2}(n + \tfrac{1}{2})^{n+1/2} \exp\left\{-(n + \tfrac{1}{2}) - \frac{1}{24(n + \frac{1}{2})} + \frac{7}{2880}\frac{1 + \phi_1}{(n + \frac{1}{2})^3}\right\}, \tag{28}$$

or

$$n! = (2\pi)^{1/2}n^{n+1/2} \exp\left\{-n + \frac{1}{12n} - \frac{1 + \phi_2}{360n^3}\right\}, \tag{29}$$

where

$$|\phi_i| < \tfrac{1}{6}, \qquad \phi_i \to 0 \quad as \quad n \to \infty. \tag{30}$$

Feller's exponential expansion of the binomial distribution (7) is given by the following

THEOREM 3. *If* $k \geqq 4$, $n - k \geqq 4$, *and* $|x_1| < \sigma_1$,

$$\begin{aligned} T_k = (2\pi)^{-1/2}\sigma_1^{-1} \exp\Bigg\{ &-\sum_2^{\infty} \frac{p^{v-1} - (-q)^{v-1}}{v(v - 1)} \frac{x_1^v}{\sigma_1^{v-2}} \\ &+ \frac{1}{24\sigma_1^2} \sum_3^{\infty} [p^{v-1} - (-q)^{v-1}] \left(\frac{x_1}{\sigma_1}\right)^{v-2} + \frac{1 + 2pq}{24\sigma_1^2} - \rho_1 \Bigg\}, \end{aligned} \tag{31}$$

where x_1 *and* σ_1 *are defined by* (8), *and*

$$\rho_1 = \frac{7}{2880}\left\{\frac{1 + \phi_1}{(k + \frac{1}{2})^3} + \frac{1 + \phi_1'}{[(n + 1) - (k + \frac{1}{2})]^3}\right\} + \frac{1 + \phi_2}{360(n + 1)^3}. \tag{32}$$

Here, as in the sequel, the subscript k on $x_{1,k}$ will be omitted when there is no chance of confusion.

In order to obtain an exponential expansion for the hypergeometric distribution (1) of the type (31), we consider the following norming. Let

$$x_{2,k} = \{k + \tfrac{1}{2} - (n + 1)p\}\sigma_2^{-1} \quad \text{and} \quad \sigma_2^2 = (N - n + 1)pq, \tag{33}$$

where p and q are defined by (15). To utilize Feller's binomial expansion, we must express the non-binomial portion of G_k as a suitable exponential series. Write

$$G_k = T_k C_k , \tag{34}$$

with

$$\begin{aligned} C_k = {} & \frac{(D+1)!(N-D+1)!(N-n+1)!}{(D-k)!(N-D-n+k)!(N+2)!} \\ & \cdot \frac{(N+1)(N+2)}{(D+1)(N-D+1)(N-n+1)} \cdot \frac{1}{p^k q^{n-k}} . \end{aligned} \tag{35}$$

In the expansion of C_k, we shall use (28) for the two factorials involving k and (29) for the four not involving k. Then, if $D - k \geqq 4$ and $N - D - n + k \geqq 4$,

$$\begin{aligned} \log C_k = {} & [(D+1) + \tfrac{1}{2}] \log (D+1) \\ & + [(N-D+1) + \tfrac{1}{2}] \log (N-D+1) \\ & + [(N-n+1) + \tfrac{1}{2}] \log (N-n+1) \\ & - [(D+1) - (k+\tfrac{1}{2})] \log [(D+1) - (k+\tfrac{1}{2})] \\ & - [(N-D+1) - (n+1) + (k+\tfrac{1}{2})] \log [(N-D+1) \\ & - (n+1) + (k+\tfrac{1}{2})] \\ & - [(N+2) + \tfrac{1}{2}] \log (N+2) \\ & + \frac{1}{12(D+1)} + \frac{1}{12(N-D+1)} + \frac{1}{12(N-n+1)} - \frac{1}{12(N+2)} \\ & + \frac{1}{24[(D+1) - (k+\frac{1}{2})]} + \frac{1}{24[(N-D+1) - (n+1) + (k+\frac{1}{2})]} \\ & + \log (N+1) + \log (N+2) - \log (D+1) \\ & - \log (N-D+1) - \log (N-n+1) \\ & - k \log p - (n-k) \log q - \rho_2 , \end{aligned} \tag{36}$$

where

$$\begin{aligned} \rho_2 = {} & \frac{7}{2880} \left\{ \frac{1+\phi_1}{[(D+1) - (k+\frac{1}{2})]^3} + \frac{1+\phi_1'}{[(N-D+1) - (n+1) + (k+\frac{1}{2})]^3} \right\} \\ & + \frac{1}{360} \left\{ \frac{1+\phi_2}{(D+1)^3} + \frac{1+\phi_2'}{(N-D+1)^3} + \frac{1+\phi_2''}{(N-n+1)^3} - \frac{1+\phi_2'''}{(N+2)^3} \right\} . \end{aligned} \tag{37}$$

We now introduce the substitutions (15) and (33). Algebraic manipulation reduces (36) to

$$
\begin{aligned}
\log C_k = & -\frac{\sigma_2^2}{p}\left(1+\frac{px_2}{\sigma_2}\right)\log\left(1+\frac{px_2}{\sigma_2}\right) \\
& -\frac{\sigma_2^2}{q}\left(1-\frac{qx_2}{\sigma_2}\right)\log\left(1-\frac{qx_2}{\sigma_2}\right) \\
& +\frac{p}{24\sigma_2^2\left(1+\frac{px_2}{\sigma_2}\right)}+\frac{q}{24\sigma_2^2\left(1-\frac{qx_2}{\sigma_2}\right)} \\
& +\frac{1-pq}{12pq(N+2)}+\frac{pq}{12\sigma^2} \\
& +\log\frac{N+1}{(N+2)^{1/2}(N-n+1)^{1/2}}-\rho_2 .
\end{aligned}
\tag{38}
$$

To expand (38) as an infinite series, we must impose the condition that $|x_2| < \sigma_2$. The combination of the resulting series, in a manner analogous to that of Feller, gives

THEOREM 4. *If* $D - k \geqq 4$, $N - D - n + k \geqq 4$, *and* $|x_2| < \sigma_2$,

$$
\begin{aligned}
C_k = \frac{N+1}{(N+2)^{1/2}(N-n+1)^{1/2}}\exp\Bigg\{ & -\sum_2^\infty \frac{q^{v-1}-(-p)^{v-1}}{v(v-1)}\frac{x_2^v}{\sigma_2^{v-2}} \\
& +\frac{1}{24\sigma_2^2}\sum_3^\infty [q^{v-1}-(-p)^{v-1}]\left(\frac{x_2}{\sigma_2}\right)^{v-2} \\
& +\frac{1+2pq}{24\sigma_2^2}+\frac{1-pq}{12pq(N+2)}-\rho_2\Bigg\},
\end{aligned}
\tag{39}
$$

where p *and* q *are defined by* (15), x_2 *and* σ_2 *by* (33), *and* ρ_2 *by* (37).

Using (8), (15), (17), and (33), we can derive

$$
\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}=\frac{1}{\sigma^2} \quad \text{and} \quad \frac{N+1}{(N+2)^{1/2}(N-n+1)^{1/2}}\frac{1}{\sigma_1}=\frac{N+1}{N+2}\frac{1}{\sigma}. \tag{40}
$$

Define

$$
\rho = \rho_1 + \rho_2 . \tag{41}
$$

Combine the expansions for T_k and C_k given in Theorems 3 and 4. Make the substitutions indicated by (40) and (41), to obtain the following

THEOREM 5. *If* $k \geqq 4$, $n - k \geqq 4$, $D - k \geqq 4$, $N - D - n + k \geqq 4$, *and* $|x| < \sigma$,

$$G_k = \frac{N+1}{N+2} (2\pi)^{-1/2} \sigma^{-1} \exp \left\{ - \sum_2^\infty \frac{[p^{v-1} - (-q)^{v-1}][t^{v-1} - (-s)^{v-1}]}{v(v-1)} \frac{x^v}{\sigma^{v-2}} \right.$$

$$(42) \qquad + \frac{1}{24\sigma^2} \sum_3^\infty [p^{v-1} - (-q)^{v-1}][t^{v-1} - (-s)^{v-1}] \left(\frac{x}{\sigma}\right)^{v-2}$$

$$\left. + \frac{1 + 2pq}{24\sigma^2} + \frac{1 - pq}{12pr(N+2)} - \rho \right\},$$

where p, q, s, *and* t *are defined by* (15), x *and* σ *by* (17), *and* ρ *by* (32), (37), *and* (41).

Except for the terms independent of x and the extra factor in each of the series, (42) has the same form as Feller's expansion (31). In the next section we prove Theorem 2 by an argument almost identical to Fellers. Only minor changes are necessary to cover the differences in the two expansions. Feller's notation is used and reference to his proof is indicated by "F".

6. Proof of theorem 2. Since we are interested in values of k which satisfy $\lambda \leqq k \leqq \nu$, hypothesis (19) implies (39 - F). Hypothesis (18) implies $k \geqq 9$, $N - D - n + k \geqq 9$, $n - k \geqq 9(1 / pt - \frac{2}{3}) - \frac{1}{2}$, and $D - k \geqq 9(1 / qs - \frac{2}{3}) - \frac{1}{2}$. In most cases this is sufficient to satisfy the hypotheses of Theorem 5. To cover the possibility of either pt or qs being close to one, we have included the extra hypotheses $n - k \geqq 4$ and $D - k \geqq 4$. In any case, the hypotheses of Theorem 5 are satisfied and the expansion (42) of G_k is valid. By using (39 - F), the remainder, ρ, in (42) can be shown to satisfy

$$(43) \qquad 0 < \rho < \frac{1}{12\sigma^6}.$$

The remaining portion of the proof is devoted to showing that the expansion (42), where ρ satisfies (43), can be written as a product of two factors—the first independent of k, and the second similar in form to the right side of (26) with argument x replaced by η_k, defined by either (22) or (23). If η_k is given by (22), we show that the second factor is less than the integral of (26) with x replaced by η_k, and if by (23), that it is greater than the corresponding integral of (26). The proof is completed by summing over all admissible k values.

For each k, define ξ_k by (38 - F), where a is defined by (16). In the sequel we shall use the fact that $|a| \leqq \frac{1}{6}$. The subscript k on ξ_k and x_k will be suppressed when convenient. By (39 - F), we have

$$(44) \quad x \leqq \xi \leqq \tfrac{10}{9}x \quad \text{for } a > 0, \qquad \text{and} \qquad \tfrac{8}{9}x \leqq \xi \leqq x \qquad \text{for } a < 0.$$

Set

$$(45) \qquad A_j = \sum_j^\infty \frac{[p^{v-1} - (-q)^{v-1}][t^{v-1} - (-s)^{v-1}]}{v(v-1)} \frac{x^v}{\sigma^{v-2}}.$$

The first series of (42) is A_2. We write

$$A_2 = \tfrac{1}{2}\xi^2 + A, \tag{46}$$

where

$$A = \left[\frac{(p^3 + q^3)(t^3 + s^3)}{12} - \frac{a^2}{2}\right]\frac{x^4}{\sigma^2} + A_5. \tag{47}$$

The function within the brackets is defined in the (p, s) unit square with absolute maxima of $\frac{5}{72}$ at (1, 1) and (0, 0) and the unique absolute minimum of $\frac{1}{192}$ at $(\frac{1}{2}, \frac{1}{2})$.

We shall need bounds on A. First, consider the case $a > 0$. In this case, all terms of A_5 are positive; so, by (47) and (44),

$$A > \frac{1}{192}\frac{x^4}{\sigma^2} > \frac{1}{300}\frac{\xi^4}{\sigma^2} \qquad \text{if } a > 0. \tag{48}$$

If $a < 0$, A_5 is an alternating series with the first term negative. Each negative term is smaller in absolute value than the preceding positive one. So, by (39-F) and (47),

$$A \geqq \left[\frac{(p^3 + q^3)(t^3 + s^3)}{12} - \frac{a^2}{2} + \frac{(p^4 - q^4)(t^4 - s^4)}{30}\right]\frac{x^4}{\sigma^2}. \tag{49}$$

The function within the brackets is defined in the (p, s) unit square and has its unique absolute min mum of $\frac{1}{192}$ at $(\frac{1}{2}, \frac{1}{2})$. Using (44),

$$A \geqq \frac{1}{192}\frac{x^4}{\sigma^2} \geqq \frac{1}{192}\frac{\xi^4}{\sigma^2} \qquad \text{if } a < 0. \tag{50}$$

The series A_5 can be majorized by a geometric series to give a uniform upper bound of $1.01x^4 / 16\sigma^2$. Thus, from (47) we obtain

$$A < \left(\frac{5}{72} + \frac{1.01}{16}\right)\frac{x^4}{\sigma^2} < \frac{2}{15}\frac{x^4}{\sigma^2}. \tag{51}$$

Let

$$B_j = \frac{1}{24\sigma^2}\sum_j^{\infty} [p^{v-1} - (-q)^{v-1}][t^{v-1} - (-s)^{v-1}]\left(\frac{x}{\sigma}\right)^{v-2}. \tag{52}$$

The second series of (42) is B_3. We write

$$B_3 = \frac{a}{4\sigma^3}\xi + B, \tag{53}$$

where

$$B = \frac{1}{2}\left[\frac{(p^3 + q^3)(t^3 + s^3)}{12} - \frac{a^2}{2}\right]\frac{x^2}{\sigma^4} + B_5. \tag{54}$$

Bounds on B are obtained in a similar manner to those obtained on A. An argument on (54) identical to that preceding (48), with A_5 replaced by B_5, gives

$$B > \frac{1}{2}\left(\frac{1}{192}\right)\frac{x^2}{\sigma^4} > 0 \quad \text{if } a > 0. \tag{55}$$

Likewise, if $A < 0$, the argument preceding (49), with A_5 replaced by B_5, applies to (54). Therefore, from (54),

$$B \geqq \frac{1}{2}\left[\frac{(p^3 + q^3)(t^3 + s^3)}{12} - \frac{a^2}{2} + \frac{(p^4 - q^4)(t^4 - s^4)}{18}\right]\frac{x^2}{\sigma^4}. \tag{56}$$

The function within the brackets has the same absolute minimum as that of (49); so, (55) is also valid if $a < 0$. While a uniform upper bound on A is sufficient for our purpose, we must consider the two cases separately on B. First, let $a < 0$, then the series B_5 can be majorized by $x^2 / 192\sigma^4$. Using (44) and the discussion following (47), we have from (54)

$$B \leqq \left[\frac{1}{2}\left(\frac{5}{72}\right) + \frac{1}{192}\right]\frac{x^2}{\sigma^4} < \frac{1}{19}\frac{\xi^2}{\sigma^4} \quad \text{if } a < 0. \tag{57}$$

For $a > 0$, B_5 is a positive term series which can be majorized by $x^2 / 12\sigma^4$. Again, by (44) and the discussion following (47), we have from (54)

$$B \leqq \left[\frac{1}{2}\left(\frac{5}{72}\right) + \frac{1}{12}\right]\frac{x^2}{\sigma^4} < \frac{1}{8}\frac{\xi^2}{\sigma^4} \quad \text{if } a > 0. \tag{58}$$

Define $\Delta\xi_k$ by (50-F). Then (51-F) and (52-F) follow. Substitution of (46), (53), and (52-F) into the expansion (42) gives

$$\begin{aligned} G_k = \frac{N+1}{N+2}(2\pi)^{-1/2}\Delta\xi \exp\Big\{ &-\frac{1}{2}\xi^2 + \frac{a}{4\sigma^3}\xi \\ &- \frac{1}{2}\log\left(1 + \frac{4a}{\sigma}\xi\right) - A + B + \frac{1+2pq}{24\sigma^2} \\ &+ \frac{1-pq}{12pq(N+2)} - \rho\Big\}. \end{aligned} \tag{59}$$

To eliminate the logarithm term, define C by (54-F). Expressing C as an infinite series, we can bound C to obtain (55-F). Define y and Δy by (56-F), where u is a parameter to be determined. We note that y as a translation of ξ also satisfies (51-F). If we define u by (57-F) and define η_k by (22), then (58-F) is valid. Likewise, if we define u by (59-F) and define η_k by (23), the identities (58-F) are still valid. Because of (58-F), our Theorem 2 will be proved if we show that with u defined by (57-F),

$$G_k \leqq \frac{N+1}{N+2}e^R\left\{\Phi\left(y_k + \frac{1}{2}\Delta y_k\right) - \Phi\left(y_k - \frac{1}{2}\Delta y_k\right)\right\}, \tag{60}$$

and that the inequality in (60) is reversed if u is defined by (59-F); R is defined by (21).

Using (54-F) and (56-F), we can transform (59) to

$$G_k = \frac{N+1}{N+2} e^R (2\pi)^{-1/2} F_k, \tag{61}$$

where

$$F_k = \Delta y \exp\left\{ -\frac{1}{2} y^2 + \frac{(\Delta y)^2}{24} (y^2 - 1) + E \right\} \tag{62}$$

and E is given by (62-F).

Let u be defined by (57-F). Noting the form of (62) and Lemma 1, with h and x replaced by Δy and y, respectively, the inequality (60) will be proved if we show that (63-F) is satisfied.

Substitution of the bounds (18), (43), (48), (50), (57), (58), and (55-F) into (62-F) gives, using (63-F),

$$\sigma^2 E_1 < \frac{1}{24\sigma} - \frac{25}{54} \xi + \frac{1}{8} \xi^2 - \frac{1}{300} \xi^4 \qquad \text{if } a > 0 \tag{63}$$

and

$$\sigma^2 E_1 < \frac{2}{9\sigma} - \frac{1}{2} \xi + \frac{1}{5} \xi^2 - \frac{1}{192} \xi^4 \qquad \text{if } a < 0. \tag{64}$$

We are interested in values of x which satisfy (39-F). For such values, $\xi \geqq 107/216\sigma$. Elementary calculations show that the quartics in (63) and (64) are negative if $\xi \geqq 107/216\sigma$; therefore, (63-F) is true. This implies (60) is also true. Summing over all k values in the interval $\lambda \leqq k \leqq \nu$, (58-F) and (60) give (20). Thus, the upper bound for $H_{\lambda,\nu}$ is valid.

A similar argument suffices to prove the lower bound is valid. Let u be defined by (59-F), then, as before, from (62) and Lemma 1, (60) with the reverse inequality will be proved if we show that

$$E_2 \equiv E - \frac{(\Delta y)^4}{264} \geqq 0. \tag{65}$$

First, we need several auxiliary bounds. By (24), (44), and (54) we have

$$\begin{aligned} A &< \frac{3M}{20\sigma} \xi \qquad \text{if } a < 0, \\ A &< \frac{2M}{15\sigma} \xi \qquad \text{if } a > 0. \end{aligned} \tag{66}$$

Also, (69-F) follows if $a < 0$, and

$$\frac{y^2}{24\sigma^2}\left(1 + \frac{4a\xi}{\sigma}\right) \leqq \left(\frac{11}{9}\right)^2 \frac{1}{24\sigma^2}\left(\xi + \frac{2a}{\sigma} - \frac{u}{\sigma}\right)^2 \tag{67}$$

if $a > 0$. Substitution of the bounds (18), (43), (55) (which is independent of a), (55-F), (66), (69-F), and (67) into (62-F) gives (70-F) if $a < 0$, and

$$E_2 \geqq \frac{u^2}{2\sigma^2}\left(1 - \frac{1}{6\sigma^2}\right)\frac{u}{18\sigma^4} - \frac{1}{30\sigma^4} + \left(-\frac{u}{\sigma} - \frac{2M}{15\sigma} - \frac{1}{18\sigma^3} + \frac{u}{6\sigma^3}\right)\xi - \frac{1}{12\sigma^2}\xi^2 \tag{68}$$

if $a > 0$. Bounding the constant term in (70-F) and (68) (constant with respect to ξ) and evaluating the coefficient of ξ by means of (59-F), we obtain

$$E_2 \geqq -\frac{6}{107\sigma^2} - \frac{1}{50\sigma^4} + \frac{293}{2268}\left(\frac{\xi}{\sigma}\right) - \frac{1}{24}\left(\frac{\xi}{\sigma}\right)^2 \tag{69}$$

if $a < 0$, and

$$E_2 \geqq -\frac{1}{500\sigma^6} - \frac{1}{30\sigma^4} + \frac{152}{1134}\left(\frac{\xi}{\sigma}\right) - \frac{1}{12}\left(\frac{\xi}{\sigma}\right)^2 \tag{70}$$

if $a > 0$. The right sides of (69) and (70) are parabolas in ξ opening downward. To show non-negativity, we need only to check at the endpoints of the ξ intervals which correspond to (39-F). These are

$$\begin{aligned} \frac{107}{216\sigma} &< \xi < \frac{2\sigma}{3} \qquad \text{if } a < 0, \\ \frac{1}{2\sigma} &< \xi < \frac{20\sigma}{27} \qquad \text{if } a > 0. \end{aligned} \tag{71}$$

Making the above substitutions, the right sides of (69) and (70) are seen to be positive. Hence, (65) is true; therefore, the lower bound, (60) with inequality reversed, is also valid. As before, summing over all admissible k values gives (20) with the inequality reversed. Q.E.D.

Acknowledgment. The author would like to express his sincere thanks to Professor W. G. Madow, who suggested the problem, for his careful checking of the numerous computations and his assistance in preparation of the manuscript.

REFERENCES

[1] WILLIAM FELLER, *Probability Theory and Its Applications*, Vol. 1, John Wiley and Sons, 1952.

[2] WILLIAM FELLER, "On the Normal Approximation to the Binomial Distribution", *Ann. Math. Stat.*, Vol. 16 (1945), pp. 319–329.