Statistics & Probability Letters 4 (1986) 313–318
North-Holland

October 1986

THE STUTTERING GENERALIZED WARING DISTRIBUTION

John PANARETOS
University of Patras, and University of Crete, Greece

Evdokia XEKALAKI
University of Crete, Greece *

Received February 1985
Revised April 1986

1. Introduction

A vast number of problems in areas as diverse as biology, economics, accident theory, linguistics, reliability and bibliographical analysis have been linked with the generalized Waring distribution defined by the probability function (p.f.)

$$P(X=x)=\frac{c_{(m)}}{(a+c)_{(m)}}\frac{a_{(x)}m_{(x)}}{(a+m+c)_{(x)}}\frac{1}{x!},\quad x=0,1,2,\ldots,\quad a,m,c>0, \tag{1.1}$$

where $\alpha_{(\beta)}$ stands for $\Gamma(\alpha+\beta)/\Gamma(\alpha)$, $\alpha>0$, $\beta\in R$. (See e.g. Xekalaki (1981, 1983a), Xekalaki and Panaretos (1983), Schubert and Glänzel (1984).) The nature of (1.1) reveals the possibility of obtaining this distribution in the context of sampling from an urn. In fact, Xekalaki (1981) has provided an inverse sampling scheme with additional replacements. This urn representation of the generalized Waring distribution can of course be associated with a variety of games of chance. There are, however, several such games that one can define associated with the generalized Waring distribution in an indirect and rather general sort of way. Motivated by a game of this nature the present authors were led to an extension of the generalized Waring distribution, the stuttering generalized Waring distribution, with properties closely related to those of the generalized Waring distribution. The derivation of this distribution resembles the derivation of Panaretos and Xekalaki's (1986) generalized binomial distribution.

So, in Section 2 a description of the model that gives rise to the stuttering generalized Waring distribution is given. In Section 3 the probability generating function of the distribution is derived in terms of Lauricella's hypergeometric series of type D and the mean and variance of the distribution are obtained.

* Now at the Athens School of Economics and Business Studies, Athens, Greece.

The case of higher order moments is also discussed. Section 4 considers an alternative derivation of the stuttering generalized Waring distribution. Finally, in Section 5 some potential applications are discussed in areas such as bibliographic research, linguistics, ecology and inventory control.

2. The model

Consider an urn that contains c white and a black balls. Let k, m_1, $m_2, \ldots, m_k$ be positive integers and define $M_r = \sum_{i=1}^{r} m_i$, $r = 1, 2, \ldots, k$. The player draws a ball at random, observes its colour and returns it to the urn along with one additional ball of the same colour before he draws the next ball. Drawings are continued in this fashion until M_k black balls have been drawn. Let X_i be the number of white balls sampled after the (M_{i-1})-th black ball has been drawn and before the (M_i)-th black ball is drawn, $i = 1, 2, \ldots, k$, $(M_0 = 0)$ and let $X = \sum_{i=1}^{k} iX_i$. The player's gain is defined to be the difference $X - h$ where h is a fee paid before the player enters the game.

Theorem 2.1. *For any value* x, $x = 0, 1, 2, \ldots,$ *the* $P(X = x)$ *is given by*

$$P(X=x) = \frac{c_{(\Sigma m_i)}}{(a+c)_{(\Sigma m_i)}} \sum_{\Sigma_{j=1}^{k} jx_j = x} \frac{a_{(\Sigma x_i)}}{(a+c+\Sigma m_i)_{(\Sigma x_i)}} \prod_{i=1}^{k} \frac{(m_i)_{(x_i)}}{x_i!}, \quad x = 0, 1, 2, \ldots. \tag{2.1}$$

Proof. From the definition of X it follows that $P(X = x) = \sum_{\Sigma i x_i = x} P(\bigcap_{i=1}^{k} \{X_i = x\})$. Xekalaki (1984b) showed that the urn scheme considered leads to a multivariate version of (1.1), the multivariate generalized Waring distribution, representing the distribution of $(X_1, X_2, \ldots, X_k)$. I.e.

$$P\left(\bigcap_{i=1}^{k} \{X_i = x_i\}\right) = \frac{c_{(\Sigma m_j)}}{(a+c)_{(\Sigma m_i)}} \frac{a_{(\Sigma x_i)} \prod_{i=1}^{k} (m_i)_{(x_i)}}{(a + \Sigma m_i + c)_{(\Sigma x_i)}} \frac{1}{\prod_{i=1}^{k} x_i!},$$
$$x = 0, 1, 2, \ldots, \quad i = 1, 2, \ldots, k. \tag{2.2}$$

The result follows then as an immediate consequence. Hence the theorem has been established.

Theorem 2.2. *Let* X *be defined as in Theorem* 2.1. *Then* (2.1) *defines a proper probability distribution on* $\{0, 1, 2, \ldots\}$.

Proof. It suffices to show that $\sum_{x=0}^{\infty} P(X = x) = 1$.

From (2.1) we have

$$\sum_{x=0}^{\infty} P(X=x) = \frac{c_{(\Sigma m_i)}}{(a+c)_{(\Sigma m_i)}} \sum_{x=0}^{\infty} \sum_{jx_j = x} \frac{a_{(\Sigma x_i)}}{(a+c+\Sigma m_i)_{(\Sigma x_i)}} \prod_{i=1}^{k} \frac{(m_i)_{(x_i)}}{x_i!}.$$

Set $x_i = r_i$, $x = r + \sum_{i=1}^{k} (i-1) r_i$, $i = 1, 2, \ldots, k$. Then

$$\begin{aligned}
\sum_{x=0}^{\infty} P(X=x) &= \frac{c_{(\Sigma m_i)}}{(a+c)_{(\Sigma m_i)}} \sum_{r=0}^{\infty} \sum_{\Sigma r_i = r} \frac{a_{(\Sigma r_i)}}{(a+\Sigma m_i + c)_{(\Sigma r_i)}} \prod_{i=1}^{k} \frac{(m_j)_{(r_j)}}{r_j!} \\
&= \frac{c_{(\Sigma m_i)}}{(a+c)_{(\Sigma m_i)}} \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} \frac{a_{(\Sigma r_i)}}{(a+\Sigma m_i + c)_{(\Sigma r_i)}} \prod_{j=1}^{k} \frac{(m_j)_{(r_j)}}{r_j!} \\
&= \frac{c_{(\Sigma m_i)}}{(a+c)_{(\Sigma m_i)}} \sum_{l=0}^{\infty} \frac{a_{(l)} (m_i)_{(l)}}{(a+\Sigma m_i + c)_{(l)}} \frac{1}{l!} = 1.
\end{aligned}$$

Theorems 2.1 and 2.2 lead to the definition of a new probability distribution which arises from an intermingling of generalized Waring streams.

Definition 2.1. Let k be a fixed positive integer. A r.v. X whose probability distribution is given by (2.1) will be said to have the *stuttering generalized Waring distribution* with positive parameters k, a, m_1, $m_2, \ldots, m_k$ and c.

3. The probability generating function and the moments of the distribution

Let X be a r.v. having the stuttering generalized Waring distribution with p.f. given by (2.1) and let $G(s)$ be its probability generating function (p.g.f.). Then

$$G(s) = c_{(\Sigma m_i)} \sum_{x=0}^{\infty} s^x \sum_{\Sigma j x_j = x} \frac{a_{(\Sigma x_i)}}{(a+c)_{(\Sigma x_i + \Sigma m_i)}} \prod_{i=1}^{k} \frac{(m_i)_{(x_i)}}{x_i!}.$$

If the transformation $x_i = r_i$, $x = r + \Sigma_{i=1}^{k}(i-1)r_i$ is applied, the above relationship becomes

$$G(s) = \frac{c_{(\Sigma m_i)}}{(a+c)_{(\Sigma m_i)}} \sum_{r_1=0}^{\infty} \sum_{r_2=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} \frac{a_{(\Sigma r_i)}}{(a + \Sigma m_i + c)_{(\Sigma r_i)}} \prod_{j=1}^{k} \frac{(m_j)_{(r_j)} s^{j r_j}}{r_j!},$$

i.e.

$$G(s) = \frac{c_{(\Sigma m_i)}}{(a+c)_{(\Sigma m_i)}} F_D\left(a;\, m_1, m_2, \ldots, m_k;\, a + \sum m_i + c;\, s, s^2, \ldots, s^k\right) \tag{3.1}$$

where F_D denotes Lauricella's hypergeometric series of type D defined by

$$F_D\left(\alpha;\, \beta_1, \beta_2, \ldots, \beta_k;\, \alpha + \sum \beta_i + \gamma;\, s_1, s_2, \ldots, s_k\right)$$
$$= \sum_{r_1=0}^{\infty} \cdots \sum_{r_k=0}^{\infty} \frac{\alpha_{(\Sigma r_i)} (\beta_1)_{(r_1)} \cdots (\beta_k)_{(r_k)}}{(\alpha + \Sigma \beta_i + \gamma)_{(\Sigma r_i)}} \frac{s_1^{r_1}}{r_1!} \cdots \frac{s_k^{r_k}}{r_k!}, \quad |s_i| \leqslant 1.$$

The moments of the stuttering generalized Waring distribution can therefore be obtained from (3.1) by differentiation. But, differentiating the p.g.f. leads to involved algebraic expressions that make even the derivation of low order moments tedious. However, obtaining the moments of the distribution in (2.1) becomes easier, if one observes that they can be associated with the moments of the random vector $(X_1, X_2, \ldots, X_k)$. From the definition of the r.v. X it follows that its mean and variance will be dependent upon the mean and covariances of the r.v.'s $X_1, X_2, \ldots, X_k$. In particular,

$$E(X) = \sum_{i=1}^{k} i E(X_i) \quad \text{and} \quad V(X) = \sum_{i=1}^{k} \sum_{j=1}^{k} ij \operatorname{Cov}(X_i, X_j).$$

As mentioned earlier, Xekalaki (1984b) showed that the random vector $(X_1, \ldots, X_k)$ follows the k-variate generalized Waring distribution with p.f. given by (2.3). As a result,

$$E(X_i) = \frac{a m_i}{c-1}, \quad V(X_i) = \frac{a m_i (c + m_i - 1)(c + a - 1)}{(c-1)^2 (c-2)}, \quad \operatorname{Cov}(X_i, X_j) = \frac{m_i m_j a (a + c - 1)}{(c-1)^2 (c-2)},$$

$i \neq j$.

Hence

$$E(X)=\frac{a}{c-1}\sum_{i=1}^{k} im_i \tag{3.2}$$

and

$$V(X)=\frac{a(c+a-1)}{(c-1)(c-2)}\left\{\sum_{i=1}^{k} i^2 m_i+\frac{1}{c-1}\left(\sum_{i=1}^{k} im_i\right)^2\right\}. \tag{3.3}$$

Since $E(X_i)$ and $V(X_i)$, $\mathrm{Cov}(X_i, X_j)$ exist only when $c>1$ and $c>2$ respectively it follows that $E(X)$ and $V(X)$ become infinite if $c \leqslant 1$ and $c \leqslant 2$ respectively.

Moments of higher order can be obtained through the usual transformation formulae from the factorial moments of the distribution. The latter can be obtained by the formula

$$\mu_{[l]}=E(X^{(l)})=\sum_{\sum l_i=l}\binom{l}{l_1,\ldots,l_k}E\left\{\prod_{i=1}^{k}\sum_{n_i=0}^{l_i}\binom{l_i}{n_i}i^{(l_i-n_i)}X_i^{(n_i)}\right\},\quad l=0,1,2,\ldots. \tag{3.4}$$

Obviously, $\mu_{[l]}$ is an expression of the first l factorial moments of the k-variate generalized Waring distribution given by

$$\mu_{[l_1,l_2,\ldots,l_k]}=\frac{a_{(\Sigma l_i)}\prod_{i=1}^{k}(m_i)_{(l_i)}}{(c-1)(c-2)\cdots(c-\Sigma l_i)},\quad \sum_{i=1}^{k} l_i \leqslant l \tag{3.5}$$

(Xekalaki, 1984b). Then, since $\mu_{(l_1,\ldots,l_k]}$ does not exist unless $c>\Sigma_{i=1}^{k} l_i$ it follows that $\mu_{[l]}$ becomes infinite for $l \geqslant c$. Hence the factorial moment generating function of the stuttering generalized Waring distribution does not exist.

For $k=1$, formulae (3.1), (3.2) and (3.3) reduce to those of the p.g.f., mean and variance of the generalized Waring distribution respectively.

4. An alternative urn scheme

Consider an urn that contains $c+\Sigma_{i=1}^{k} m_i$ chips, c blank and m_i marked with the number i, $i=1, 2,\ldots, k$ $(k \geqslant 1)$. The player draws a chip at random, records its number and returns it to the urn along with an identical chip before he draws the next chip. Draws are continued in this manner until a blank chips are drawn. The player's gain or loss is defined to be the difference $X-d$ where d is a fee payed by the player before the enters the game and X is the sum of the recorded numbers of the drawn chips.

Theorem 4.1. *For any value x, $x=0, 1, 2,\ldots$ the $P(X=x)$ is given by* (2.1) *and hence it defines the stuttering generalized Waring distribution with parameters* $(k, a, m_1,\ldots, m_k, c)$.

This model is pertinent to Steyn's (1956) urn scheme which led to a distribution of a more general form than that of (2.1). However, there is a flaw in the argument used to derive the moments as this presupposes the existence of the factorial moment generating function (which in fact does not exist).

5. Applications

The representation of a stuttering generalized Waring variable X by the sum $\Sigma_{i=1}^{k} iX_i$ where $(X_1, X_2,\ldots, X_k)$ is a k-variate generalized Waring vector, suggests the possibility of utilizing the stuttering generalized Waring distribution in practical situations where single events, pairs of events, triplets of events,..., k-tuples of events can be assumed to be jointly distributed according to the k-variate generalized Waring distribution.

The stuttering generalized Waring distribution, for example could be applied in bibliographic research which is an area where the generalized Waring distribution with parameters a, $m = 1$ and c has found applications (see e.g. Schubert and Glänzel (1984) and the references therein). In this context X_i may represent the number of authors who authored i articles in a given time period whence X will represent the total number of articles written by the $X_1 + \ldots + X_k$ authors. An alternative type of problem in the same area may arise in situations where scientific articles are counted by the name of the author. In such cases articles by i authors will be counted i times. Hence if X_i is thought of as representing the number of articles authored by i authors, X will represent the observed total number of articles which is greater than or equal to $X_1 + X_2 + \ldots + X_k$, the actual total number of articles. So, suppose that in such a situation it is reasonable to assume that $(X_1,\ldots X_k)$ follows the k-variate generalized Waring distribution with parameters a, $m_1 = m_2 = \ldots m_k = 1$ and c. Then if the stuttering generalized Waring distribution with parameters k, a, $m_1 = m_2 = \ldots = m_k = 1$ and c is found to describe satisfactorily the observed distribution of X one can use the estimates of the parameters a and c to infer about the distribution of $X_1 + X_2 + \ldots + X_k$, the unobserved actual number of articles (generalized Waring with parameters a, k, and c (Xekalaki (1984b))).

The generalized Waring distribution has also been applied to problems in the area of linguistics (see e.g. Herdan (1964)). Therefore, the stuttering generalized Warring distribution can possibly be used to describe the distribution of the total number X of sounds occurring in a spoken utterance when the numbers X_1, $X_2,\ldots, X_k$ of groups of 1, 2,..., k distinctive sounds have the k-variate generalized Waring distribution. Alternatively, the stuttering generalized Waring distribution can be used as a potential model for the description of the distribution of the number X of morphemes in a text with X_i representing the number of groups of i morphemes, $i = 1, 2,\ldots, k$.

The stuttering generalized Waring distribution may also be used as a stochastic abundance model in the ecological context where the generalized Waring distribution arises in connection with the description of features of an animal or plant population (see, e.g. Engen (1978)). Suppose that the structure of a population of organisms allows stratification into species according to some taxonomic classification. Then X can be thought of as representing the total number of individual organisms of a taxonomic group living in a particular area and the r.v. X_i may denote the number of species represented by i organisms in the population under consideration. So inferences can be drawn about the distribution of the size of the taxonomic group from information concerning the distribution of the vector $(X_1, X_2,\ldots, X_k)$, the species structure of the group. (Problems of similar nature have been examined by Panaretos (1983).)

Finally, inventory control problems may provide an interesting case for the application of the stuttering generalized Waring distribution since the ordinary generalized Waring distribution arises as a demand distribution (Xekalaki (1983b)). In this context X_i may represent the number of orders for i units of a particular product and X the total number of ordered units.

Acknowledgement

This work was partly supported by a grant from the ministry of Research and Technology of Greece.

References

Engen, S. (1978), *Stochastic Abundance Models* (Chapman and Hall, London).

Herdan, G. (1964), *Quantitative Linguistics* (Butterworths, London).

Panaretos, J. (1983), A generating model involving Pascal and logarithmic series distributions, *Commun. Statist. (Theory and Methods)* **12**, 841–848.

Panaretos, J. and E. Xekalaki (1986), On generalized binomial distributions and their relation to generalized Poisson distributions, *Ann. Inst. Statist. Math. A.* (to appear).

Schubert, A. and W. Glänzel (1984), A dynamic look at a class of skew distributions. A model with scientometric applications, *Scientometrics* **6**, 149–167.

Steyn, H.S. (1956), On the univariable series $F(t)F(a;\ b_1, b_2,\ldots, b_k;\ c;\ t,\ t^2,\ldots, t^k)$ and its applications in probability theory, *Proc. Kon. Ned. Akad. V. Wetensch Ser. A* **59**, 190–197.

Xekalaki, E. (1981), Chance mechanisms for the univariate generalized Waring distribution and related characterizations. Statistical Distributions in Scientific Work, in: C. Taillie, C.P. Patil and B. Baldessari, eds., *Models, Structures and Characterizations* (D. Reidel, Holland) 4, 157–171.

Xekalaki, E. (1983a), The univariate generalized Waring distribution in relation to accident theory: Proneness, spells or contagion?, *Biometrics* **39**, 887–895.

Xekalaki, E. (1983b), A property of the Yule distribution and its application, *Commun. Statist. (Theory and Methods)* **12**, 1181–1189.

Xekalaki, E. (1984a), The bivariate generalized Waring distribution and its application to accident theory, *J. Roy Statist. Soc. Ser. A* **147** (3), 488–498.

Xekalaki, E. (1984b), The multivariate generalized Waring distribution, *Commun. Statist. (Theory and Methods) A* **15** (special Issue on Discrete Distributions), to appear.

Xekalaki, E. and J. Panaretos (1983), Identifiability of compound Poisson distributions, *Scand. Actuarial J.*, 39–45.