ON HYPERGEOMETRIC GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION

M. E. GHITANY, S. A. AL-AWADHI, and S. L. KALLA

It is shown that the hypergeometric generalized negative binomial distribution has moments of all positive orders, is overdispersed, skewed to the right, and leptokurtic. Also, a three-term recurrence relation for computing probabilities from the considered distribution is given. Application of the distribution to entomological field data is given and its goodness-of-fit is demonstrated.

1. Introduction. A certain mixture distribution arises when all (or some) parameters of a distribution vary according to some probability distribution, called the *mixing distribution*. A well-known example of discrete-type mixture distribution is the negative binomial distribution which can be obtained as a Poisson mixture with gamma mixing distribution.

Let X has a conditional Poisson distribution with parameter λ, that is, X has a conditional probability mass function (pmf)

$$f(x \mid \lambda) = P(X = x \mid \Lambda = \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, \ldots, \lambda > 0. \tag{1.1}$$

Now suppose that Λ is a continuous random variable with probability density function (pdf)

$$g(\lambda) = \frac{\alpha^a (\alpha + 1)^{p-a}}{\Gamma(p)} \lambda^{p-1} e^{-(\alpha+1)\lambda} {}_1F_1(a; p; \lambda), \quad \lambda > 0, \ a, p, \alpha > 0, \tag{1.2}$$

where

$$_1F_1(a; b; z) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \frac{z^n}{n!} \tag{1.3}$$

is the confluent hypergeometric function, also denoted by $M(a, b, z)$, see Abramowitz and Stegun [1]. Here, $(a)_n$ denotes the *Pochhammer's symbol*:

$$(a)_0 = 1, \quad (a)_n = a(a+1) \cdots (a+n-1), \quad n = 1, 2, \ldots \tag{1.4}$$

and $(a)_n = \Gamma(a+n)/\Gamma(a)$ for $a > 0$ where $\Gamma(\cdot)$ denotes the gamma function.

Bhattacharya [2] showed that the unconditional pmf of X, that is,

$$f(x) = P(X = x) = \int_0^\infty f(x \mid \lambda) g(\lambda) d\lambda \tag{1.5}$$

is given by

$$f(x) = \frac{\alpha^a (\alpha+1)^{p-a} (p)_x}{x!(\alpha+2)^{x+p}} \, {}_2F_1\left(a, x+p; p; \frac{1}{\alpha+2}\right), \quad x = 0, 1, 2, \ldots, \tag{1.6}$$

where

$$ {}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n \, (b)_n}{(c)_n} \frac{z^n}{n!} \tag{1.7}$$

is the *hypergeometric function*, also denoted by $F(a, b; c; z)$, see Abramowitz and Stegun [1].

We use the notation HGNB(α, a, p) to denote the *hypergeometric generalized negative binomial* distribution with pmf (1.6).

SPECIAL CASES. (i) If $a = p$, then, using [1, formula (15.1.8), page 556],

$$ {}_2F_1(a, b; a; z) = {}_2F_1(b, a; a; z) = (1-z)^{-b}, \tag{1.8}$$

with $b = x + a$ and $z = 1/(\alpha+2)$, (1.6) reduces to

$$f(x) = \frac{(a)_x}{x!}\left(\frac{\alpha}{\alpha+1}\right)^a \left(1 - \frac{\alpha}{\alpha+1}\right)^x, \quad x = 0, 1, 2, \ldots \tag{1.9}$$

which is the pmf of negative binomial distribution with parameters $a > 0$ and success probability $\theta = \alpha/(\alpha+1) \in (0, 1)$.

(ii) If $a = 2$, $p = 1$, then, see the appendix, using the relation

$$ {}_2F_1(2, b; 1; z) = (1-z)^{-b}\left(1 - \frac{bz}{z-1}\right), \tag{1.10}$$

with $b = x + 1$ and $z = 1/(\alpha+2)$, (1.6) reduces to

$$f(x) = \alpha^2 \frac{x+\alpha+2}{(\alpha+1)^{x+3}}, \quad x = 0, 1, 2, \ldots \tag{1.11}$$

which is the pmf of Poisson-Lindley distribution with parameter α considered by Sankaran [7].

(iii) If $a = 1, p = 2$, then, see the appendix, using the relation

$$ {}_2F_1(1, n+2; 2; z) = \frac{1}{(n+1)z}\left[(1-z)^{-(n+1)} - 1\right], \quad n = 0, 1, 2, \ldots, \tag{1.12}$$

with $n = x$ and $z = 1/(\alpha+2)$, (1.6) reduces to

$$f(x) = \alpha(\alpha+1)\left[\frac{1}{(\alpha+1)^{x+1}} - \frac{1}{(\alpha+2)^{x+1}}\right], \quad x = 0, 1, 2, \ldots \tag{1.13}$$

which is the pmf of a generalized mixture of two geometric distributions with success probabilities $\theta_1 = 1/(\alpha+1)$ and $\theta_2 = 1/(\alpha+2)$, respectively.

The negative binomial distribution provides a more flexible alternative to the Poisson distribution particularly when the variance of the data is significantly larger than the mean. Johnson et al., [4, Chapter 5], provides a comprehensive survey of the applications and generalizations/extensions of the negative binomial distributions.

The discrete Poisson-Lindley distribution was shown by Sankaran [7] to provide, for particular data sets, better fit than other discrete distributions such as negative binomial, Poisson and Hermite distributions. Yet, no attempt has been made to study the properties of this distribution analytically.

The aim of this paper is to investigate some important properties of the hypergeometric generalized negative binomial distribution. These include existence of moments as well as properties of statistical measures such as the index of dispersion, skewness, and kurtosis. Also, a recurrence relation for calculating probabilities from the considered distribution is given. Finally, the distribution is fitted to entomological field data and its goodness-of-fit is demonstrated.

2. Moments and associated measures. We start this section by showing that the HGNB distribution has moments of all positive orders.

THEOREM 2.1. *For all $\alpha, a, p > 0$, the* HGNB(α, a, p) *distribution has moments of all positive orders*

$$\mu'_r = E(X^r) = \sum_{n=0}^{r} S(r,n) \frac{(p)_n}{(\alpha+1)^n} {}_2F_1\left(a,-n;p;-\frac{1}{\alpha}\right) \quad r = 1,2,\ldots, \tag{2.1}$$

where $S(r,n)$ are the Stirling numbers of the second kind

$$S(r,n) = \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} (-1)^i (n-i)^r. \tag{2.2}$$

PROOF. Since $X \mid \Lambda = \lambda$ has a Poisson distribution with parameter λ, then

$$E[X(X-1)\cdots(X-n+1) \mid \Lambda = \lambda] = \lambda^n. \tag{2.3}$$

Hence, the factorial moments of X, that is, $\mu'_{[n]} = E[X(X-1)\cdots(X-n+1)]$, are given by

$$\mu'_{[n]} = E\{E[X(X-1)\cdots(X-n+1) \mid \Lambda]\} = E\{\Lambda^n\}. \tag{2.4}$$

Making use of the following integral, see Erdély [3],

$$\int_0^\infty e^{-st} t^{b-1} {}_1F_1(a;c;qt) dt = \frac{\Gamma(b)}{s^b} {}_2F_1\left(a,b;c;\frac{q}{s}\right), \tag{2.5}$$

provided $\operatorname{Re} b, \operatorname{Re} s > 0$, $\operatorname{Re} s > \operatorname{Re} q$, $|s| > |q|$, we obtain

$$\begin{aligned} E\{\Lambda^n\} &= \frac{\alpha^a (\alpha+1)^{p-a}}{\Gamma(p)} \int_0^\infty \lambda^{n+p-1} e^{-(\alpha+1)\lambda} {}_1F_1(a;p;\lambda) d\lambda \\ &= \frac{\alpha^a (p)_n}{(\alpha+1)^{n+a}} {}_2F_1\left(a,n+p;p;\frac{1}{\alpha+1}\right). \end{aligned} \tag{2.6}$$

Now using [1, formula (15.3.4), page 559]:

$$_2F_1(a,b;c;z) = (1-z)^{-a} {_2F_1}\left(a,c-b;c;\frac{z}{z-1}\right),\tag{2.7}$$

with $b = n+p$, $c = p$, $z = 1/(\alpha+1)$, and the definition of hypergeometric function, respectively, we obtain

$$\mu'_{[n]} = \frac{(p)_n}{(\alpha+1)^n} {_2F_1}\left(a,-n;p;-\frac{1}{\alpha}\right) = \frac{(p)_n}{(\alpha+1)^n} \sum_{k=0}^{n} \frac{(a)_k(-n)_k}{(p)_k} \frac{(-1/\alpha)^k}{k!},\tag{2.8}$$

which is finite.

Finally, since $S(r,n)$, $n = 0,1,\ldots,r$, are finite and $\mu'_r = \sum_{n=0}^{r} S(r,n)\mu'_{[n]}$, the theorem follows. □

SPECIAL CASES. (i) If $a = p$, then, using (1.8) with $b = -n$ and $z = -1/\alpha$, (2.1) reduces to

$$\mu'_r = \sum_{n=0}^{r} S(r,n)\frac{(a)_n}{\alpha^n}\tag{2.9}$$

which is the rth moment of the negative binomial distribution with pmf (1.9).

(ii) If $a = 2$, $p = 1$, then, using (1.10) with $b = -n$ and $z = -1/\alpha$, (2.1) reduces to

$$\mu'_r = \sum_{n=0}^{r} S(r,n)n!\frac{n+\alpha+1}{(\alpha+1)\alpha^n}\tag{2.10}$$

which is the rth moment of the Poisson-Lindley distribution with pmf (1.11).

(iii) If $a = 1$, $p = 2$, then, see the appendix, using the relation

$$_2F_1(-n,1;2;z) = -\frac{n!}{(2)_n z}[(1-z)^{n+1}-1], \quad n = 0,1,2,\ldots\tag{2.11}$$

with $z = -1/\alpha$, (2.1) reduces to

$$\mu'_r = \sum_{n=0}^{r} S(r,n)n!\left[\frac{\alpha+1}{\alpha^n} - \frac{\alpha}{(\alpha+1)^n}\right]\tag{2.12}$$

which is the rth moment of the generalized mixture of geometric distributions with pmf (1.13).

THEOREM 2.2. *For all $a,p,\alpha > 0$, the HGNB(α,a,p) distribution is overdispersed, skewed to the right, and leptokurtic.*

PROOF. The characteristic function of $X \sim$ HGNB(α,a,p), see [2, page 28], is given by

$$\psi_X(t) = E\{e^{itX}\} = \frac{\left(1-(e^{it}-1)/(\alpha+1)\right)^{a-p}}{\left(1-(e^{it}-1)/\alpha\right)^a}, \quad i = \sqrt{-1}, \ -\infty < t < \infty.\tag{2.13}$$

Using the cumulant generating function $K_X(t) = \ln[\psi_X(t)]$, the rth cumulant of X is given by $\kappa_r = i^{-r}(d^r/dt^r)K_X(0)$. Therefore, the first four cumulants of X, respectively,

are given by

$$
\begin{aligned}
\kappa_1 &= \frac{a + p\alpha}{\alpha(\alpha+1)}, \\
\kappa_2 &= \frac{a(\alpha+1)^3 - (a-p)\alpha^2(\alpha+2)}{\alpha^2(\alpha+1)^2}, \\
\kappa_3 &= \frac{a(\alpha+1)^4(\alpha+2) - (a-p)\alpha^3(\alpha+2)(\alpha+3)}{\alpha^3(\alpha+1)^3}, \\
\kappa_4 &= \frac{a(\alpha+1)^5[(\alpha+3)^2-3] - (a-p)\alpha^4(\alpha+2)[(\alpha+4)^2-3]}{\alpha^4(\alpha+1)^4}.
\end{aligned}
\tag{2.14}
$$

Recall that the index of dispersion (ID), skewness ($\sqrt{\beta_1}$), and kurtosis (β_2) of X, respectively, are given by

$$
\mathrm{ID} = \frac{\kappa_2}{\kappa_1}, \qquad \sqrt{\beta_1} = \frac{\kappa_3}{\sqrt{\kappa_2^3}}, \qquad \beta_2 = 3 + \frac{\kappa_4}{\kappa_2^2}.
\tag{2.15}
$$

It follows that

$$
\begin{aligned}
\mathrm{ID} &= \frac{a(\alpha+1)^3 - (a-p)\alpha^2(\alpha+2)}{\alpha(\alpha+1)(a+p\alpha)} > 1, \\
\sqrt{\beta_1} &= \frac{a(\alpha+1)^4(\alpha+2) - (a-p)\alpha^3(\alpha+2)(\alpha+3)}{\left[a(\alpha+1)^3 - (a-p)\alpha^2(\alpha+2)\right]^{3/2}} > 0, \\
\beta_2 &= 3 + \frac{a(\alpha+1)^5[(\alpha+3)^2-3] - (a-p)\alpha^4(\alpha+2)[(\alpha+4)^2-3]}{\left[a(\alpha+1)^3 - (a-p)\alpha^2(\alpha+2)\right]^2} > 3,
\end{aligned}
\tag{2.16}
$$

proving the theorem. □

REMARKS. (i) If $a = p$, the index of dispersion does not depend on a while the skewness and kurtosis depend on a.

(ii) Expressions for the index of dispersion, skewness and kurtosis for the negative binomial distribution with pmf (1.9), Poisson-Lindley with pmf (1.11) and generalized mixture of geometric distributions with pmf (1.13), respectively, are obtained when $(a,p) \equiv (a,a), (2,1), (1,2)$.

Figures 3.1, 3.2, and 3.3, respectively, show the index of dispersion, skewness, and kurtosis of the HGNB$(1,a,p)$ distribution for selected values of a and p.

3. Recurrence relation. The following theorem provides a recurrence relation for calculating probabilities.

THEOREM 3.1. *For all $\alpha, a, p > 0$, the HGNB(α, a, p) distribution satisfies the recurrence relation*

$$
(x+1)f(x+1) = \left(\frac{a+x}{\alpha+1} + \frac{p+x-a}{\alpha+2}\right)f(x) - \frac{p+x-1}{(\alpha+1)(\alpha+2)}f(x-1),
\tag{3.1}
$$

where

$$
f(-1) = 0, \qquad f(0) = \left(\frac{\alpha}{\alpha+1}\right)^a \left(\frac{\alpha+1}{\alpha+2}\right)^{p-a}.
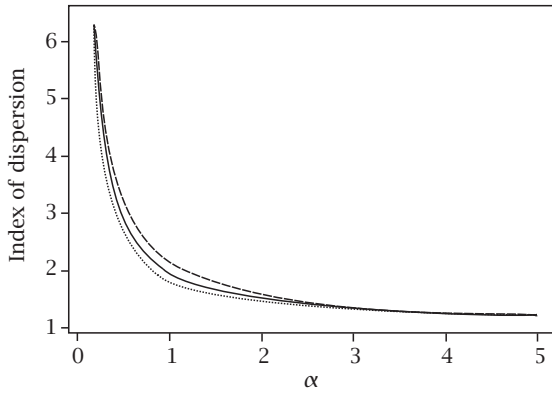\tag{3.2}
$$

FIGURE 3.1. The index of dispersion of the HGNB(α, a, p) distribution when $(a, p) \equiv (a, a)$ ($-$), $(1, 2)$ ($\cdots$), $(2, 1)$ ($---$).
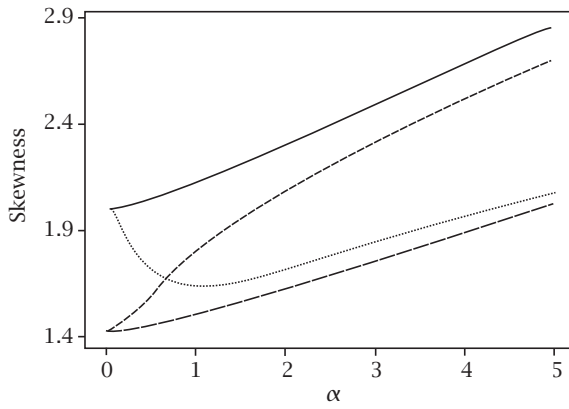

FIGURE 3.2. The coefficient of skewness of the HGNB(α, a, p) distribution when $(a, p) \equiv (1, 1)$ ($-$), $(1, 2)$ ($\cdots$), $(2, 1)$ ($---$), $(2, 2)$ ($-.-$).
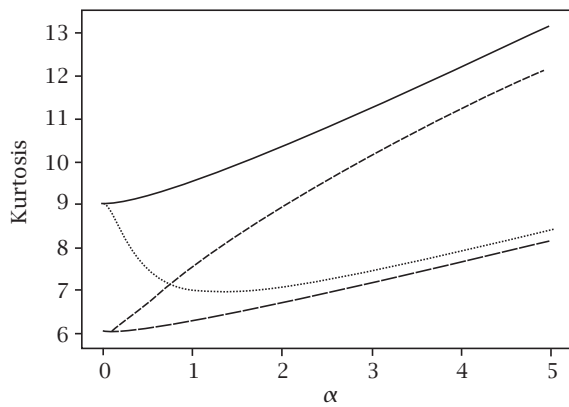

FIGURE 3.3. The coefficient of kurtosis of the HGNB(α, a, p) distribution when $(a, p) \equiv (1, 1)$ ($-$), $(1, 2)$ ($\cdots$), $(2, 1)$ ($---$), $(2, 2)$ ($-.-$).

Proof. Using [1, formula (15.2.11), page 558]:

$$(c-b)\,_2F_1(a,b-1;c;z)+(2b-c-bz+az)\,_2F_1(a,b;c;z)+b(z-1)\,_2F_1(a,b+1;c;z)=0. \tag{3.3}$$

Hence, for $b = p+x$, $c = p$ and $z \neq 1$, we obtain

$$(p+x)\,_2F_1(a,p+x+1;p;z) = \left(\frac{a+x}{1-z}+p+x-a\right)\,_2F_1(a,p+x;p;z) \\ -\frac{x}{1-z}\,_2F_1(a,p+x-1;p;z). \tag{3.4}$$

Rewrite $f(x)$, given by (1.6), as

$$f(x) = v(x)\,_2F_1(a,p+x;p;y), \quad x = 0,1,2,\ldots, \tag{3.5}$$

where

$$v(x) = (1-2y)^a(1-y)^{p-a}\frac{(p)_x y^x}{x!}, \quad y = \frac{1}{\alpha+2}. \tag{3.6}$$

It follows that

$$f(1) = \left(\frac{a}{\alpha+1}+\frac{p-a}{\alpha+2}\right)f(0). \tag{3.7}$$

Using the relations

$$\frac{x+1}{p+x}v(x+1) = yv(x) = y^2\frac{p+x-1}{x}v(x-1), \quad x = 1,2,\ldots \tag{3.8}$$

and (3.4), (3.5), and (3.6), we obtain, for $x = 1,2,\ldots,$

$$(x+1)f(x+1) = (x+1)v(x+1)\,_2F_1(a,p+x+1;p;y) \\ = \left(\frac{a+x}{1-y}+p+x-a\right)yv(x)\,_2F_1(a,p+x;p;y) \\ -\frac{y^2}{1-y}(p+x-1)v(x-1)\,_2F_1(a,p+x-1;p;y) \\ = \left(\frac{a+x}{1-y}+p+x-a\right)yf(x)-\frac{y^2}{1-y}(p+x-1)f(x-1). \tag{3.9}$$

Since $y = 1/(\alpha+2)$ and $1-y = (\alpha+1)/(\alpha+2)$, the theorem is proved. □

Table 4.1 shows the values of $f(x)$ of HGNB$(1,a,p)$ using the above recurrence relation when $(a,p) \equiv (1,1),(1,2),(2,1),(2,2)$.

4. Application. Table 4.2 shows the frequency distribution of European corn borer larvae *Pyrausta nubilalis* (Hnb.) in field corn, reported by McGuire et al. [5]. They showed that this frequency distribution is best fitted by the negative binomial distribution as compared to Neyman type A distribution and Poisson binomial distribution.

TABLE 4.1. Calculating $f(x)$ of HGNB$(1,a,p)$ using the recurrence relation (3.1).

x	$(a,p) \equiv$	$f(x)$			
		(1,1)	(1,2)	(2,1)	(2,2)
0		0.500000	0.333333	0.375000	0.250000
1		0.250000	0.277778	0.250000	0.187500
2		0.125000	0.175926	0.156250	0.125000
3		0.062500	0.100309	0.093750	0.078125
4		0.031250	0.054270	0.054688	0.046875
5		0.015625	0.028565	0.031250	0.027344
6		0.007813	0.014711	0.017578	0.015625
7		0.003906	0.007508	0.009766	0.008789
8		0.001953	0.003805	0.005371	0.004883
9		0.000977	0.001919	0.002930	0.002686
10		0.000488	0.000965	0.001587	0.001465

In the following, the HGNB(α,a,p) distribution is fitted to this data. The method of moments estimators of α, a, p, respectively, are given by $\hat{\alpha} = 5.550266 \times 10^{-4}$, $\hat{a} = 8.46876 \times 10^{-8}$, $\hat{p} = 5.2342$.

In calculating the chi-square statistic $\chi^2 = \sum_{i=1}^{m}(o_i - e_i)^2/e_i$, where $o_i(e_i)$ are the observed (expected) frequencies, $m = 13$ after combining the observed (expected) frequencies corresponding to counts 12 to 25, as did McGuire et al. [5], that is, $o_{13} = 15$ and $e_{13} = 14.87(14.55)$ for NB (HGNB). Also, the degrees of freedom are given by $m - t - 1$ where $t = 2(3)$ is the number of estimated parameters for NB (HGNB).

From Table 4.2, we observe that fitting the HGNB distribution gives an improvement over fitting the NB distribution as judged by the chi-square value.

Appendix. In the following, we make use of the following well-known relations of hypergeometric functions:

$$_2F_1(a,b;c;z) = (1-z)^{-b}\, _2F_1\left(c-a,b;c;\frac{z}{z-1}\right). \tag{A.1}$$

(see [1, formula (15.3.5), page 559])

$$_2F_1(-n,1;m;z) = -\frac{n!(z-1)^{m-2}}{(m)_n z^{m-1}}\left[(1-z)^{n+1} - \sum_{k=0}^{m-2}\frac{(n+1)_k}{k!}\left(\frac{z}{z-1}\right)^k\right], \tag{A.2}$$

where $n = 0,1,2,\ldots$, $m = 1,2,\ldots$, and for $m = 1$ the sum on the right-hand side is 0 (see [6, formula (179), page 466]).

PROOF OF (1.10). Using (A.1) with $a = 2$, $c = 1$, and the definition of the hypergeometric function, respectively, we obtain

$$_2F_1(2,b;1;z) = (1-z)^{-b}\, _2F_1\left(-1,b;1;\frac{z}{z-1}\right) = (1-z)^{-b}\left(1 - \frac{bz}{z-1}\right). \tag{A.3}$$

$\square$

TABLE 4.2. Fitting negative binomial (NB) and hypergeometric generalized negative binomial (HGNB) distributions to the frequency distribution of European corn borer larvae *Pyrausta nubilalis* (Hnb.) in field corn.

Count per plot x_i	Observed frequency o_i	Expected frequency e_i NB	Expected frequency e_i HGNB
0	10	8.92	8.62
1	18	22.96	22.54
2	39	35.37	35.14
3	33	42.32	42.36
4	42	43.34	43.59
5	56	39.92	40.24
6	36	34.01	34.31
7	26	27.31	27.52
8	19	20.92	21.04
9	19	15.42	15.46
10	7	11.02	11.00
11	4	7.66	7.62
12	4	5.21	5.15
13	4	3.47	3.41
14	2	2.27	2.22
15	1	1.47	1.42
16	2	0.94	0.90
17	1	0.59	0.56
18	0	0.20	0.35
19	0	0.12	0.21
20	0	0.08	0.13
21	0	0.05	0.08
22	0	0.03	0.05
23	0	0.02	0.03
24	0	0.01	0.02
25	1	0.41	0.02
Total	324	324	324
χ^2		14.55	14.23
df		10	9
P-value		0.149	0.114

PROOF OF (1.12). Using (A.1) with $a = n + 2$, $b = 1$, $c = 2$, and (A.2) with $m = 2$, respectively, we obtain

$$_2F_1(1, n+2; 2; z) = {_2F_1}(n+2, 1; 2; z)$$

$$= (1-z)^{-1} {_2F_1}\left(-n, 1; 2; \frac{z}{z-1}\right)$$

$$= (1-z)^{-1} \frac{-1}{(n+1)(z/(z-1))} \left[\left(1 - \frac{z}{z-1} \right)^{n+1} - 1 \right]$$

$$= \frac{1}{(n+1)z} \left[(1-z)^{-(n+1)} - 1 \right].$$

(A.4)

□

PROOF OF (2.11). Using (A.2) with $m = 2$, we obtain

$$_2F_1(-n,1;2;z) = -\frac{n!}{(2)_n z} \left[(1-z)^{n+1} - 1 \right].$$

(A.5)

□

REFERENCES

[1]   M. Abramowitz and I. Stegun, *Handbook of Mathematical Function*, 2nd ed., Dover, New York, 1972.

[2]   S. K. Bhattacharya, *Confluent hypergeometric distributions of discrete and continuous type with applications to accident proneness*, Calcutta Statist. Assoc. Bull. **15** (1966), 20–31.

[3]   A. Erdély, *Higher Transcendental Functions*, vol. I, McGraw-Hill, New York, 1953.

[4]   N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*, 2nd ed., John Wiley & Sons, New York, 1992.

[5]   J. U. McGuire, T. A. Brindley, and T. A. Bancroft, *The distribution of European corn borer larval Pyrausta nubilalis (Hbn.) in field corn*, Biometrics **13** (1957), 65–78.

[6]   A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series*, vol. 4, Gordon and Breach Science Publishers, New York, 1992.

[7]   M. Sankaran, *The discrete Poisson-Lindley distribution*, Biometrics **26** (1970), 145–149.

M. E. GHITANY AND S. A. AL-AWADHI: DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, FACULTY OF SCIENCE, KUWAIT UNIVERSITY, P.O. BOX 5969, SAFAT 13060, KUWAIT

S. L. KALLA: DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, FACULTY OF SCIENCE, KUWAIT UNIVERSITY, P.O. BOX 5969, SAFAT 13060, KUWAIT