

CHARACTERIZATION AND STATISTICAL TEST USING TRUNCATED EXPECTATIONS FOR A CLASS OF SKEW DISTRIBUTIONS

A. TELCS, W. GLÄNZEL and A. SCHUBERT

Library of the Hungarian Academy of Sciences, Department of Informatics and Science Analysis, Budapest, POB 7, H-1361 Hungary

Communicated by R.M. Frumkina

Received 7 December 1983

Revised 28 September 1984

The expectation of left truncated Waring and Pareto distributions is a linear function of the point of truncation. Based on this property, a characterization theorem and statistical tests can be constructed.

Key words: Characterization; truncated expectation; Waring distribution; Pareto distribution; word frequency distribution.

1. Introduction

1.1. History

The Waring distribution was originally proposed by Irwin (1963) for modelling biological phenomena. Later on, it has found applications in accident statistics (Irwin, 1968), quantitative linguistics (Herdan, 1964) and information science (Tague, 1981) as well. A remarkable direction in its theoretical development was that oriented toward identifying the 'Generalized Waring Distribution', a family of distributions, including a number of well-known and widely used distributions like Poisson, negative binomial, etc. (Irwin, 1975a, 1975b, 1975c; Xekalaki, 1981). A characteristic feature of the Waring distribution is that with suitably chosen parameters it might be highly skewed, namely it may have an extremely long tail. This feature enables it to model all the above-mentioned phenomena; however, it also has some undesirable consequences in applying usual statistical tests, particularly the chi-squared test for establishing goodness of fit. On the practical side, the long tail manifests in empirical samples as a long sequence of small non-negative integers, which needs pooling before applying the test. Pooling is, however, always somewhat arbitrary and may lead to inconsequential results. (Some inconsistencies of the chi-squared test as applied to a Waring sample have already been pointed out by Herdan (1964, p. 88).) On theoretical side, depending on its

$$r(n) = \frac{e(n+1) - e(n)}{e(n+1) - n}, \quad n \in \mathbb{N}_0, \quad (1.5)$$

provided that both sides make sense.

2. Properties of the Waring distribution

2.1. Definition

We say that X has a Waring distribution with parameters $\alpha > 1$ and $N > 0$ (further on $\text{Wd}(\alpha, N)$) if

$$p(k) = p_{\text{W}}(k; \alpha, N) = \frac{\alpha}{\alpha + N} \prod_{i=1}^k \frac{N+i-1}{N+i+\alpha}, \quad k \in \mathbb{N}_0, \quad (2.1)$$

or, equivalently,

$$P(k) = P_{\text{W}}(k; \alpha, N) = \prod_{i=0}^k \frac{N+i-1}{N+\alpha+i-1}, \quad k \in \mathbb{N}_0, \quad (2.2)$$

2.2. Remarks

(i) Function $p_{\text{W}}(k; \alpha, N)$ obeys the recursion relation

$$p_{\text{W}}(k+1; \alpha, N) = \frac{N+k}{N+\alpha+k+1} p_{\text{W}}(k; \alpha, N), \quad k \in \mathbb{N}_0, \quad (2.3)$$

(ii) Functions $p_{\text{W}}(k; \alpha, N)$ and $P_{\text{W}}(k; \alpha, N)$ are related by the equation

$$P_{\text{W}}(k; \alpha, N) = \frac{N+k-1}{\alpha} p_{\text{W}}(k-1; \alpha, N), \quad k \in \mathbb{N}_0. \quad (2.4)$$

2.3. Moments

The expectation and the variance of $\text{Wd}(\alpha, N)$ are:

$$E(X; \alpha, N) = \frac{N}{\alpha - 1}, \quad (2.5)$$

provided that $\alpha > 1$, otherwise it does not exist; and

$$D^2(X; \alpha, N) = \frac{\alpha N(\alpha + N - 1)}{(\alpha - 2)(\alpha - 1)^2} = \frac{\alpha}{\alpha - 2} E(x; \alpha, N)(E(x; \alpha, N) + 1), \quad (2.6)$$

provided that $\alpha > 2$, otherwise it does not exist.

Lemma. *Truncating $\text{Wd}(\alpha, N)$ from the left by n results in $\text{Wd}(\alpha, N+n)$.*

parameters, the Waring distribution may exhibit a non-Gaussian character, rendering Gaussian moment statistics inapplicable (Haitun, 1982a, 1982b, 1982c).

1.2. Aim of the paper

In this paper a characterization theorem for the Waring distribution is presented. The theorem is based on truncated expectations and belongs to the family of representation theorems of Kotz and Shanbhag (1980). More specifically, the statement of the theorem is the discrete analog of the characterization of the Pareto distribution given by Morrison (1978) (in the proof we proceed on a completely different tack). The theorem may serve as a basis for a simple test to decide whether Waring distribution fits to a given empirical sample or to a set of such samples.

1.3. Basic concepts and notations

Let X be a non-negative integer valued random variable. The following notations will be used:

$$p(k) = \Pr(X = k), \quad k \in \mathbb{N}_0,$$

$$P(k) = \Pr(X \geq k), \quad k \in \mathbb{N}_0;$$

$$r(k) = \Pr(X = k | X \geq k), \quad k \in \mathbb{N}_0.$$

The function $r(k)$ so defined is called the *hazard rate* of X . Functions $P(k)$ and $r(k)$ uniquely determine each other:

$$r(k) = 1 - \frac{P(k+1)}{P(k)}, \quad k \in \mathbb{N}_0, \quad (1.1)$$

$$P(k) = \prod_{i=0}^{k-1} (1 - r(i)), \quad k \in \mathbb{N}_0. \quad (1.2)$$

A given distribution can be truncated from the left by n resulting in the distribution:

$$p^n(k) = \Pr(X = k | X \geq n), \quad k \in [n, \infty).$$

The expectation of this truncated distribution will be denoted by

$$e(n) = E(X | X \geq n), \quad n \in \mathbb{N}_0.$$

Obviously,

$$p^n(k) = \frac{p(k)}{P(n)}, \quad k \in [n, \infty), \quad (1.3)$$

$$e(n) = \frac{\sum_{k=n}^{\infty} kp(k)}{P(n)}, \quad n \in \mathbb{N}_0. \quad (1.4)$$

It can easily be shown from the above formulae that

Proof. Equation (1.3) together with equations (2.1) through (2.4) leads to

$$p_{\text{W}}^n(k; \alpha, N) = p_{\text{W}}(k - n; \alpha, N + n). \quad (2.7)$$

Corollary. From equation (2.7) it follows that

$$E(X - n | X \geq n; \alpha, N) = E(X; \alpha, N + n). \quad (2.8)$$

Characterization theorem. Let X be a non-negative integer valued random variable and assume that the range of X is the whole \mathbb{N}_0 . Then, X has $\text{Wd}(\alpha, N)$ if and only if $e(n)$ is a linear function of n , namely

$$e(n) = \frac{\alpha}{\alpha - 1} n + \frac{N}{\alpha - 1}, \quad \alpha > 1, N > 0, n \in \mathbb{N}_0. \quad (2.9)$$

Proof. Substituting equation (2.9) into equation (1.5) and the result into equation (1.2) we get

$$P(k) = \prod_{i=0}^{k-1} \frac{N + i}{\alpha + N + i},$$

in accordance with equation (2.2). For the reverse case, from equations (2.5) and (2.8) we have

$$\begin{aligned} e(n) &= E(X | X \geq n; \alpha, N) \\ &= n + E(X; \alpha, N + n) \\ &= \frac{\alpha}{\alpha - 1} n + \frac{N}{\alpha - 1}, \end{aligned}$$

as stated by the theorem.

2.4. Generalizations

(i) The theorem is also valid in the case $\alpha \leq -1$, provided that N is a negative integer. In this case a finite Waring distribution is obtained. Particularly, $\alpha = -1$ leads to the discrete uniform distribution.

(ii) Let Y be a non-negative real valued random variable. The equation

$$E(Y | Y \geq y) = \frac{\alpha}{\alpha - 1} y + \frac{\beta}{\alpha - 1}, \quad \alpha > 1, \beta > 0, y \in \mathbb{R}_0,$$

analogous to equation (2.9), characterizes now the Pareto distribution, $\text{Pd}(\alpha, \beta)$:

$$\text{Pr}(Y \geq y) = P_p(y; \alpha, \beta) = \left(1 + \frac{y}{\beta}\right)^{-\alpha}, \quad y \in \mathbb{R}_0$$

(see, for example, Johnson and Kotz, 1970), which can thus be regarded as the continuous analog to $\text{Wd}(\alpha, N)$.

2.5. Asymptotic behaviour

The tail of both $Wd(\alpha, N)$ and $Pd(\alpha, \beta)$ is asymptotically proportional to $x^{-\alpha}$, i.e.

$$\lim_{x \rightarrow \infty} x^\alpha P_W(x; \alpha, N) = c,$$

$$\lim_{x \rightarrow \infty} x^\alpha P_p(x; \alpha, \beta) = c',$$

c and c' being constants depending only on the parameters but independent of x . It can thus be seen that both distributions are Gaussian or non-Gaussian according as $\alpha > 2$ or $\alpha \in (0, 2]$, respectively (see, for example, Feller, 1957).

3. Statistical applications

3.1. Single-sample test

Consider a sample of size N , whose elements take the values $0, 1, 2, \dots$ of a random variable with absolute frequencies $N(0), N(1), N(2), \dots$, respectively. Obviously,

$$N = \sum_{k=0}^{\infty} N(k), \tag{3.1}$$

and the relative frequencies are

$$f(k) = \frac{N(k)}{N}, \quad k \in \mathbb{N}_0. \tag{3.2}$$

We shall also use the notation

$$F(k) = \sum_{i=k}^{\infty} f(i), \quad k \in \mathbb{N}_0. \tag{3.3}$$

The n th truncated sample mean is defined as

$$x(n) = \frac{\sum_{k=n}^{\infty} kf(k)}{F(n)}, \quad n \in \mathbb{N}_0. \tag{3.4}$$

According to the Characterization Theorem given in the preceding section, in order to test whether the sample is from a Waring population, it is enough to check if a straight line can be fitted to the set of points $\{(n, x(n))\}$. It should be taken into account that the variance of $x(n)$ is increasing with n , namely,

$$D^2(x(n)) = \frac{\alpha}{\alpha - 2} \frac{e(n)(e(n) + 1)}{NF(n)}, \quad n \in \mathbb{N}_0, \alpha > 2 \tag{3.5}$$

(cf. equation (2.6)). Thus, it appears that

$$w(n) = \left(\frac{NF(n)}{x(n)(x(n) + 1)} \right)^{1/2} \quad n \in \mathbb{N}_0, \tag{3.6}$$

are proper weights in applying a weighted least squares fitting procedure. These weights can be calculated directly from the sample; in case of $\alpha > 2$, they are inversely proportional to the sample estimator of the standard error of the respective truncated mean (the unknown common factor $\alpha/(\alpha - 2)$ can be cancelled), and even in the non-Gaussian case of $\alpha \in (1, 2]$ they properly assess the relative weights of the data points.

The correlation coefficient provides information about the goodness of fit. Denoting the slope and the intercept of the fitted straight line by a and b , respectively, the expressions

$$\hat{\alpha} = \frac{a}{a-1} \quad \text{and} \quad \hat{N} = \frac{b}{a-1} \quad (3.7)$$

are unbiased estimators of the parameters, but no error estimation can be offered.

3.2. Multi-sample test

Consider a set of samples which are assumed to have a similar nature, namely their distribution is of the same type possibly with different parameters. It is always useful to have some information about the validity of this assumption without testing the samples one-by-one. If the assumed distribution is Waring, the following test is suggested by the Characterization Theorem.

Let J be the number of samples and denote the characteristics of the single samples by indices $j \in [1, J]$; $x_j(k)$ is thus the k th truncated sample mean of the j th sample. We shall also use the notation

$$X(k) = \sum_{j=1}^J x_j(k), \quad k \in \mathbb{N}_0.$$

If the conditions of the Characterization Theorem hold for all $j \in [1, J]$, the set of points $\{k, X(k)\}$ should lie on a straight line if all samples are from Waring distribution. Of course, the 'only if' part of the theorem cannot be saved. One can only say that if the points do not fit to a straight line, then the assumption that all samples are from Waring distributions does not hold.

If the $(2 + \varepsilon)$ th moments exist (i.e. if $\alpha_j > 2 + \varepsilon$) for all $j \in [1, J]$ and a fixed $\varepsilon > 0$, then the limiting distribution of the $X(k)$'s is normal for each k and Gaussian statistics can be applied for hypothesis testing.

3.3. Empirical results

Our results will be illustrated with the example of word frequency distributions. The samples tested are well known from the literature: the frequency distribution of nouns in Macaulay's essay on Bacon (four disjoint subsamples a , b , c and d) was originally reported by Yule (1944) and was analyzed among others by Herdan (1960); the word frequency statistics of Pushkin's story *The Captain's Daughter* was compiled by Epstein and Josselson (1953) and was referred to in Herdan (1974).

Herdan (1960, p. 53) plotted the Macaulay data on a lognormal grid and got what he has called 'a sensibly straight line'. Passing his findings unchallenged, we only assert that the multi-sample test described in Section 3.2 results in a line (Fig. 1) likewise straight; thus, the single samples seem worth analyzing separately. Using the method of Section 3.1, the parameters of the Waring distribution were estimated for the four subsamples. The results are presented in Table 1 (parameters a and b are the slope and intercept of the fitted straight line; \hat{a} and \hat{N} were calculated according to equation (3.7)). As was expected, the parameters of the four subsamples are rather similar.

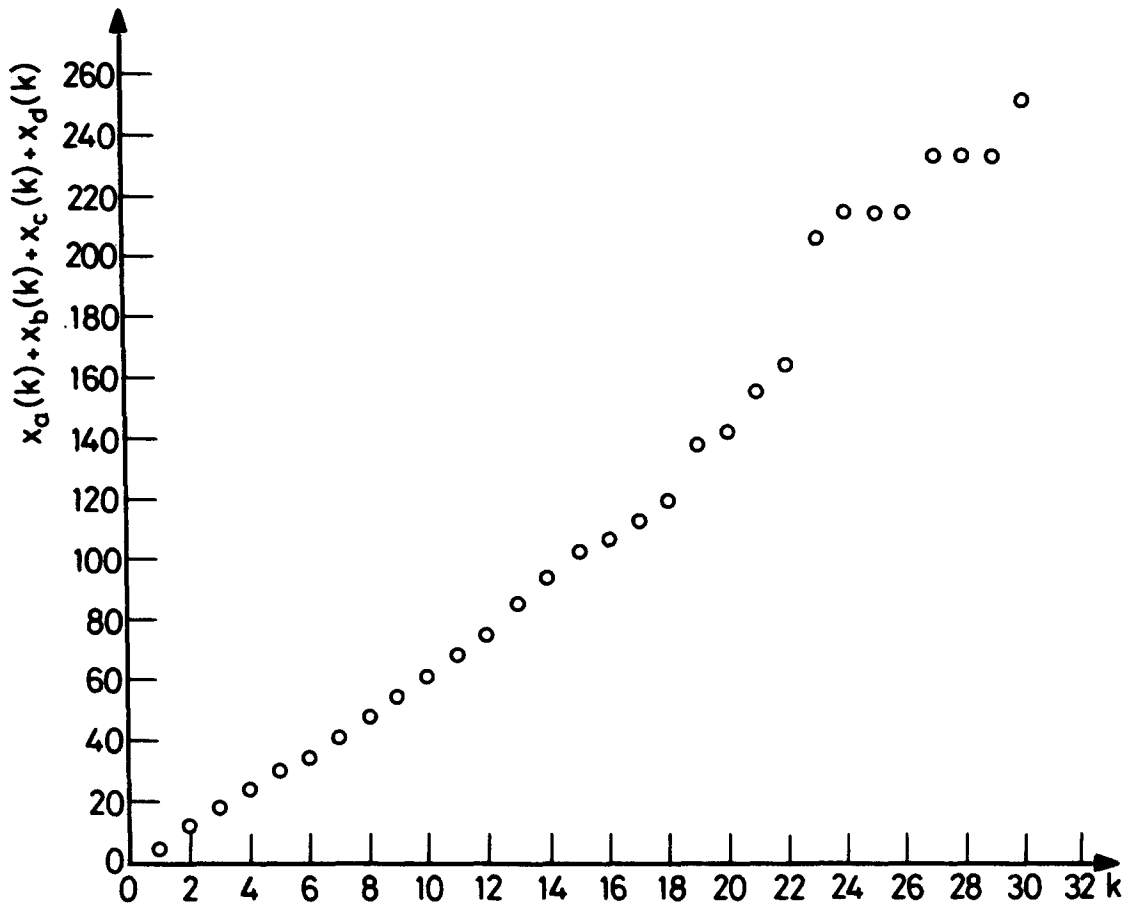


Fig. 1. Multi-sample test for the Macaulay data.

In order to check the fit, the residuals,

$$\text{res}(k) = x(k) - ak - b,$$

were plotted against k (Fig. 2). In case of proper fit, the residuals have to fluctuate randomly around zero, with an amplitude of about the standard error of the truncated mean, $x(k)$. The standard error 'cornet' of Fig. 2 was calculated using the average parameters $\bar{a} = 1.606$, $\bar{b} = 0.582$ ($\bar{\alpha} = 2.650$, $\bar{N} = 0.960$). The residuals closely follow the required pattern.

The Pushkin sample is just the one that Herdan (1964) modelled by the Waring

Table 1
Waring parameters for subsamples *a*, *b*, *c* and *d* of Macaulay's essay on Bacon

Parameter	<i>a</i>	<i>b</i>	α	<i>N</i>
Sample				
<i>a</i>	1.586	0.663	2.706	1.167
<i>b</i>	1.660	0.508	2.515	0.770
<i>c</i>	1.575	0.624	2.739	1.085
<i>d</i>	1.603	0.531	2.658	0.881

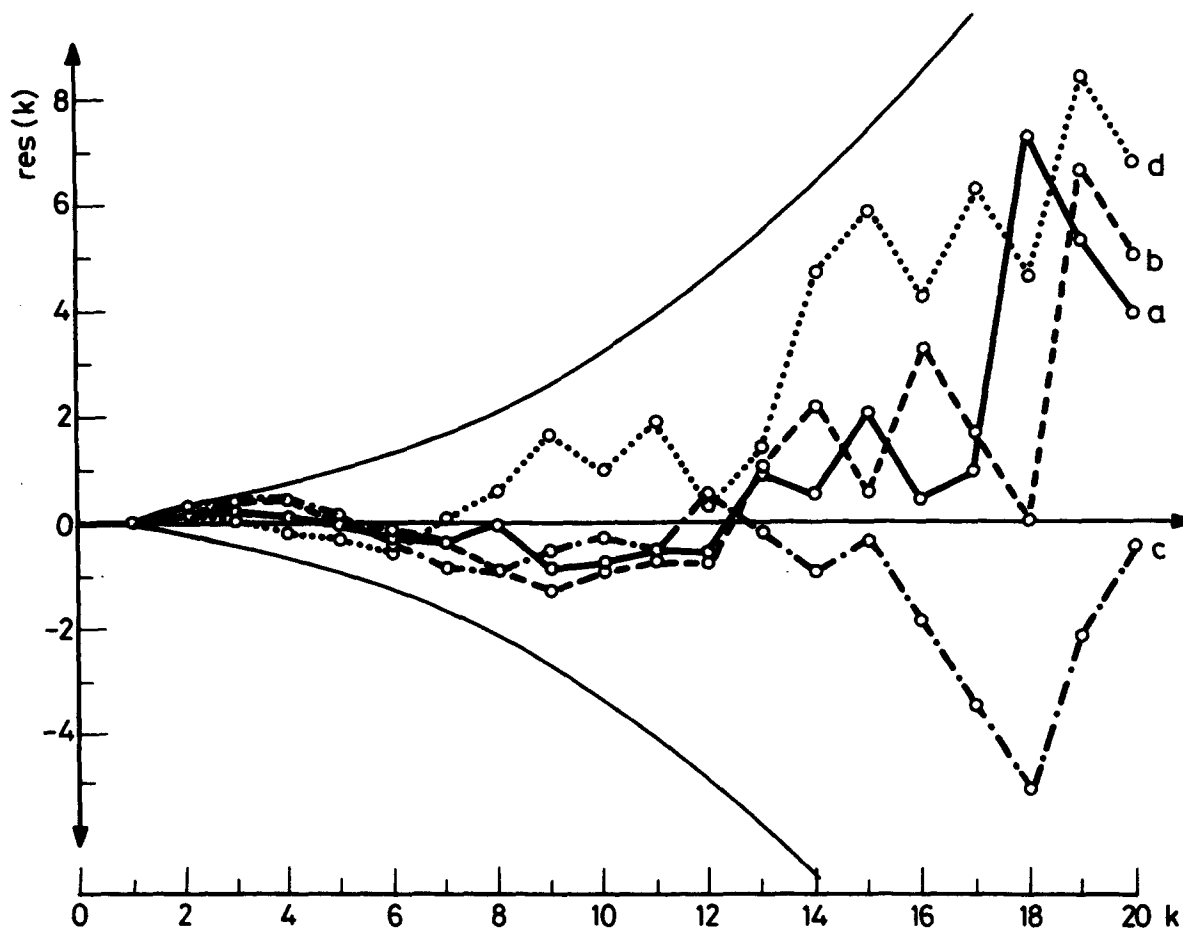


Fig. 2. Plot of residuals vs. frequency values for the Macaulay samples.

distribution (though even he had some reservations about the goodness of fit at higher frequencies ($k > 30$)). The parameters estimated by the weighted least squares method are $a=4.064$, $b=2.211$ ($\hat{a}=1.326$, $\hat{N}=0.722$; Herdan's estimates were $\hat{a}=1.213$, $\hat{N}=1.211$). The $\text{res}(k)$ vs. k plot in Fig. 3 clearly demonstrates that there is a tendentious deviation from the straight line in this case; the residuals do not fluctuate randomly, but follow a concave pattern even for frequencies smaller than 20. As a result, the postulate of a Waring distribution has to be refuted for this sample.

It should be noted again that the Macaulay sample is based on the *nouns* only, while the Pushkin data refers to the complete text; this might be one source of the difference in the nature of the two samples.

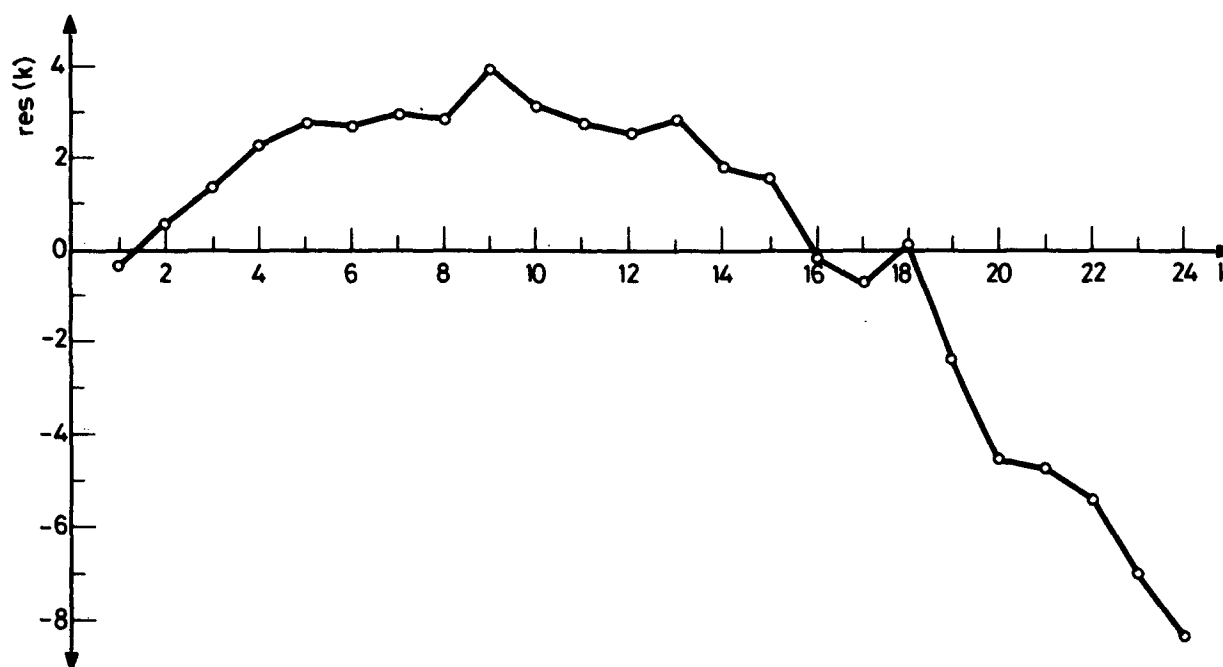


Fig. 3. Plot of residuals vs. frequency values for the Pushkin data.

3.4. Efficiency of parameter estimation

Maximum-likelihood estimation of the Waring parameters is well known from the literature (Irwin, 1975). There is no exact formula, however, for the error of the estimated parameters, therefore efficiency of parameter estimation based on weighted least squares cannot be determined.

We assessed the efficiency of estimation of parameters for the Macaulay samples with the aid of a simulation experiment. One hundred Waring samples, each containing 1000 elements, were generated via a simulated urn model (Xekalaki, 1981) (the average parameters $\bar{a}=1.606$, $\bar{b}=0.582$ were used), and parameters were estimated both by maximum-likelihood and weighted least squares methods. Empirical mean and standard error of the estimators over the 100 samples were then calculated and efficiencies of weighted least squares estimators relative to maximum-likelihood estimators were determined. The results are presented in Table 2.

Table 2
Comparison of weighted least squares and maximum-likelihood parameter estimation

Method	Parameter	a		b	
		Mean	Std.err.	Mean	Std.err.
Maximum-likelihood		1.604	0.0162	0.578	0.0109
Weighted least squares		1.582	0.0400	0.605	0.0286
Relative efficiency			0.636		0.618

Note added in proof

A result similar to our Characterization Theorem was published recently by Xekalaki (1983) in terms of the hazard function.

References

- B. Epstein and H.H. Josselson, *The Russian World Count, V* (Wayne University Press, Detroit, 1973).
- W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1957).
- S.D. Haitun, Stationary scientometric distributions. Part I. Different approximations, *Scientometrics* 4 (1) (1982a) 5–25; Part II. Non Gaussian nature of scientific activities, *Scientometrics* 4 (2) (1982b) 89–104; Part III. The role of the Zipf distribution, *Scientometrics* 4 (3) (1982c) 181–194.
- G. Herdan, *Type-Token Mathematics* (Mouton, The Hague, 1960).
- G. Herdan, *Quantitative Linguistics* (Butterworths, London, 1964).
- J.O. Irwin, The place of mathematics in medical and biological statistics, *J. Roy. Stat. Soc. A* 126 (1963) 1–45.
- J.O. Irwin, The generalized Waring distribution applied to accident theory, *J. Roy. Stat. Soc. A* 131 (1968) 205–225.
- J.O. Irwin, The generalized Waring distribution, *J. Roy. Stat. Soc. A* 138 Part 1, 18–31; Part 2, 204–227; Part 3, 374–384 (1975).
- N.L. Johnson and S. Kotz, *Distributions in Statistics, Continuous Univariate Distributions, Vol. 2* (Houghton Mifflin Co., Boston, 1970).
- S. Kotz and D.N. Shanbag, Some new approaches to probability distributions, *Adv. Appl. Prob.* 12 (1980) 903–921.
- D.G. Morrison, On linearly increasing mean residual lifetimes, *J. Appl. Prob.* 15 (1978) 617–620.
- J. Tague, The success-breeds-success phenomenon and bibliometric processes, *J. Am. Soc. Inf. Sci.* 32 (4) (1981) 280–286.
- G.U. Yule, *The Statistical Study of Literary Vocabulary* (Cambridge University Press, 1944).
- E. Xekalaki, Chance mechanism for the univariate generalized Waring distribution and related characterizations, in: C. Taillie and G.P. Patil, eds., *Statistical Distributions in Scientific Work. Vol. 4. Models, Structures, and Characterizations* (D. Reidel Publ. Co., Dordrecht, 1981) pp. 157–171.
- E. Xekalaki, Hazard functions and life distributions in discrete time, *Commun. Statist. – Theor. Meth.* 12 (21) (1983) 2503–2509.