The Multivariate Waring Distribution and Its Application

SHAN Shi[1], JIANG Guohua[2], JIANG Lan[1]

[1]*Department of Management and Information Engineering, Shanghai University, Shanghai 201800, China*
[2]*China National Institute for Educational Research, Beijing 100088, China*

The multivariate Waring distribution is developed and investigated. A special case, namely the bivariate Waring distribution, is considered. It is shown that this multivariate distribution enjoys nice statistical properties. An application to the distribution of scientific productivity is discussed.

1. Introduction

The article presents a theoretical generalization of the size-frequency form of Zipf's law based on a multivariate extension of a well known Zipf-like distribution, namely the Waring distribution.

By the size-frequency form of Zipf's law is meant here any classification of units such that the proportion of classes with k units is approximately proportional to $k^{-(1+a)}$, for some constant $a > 0$. In informetrics this regularity is also known as Lotka's law. It is a well known fact that such a relationship holds to a surprisingly good approximation in a variety of empirical areas, including linguistics (Simon, 1955), the distribution of scientific productivity (Price, 1976), the distributions of the number of pages visited within a given web site (Huberman et al., 1998) and (in) links pointing to web pages (Bormboldt & Ebel, 2001). Let X be a random variable defined on the set of positive (or non-negative) integers and let P(X=k) be the probability that X=k. The (univariate) Waring distribution with parameters a and b has the form

$$P(X = k) = a\frac{b^{(k-1)}}{(a+b)^{(k)}}, \quad k=1, 2, \ldots, \tag{1}$$

where $a^{(b)} = \Gamma(a+b)/\Gamma(a)$, $\Gamma(\cdot)$ is the gamma function, and both a and b are positive real numbers. It is easy to verify that the Waring distribution is Zipf-like, i.e., $P(X = k) = \varphi(k)/k^{a+1}$ with $\lim \varphi(k) = c > 0$ as k tends to infinity.

One of the most important factors in managerial planning is the ability to predict, if even roughly accurate, the future value of a variable of interest. Letting X and Y be

Shi Shan, Guohua Jiang and Lan Jiang, *The multivariate Waring distribution and its applications*, In: Guohua Jiang, Ronald Rousseau, Yishan Wu (Eds), Proceedings of the 9th International Conference on Scientometrics and Informetrics – ISSI 2003, Dalian University of Technology, Dalian, China, 2003, 289-296.

variables denoting the numbers of units from some sources during two successive equal time periods, then the joint probability of (X, Y), i.e., the bivariate theoretical model, will play an important role in predicting Y based upon observed X. By using regression functions, correlation coefficients and conditional probabilities, the prediction model is well formulated in an adaptive manner in order to update the information, which is the base for the prediction.

In the following sections, we will develop a multivariate Zipf-like distribution, namely the multivariate Waring distribution, which describes multi-dimensional Zipfian phenomena. We will further present some examples for its application. Moreover, we will show that this multivariate distribution has some nice properties.

2. Multivariate Waring Distribution

Let $P(X_1 = k_1, \cdots, X_n = k_n; a; b_1, \cdots, b_n)$ be the probability of the event "$X_1 = k_1, \cdots, X_n = k_n$" with parameters a and b_i, i = 1,…,n. The multivariate Waring distribution is defined by

$$P(X_1 = k_1, \cdots, X_n = k_n; a; b_1, \cdots, b_n)$$

$$= a \frac{\Gamma(\sum_{i=1}^{n} k_i - n + 1)\Gamma(\sum_{i=1}^{n} b_i + a)}{\Gamma(\sum_{i=1}^{n} k_i + \sum_{i=1}^{n} b_i - n + a + 1))} \prod_{i=1}^{n} \frac{\Gamma(k_i + b_i - 1)}{\Gamma(k_i)\Gamma(b_i)} \qquad (2)$$

$$(k_i = 1,2,\cdots; i = 1,2,\cdots, n)$$

where $\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$, and the parameters a and b_i, i=1,…n are positive real numbers. For n=1, Eq.(2) reduces to the following univariate form

$$P(X = k; a; b) = \frac{a\Gamma(b + k - 1)\Gamma(a + b)}{\Gamma(b)\Gamma(a + b + k)}, \qquad (3)$$

which is called the Waring distribution. Letting $a^{(b)} = \Gamma(a + b)/\Gamma(a) \quad (a > 0, b \geq 0)$, Eq.(3) becomes Eq.(1).

THEOREM I. Any marginal distribution of multivariate Waring distribution is also a (multivariate) Waring distribution:

$$\sum_{k_{s+1}=1}^{\infty}\cdots\sum_{k_n=1}^{\infty}P(X_1=k_1,\cdots,X_s=k_s,X_{s+1}=k_{s+1},\cdots,X_n=k_n;a;b_1,\cdots,b_s,b_{s+1},\cdots,b_n)$$

$$=P(X_1=k_1,\cdots,X_s=k_s;a;b_1,\cdots,b_s)\,. \qquad (4)$$

PROOF: See the Appendix in Section 2.

It is obvious that the marginal distribution of $X_i\,(i=1,2,\cdots,n)$ is the Waring distribution with parameters a and b_i.

THEOREM II. Let $(X_1,\cdots,X_n)$ follow a multivariate Waring distribution. Then its expectation $E(X_1,\cdots,X_n)$ can be expressed as

$$E(X_1,\cdots,X_n)=a\int_0^1(1-\theta)^{a-n-1}\prod_{i=1}^{n}(1-\theta+b_i\theta)d\theta \qquad (5)$$

It exists if and only if $a>n$.

PROOF: See the Appendix in Section 2.

3. Bivariate Waring Distribution

The bivariate Waring distribution is defined by

$$P(X=k,Y=j;a;b,c)$$

$$=a\frac{\Gamma(k+j-1)\Gamma(b+k-1)\Gamma(c+j-1)\Gamma(a+b+c)}{\Gamma(a+b+c+k+j-1)\Gamma(b)\Gamma(c)(k-1)!(j-1)!}$$

$$=a\frac{(k+j-2)!b^{(k-1)}c^{(j-1)}}{(a+b+c)^{(k+j-1)}(k-1)!(j-1)!}\,. \qquad (6)$$

$(k,j=1,2,\cdots)$

PROPERTY I. If (X,Y) follows the bivariate Waring distribution, then

$$E(X,Y)=1+\frac{b+c}{a-1}+\frac{2bc}{(a-1)(a-2)}$$

and

$$Cov(X,Y)=1+\frac{b+c}{a-1}+\frac{2bc}{(a-1)(a-2)}-(1+\frac{b}{a-1})(1+\frac{c}{a-1})\,.$$

PROOF: See the Appendix in Section 3.

Let "$X = k \mid Y = m$" denote the event "X=k conditional on Y=m" and let the probability of "$X = k \mid Y = m$" be P(X=k|Y=m). Then we have the following property.

PROPERTY II. Let (X,Y) follow the bivariate Waring distribution $P(X = k, Y = j; a; b, c)$, then

$$P(X = k \mid Y = m) = q(X = k; a + c; b; m)$$
$$= \frac{(a+c)^{(b)}}{(a+c+m)^{(b)}} \cdot \frac{b^{(k-1)} m^{(k-1)}}{(a+b+c+m)^{(k-1)}} \cdot \frac{1}{(k-1)!}, \qquad (7)$$

where m is an integer and k=1,2,….

PROOF: See the Appendix in Section 3.

Eq.(7) is a three parameters discrete distribution and is called the generalized Waring distribution with parameters $a + c$, b and m. Symmetrically, it follows that $P(Y = j \mid X = h) = q(Y = j; a + b; c; h)$ (see Eq.(7)).

REMARK. The generalized Waring distribution received its name from Irwin (in his address for the chairmanship of the Royal Statistical Society, 1963) who based its generalization on an inverse factorial series developed by Waring in 18th century. The distribution has the following form

$$q(X = k; \alpha; \beta; \rho) = \frac{\alpha^{(\beta)}}{(\alpha+\rho)^{(\beta)}} \cdot \frac{\beta^{(k-1)} \rho^{(k-1)}}{(\alpha+\beta+\rho)^{(k-1)}} \cdot \frac{1}{(k-1)!}, \qquad (8)$$

(k=1, 2, …)

and is called the generalized Waring with parameters α, β and ρ. In the particular case that $\rho = 1$ it becomes the (simple) Waring with parameters α and β. It is interesting to note that another special case of the distribution, namely the case $\beta = \rho = 1$ was called the Yule distribution by Kendall (also in a presidential address to the Royal Statistical Society, 1961). Kendall suggests using the Yule distribution for bibliographic and economic applications.

PROPERTY III. Let (X,Y) follow the bivariate Waring distribution $P(X = k, Y = j; a; b, c)$. The regression function of X on Y=m is

$$E(X \mid Y=m) = 1 + \frac{b}{a+c-1} m . \qquad (9)$$

PROOF: See the Appendix in Section 3.

THEOREM III. Let X and Y be positive, integer-valued random variables such that the conditional distribution of X given Y=m is the generalized Waring distribution with parameters $a+c$, b and m as given by Eq.(7). Then the distribution of X is the Waring distribution with parameters a and b as given by Eq.(1) if and only if (iff) the distribution of Y is the Waring distribution with parameters a and c, i.e., the case for b = c in Eq.(3).

PROOF: See the Appendix in Section 3.

Equivalently, the above theorem can be expressed as follows. If $P(X=k \mid Y=m) = q(X=k; a+c; b; m)$ (see Eq.(7)) holds for positive, integer-valued random variable X and Y, then X follows $P(X=k; a; b)$ (see Eq.(1)) if and only if Y follows $P(Y=j; a; c)$ (also see Eq.(1)).

4. Applications

Bai (1996) reported an actual distribution of scientific productivity among authors in six main Chinese magazines of information science during the first three-year period (1987-1989) and the second three-year period (1990-1992). Let $\hat{f}(k, j)$ denote the number of authors who have produced k papers in the first period and j papers in the second period, $\hat{f}(k)$ (the marginal form for k) is equal to $\sum_j \hat{f}(k, j) = \hat{f}(k, \cdot)$, and $\hat{f}(j)$ (the marginal form for j) equals $\sum_k \hat{f}(k, j) = \hat{f}(\cdot, j)$. Table 1 shows a joint distribution of scientific productivity during the two successive periods.

TABLE 1. Observed number of authors who produced at least one paper in both the two successive periods (1987-1989, 1990-1992) in six main Chinese magazine of information science (k and j are the no. of papers produced within the first and second periods respectively).

k \ j	1	2	3	4	5	6	7	8	9	10	12	13	15	18	$\hat{f}(j)$
1	100	33	15	5	3	2	2	1	0	0	0	0	0	0	161
2	34	12	8	6	4	0	0	1	0	0	0	1	0	0	66
3	11	6	7	3	2	0	0	1	0	1	0	0	0	0	31
4	6	5	4	3	2	1	0	0	0	0	0	0	0	0	21
5	2	0	0	2	2	0	0	0	0	0	0	0	1	0	7
6	0	1	1	2	2	0	2	0	0	0	0	0	0	1	9
7	0	0	0	0	0	0	0	0	1	0	1	1	0	0	3
8	0	0	2	1	1	0	0	0	0	0	0	0	0	0	4
10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
$\hat{f}(k)$	153	57	37	22	16	3	5	3	1	1	1	2	1	1	303

The parameters a, b and c can be estimated from $\hat{f}(1,\cdot)$ and $\hat{f}(\cdot,1)$, each of which is the first fitted frequency of its corresponding marginal form, and sample mean $\overline{X}$ or sample mean $\overline{Y}$, where $\overline{X} = \sum_k k\hat{f}(k)$ and $\overline{Y} = \sum_j j\hat{f}(j)$. This estimation, which seems to be by far the most convenient manner, is only for exploratory and illustrational purposes. The theoretical distribution can be calculated by the relations

$$\begin{cases} P(X = k+1, Y = j) = \dfrac{(k+j-i)(b+k-1)}{(a+b+c+k+j-1)\,j} P(X = k, Y = j) \\ P(X = k, Y = j+1) = \dfrac{(k+j-1)(c+j-1)}{(a+b+c+k+j-1)k} P(X = k, Y = j) \end{cases},$$

when parameters a, b and c are known.

TABLE 2. Expected number of authors who produced at least one paper in both the two successive periods (1987-1989, 1990-1992) in six main Chinese magazines of information science (k and j are the no. of papers produced within the first and second periods respectively) (calculated from Table 1).

j \ k	1	2	3	4	5	6	7	8	9	10	12	13	15	18	$\hat{f}(j)$
1	105.86	33.07	12.17	5.05	2.30	1.13	0.59	0.32	0.19	0.11	0.04	0.03	0.01	0.00	160.88
2	29.75	17.09	8.73	4.50	2.39	1.33	0.76	0.45	0.28	0.17	0.07	0.05	0.02	0.01	65.61
3	10.09	8.04	5.10	3.07	1.84	1.12	0.70	0.44	0.29	0.19	0.09	0.06	0.03	0.01	31.06
4	3.92	3.88	2.87	1.95	1.29	0.85	0.56	0.37	0.25	0.18	0.09	0.06	0.03	0.01	16.31
5	1.69	1.95	1.63	1.22	0.87	0.61	0.43	0.30	0.21	0.15	0.08	0.06	0.03	0.01	9.23
6	0.79	1.03	0.94	0.76	0.58	0.43	0.32	0.23	0.17	0.12	0.07	0.05	0.03	0.01	5.54
7	0.40	0.57	0.56	0.48	0.39	0.30	0.23	0.18	0.13	0.10	0.06	0.05	0.03	0.01	3.48
8	0.21	0.32	0.34	0.31	0.26	0.21	0.17	0.13	0.10	0.08	0.05	0.04	0.02	0.01	2.27
9	0.12	0.19	0.21	0.20	0.18	0.15	0.12	0.10	0.08	0.06	0.04	0.03	0.02	0.01	1.53
10	0.07	0.12	0.14	0.14	0.12	0.11	0.09	0.08	0.06	0.05	0.03	0.03	0.02	0.01	1.05
15	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.00	0.22
$\hat{f}(k)$	152.89	66.28	32.72	17.70	10.25	6.26	3.98	2.62	1.78	1.23	0.63	0.46	0.26	0.12	297.19
$\hat{a}=3.63021, \hat{b}=3.55903, \hat{c}=3.20179.$															

As shown in Table 2, the expected frequencies have not been grouped and only the individual values are presented for comparative purposes. The fit is quite good, considering that we are approximating a very skew joint distribution and that we are not using the most efficient method of fitting. With a more efficient estimation procedure, e.g. maximum-likelihood estimation, a significant improvement can probably be obtained.

We have partitioned the observed frequencies in Table 1 into $4 \times 5 = 20$ groups (cells) so that each group would hold at least 5 authors and the percent of these groups would be no less than 80. Let $\hat{F}_{kj}$ be the observed frequency in (k, j)-group, F_{kj} the expected frequency in the corresponding group, (where k=1, 2, 3, 4; j=1, 2, 3, 4, 5), and $\chi^2 = \sum_{j=1}^{5}\sum_{k=1}^{4}(F_{kj} - \hat{F}_{kj})^2 / \hat{F}_{kj}$. The bivariate Waring has been fitted to Bai's data based upon the estimated parameters shown on the bottom of Table 2. The results are as follows:

$$\hat{\chi}^2 = 10.6472\,,\quad \chi_p^2[(m-1)\times(n-1)] = \chi_{0.05}^2(3\times 4) = 21.026\,.$$

The χ^2 value does not rule out the hypothesis.

APPENDIX

Section 2

We consider an alternative genesis of Eq.(A2). If the conditional distribution of $(X_1, \cdots, X_n)$, given θ, is of the form

$$P(X_1 = k_1, \cdots, X_n = k_n \mid \theta) = \prod_{i=1}^{n} P(X_i = k_i \mid \theta)$$

where

$$P(X_i = k_i \mid \theta) = \frac{\Gamma(k_i + b_i - 1)}{\Gamma(k_i)\Gamma(b_i)} \theta^{b_i} (1-\theta)^{k_1 - 1},$$

$$(k_i = 1,2,\cdots; i = 1,2,\cdots,n)$$

which is a negative binomial distribution with parameters b_i $(b_i > 0, i = 1,2,\cdots,n)$ and θ $(0 < \theta < 1)$, and the parameter θ is also a random variable having a distribution $F(\theta)$ with $F'(\theta) = f(\theta) = a(1-\theta)^{a-1}$ $(a > 0)$, which is a special case of the Beta distribution with a density function $f(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{b-1}(1-\theta)^{a-1}$ for b=1, then the unconditional distribution of $(X_1, \cdots, X_n)$ becomes

$$\begin{aligned} & P(X_1 = k_1, \cdots, X_n = k_n) \\ &= \int_0^1 P(X_1 = k_1, \cdots, X_n = k_n \mid \theta) f(\theta) d\theta \\ &= a \prod_{i=1}^{n} \frac{\Gamma(k_i + b_i - 1)}{\Gamma(k_i)\Gamma(b_i)} \int_0^1 \theta^{\sum_{i=1}^{n} k_i - n} (1-\theta)^{\sum_{i=1}^{n} k_i + a - 1} d\theta \qquad \text{(A1)} \\ &= P(X_1 = k_1, \cdots, X_n = k_n; a; b_1, \cdots, b_n). \end{aligned}$$

Eq.(A1), i.e., Eq.(2) is a zero-modified form of the following formula

$$\begin{aligned} & P(X_1 = k_1, \cdots, X_n = k_n) \\ &= \frac{\Gamma(\sum_{i=1}^{n} k_i + 1)}{\prod_{i=1}^{n} \Gamma(k_i + 1)} \cdot \frac{\alpha \prod_{i=1}^{n} \beta_i^{(k_i)}}{(\sum_{i+1}^{n} \beta_i + \alpha)^{(\sum_{i=1}^{n} k_i + 1)}}, \qquad \text{(A2)} \end{aligned}$$

$$(k_i = 0,1,2,\cdots; i = 1,2,\cdots,n)$$

which is a special case of the multivariate inverse Polya-Eggenberger distribution (Johnson, Kotz and Balakrishnan, 1996). It should be noted that although the transformation (from Eq.(A2) to Eq.(A1)) is natural, the derivation of the corresponding moments of Eq.(A1) requires careful consideration of the regions of variation of both restricted and originally unrestricted variables. Some further results have been obtained and will be reported elsewhere.

PROOF OF THEOREM I. It is obvious from Eq.(A1) that

$$\sum_{k_{s+1}=1}^{n}\cdots\sum_{k_n=1}^{\infty}P(X_1=k_1,\cdots,X_s=k_s,X_{s+1}=k_{s+1},\cdots,X_n=k_n;a;b_1,\cdots,b_s,b_{s+1},\cdots,b_n)$$

$$=\int_0^1\left\{\sum_{k_{s+1}=1}^{\infty}\cdots\sum_{k_n=1}^{\infty}\prod_{i=s+1}^{n}P(X_i=k_i\mid\theta)\right\}\prod_{i=1}^{s}P(X_i=k_i\mid\theta)f(\theta)d\theta$$

$$=\int_0^1\prod_{i=1}^{s}P(X_i=k_i\mid\theta)f(\theta)d\theta$$

$$=\int_0^1 P(X_1=k_1,\cdots X_s=k_s\mid\theta)f(\theta)d\theta$$

$$=P(X_1=k_1,\cdots,X_s=k_s;a;b_1,\cdots,b_s).$$

PROOF OF THEOREM II. Seeing

$$\sum_{i=1}^{\infty}k_i\frac{\Gamma(k_i+b_i-1)}{\Gamma(k_i)\Gamma(b_i)}\theta^{k_i-1}(1-\theta)^{k_i}=1+\theta b_i/(1-\theta),$$

which is expectation of X_i of the negative binomial with b_i and θ, we have

$$E(X_1,\cdots,X_n)=\sum_{k_n=1}^{\infty}\cdots\sum_{k_1=1}^{\infty}k_1\cdots k_nP(X_1=k_1,\cdots,X_n=k_n;a;b_1,\cdots b_n)$$

$$=a\int_0^1(1-\theta)^{a-1}\left[\sum_{k_n=1}^{\infty}\cdots\sum_{k_1=1}^{\infty}k_1\cdots k_n\prod_{i=1}^{n}\frac{\Gamma(k_i+b_i-1)}{\Gamma(k_i)\Gamma(b_i)}\theta^{k_i-1}(1-\theta)^{b_i}\right]d\theta$$

$$=a\int_0^1(1-\theta)^{a-1}\prod_{i=1}^{n}(1+b_i\frac{\theta}{1-\theta})d\theta$$

$$=a\int_0^1(1-\theta)^{a-n-1}\prod_{i=1}^{n}(1-\theta+\theta b_i)d\theta \qquad (A3)$$

Thus, the first part of Theorem II has been proved. The following relation

$$a\int_0^1(1-\theta)^{a-1}\prod_{i=1}^{n}(1+b_i\frac{\theta}{1-\theta})d\theta$$

$$\geq a\prod_{i=1}^{n}b_i\int_0^1(1-\theta)^{a-1}\theta^n(1-\theta)^{-n}d\theta$$

$$=a\frac{\Gamma(n+1)\Gamma(a-n)}{\Gamma(a+1)}\prod_{i=1}^{n}b_i$$

$$= \frac{\Gamma(n+1)}{(a-1)\cdots(a-n)} \prod_{i=1}^{n} b_i$$

shows that $E(X_1,\cdots,X_n)$ will not exist if $a \le n$. If $a > n$,

$$a\int_0^1 (1-\theta)^{a-1} \prod_{i=1}^{n} (1+b_i \frac{\theta}{1-\theta})d\theta$$

$$\le a\int_0^1 (1-\theta)^{a-n-1}(1-\theta+\theta\bar{b})^n d\theta$$

$$= a\int_0^1 (1-\theta)^{\delta-1} \sum_{i=0}^{n} \binom{n}{i} \theta^i (1-\theta)^{n-i} \bar{b}^i d\theta$$

$$< \infty$$

for $a - n = \delta > 0$ and $\bar{b} = \underset{i}{Max}\{b_i\}$. The last inequality holds because

$$\int_0^1 (1-\theta)^{\delta-1} \theta^i (1-\theta)^{n-i} d\theta$$

$$= \int_0^1 \theta^i (1-\theta)^{n+\delta-i-1} d\theta$$

$$= \frac{\Gamma(i+1)\Gamma(n+\delta-i)}{\Gamma(n+\delta+1)} < \infty$$

for $i \in \{0,1,\cdots,n\}$.

Section 3

PROOF OF PROPERTY I. From Eq.(5), it holds that

$$E(XY) = a\int_0^1 (1-\theta)^{a-2-1}(1-\theta+b\theta)(1-\theta+c\theta)d\theta$$

$$= a\int_0^1 (1-\theta)^{a-3}\left[(1-\theta)^2 + bc\theta^2 + (1-\theta)\theta(b+c)\right]d\theta$$

$$= a\int_0^1 \left[(1-\theta)^{a-1} + bc(1-\theta)^{a-3}\theta^2 + (b+c)(1-\theta)^{a-2}\theta\right]d\theta$$

$$= 1 + abc\frac{\Gamma(a-2)\Gamma(3)}{\Gamma(a+1)} + a(b+c)\frac{\Gamma(a-1)\Gamma(2)}{\Gamma(a+1)}$$

$$= 1 + \frac{b+c}{a-1} + \frac{2bc}{(a-1)(a-2)}$$

where the following formula for the beta distribution $\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ has been used. Seeing from Eq.(5) that (for n=1)

$$E(X_i) = a\int_0^1 (1-\theta)^{a-1-1}(1-\theta+b_i\theta)d\theta$$

$$= a\int_0^1 \left[(1-\theta)^{a-1} + (1-\theta)^{a-2}\theta b_i\right]d\theta$$

$$= 1 + \frac{\Gamma(a-1)\Gamma(2)}{\Gamma(a+1)}ab_i = 1 + \frac{b_i}{a-1}$$

$$(i = 1,2,\cdots,n)$$

and noting that $E(X)E(Y) = (1+\frac{b}{a-1})(1+\frac{c}{a-1})$, we have

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

$$= 1 + \frac{b+c}{a-1} + \frac{2bc}{(a-1)(a-2)} - (1+\frac{b}{a-1})(1+\frac{c}{a-1}).$$

PROOF OF PROPERTY II. Substituting

$$P(Y = m; a; c) = a\frac{c^{(m-1)}}{(a+c)^{(m)}} \tag{A4}$$

and Eq.(6) for Y=m into

$$P(X = k \mid Y = m) = \frac{P(X = k, Y = m; a; b, c)}{P(Y = m; a; c)}$$

yields

$$P(X = k \mid Y = m) = \frac{b^{(k-1)}(a+c)^{(m)}(m+k-2)!}{(m-1)!(a+b+c)^{(k+m-1)}(k-1)!}$$

$$= \frac{\Gamma(a+c+b)\Gamma(m+k-1)\Gamma(a+c+m)b^{(k-1)}}{\Gamma(a+c)\Gamma(m)\Gamma(a+b+c+m+k-1)(k-1)!}$$

$$= \frac{(a+c)^{(b)}}{(a+c+m)^{(b)}}\frac{m^{(k-1)}b^{(k-1)}}{(a+b+c+m)^{(k-1)}(k-1)!}. \tag{A5}$$

PROOF OF PROPERTY III. The property follows directly from the first moment of the generalized Waring distribution with parameters $a+c$, b and m.

PROOF OF THEOREM III. We substitute Eq.(A4) and Eq.(A5) in the well-known formula

$$P(X = k) = \sum_j P(Y = j)P(X = k \mid Y = j) \tag{A6}$$

and obtain $P(X = k) = \sum_{j=1}^{\infty} P(X = k, Y = j; a; b, c)$

$$= a\frac{b^{(k-1)}}{(a+b)^{(k)}} \qquad (A7)$$

$$= P(X = k; a; b)$$

The converse of this result is also true. Using a method used by Xekalaki (1981), one can show that the functional equation (A6) in P(Y=j) has a unique solution under the conditions that P(X=k) and P(X=k|Y=j) have an explicit form such as Eq.(A5) and Eq.(A7) respectively (see Xekalaki's paper for details). Though a further discussion will be of considerable interest, it seems to be far beyond the scope of the present paper.

Acknowledgements. We are indebted to two referees for valuable comments and suggestions and, especially, for a referee's recommendation of a book on multivariate distributions, and to Mr. Bai Pencheng for his unpublished work. We are further deeply indebted to Dr. R. Rousseau for his helpful comments and encouragement.

This research was supported by the Shanghai Natural Sciences Foundation, grant 02ZD14032.

Reference

1. BAI PENGCHENG. "Some Empirical Studies about the Distribution of Scientific Productivity in the Field of Information Science in China (in Chinese)." *Technical Report, 1996, Liberal Arts College, Shanghai University.*
2. S.BORNHOLDT and H.EBEL. "World Wide Web Scaling Exponent from Simon's 1955 Model." *Physical Review E,* 65, pp.1-4 (035104), 2001.
3. B.A.HUBERMAN, L.T.PIROLLI, J.E.PITKOW and R.M.LUKOSE. "Strong Regularities in World Wide Web Surfing." *Science, 280,* pp.95-98,1998.
4. W.GLANZEL, A.TELCS and A.SCHUBERT. "Characterization by Truncated Moments and Its Application to Pearson-Type Distributions." *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete,* 66, pp.173-183, 1984.
5. J.O.IRWIN. "The Place of Mathematics in Medical and Biological Statistics." *Journal of the Royal Statistical Society,* A 126, pp.1-45, 1963.
6. N.L. JOHNSON, S.KOTZ and N.BALAKRISHNAN. *Discrete Multivariate Distributions.* John Wiley & Sons, New York, 1997.
7. R. KENDALL. "Natural Law in the Social Sciences." *Journal of the Royal Statistical Society,* A 124, pp.1-19, 1961.
8. D. de Solla PRICE. "A General Theory of Bibliometric and Other Cumulative Advantage

Processes." *Journal of the American Society for Information Science,* 27, pp.292-306, 1976.
9. H.A.SIMON. "On a Skew Distribution Function." *Biometrika*, 42, pp.425-440, 1955.
10. K. TAKAHASI, "Note on The Multivariate Burr's Distribution." *Ann. Inst. Statist. Math.*, 17, pp.257-260, 1965.
11. E. XEKALAKI. "Chance Mechanisms for The Univariate Generalized Waring Distribution And Related Characterization." In: C. TAILLIE, G. PATIL (Eds). *Statistical Distributions in Scientific Work, Vol.-4-Model, Structures, and Characterization.* pp.157-171. D. Reidel Publ. Co., Dordrecht, 1981.

A revised version of this contribution has been published as:
Shi Shan, Huohua Jiang and Lan Jiang (2004). The multivariate Waring distribution and its application. Scientometrics, 60(3), 523-535.