ANNALES SOCIETATIS MATHEMATICAE POLONAE
Series I: COMMENTATIONES MATHEMATICAE XXI (1979)
ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria I: PRACE MATEMATYCZNE XXI (1979)

TADEUSZ GERSTENKORN (Łódź)

On the recurrence relation for the moments about an arbitrary point of a class of discrete inflated distributions

Introduction. The studies of discrete inflated distributions initiated by Indian investigators some dozen years ago were developed very rapidly. This has been treated of in Section 1. The aim of the paper is to state the conditions by occurrence of which it is possible to give the recurrence relation for the incomplete (and complete) moments about an arbitrary point in the case of integer random variables with their limited range of variability and with inflated probability distribution. This task has been performed in Section 2. In Section 3 we shall give an example of application of the formula obtained to the Pólya distribution which can also serve as a starting point towards obtaining those formulae for other, more particular cases, such as, for instance, binomial and hypergeometric distributions.

1. One-dimensional discrete inflated distributions. In 1963 S. N. Singh [8] proposed to consider the inflated Poisson distribution in statistical problems. This distribution was to serve as the probabilistic description of such experiments that were in substance well described by the Poisson distribution, yet with some "inflation" of the probability at the point zero.

In 1966, M. P. Singh [7] indicated that there exist analogous situations virtually well described by the binomial distribution, except for zero in which some inflation of the theoretically expected value of probability occurs. To be exact, at this point one observes a greater frequency of events than one could expect on the basis of the assumed probabilistic simple binomial model. The cited author gave an example of such a situation in which a population of four-person families was investigated. The set of the examined families consisted of two groups: the families 1° exposed, 2° not exposed to the risk of a certain disease in a given locality. The morbidity in the families of both groups being a mixture was noted then. There were zero, one, two, three or four cases of sick individuals in each family. It turned out that the situation investigated here was sufficiently well described by the binomial distribution at every point except the morbidity taking the value zero, but on the other hand, the example considered was completely well modelled by the inflated binomial. For the morbidity equal to zero, that is, in the case of no member of the family being ill, a greater frequency was observed than one could expect from the simple binomial distribution.

The inflated binomial distribution is defined as follows:

A random variable X is said to have the inflated binomial distribution if its probability function is given by

$$P(X = k) = \begin{cases} \beta+\alpha q^n & \text{for } k = 0, \\ \alpha \binom{n}{k} p^k q^{n-k} & \text{for } k = 1, 2, \ldots, n, \end{cases} \tag{1.1}$$

where α is a parameter assuming arbitrary values from the interval $(0, 1]$ and $\beta = 1-\alpha$, $0 < p < 1$ and $p+q = 1$. If $\alpha = 1$, then the above distribution reduces to the simple (uninflated) binomial one.

In paper [6] published a little later the same author generalized distribution (1.1) in the sense that he offered the inflation of the distribution at an arbitrary point k being the value of the random variable X. Thus we say that:

The random variable X has the generalized inflated binomial distribution if its probability function is expressed by

$$P(X = k) = \begin{cases} \beta+\alpha \binom{n}{l} p^l q^{n-l} & \text{for } k = l, \\ \alpha \binom{n}{k} p^k q^{n-k} & \text{for } k = 0, \ldots, l-1, l+1, \ldots, n, \end{cases} \tag{1.2}$$

where $0 < \alpha \leqslant 1$, $\alpha+\beta = 1$, $0 < p < 1$, $p+q = 1$.

It is easy to show that both formulae (1.1) and (1.2) present probability distributions.

In a similar way one defines the generalized inflated Poisson distribution. This distribution was examined by K. N. Pandey [5].

Some other discrete inflated distributions have been introduced to investigations and examined. Here we shall mention the negative binomial, geometric, Pólya and power series distributions.

Having considered all those remarks, we can formulate the following definition in a natural way.

The random variable X with the discrete probability distribution P has the inflated distribution if its probability function is given by

$$\bar{P}(X = k) = \begin{cases} \beta+\alpha P(X = l; \Theta) & \text{for } k = l, \\ \alpha P(X = k; \Theta) & \text{for } k \neq l, \end{cases} \tag{1.3}$$

where $\alpha \in (0, 1]$, $\beta = 1-\alpha$ and Θ is a parameter of the distribution.

To simplify the whole matter we shall not make any terminological distinction used by M. P. Singh, i.e. we shall not deal with the inflated distribution apart from the generalized inflated one.

Let us note that the inflated distribution is a particular case of the mixture of two distributions P_1 and P_2, the former being degenerate at $k = l$ and given in the form

$$P_1(X = k) = \begin{cases} 1 & \text{if } k = l, \\ 0 & \text{if } k \neq l, \end{cases}$$

and the latter – any discrete distribution. We define the mixture of two distributions P_1 and P_2 as such a distribution $P(X = k)$ that is given by

$$P(X = k) = (1-\alpha)P_1(X = k)+\alpha P_2(X = k),$$

α being a coefficient (often unknown and to be estimated from a random sample) of the share of the distribution P_2 in the mixture. One can see that the investigations of inflated distributions represent a particular case of the studies of mixed distributions. The importance of these researches in practice is justified by the significance of the mixtures in statistical problems.

The essential problems considered in examination of inflated distributions are:

(1) estimation of the coefficient α (fraction of the share of a non-degenerate distribution in the mixture);

(2) estimation of the distribution parameter Θ (e.g. of the parameter p in the case of a binomial distribution);

(3) calculation of the moments of the distribution, possibly of other characteristics or their estimators.

The review of the papers on the above-mentioned topics has been given in [2].

2. The recurrence relation for the incomplete moments about an arbitrary point of a class of discrete inflated distributions. Let us consider a discrete distribution in which the random variable X assumes the values $-\infty < k < +\infty$ with the probability $p_k = P(X = k)$.

The sum defined by

$$\mu_r^c(s) = \sum_{k=s}^{\infty} (k-c)^r p_k, \quad r = 1, 2, \ldots, \tag{2.1}$$

where s is any integer from the range of variability of the random variable X, will be called a *right-hand incomplete moment of order* r about an arbitrary constant c.

Let us note that when $s = -\infty$, formula (2.1) transforms into the complete moment μ_r^c about an arbitrary constant c. However, if $c = E(X)$, then we obtain an incomplete central moment on the right, and if, moreover,

$s = -\infty$, then we obtain the most common of the afore-mentioned cases — the moment about the mean μ_r.

The probabilistic and statistical importance of incomplete moments has been brought out in [1] and [3]. The theorem given below formulates the conditions under which it is possible to state the recurrence relation for the incomplete moments about an arbitrary point of the inflated distributions discussed in the previous section.

THEOREM. *Let the random variable X assume the integer values $k = m, m+1, \ldots, l, \ldots, n$. Let $\bar{p}_k = \bar{P}(X = k)$ be its inflated probability function dependent on the parameter $\Theta = (\Theta_1, \ldots, \Theta_s)$ and given by* (1.3). *Let $\bar{\mu}_r^c(s)$ denote the moment of distribution* (1.3), *defined by* (2.1), *which is incomplete on the right and calculated with respect to a constant c (c may also depend upon the above-mentioned parameter). We assume that for $k \neq l$ the following conditions are satisfied*:

$$(k-c)\bar{p}_k = U_k - U_{k+1}, \tag{A}$$

and for $k = l$

$$(l-c)\bar{p}_l = \beta(l-c) + U_l - U_{l+1}, \tag{A'}$$

where

$$U_{k+1} = \alpha p_k [A_1 (k-c)^2 + A_2 (k-c) + A_3], \tag{B}$$

$p_k = P(X = k)$ denotes the probability of the distribution without inflation, and the coefficients A_1, A_2, A_3 may also depend upon the parameter Θ; besides

$$(r-1) A_1 \neq 1, \quad r = 2, 3, \ldots, \tag{C}$$

and

$$U_{n+1} = 0. \tag{D}$$

Under the above conditions the following relation holds:

$$\begin{aligned} \bar{\mu}_r^c(s) = {} & [1-(r-1)A_1]^{-1} \Big\{ (s-c)^{r-1} U_s + \\ & + \sum_{i=0}^{r-3} \binom{r-1}{i} [A_1 \bar{\mu}_{i+2}^c(s) + A_2 \bar{\mu}_{i+1}^c(s) + A_3 \bar{\mu}_i^c(s)] + \\ & + (r-1)[A_2 \bar{\mu}_{r-1}^c(s) + A_3 \bar{\mu}_{r-2}^c(s)] - \\ & - \beta \sum_{i=0}^{r-2} \binom{r-1}{i} [A_1 (l-c)^{i+2} + A_2 (l-c)^{i+1} + A_3] \Big\} \end{aligned} \tag{2.2}$$

or in another notation

(2.2a) $$\bar{\mu}_r^c(s) = [1+(r-1)A_1]^{-1}\Big\{(s-c)^{r-1}U_s+A_1\sum_{i=0}^{r-3}\binom{r-1}{i}\bar{\mu}_{i+2}^c(s)+$$
$$+\sum_{i=0}^{r-2}\binom{r-1}{i}\big[(A_1\bar{\mu}_{i+1}^c(s)+A_2\bar{\mu}_{i+1}^c(s)+A_3\bar{\mu}_i^c(s))-$$
$$-\beta\big(A_1(l-c)^{i+2}+A_2(l-c)^{i+1}+A_3\big)\big]\Big\}.$$

Proof. In accordance with the definition adopted and with the assumption in the form of (A), we have:

$$\bar{\mu}_r^c(s) = \sum_{\substack{k=s\\k\neq l}}^{n}(k-c)^{r-1}\bar{p}_k(k-c)+\bar{p}_l(l-c)^r$$
$$= \sum_{\substack{k=s\\k\neq l}}^{n}(U_k-U_{k+1})(k-c)^{r-1}+\bar{p}_l(l-c)^r$$
$$\sum_{\substack{k=s\\k\neq l}}^{n}[(k-c)^{r-1}U_k-(k+1-c)^{r-1}U_{k+1}]+$$
$$+\sum_{\substack{k=s\\k\neq l}}^{n}[(k+1-c)^{r-1}-(k-c)^{r-1}]U_{k+1}+\bar{p}_l(l-c)^r.$$

After condition (D) has been taken into account, the first sum takes the form

(2.3) $$(s-c)^{r-1}U_s+(l+1-c)^{r-1}U_{l+1}-(l-c)^{r-1}U_l.$$

By using the Newton binomial and assumption (A′) we have, in the case of the other sum,

(2.4) $$\sum_{\substack{k=s\\k\neq l}}^{n}\sum_{i=0}^{r-2}\binom{r-1}{i}(k-c)^iU_{k+1}+\beta(l-c)^r+U_l(l-c)^{r-1}-U_{l+1}(l-c)^{r-1}.$$

Taking into consideration results (2.3) and (2.4) and using again the Newton formula, we have, by means of simple transformations

$$\bar{\mu}_r^c(s) = (s-c)^{r-1}U_s+\sum_{i=0}^{r-2}\binom{r-1}{i}\sum_{k=s}^{n}(k-c)^iU_{k+1}.$$

The use of condition (B) gives, after easy calculations, equality (2.2a).

3. Example of application. We are going to show that the Pólya distribution with the inflation at the point $k = l$ ($l \geqslant s$) for c equal to the expected value of the variable having the Pólya uninflated distribution, i.e. for $c = E(X) = np$, satisfies the assumptions of the presented theorem.

In accordance with the definition given in Section 1 we shall say that the random variable X has the inflated Pólya distribution if its probability function is given by (1.3), p_k being the probability function of the Pólya distribution in the form

$$p_k = P(X = k) = \binom{n}{k} \frac{p^{[k,-a]} q^{[n-k,-a]}}{1^{[n,-a]}}, \tag{3.1}$$

where $0 < p < 1$, $q = 1-p$, $k = 0, 1, \ldots, s, \ldots, l, \ldots, n$ and the conditions: $-ka \leqslant p$, $-(n-k)a \leqslant q$ are fulfilled.

In the notation of formula (3.1) we have used the so-called *factorial polynomials* defined as follows:

$$x^{[0,a]} = 1, \quad x^{[k,a]} = x^{[k-1,a]}\big(x-(k-1)a\big), \quad k = 0, 1, \ldots, \tag{3.2}$$

where a is any number. In view of relation (3.2) we get

$$x^{[k,a]} = x(x-a)(x-2a)\ldots\big(x-(k-1)a\big).$$

In the case of $a = 1$ we apply the notation

$$x^{[k,1]} = x^{[k]}.$$

For the probability function (3.1) the following recurrence relation takes place

$$p_k = \frac{\big(n-(k-1)\big)\big(p+(k-1)a\big)}{k\big(q+(n-k)a\big)} p_{k-1} \tag{3.3}$$

([3], p. 34, (3.4)). We transform the difference $k-np$ in the following way

$$k-np = k\big(q+(n-k)a\big)-(n-k)(p+ka)$$

([3], p. 34, (3.1)). By using (3.3) and denoting

$$U_{k+1} = (n-k)(p+ka)\alpha p_k, \tag{3.4}$$

we may write down, for $k \neq l$,

$$(k-np)\bar{p}_k = U_k - U_{k+1},$$

which means the fulfilment of condition (A) of the theorem in Section 2. In the case of $k = l$ we have, by (3.3),

$$(l-np)\bar{p}_l = (l-np)(\beta+\alpha p_l) = \beta(l-np)+U_l-U_{l+1},$$

which shows that (A') holds.

Developing the product $(n-k)(p+ka)$ into a sum with respect to the powers $k-np$, we get

$$(n-k)(p+ka) = npq(1+na)-\big(p-na(q-p)\big)(k-np)-a(k-np)^2$$

([3], p. 34, (3.2)). So, one can see that condition (B) is satisfied because, by denoting

(3.5) $$A_1 = -a, \quad A_2 = na(q-p)-p, \quad A_3 = npq(1+na)$$

we may write down U_{k+1}, according to (3.4), in the following way:

$$U_{k+1} = \alpha p_k [A_1(k-np)^2 + A_2(k-np) + A_3],$$

assuming, however, that $(r-1)A_1 \neq 1$.

Now let us proceed to assumption (D). Let n denote the maximum value of the random variable. Considering the fact that $p_{n+1} = 0$, we may write

$$\mu_r^c(s) = \sum_{k=s}^{n+1} (k-c)^r p_k.$$

Then expression (2.3) takes the form

$$(s-c)^{r-1} U_s + (l+1-c)^{r-1} U_{l+1} - (l-c)^{r-1} U_l - (n+2-c)^{r-1} U_{n+2}.$$

It is obvious that $U_{n+2} = 0$, so assumption (D) of the theorem may be replaced by the condition $U_{n+2} = 0$.

We have shown that the Pólya distribution inflated at the point $k = l$ satisfies all the assumptions of the theorem considered.

From (3.3) and (3.4) we have

(3.6) $$U_s = sp_s(q+(n-s)a)\alpha.$$

By using formulae (2.2) (or (2.2a)), (3.5) and (3.6) it is easy to write down the recurrence formula for the moments about $c = np$ of the inflated Pólya distribution as well as to obtain various particular cases in the way discussed in [1].

References

[1] T. Gerstenkorn, *Bemerkungen über die zentralen unvollständingen und absoluten Momente der Pólya-Verteilung*, Appl. Math. (Zast. Mat.), 14, 4 (1975), p. 579–597.

[2] —, *Jednowymiarowe rozkłady dyskretne ze zniekształceniem*, Materiały konferencji „Metody statystyczne w sterowaniu jakością produkcji", Warszawa Jabłonna 24–28 listopada 1975. (*One-dimensional discrete inflated distributions*, Proceedings of the conference "Statistical methods in quality control", Warsaw-Jabłonna November 24–28, 1975, Published by Ossolineum, Wrocław 1977.)

[3] —, *The recurrence relations for the moments of the discrete probability distributions*, Diss. Math. (Rozprawy Mat.) 83 (1971), PWN, Warszawa.

[4] A. R. Kamat, *Incomplete and absolute moments of some discrete distributions*, Proceedings of the International Symposium "Classical and contagious distributions", McGill University, Montreal, Canada, August 15th–20th 1963, edited by G. P. Patil, Statistical Publishing Society, Calcutta 1965, p. 45–64.

[5] K. N. Pandey, *On generalized inflated Poisson distribution*, J. Sci. Res. Banaras Hindu Univ. 15, 2 (1964–1965), p. 157–162.

[6] M. P. Singh, *A note on generalized inflated binomial distribution*, Sankhyā Ser. A. 28, 1 (1966), p. 99.

[7] —, *Inflated binomial distribution*, J. Sci. Res. Banaras Hindu Univ. 16, 1 (1965–1966), p. 87–90.

[8] S. N. Singh, *A note on inflated Poisson distribution*, J. Indian Statist. Assoc. 1, 3 (1963), p. 140–144.

ŁÓDŹ UNIVERSITY, INSTITUTE OF MATHEMATICS