

ASYMPTOTIC FACTORIAL POWERS EXPANSIONS FOR BINOMIAL AND NEGATIVE BINOMIAL RECIPROCAL

GRZEGORZ A. REMPALA

(Communicated by Richard A. Davis)

To my parents

ABSTRACT. By considering the variance formula for a shifted reciprocal of a binomial proportion, the asymptotic expansions of any order for first negative moments of binomial and negative binomial distributions truncated at zero are obtained. The expansions are given in terms of the factorial powers of the number of trials n . The obtained formulae are more accurate than those of Marciniak and Wesolowski (1999) and simpler, as they do not involve the Eulerian polynomials.

1. INTRODUCTION

The problem of computing a first inverse moment for binomial or negative binomial distributions arises naturally in many statistical estimation schemes in which the inference is based on the samples of random length. For instance, suppose that one wants to estimate the mean of a population X on the basis of a sample X_1, \dots, X_K of random length K , which is a natural-valued, nonzero random variable independent of the X_i 's. The usual approach is then to consider the unbiased estimator

$$T = \frac{1}{K} \sum_{i=1}^K X_i,$$

which has the variance (only if X is square integrable) given by

$$\text{Var}(T) = \sigma^2 E(1/K),$$

where σ^2 is the variance of X . For some general approaches to the problem of computing inverse and inverse factorial moments, see, for instance, Jones (1987) or Rockower (1988), and references therein.

In this paper we treat only two special distributions of K : binomial and negative binomial. As pointed out in Marciniak and Wesolowski (1999), they appear in this context in a natural way if one takes a sample with replacement from a given finite population of size K and is interested in estimation of some parameters in a sub-population A of size K_A . Then two cases are of potential interest: (i) the size of the

Received by the editors March 1, 2001 and, in revised form, August 1, 2002.

2000 *Mathematics Subject Classification*. Primary 60E05 62E20; Secondary 11B15, 05A16.

Key words and phrases. Factorial power, asymptotic expansions, indirect estimator, inverse moments, elementary symmetric polynomial, positive binomial distribution, truncated negative binomial distribution.

sample is k and then the size k_A of the subsample falling in A is a random variable with the binomial $b(k, K_A/K)$ distribution; (ii) elements are drawn up to k -th draw from outside of A and then k_A has the negative binomial $nb(k, K_A/K)$ distribution (this is closely related to the inverse sampling technique used in survey sampling methodology as described, for instance, in chapter 4.5 of Cochran 1977). If in both situations only samples with nonzero k_A 's are considered, then the distributions are truncated at zero, i.e., we consider the case when K follows either a positive binomial or a truncated negative binomial distribution. Generally, especially for large sample sizes, $E(1/K)$ is then a complicated function of k and thus it would be very useful to find its approximation for large k .

In their work, Marciniak and Wesolowski (1999) (referred to as MW hereafter) have obtained the following asymptotic expansion formulae for the first inverse moments of binomial and negative binomial distributions truncated at zero. In the sequel, let $p \in (0, 1)$ and $q = 1 - p$.

(i) If M is a positive binomial $pb(n, p)$ random variable, i.e.,

$$P(M = i) = \frac{1}{1 - q^n} \binom{n}{i} p^i q^{n-i}, \quad i = 1, \dots, n,$$

then for any $r = 1, 2, \dots$,

$$(1.1) \quad (1 - q^n) E M^{-1} = \sum_{l=0}^{r-1} \frac{A_l(q)}{p^{l+1} n^{l+1}} + o(n^{-r}).$$

(ii) If N is a truncated negative binomial $tnb(n, p)$ random variable, i.e.,

$$P(N = i) = \frac{1}{1 - q^n} \binom{n+i-1}{i} p^i q^n, \quad i = 1, 2, \dots,$$

then for any $r = 1, 2, \dots$,

$$(1.2) \quad (1 - q^n) E N^{-1} = \frac{q}{pn} + \sum_{l=1}^{r-1} \frac{A_l(q)}{p^{l+1} n^{l+1}} + o(n^{-r}).$$

Here the $A_l(q)$'s are the Eulerian polynomials in q defined by the generating function

$$\frac{1 - t}{1 - t \exp[(1 - t)s]} = \sum_{l=0}^{\infty} A_l(t) \frac{s^l}{l!}.$$

As pointed out in MW, the formula (1.2) follows from (1.1), since we have (see Lemma 1 in MW)

$$(1.3) \quad (1 - q^n) E N^{-1} = (1 - q^n) E M^{-1} - 1/n - q^n (\log(q) - 1/n).$$

The purpose of the current paper is to establish a formula for the asymptotic expansion of $(1 - q^n) E M^{-1}$ different from (1.1) and not involving the quantities $A_l(q)$. The key observation leading to the result is as follows. If S_{n-1} is a binomial random variable $b(n - 1, p)$, then

$$\begin{aligned} E \left(\frac{n}{S_{n-1} + 1} \right)^2 &= n \sum_{i=0}^{n-1} \frac{n}{(i+1)^2} \binom{n-1}{i} p^i q^{n-1-i} \\ &= \frac{n}{p} (1 - q^n) E M^{-1}. \end{aligned}$$

Setting $U_{n-1} = n/(S_{n-1} + 1)$ we obtain thus the relation

$$(1.4) \quad (1 - q^n) E M^{-1} = \frac{p}{n} E U_{n-1}^2.$$

In view of the above, in order to obtain the asymptotic expansion for $(1 - q^n) E M^{-1}$ (and thus via (1.3) also for $(1 - q^n) E N^{-1}$) it suffices to find the expansion for $E U_{n-1}^2$. It turns out that this approach leads to somewhat different expansions than those offered in MW.

The paper is organized as follows. Beyond the current Introduction, the paper is subdivided into two additional sections, with Section 2 containing the main results and Section 3 containing the proofs along with some auxiliary technical lemmas. The key result allowing us to expand $E U_{n-1}^2$ is given in Proposition 1 of Section 2 and leads to Theorems 1 and 2 therein, in which we state the main results of the paper, namely the asymptotic factorial power expansion formulae for the first negative moments of the positive binomial and truncated negative binomial distributions. In the discussion following the theorems we compare our factorial expansions to the expansions given in MW.

The result of Proposition 1 is obtained by applying the representation of some random variable equidistributed with U_n in terms of the linear combination of certain symmetric functions. Both the representation and the second moment formula are special cases of the general results for the so-called *indirect estimators* introduced in Rempala and Székely (1998)[8]. In the course of the proof of Proposition 1 (in Section 3) we discuss briefly the approach of Rempala and Székely (1998) to calculating the variance of U_n , since, in principle, it could be used also to establish the expansions for higher negative moments. Some brief discussion of this issue is also provided at the end of Section 2.

2. FACTORIAL POWERS EXPANSIONS

Consider again the random variable

$$U_n = \frac{n+1}{S_n+1},$$

where S_n is distributed as a binomial random variable $b(n, p)$. The main auxiliary result leading to the factorial expansions described in this section is the identity contained in the following.

Proposition 1. *For any $n \geq 2$,*

$$(1 - q^n) E M^{-1} = \frac{p}{n} E U_{n-1}^2 = \frac{p}{n} \left\{ \sum_{l=0}^{n-1} \frac{(pq)^l}{\binom{n-1}{l}} \left[\sum_{k=l}^{n-1} q^{k-l} \binom{k}{l} \right]^2 \right\}.$$

The method of verification of the above identity relies on the fact that the distribution of U_n is equal to that of the so-called indirect estimator introduced in Rempala and Székely (1998) which in turn can be analyzed via standard methods of the theory of symmetric functions (see Rempala and Székely 1998 and also Frees 1989[4] for further discussion). Detailed derivation of Proposition 1 is given in Section 3.

Note that the formula stated in Proposition 1 seems to be an interesting combinatorial identity in its own right. For instance, for $p = q = 1/2$ it reveals that

$$(2.1) \quad \sum_{l=1}^n \frac{1}{l} \binom{n}{l} = \frac{2^{n-1}}{n} \sum_{l=0}^{n-1} \binom{n-1}{l}^{-1} \left\{ \sum_{k=l}^{n-1} \binom{k}{l} 2^{-k} \right\}^2,$$

which, otherwise, would be quite challenging to verify directly. However, it seems that perhaps the biggest importance of Proposition 1 lies in the fact that it provides immediate insight into the asymptotic behavior of the quantity $(1 - q^n) E M^{-1}$. Namely, it gives the basis for the following.

Theorem 1. *For any $r = 1, 2, \dots$,*

$$\begin{aligned} (1 - q^n) E M^{-1} &= \frac{1}{np} \left\{ \sum_{l=0}^{r-1} \frac{(q/p)^l}{\binom{n-1}{l}} + O\left(\binom{n-1}{r}^{-1}\right) \right\} \\ &= \frac{1}{np} \left\{ \sum_{l=0}^{r-1} \frac{l! (q/p)^l}{(n-1)^{[l]}} + O\left(1/n^{[r]}\right) \right\} \\ &= \sum_{l=0}^{r-1} \frac{l! q^l}{p^{l+1} n^{[l+1]}} + o\left(1/n^{[r]}\right), \end{aligned}$$

where for any $k \geq 1$ we denote $s^{[k]} = s(s-1) \cdots (s-k+1)$ and $s^{[0]} = 1$.

Observe that the form of the last expression resembles that of (1.1) with $A_l(q)$ replaced by $l! q^l$ and n^l replaced by $n^{[l]}$ (and, in fact, it seems that (1.1) could now be derived from our result via some tedious combinatorial calculation involving the recursion relation for the $A^l(q)$'s, but we shall not pursue it here). The above formula for $r = 3$ appears in Rempala and Székely (1998) in Example 3.3; however, due to an error, the final form of the expression there has an incorrect coefficient at the second factorial power of n . Indeed, an inspection of the first expression at the right-hand side of the final approximation formula in Example 3.3 of Rempala and Székely (1998) reveals that the polynomial inside the square bracket there should be in q rather than in p . Replacing p with q gives the correct second-term coefficient q/p^2 (and not p^2/q^3 as mistakenly reported). This agrees with our result above.

In view of Theorem 1, for instance, the following asymptotic expansions for the right-hand side of (2.1) follow immediately upon taking $p = q = 1/2$:

$$\begin{aligned} \sum_{l=1}^n \frac{1}{l} \binom{n}{l} &= \frac{2^{n+1}}{n} \{1 + O((n-1)^{-1})\} \\ &= \frac{2^{n+1}}{n} \left\{ 1 + (n-1)^{-1} + O\left(\binom{n-1}{2}^{-1}\right) \right\} \\ &= \frac{2^{n+1}}{n} \left\{ 1 + (n-1)^{-1} + \binom{n-1}{2}^{-1} + O\left(\binom{n-1}{3}^{-1}\right) \right\} \\ &= \frac{2^{n+1}}{n} \left\{ 1 + (n-1)^{-1} + \binom{n-1}{2}^{-1} + \binom{n-1}{3}^{-1} + O\left(\binom{n-1}{4}^{-1}\right) \right\} \\ &= 2^{n+1} \left\{ \frac{0!}{n^{[1]}} + \frac{1!}{n^{[2]}} + \frac{2!}{n^{[3]}} + \frac{3!}{n^{[4]}} + o\left(1/n^{[4]}\right) \right\}, \text{ etc.} \end{aligned}$$

The expansions are simpler than the ones given in Section 2 of MW since they do not require the recursive calculation of the values $A_l(1/2)$ in order to obtain the coefficients at the appropriate powers of n . Moreover, as will be seen from the discussion at the end of this section, for sufficiently large n they also seem to be more accurate than the Eulerian ones. For some related expansions, see also Chapter 9 of Graham, Knuth, and Patashnik (1994) [5].

The proof of Theorem 1 is presented in Section 3. Unlike MW who exploited the Stolz theorem and the recurrence relation between the Eulerian numbers in order to establish their result, we shall only rely on the relation (1.4) and the representation of Proposition 1 along with the Stirling approximation formula and some basic properties of a binomial distribution.

Note that in view of (1.3) we immediately obtain from Theorem 1 the following corresponding result for the truncated negative binomial case.

Theorem 2. For any $r = 1, 2, \dots$,

$$\begin{aligned} (1 - q^n) E N^{-1} &= \frac{1}{np} \left\{ q + \sum_{l=1}^{r-1} \frac{(q/p)^l}{\binom{n-1}{l}} + O\left(\binom{n-1}{r}^{-1}\right) \right\} \\ &= \frac{1}{np} \left\{ q + \sum_{l=1}^{r-1} \frac{l! (q/p)^l}{(n-1)^{[l]}} + O\left(1/n^{[r]}\right) \right\} \\ &= \frac{q}{np} + \sum_{l=1}^{r-1} \frac{l! q^l}{p^{l+1} n^{[l+1]}} + o\left(1/n^{[r]}\right). \end{aligned}$$

As before, note the similarity between the last expression and (1.2).

Taking $p = q = 1/2$ in the above theorem we obtain, for instance, the following expansions:

$$\begin{aligned} \sum_{l=1}^{\infty} \frac{1}{l 2^l} \binom{n+l-1}{l} &= \frac{2^{n+1}}{n} \left\{ \frac{1}{2} + O((n-1)^{-1}) \right\} \\ &= \frac{2^{n+1}}{n} \left\{ \frac{1}{2} + (n-1)^{-1} + O\left(\binom{n-1}{2}^{-1}\right) \right\} \\ &= \frac{2^{n+1}}{n} \left\{ \frac{1}{2} + (n-1)^{-1} + \binom{n-1}{2}^{-1} + O\left(\binom{n-1}{3}^{-1}\right) \right\} \\ &= \frac{2^{n+1}}{n} \left\{ \frac{1}{2} + (n-1)^{-1} + \binom{n-1}{2}^{-1} + \binom{n-1}{3}^{-1} \right\} \\ &\quad + O\left(\binom{n-1}{4}^{-1}\right) \\ &= 2^{n+1} \left\{ \frac{0!}{2 n^{[1]}} + \frac{1!}{n^{[2]}} + \frac{2!}{n^{[3]}} + \frac{3!}{n^{[4]}} + o\left(1/n^{[4]}\right) \right\}, \text{ etc.} \end{aligned}$$

Factorial vs. Eulerian expansions. At this point it would be perhaps of interest to compare the accuracy of approximations offered by Eulerian and factorial expansions. Whereas the detailed study of this issue goes beyond the scope of the present work, we shall offer some insight into it, with the help of Proposition 1, for approximations up to order $r = 5$, which was the order discussed in detail by the

$q=.1$	$r=2$	$r=3$	$r=4$	$r=5$
$n=50$				
$(1 - P_r/S) * 10^3$	0.0557975677	0.0016000410	0.0000562419	0.0000024458
$(1 - Q_r/S) * 10^3$	0.0105491762	0.0000750116	0.0000007267	0.0000000090
$n=100$				
$(1 - P_r/S) * 10^3$	0.0137616267	0.0001966389	0.0000034406	0.0000000745
$(1 - Q_r/S) * 10^3$	0.0025508929	0.0000087765	0.0000000407	0.0000000002
$q=.5$	$r=2$	$r=3$	$r=4$	$r=5$
$n=50$				
$(1 - P_r/S) * 10^3$	1.2905975183	0.1156452801	0.0138160861	0.0020665637
$(1 - Q_r/S) * 10^3$	0.8909538998	0.0583630281	0.0052189299	0.0005977040
$n=100$				
$(1 - P_r/S) * 10^3$	0.3106057899	0.0136683461	0.0008010569	0.0000587132
$(1 - Q_r/S) * 10^3$	0.2106268526	0.0065882050	0.0002777314	0.0000147949
$q=.9$	$r=2$	$r=3$	$r=4$	$r=5$
$n=50$				
$(1 - P_r/S) * 10^3$	60.271144388	5.7987259781	n/a ¹	n/a
$(1 - Q_r/S) * 10^3$	57.345665419	2.4929347427	n/a	n/a
$n=100$				
$(1 - P_r/S) * 10^3$	22.632333818	7.2993181371	2.9334436732	1.2759446781
$(1 - Q_r/S) * 10^3$	21.817181470	6.8449954718	2.6774797817	1.1146613978

¹ n is not sufficiently large for valid approximation.

TABLE 1. Comparison between the relative errors of the approximations of $S = (1 - q^n)EM^{-1}$ with $q = 0.1, 0.5, 0.9$ by the first r (≤ 5) terms of the Eulerian expansion (P_r) given in (1.1) and the factorial powers expansion (Q_r) of Theorem 1.

celebrated paper of David and Johnson (1965-1967). In our discussion below we focus on the positive binomial case, but note that a similar argument applies to the truncated negative binomial distribution. Let $S = (1 - q^n)EM^{-1}$ and denote the first r terms of the Eulerian expansion given in (1.1) by P_r and the first r terms of the factorial powers expansion of Theorem 1 by Q_r . That is, we define

$$P_r = \sum_{l=0}^{r-1} \frac{A_l(q)}{n^{l+1} p^{l+1}}, \quad Q_r = \sum_{l=0}^{r-1} \frac{l! q^l}{n^{[l+1]} p^{l+1}}.$$

In David and Johnson (1956-1957) an approximation formula for EM^{-1} was given as

(2.2) $EM^{-1} \approx \frac{1}{np} + \frac{q}{n^2 p^2} + \frac{q(q+1)}{n^3 p^3} + \frac{q(q^2 + 4q + 1)}{n^4 p^4} + \frac{q(q^3 + 11q^2 + 11q + 1)}{n^5 p^5} = P_5.$

Note that the corresponding approximation via factorial powers is then

(2.3) $EM^{-1} \approx \frac{1}{np} + \frac{q}{n^{[2]} p^2} + \frac{2q^2}{n^{[3]} p^3} + \frac{6q^3}{n^{[4]} p^4} + \frac{24q^4}{n^{[5]} p^5} = Q_5.$

By direct computation we find that the difference $\Delta_5 = Q_5 - P_5$ is

$$(2.4) \quad \Delta_5 = \frac{q [(119q^3 + 93q^2 + 27q + 1)n^3 - (109q^3 + 511q^2 + 211q + 9)n^2]}{n^{[5]} n^4 p^5} + \frac{q [2(37q^3 + 311q^2 + 239q + 13)n - 24(q^3 + 11q^2 + 11q + 1)]}{n^{[5]} n^4 p^5}$$

and hence $\Delta_5 \geq 0$, at least for sufficiently large n . Via Proposition 1 it follows thus that (for sufficiently large n) we have

$$S \geq Q_5 \geq P_5$$

and hence the factorial powers approximation (2.3) has an edge over (2.2). Moreover, since by the definition of the Eulerian numbers we have that for any integer $l \geq 1$,

$$A_l(q) \geq l! q^l,$$

then for fixed but large n the difference Δ_r is a decreasing function of $r = 1, 2, 3, \dots$ and from (2.4) it follows that we also have

$$S \geq Q_r \geq P_r,$$

for sufficiently large n and any $r \leq 5$. Hence it follows that the truncations of (2.3) also have an edge over the corresponding truncations of (2.2). This is numerically illustrated in Table 1 above where we have compared the relative errors of approximations (2.2) and (2.3) and their truncations for $q = .1, .5, .9$.

Higher moments. It is perhaps worth noticing that, in general, the method used in the proof of Proposition 1, which is based on the representation (3.4) given in Section 3, should allow us to write out the explicit (albeit complicated) formulae for the higher negative moments of M . For instance, at least in principle, with the help of (3.4) the exact formula for EU_n^s ($s > 2$) as the sum of decreasing factorial powers of n , contributed by the uncorrelated components U_{nl}^* , can also be obtained, similarly to the formula for EU_n^2 presented in Proposition 1. Then, as indicated in the Introduction, the formula for EM^{1-s} would follow, since the same argument which yields (1.4) gives also

$$(1 - q^n) EM^{1-s} = \frac{p}{n^s} EU_{n-1}^s$$

for any integer $s \geq 2$. With the exact formula for EM^{1-s} at hand, the method similar to the one presented in the proof of Theorem 1 could perhaps be applied again to obtain the asymptotic expansion. The approach seems to be quite challenging, however, since the formulae for EU_{n-1}^s with $s > 2$ are highly complicated, with their complexity increasing with the value of s .

3. PROOFS

Proof of Proposition 1. First, we show that $EU_n = 1/p - q^{n+1}/p$. This can be easily verified by arguing that U_n has the same distribution as a sum of $n + 1$ random variables with expectations $1, q, \dots, q^n$, respectively. To this end, consider the random sample of size n of Bernoulli $B(q)$ variables, i.e., the i.i.d. sequence Y_1, \dots, Y_n of random variables such that $P(Y_1 = 1) = q = 1 - P(Y_1 = 0)$. For

any integer $1 \leq k \leq n$, the *elementary symmetric polynomial* corresponding to Y_1, \dots, Y_n is defined as

$$(3.1) \quad U_{nk} = \sum_{1 \leq i_1 < \dots < i_k \leq n} Y_{i_1} \cdots Y_{i_k} = \binom{\sum Y_i}{k}$$

and $U_{n0} = 1$. Clearly, $EU_{nk} = \binom{n}{k} q^k$. Denoting $S_n = \sum_{i=1}^n (1 - Y_i)$, we have

$$\begin{aligned} \frac{n+1}{S_n+1} &= \binom{n+1}{S_n+1} \binom{n}{S_n}^{-1} = \sum_{k=0}^{n-S_n} \binom{n-k}{S_n} \binom{n}{S_n}^{-1} \\ &= \sum_{k=0}^{\sum Y_i} \binom{\sum Y_i}{k} \binom{n}{k}^{-1} = \sum_{k=0}^n \binom{n}{k}^{-1} U_{nk} \end{aligned}$$

(since $U_{nk} = 0$ for $k > \sum Y_i$), and thus the relation

$$(3.2) \quad \frac{n+1}{S_n+1} = \sum_{k=0}^n \binom{n}{k}^{-1} U_{nk}$$

holds true. Since the distribution functions of U_n and $\frac{n+1}{S_n+1}$ coincide, the above implies, in particular, that

$$(3.3) \quad EU_n = E \left(\frac{n+1}{S_n+1} \right) = \sum_{k=0}^n q^k = (1 - q^{n+1})/p.$$

In order to obtain the formulae for EU_n^2 it is now enough to find the expression for the variance of U_n . To this end note that by applying to the products in (3.1) the identity

$$\prod_{i=1}^k a_i = \prod_{i=1}^k (a_i - a + a) = a^k + \sum_{l=1}^k a^{k-l} \sum_{1 \leq i_1 < \dots < i_l \leq k} (a_{i_1} - a) \cdots (a_{i_l} - a)$$

(valid for any real numbers $a, a_i, i = 1, 2, \dots, k$) with $a = q$ we obtain

$$\begin{aligned} U_{nk} &= \binom{n}{k} \left\{ q^k + \sum_{l=1}^k q^{k-l} \binom{n-l}{k-l} \binom{n}{k}^{-1} \sum_{1 \leq i_1 < \dots < i_l \leq n} (Y_{i_1} - q) \cdots (Y_{i_l} - q) \right\} \\ &= \binom{n}{k} \left\{ q^k + \sum_{l=1}^k q^{k-l} \binom{k}{l} \binom{n}{l}^{-1} \sum_{1 \leq i_1 < \dots < i_l \leq n} (Y_{i_1} - q) \cdots (Y_{i_l} - q) \right\} \\ &= \binom{n}{k} \left\{ q^k + \sum_{l=1}^k q^{k-l} \binom{k}{l} \binom{n}{l}^{-1} U_{nl}^* \right\} \end{aligned}$$

where the last equality defines the random variables U_{nl}^* for $1 \leq l \leq n$. Applying the above relation (which is a special case of the famous Hoeffding decomposition for symmetric statistics, see e.g., Serfling 1980) [10] to (3.2) we obtain

$$(3.4) \quad \begin{aligned} \frac{n+1}{S_n+1} - E \left(\frac{n+1}{S_n+1} \right) &= \sum_{k=1}^n \binom{n}{k}^{-1} (U_{nk} - q^k) = \sum_{k=1}^n \sum_{l=1}^k q^{k-l} \binom{k}{l} \binom{n}{l}^{-1} U_{nl}^* \\ &= \sum_{l=1}^n \binom{n}{l}^{-1} U_{nl}^* \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}. \end{aligned}$$

Note that, by construction, the random variables U_{nl}^* are uncorrelated for the different l 's and that $Var U_{nl}^* = \binom{n}{l}(pq)^l$. Therefore,

$$Var U_n = Var \left(\frac{n+1}{S_n+1} \right) = \sum_{l=1}^n \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2,$$

which is a special case of the formula (3.3) in Rempala and Székely (1998).

Now, in view of the above and our relation (3.3) we have that

$$\begin{aligned} (3.5) \quad EU_n^2 &= \sum_{l=1}^n \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 + (EU_n)^2 \\ &= \sum_{l=0}^n \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 \end{aligned}$$

and thus, via the relation (1.4), that

$$(3.6) \quad (1 - q^n) EM^{-1} = \frac{p}{n} EU_{n-1}^2 = \frac{p}{n} \left\{ \sum_{l=0}^{n-1} \frac{(pq)^l}{\binom{n-1}{l}} \left[\sum_{k=l}^{n-1} q^{k-l} \binom{k}{l} \right]^2 \right\},$$

which completes the proof.

Proof of Theorem 1. In order to establish the main result we shall obtain the asymptotic expansion for (3.6) or, which is the equivalent, (3.5). Before we proceed with the proof, let us present two auxiliary results which shall be exploited in the sequel. The first of them is the following lemma of Rempala and Székely (1998).

Lemma 1. *For any real number a and integers c, n satisfying $0 \leq c \leq n$ we have*

$$\sum_{k=c}^n \binom{k}{c} a^k = \begin{cases} \binom{n+1}{c+1} & \text{if } a = 1; \\ \frac{a^c}{(1-a)^{c+1}} \left[\sum_{k=c+1}^{n+1} \binom{n+1}{k} a^{n+1-k} (1-a)^k \right] & \text{if } a \neq 1. \end{cases}$$

In particular, for $0 \leq a < 1$,

$$(3.7) \quad \sum_{k=c}^n \binom{k}{c} a^k \leq \frac{a^c}{(1-a)^{c+1}}.$$

We shall also use the following estimate on the tail of a binomial random variable due to Bahadur (1960) [1].

Lemma 2. *Let X be a random variable having a binomial $b(n, p)$ distribution with $0 < p < 1$. Then for $k \geq np$ and $q = 1 - p$,*

$$P(X \geq k) \leq \binom{n}{k} p^k q^{n-k} \frac{q(k+1)}{k+1 - (n+1)p}.$$

For the proof of the theorem, consider an arbitrary integer $r \geq 1$ and rewrite the expression (3.5) as

$$(3.8) \quad EU_n^2 = \sum_{l=0}^n \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 = \sum_{l=0}^{r-1} \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 + R(r, n).$$

First, we shall argue that

$$(3.9) \quad \sup_n \binom{n}{r} R(r, n) < \infty.$$

To this end, write

$$\begin{aligned} \binom{n}{r} R(r, n) &= \binom{n}{r} \sum_{l=r}^{[(n+1)p]-1} \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 \\ &\quad + \binom{n}{r} \sum_{l=[(n+1)p]}^n \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 \\ &= A(r, n) + B(r, n) \end{aligned}$$

Apropos $B(r, n)$, note that

$$\begin{aligned} B(r, n) &= \binom{n}{r} \sum_{l=[(n+1)p]}^n \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 \\ &= \binom{n}{r} \sum_{l=[(n+1)p]}^n \frac{(pq)^l}{\binom{n}{l}} \frac{1}{p^{2l+2}} \left\{ \sum_{k=l+1}^{n+1} \binom{n+1}{l} p^k q^{n+1-k} \right\}^2 \\ &\quad \text{by Lemma 1 with } a = q \\ &\leq \binom{n}{r} \sum_{l=[(n+1)p]}^n \frac{(pq)^l}{\binom{n}{l}} \frac{1}{p^{2l+2}} \left\{ \binom{n+1}{l+1} p^{l+1} q^{n+1-l} \frac{l+2}{l+2-(n+2)p} \right\}^2 \\ &\quad \text{by Lemma 2} \\ &\leq q^n \binom{n}{r} \sum_{l=[(n+1)p]}^n \frac{p^l q^{n-l}}{\binom{n}{l}} \left\{ \binom{n+1}{l+1} q \frac{l+2}{q} \right\}^2 \\ &\quad \text{since } l+2 \geq (n+1)p+1 \\ &= q^n \binom{n}{r} \sum_{l=[(n+1)p]}^n \binom{n+1}{l+1} p^l q^{n-l} \frac{n+1}{l+1} (l+2)^2 \\ &= q^n \binom{n}{r} \sum_{l=[(n+1)p]}^n \binom{n}{l} p^l q^{n-l} \left(\frac{n+1}{l+1} \right)^2 (l+2)^2 \\ &\leq q^n \frac{(n+2)^2}{p^2} \binom{n}{r} \quad \text{since } \frac{n+1}{l+1} \leq 1/p. \end{aligned}$$

Thus,

$$B(r, n) = o\left(\binom{n}{r}^{-1}\right),$$

and in order to establish (3.9), it is now enough to show that

$$(3.10) \quad \sup_n A(r, n) < \infty.$$

To this end observe that for sufficiently large integers l, n satisfying $l \leq np$ we have, via the Stirling formula,

$$\begin{aligned} \binom{n}{l}^{-1} &\leq C \left(\frac{l}{n-l}\right)^l \left(\frac{n-l}{n}\right)^n \sqrt{l} \leq C \left(\frac{np}{n-np}\right)^l \left(\frac{n-l}{n}\right)^n \sqrt{l} \\ &\leq C \left(\frac{p}{q}\right)^l \left(\frac{n-l}{n}\right)^n \sqrt{l} \end{aligned}$$

for some universal constant $C > 1$ that does not depend upon n or l . Since for $0 \leq l \leq n$ we have $\left(\frac{n-l}{n}\right)^n \leq e^{-l}$, the above entails

$$(3.11) \quad \binom{n}{l}^{-1} \leq C \left(\frac{p}{q}\right)^l e^{-l} \sqrt{l} \quad \text{for sufficiently large } l \leq np.$$

Consider s, n large enough for (3.11) to hold for all l such that $s+r \leq l \leq np$ and observe that

$$\begin{aligned} A(r, n) &= \binom{n}{r} \left[\sum_{l=r}^{[(n+1)p]-1} \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 \right] \\ &\leq \binom{n}{r} \sum_{l=r}^{[(n+1)p]-1} \frac{(pq)^l}{\binom{n}{l}} \frac{1}{p^{2l+2}} \text{ by (3.7) of Lemma 1 with } a = q \\ &= p^{-2} \sum_{l=r}^{s-1} \left(\frac{q}{p}\right)^l \frac{\binom{l}{r}}{\binom{n-r}{l-r}} + p^{-2} \sum_{l=s}^{[(n+1)p]-1} \left(\frac{q}{p}\right)^l \frac{\binom{l}{r}}{\binom{n-r}{l-r}} = (I) + (II). \end{aligned}$$

The expression (I) above tends to q^r/p^{r+2} as $n \rightarrow \infty$. Regarding the expression (II), we have by (3.11)

$$\begin{aligned} (II) &\leq C e^r \left(\frac{q}{p}\right)^r p^{-2} \sum_{l=s}^{[(n+1)p]-1} \binom{l}{r} \sqrt{l} e^{-l} \\ &\leq \frac{C q^r e^r}{p^{r+2} r!} \sum_{l=s}^{[(n+1)p]-1} l^{r+1} e^{-l} \leq \frac{C q^r e^r}{p^{r+2} r!} \Gamma(r+2) = \frac{C q^r (r+1) e^r}{p^{r+2}}. \end{aligned}$$

Here $\Gamma(\cdot)$ stands for Euler’s gamma function. Hence, for n sufficiently large,

$$A(r, n) \leq (I) + (II) \leq \frac{C q^r e^r}{p^{r+2}} (r+2) < \infty,$$

which entails (3.10) and thus also (3.9). Now, in view of (3.8) and Lemma 1, we may write for any $r \geq 1$,

$$\begin{aligned} EU_n^2 &= \sum_{l=0}^{r-1} \frac{(pq)^l}{\binom{n}{l}} \left\{ \sum_{k=l}^n q^{k-l} \binom{k}{l} \right\}^2 + R(r, n) \\ &= \sum_{l=0}^{r-1} \frac{(pq)^l}{\binom{n}{l}} \frac{1}{p^{2l+2}} \left\{ 1 - o\left(\binom{n}{r}^{-1}\right) \right\}^2 + R(r, n) \\ &= \sum_{l=0}^{r-1} \binom{n}{l}^{-1} \frac{q^l}{p^{l+2}} + R(r, n) + o\left(\binom{n}{r}^{-1}\right), \end{aligned}$$

and the final result follows via (1.4).

Acknowledgement. The author is indebted to Professor Jacek Wesolowski for his encouragement and helpful comments on the early draft of the paper. Thanks are also due to the anonymous referee for his valuable suggestions that have improved the presentation of the material.

REFERENCES

1. R. R. Bahadur (1960). Some approximations to the binomial distribution function. *Ann. Math. Statist.*, 31:43–54. MR **22**:11424
2. W. G. Cochran (1977). *Sampling Techniques*. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York. MR **57**:14212
3. F. N. David and N. L. Johnson (1956-1957). Reciprocal Bernoulli and Poisson variables. *Metron* 18: 77–81. MR **18**:520g
4. E. W. Frees (1989). Infinite order U -statistics. *Scand. J. Statist.*, 16(1):29–45. MR **90h**:62047
5. R. L. Graham, D. E. Knuth and O. Patashnik (1994). *Concrete Mathematics*. Addison-Wesley Publishing Company, Reading, MA, second edition. MR **97d**:68003
6. M. C. Jones (1987). Inverse factorial moments. *Statistics and Probability Letters* 6: 37–42. MR **90e**:62025a
7. E. Marciniak and J. Wesolowski (1999). Asymptotic Eulerian expansions for binomial and negative binomial reciprocals. *Proc. Amer. Math. Soc.*, 127(11):3329–3338. MR **2000b**:60033
8. G. Rempala and G. Székely (1998). On estimation with elementary symmetric polynomials. *Random Oper. Stochastic Equations*, 6(1):77–88. MR **99e**:62025
9. E. B. Rockower (1988). Integral identities for random variables. *American Statistician* 42: 68–72. MR **89a**:60041
10. R. J. Serfling (1980). *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York. MR **82a**:62003

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF LOUISVILLE, LOUISVILLE, KENTUCKY 40292
E-mail address: grzes@louisville.edu