ELSEVIER

Insurance: Mathematics and Economics 16 (1995) 255-260

INSURANCE
MATHEMATICS & ECONOMICS

Quadratic distance estimators for the zeta family ☆

Louis G. Doray [a], Andrew Luong [b]

[a] *Department of Mathematics and Statistics, University of Montreal, P.O. Box 6128, Succursale Centre-ville, Montreal, Canada, H3C 3J7*

[b] *Ecole d'actuariat, Université Laval, Cité Universitaire, Quebec, Que. G1K 7P4, Canada*

Received July 1994; revised January 1995

Abstract

The zeta distribution is a discrete distribution which has been relatively little used in actuarial science and statistics, a reason being that most estimators proposed in the literature for the parameter of this distribution require iterative methods or the extensive use of tables for its calculation, due to the complicated form of its probability mass function. In this paper, we propose a new estimator, based on quadratic distance, asymptotically fully efficient for parameter values greater than 2 and highly efficient for smaller values, but computationally more appealing than the maximum likelihood estimator; we also compare its asymptotic variance with that of the moment estimator and the estimator based on the ratio of the observed frequencies of the first two classes.

Keywords: Quadratic distance estimator; Weighted least-squares estimator; Asymptotic efficiency; Discrete Pareto; Zipf

1. Introduction

The zeta distribution is a one-parameter discrete distribution with probability mass function

$$P[N=i]=p_i=\frac{i^{-(\rho+1)}}{\zeta(\rho+1)},\quad \rho>0,\ i=1,2,\ldots,$$

where $\zeta(\rho+1)$ is the Riemann zeta function

$$\zeta(\rho+1)=\sum_{i=1}^{\infty}\frac{1}{i^{\rho+1}},$$

which converges for all $\rho>0$. This function has been tabulated for integral values of ρ (see Abramowitz and Stegun (1972), p. 811). The probability function is a strictly decreasing function of i, for all $\rho>0$, so that the mode is always 1. The values of $\rho>1$ are useful for most applications. The zeta distribution belongs to the exponential family.

Properties of the zeta distribution are discussed in Johnson et al. (1992), pp. 465-471. We should note that it is a long-tail distribution with rth moment about 0,

$$E(N^r)=\frac{\zeta(\rho-r+1)}{\zeta(\rho+1)},\quad r<\rho.$$

If $r\geq\rho$, this expectation does not exist.

☆ The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada and of the Insurance Chair of Laval University.

SSDI 0167-6687(95)00008-9

The zeta distribution has found applications in insurance. Seal (1947) has modeled the number of policies per person insured in an insurance portfolio with this distribution. Recalling that the Pareto law has been used to model the distribution of individual incomes in a population, he assumed that the number of insurance contracts a policyholder had was proportional to that individual's income, to derive the zeta distribution, which he called the discrete Pareto distribution. The zeta distribution is also known as the Zipf distribution, although some authors reserve this name for the case $\rho = 1$. It models well phenomena ranked by size or order which follow 'the higher-the fewer' rule (see Olkin et al., 1980; Simon and Bonini, 1958). For example, it was used to model occurrences of words found in texts (Parunak, 1979), the distribution of surname frequencies in a sample of people married in England (Fox and Lasker, 1983), or the number of publications per author in scientific journals.

In Section 2, we review the estimators which have appeared in the literature to estimate the parameter ρ, while in Section 3, we propose quadratic distance estimators, easy to compute and asymptotically very efficient.

2. Traditional estimators of ρ

Various estimators for the parameter ρ have been proposed. The maximum likelihood estimator $\hat{\rho}$ is the value of ρ solving the equation

$$h(\rho) = -\frac{\zeta'(\rho+1)}{\zeta(\rho+1)} = \frac{\Sigma_{i=1}^{n} \ln N_i}{n},$$

where $N_1, N_2, \ldots, N_n$ is the sample; its asymptotic variance is

$$\operatorname{Var}(\hat{\rho}) = -\frac{1}{nh'(\rho)} = \frac{[\zeta(\rho+1)]^2}{n\{\zeta(\rho+1)\zeta''(\rho+1) - [\zeta'(\rho+1)]^2\}}.$$

Both the functions $h(\rho)$ and $-h'(\rho)$ have been tabulated, for some values of ρ (see Johnson et al., 1992, pp. 468–469). However, Moore (1956) warns the reader that the function $-h'(\rho)$ (of which Johnson's table is an adaptation) was evaluated by numerical differentiation and 'it is thought that the table is accurate to the number of significant figures that are given.'

Table 1
Asymptotic variance of $\hat{\rho}$

ρ	$n\operatorname{Var}(\hat{\rho})$
0.5	0.259406
1.0	1.13061
1.5	2.85345
2.0	5.80448
2.5	10.5352
3.0	17.8393
3.5	28.8525
4.0	45.1939
4.5	69.1681
5.0	104.051

With the power of symbolic programming languages, we might expect some computational difficulties for obtaining the MLE might be alleviated. Using MATHEMATICA, we recalculated the values of $-h'(\rho)$ for $\rho = 0.5, 1.0, \ldots, 5.0$. Table 1 contains the exact values of $n\operatorname{Var}(\hat{\rho})$ to 6 significant figures. Moore's table was not as accurate as thought.

Since $\rho = (\ln(p_1/p_2)/\ln 2) - 1$ for the zeta distribution, the ratio estimator of ρ, suggested by Seal (1947), is obtained directly by using the maximum likelihood estimator (MLE) f_i/n for p_i, where f_i is the observed frequency of i's in the sample, giving

$$\rho^* = \frac{\ln(f_1/f_2)}{\ln 2} - 1.$$

Proposition 1. *The asymptotic variance of ρ^* is:*

$$\operatorname{Var}(\rho^*) = \frac{\zeta(\rho+1)(1+2^{\rho+1})}{n(\ln 2)^2}. \tag{2.1}$$

Table 2
Optimal choice of (i, j)

ρ	(i, j)	$\operatorname{Var}(\rho^{**})/\operatorname{Var}(\rho^*)$
0.5	(1, 4)	0.588
1.0	(1, 3)	0.796
1.5	(1, 3)	0.992
≥ 2.0	(1, 2)	1.000

Proof. Since $f_i \sim \text{Bin}(n, p_i)$ and $(f_i, f_j) \sim \text{Trinomial}(n, p_i, p_j)$, $i \neq j$, it follows that the approximate variance

$$\text{Var}(\ln f_i) = \left[\frac{\mathrm{d} \ln x}{\mathrm{d} x}\Big|_{x=\mathrm{E}(f_i)}\right]^2 \times \text{Var}(f_i)$$

$$= \frac{np_i(1-p_i)}{(np_i)^2} = \frac{1-p_i}{np_i},$$

and the approximate covariance

$$\text{Cov}(\ln f_i, \ln f_j)$$

$$= \left[\frac{\mathrm{d} \ln x}{\mathrm{d} x}\Big|_{x=\mathrm{E}(f_i)}\right] \times \left[\frac{\mathrm{d} \ln y}{\mathrm{d} y}\Big|_{y=\mathrm{E}(f_j)}\right]$$

$$\times \text{Cov}(f_i, f_j) = \frac{1}{np_i}\frac{1}{np_j}(-np_i p_j) = -1/n.$$

Then, with $i = 1$, $j = 2$, we obtain

$$\text{Var}(\rho^*) = \frac{1}{(\ln 2)^2}\left[\frac{1-p_1}{np_1} + \frac{1-p_2}{np_2} + \frac{2}{n}\right],$$

giving (2.1) upon simplification.

Table 3 in the next section contains the reciprocals of the asymptotic efficiencies of ρ^* and of other estimators presented in Sections 2 and 3 for various values of the parameter ρ. Note that Johnson et al. (1992) contain an error on p. 468 in their expression for $\text{Var}(\rho^*)$, explaining the difference between the values in the first column of Table 3 and their corresponding values. The estimator ρ^* has very low efficiency, especially for small values of ρ.

The moment estimator $\bar{\rho}$ (ref. Seal, 1952) satisfies the equation

$$\bar{N} = \frac{\zeta(\bar{\rho})}{\zeta(\bar{\rho}+1)},$$

where $\bar{N}$ is the sample average. Moore (1956) contains a table of $\bar{\rho}$ corresponding to various values of $\bar{N}$.

Proposition 2. *The asymptotic variance of* $\bar{\rho}$ *is*:

$$\text{Var}(\bar{\rho})$$

$$= \frac{\left[\zeta(\rho+1)\zeta(\rho-1) - \zeta(\rho)^2\right]\left[\zeta(\rho+1)\right]^2}{n\left[\zeta'(\rho)\zeta(\rho+1) - \zeta'(\rho+1)\zeta(\rho)\right]^2}$$

Proof. Let $f(\rho) = \zeta(\rho)/(\zeta(\rho+1))$.

$$\text{Var}(f(\rho))\,|_{\rho=\bar{\rho}}$$

$$= \text{Var}(\bar{\rho})\left[\frac{\mathrm{d}f(\rho)}{\mathrm{d}\rho}\right]^2 = \frac{1}{n}\text{Var}(N),$$

leading to the above expression for $\text{Var}(\bar{\rho})$.

The moment estimator has a smaller asymptotic variance than ρ^*, but requires an iterative procedure for its calculation. Note also that for $\rho \leq 1$, $\mathrm{E}(N)$ does not exist, and for $\rho \leq 2$, the moment estimator does not have a finite asymptotic variance, suggesting it might not be stable.

The relation $\ln(p_j/p_i) = (\rho+1)\ln(i/j)$ for the zeta distribution also suggests the estimator ρ^{**} for ρ,

$$\rho^{**} = \frac{\ln(f_j/f_i)}{\ln(i/j)} - 1.$$

It is easily shown that, as $n \to \infty$, ρ^{**} is unbiased and

$$\text{Var}(\rho^{**}) = \frac{\zeta(\rho+1)(i^{\rho+1} + j^{\rho+1})}{n[\ln(i/j)]^2}.$$

The (i, j) that will minimize $\text{Var}(\rho^{**})$ depends on the value of the parameter ρ. Table 2 contains the optimal choice of (i, j) for various values of ρ, along with the ratio $\text{Var}(\rho^{**})/\text{Var}(\rho^*)$.

The true value of the parameter ρ is unknown, but a quick estimate can be obtained by solving for ρ, $\hat{p}_1 = f_1/n = 1/\zeta(\rho+1)$. Johnson et al. (1992) comment on p. 466, that the value of ρ found in many applications is slightly in excess of 1. The estimator ρ^* using $(i, j) = (1, 2)$ is recommended over the use of ρ^{**}, with $(i, j) = (1, 3)$, because the gain in relative efficiency is small compared to the danger of very poor efficiency if ρ was in fact larger than 2. For example, if $\rho = 1$, the relative efficiency of the estimator ρ^{**} using $(i, j) = (1, 3)$ compared to ρ^*, is 0.80; but for $\rho = 2$, this ratio is 1.24; for $\rho = 3$, it is 1.92 and it increases to 4.47 for $\rho = 5$.

In the next section, we consider quadratic distance estimators which make use of all the sample, not only the first two observed frequencies, as does ρ^*, which can be viewed as a special case of this class of estimators.

3. Quadratic distance estimators of ρ

Fox and Lasker (1983) observed that, if data followed the discrete Pareto distribution, a plot of the log of the number of events occurring i times against log i should give a straight line of slope $-(\rho+1)$. The following relation also holds,

$$\ln\frac{p_{i+1}}{p_i} = (\rho+1)\ln\frac{i}{i+1}, \quad i=1,2,\dots.$$

Using the estimator f_i/n for p_i, we can consider the analogous of a linear regression model

$$\ln\frac{f_{i+1}}{f_i} = (\rho+1)\ln\frac{i}{i+1} + \epsilon_i, \quad i=1,2,\dots,$$

where ϵ_i is a random error. This representation is similar to the one given by Luong and Garrido (1993). It can easily be seen that asymptotically, ϵ_i has a mean of 0.

Let Y be the vector

$$Y = \left[\ln\frac{f_2}{f_1}, \ln\frac{f_3}{f_2}, \dots, \ln\frac{f_{k+1}}{f_k}\right]',$$

X the vector

$$X = \left[\ln\frac{1}{2}, \ln\frac{2}{3}, \dots, \ln\frac{k}{k+1}\right]',$$

and ϵ the error vector $\epsilon = [\epsilon_1, \dots, \epsilon_k]'$, where it is assumed that $f_1, \dots, f_{k+1}$ are non-zero. Then the model can be rewritten as

$$Y = (\rho+1)X + \epsilon.$$

Proposition 3. The asymptotic variance-covariance matrix of vector ϵ, Σ, is equal to

$$\Sigma = \frac{1}{n}\begin{pmatrix} \frac{p_1+p_2}{p_1p_2} & -\frac{1}{p_2} & 0 & 0 & \cdots & 0 \\ -\frac{1}{p_2} & \frac{p_2+p_3}{p_2p_3} & -\frac{1}{p_3} & 0 & \cdots & 0 \\ 0 & -\frac{1}{p_3} & \frac{p_3+p_4}{p_3p_4} & -\frac{1}{p_4} & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & -\frac{1}{p_{k-1}} & \frac{p_{k-1}+p_k}{p_{k-1}p_k} & -\frac{1}{p_k} \\ 0 & 0 & \cdots & 0 & -\frac{1}{p_k} & \frac{p_k+p_{k+1}}{p_kp_{k+1}} \end{pmatrix}$$

Proof. From the proof of Proposition 1, it follows that

$$\begin{aligned} \operatorname{Var}(\epsilon_i) &= \operatorname{Var}(\ln f_{i+1}) + \operatorname{Var}(\ln f_i) \\ &\quad - 2\operatorname{Cov}(\ln f_{i+1}, \ln f_i) \\ &= \frac{1-p_{i+1}}{np_{i+1}} + \frac{1-p_i}{np_i} + \frac{2}{n} = \frac{p_i+p_{i+1}}{np_ip_{i+1}}, \end{aligned}$$

$$\begin{aligned} \operatorname{Cov}(\epsilon_i, \epsilon_{i+1}) &= \operatorname{Cov}(\ln f_{i+1}, \ln f_{i+2}) - \operatorname{Var}(\ln f_{i+1}) \\ &\quad - \operatorname{Cov}(\ln f_i, \ln f_{i+2}) + \operatorname{Cov}(\ln f_i, \ln f_{i+1}) \\ &= -\frac{1}{n} - \frac{1-p_{i+1}}{np_{i+1}} + \frac{1}{n} - \frac{1}{n} = -\frac{1}{np_{i+1}}. \end{aligned}$$

If $|i-j|>1$, $\operatorname{Cov}(\epsilon_i, \epsilon_j) = 0$, so that $\Sigma = \bar{E}(\epsilon\epsilon')$ is the above tridiagonal matrix.

The most efficient quadratic distance estimator $\tilde{\rho}$, which can also be viewed as a form of weighted least-squares estimator of ρ, minimizes

$$[Y-(\rho+1)X]'\Sigma^{-1}[Y-(\rho+1)X].$$

Explicitly,

$$\tilde{\rho} = -1 + (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y;$$

it is consistent and asymptotically normal with asymptotic variance

$$\operatorname{Var}(\tilde{\rho}) = (X'\Sigma^{-1}X)^{-1}.$$

An asymptotically equivalent estimator can be obtained by choosing an initial consistent estimator such as

$$\tilde{\rho}_i = -1 + (X'X)^{-1}X'Y,$$

and using $\tilde{\rho}_i$ to obtain an estimate for Σ^{-1}, $\tilde{\Sigma}^{-1}$. The asymptotically equivalent quadratic distance estimator minimizes $[Y-(\rho+1)X]'\tilde{\Sigma}^{-1}[Y-(\rho+1)X]$. The procedure can be repeated until convergence. It often takes just a few iterations for convergence to be reached.

Inverting matrices of a large order might present some numerical difficulties with some software; with MATHEMATICA, we were able to use $k=270$ to evaluate Σ^{-1} in the expression $(X'\Sigma^{-1}X)^{-1}$ and form Table 3, which contains the ratios $\mathrm{Var}(\tilde{\rho})/\mathrm{Var}(\hat{\rho})$, for $\rho = 1.0, 1.5, \ldots, 5.0$. We see that $\tilde{\rho}$ is asymptotically efficient for $\rho \geq 2$ and highly efficient for $1 \leq \rho < 2$. It is also the best estimator proposed so far, which does not involve the use of tables. As k is increased, the variance of $\tilde{\rho}$ keeps decreasing.

We can also consider the estimator ρ^{Δ} obtained when Σ is replaced by the diagonal matrix with diagonal element $\sigma_{ii} = (p_i + p_{i+1})/np_i p_{i+1}$ and off diagonal elements equal to 0. The estimator ρ^{Δ} equals

$$\rho^{\Delta} = -1 + \frac{\sum_{i=1}^{k} \frac{p_i p_{i+1}}{p_i + p_{i+1}} \left[\ln\left(\frac{i}{i+1}\right)\right]\left[\ln\left(\frac{f_{i+1}}{f_i}\right)\right]}{\sum_{i=1}^{k} \frac{p_i p_{i+1}}{p_i + p_{i+1}} \left[\ln\left(\frac{i}{i+1}\right)\right]^2}.$$

It is asymptotically unbiased, with variance

$$\mathrm{Var}(\rho^{\Delta}) = \frac{1}{n\sum_{i=1}^{k} \frac{p_i p_{i+1}}{p_i + p_{i+1}} \left[\ln\left(\frac{i}{i+1}\right)\right]^2}.$$

The estimator ρ^{Δ} performs almost as poorly as Seal's estimator (ρ^* is a special case of ρ^{Δ}, with $k = 1$). We calculated $\mathrm{Var}(\rho^{\Delta})$ with $k \to \infty$ in Table 3.

We considered two other quadratic distance estimators of ρ. The first one is derived from the model

$$Y = (\rho + 1)X + \epsilon, \quad \text{with}$$

$$Y = \left[\ln\frac{f_2}{f_1}, \ln\frac{f_4}{f_3}, \ldots, \ln\frac{f_{k+1}}{f_k}\right]',$$

$$X = \left[\ln\frac{1}{2}, \ln\frac{3}{4}, \ldots, \ln\frac{k}{k+1}\right]'.$$

The variance-covariance matrix Σ of ϵ is a diagonal matrix, which facilitates the calculation of the least-squares estimator and its variance.

The second estimator uses

$$Y = \left[\ln\frac{f_2}{f_1}, \ln\frac{f_3}{f_1}, \ldots, \ln\frac{f_{k+1}}{f_1}\right]',$$

$$X = \left[\ln\frac{1}{2}, \ln\frac{1}{3}, \ldots, \ln\frac{1}{k+1}\right]'.$$

Table 3
Ratios of asymptotic variances

ρ	$\frac{\mathrm{Var}(\rho^*)}{\mathrm{Var}(\hat{\rho})}$	$\frac{\mathrm{Var}(\bar{\rho})}{\mathrm{Var}(\hat{\rho})}$	$\frac{\mathrm{Var}(\tilde{\rho})}{\mathrm{Var}(\hat{\rho})}$	$\frac{Var(\rho^{\Delta})}{Var(\hat{\rho})}$
1.0	15.14	–	1.105	12.70
1.5	6.514	–	1.012	5.684
2.0	3.879	–	1.001	3.484
2.5	2.741	1.701	1.000	2.515
3.0	2.147	1.272	1.000	2.003
3.5	1.798	1.143	1.000	1.700
4.0	1.576	1.087	1.000	1.506
4.5	1.427	1.056	1.000	1.375
5.0	1.323	1.039	1.000	1.283

The asymptotic variance of the least-squares estimator for these two models was high, between that of ρ^{Δ} and ρ^*.

The expression $[Y - (\rho + 1)X]'\Sigma^{-1}[Y - (\rho + 1)X]$ defines naturally a measure of distance between the empirical cdf F_n and the parametric family F_ρ. We might expect to make use of such distances to construct test statistics for goodness-of-fit for the zeta parametric family. This question will be addressed in a subsequent paper which emphasizes the testing aspects. We will also consider the problem of modelling with the zeta distribution when covariates are present. Finally, as pointed out by a referee, a Bayesian approach for estimation might lead to consider $P[n = i \mid \rho] = i^{-\rho+1}/\zeta(\rho + 1)$ and introduce a suitable prior density for ρ.

References

Abramowitz, M. and I.A. Stegun (1972). *Handbook of Mathematical Functions.* Dover Publications, Inc., New York.

Fox, W.R. and G.W. Lasker (1983). The Distribution of Surname Frequencies. *International Statistical Review* 51, 81–87.

Johnson, N.L., S. Kotz and A.W. Kemp (1992). *Univariate Discrete Distributions.* John Wiley and Sons, Inc., New York.

Luong, A. and J. Garrido (1993). Minimum quadratic distance estimation for a parametric family of discrete distributions defined recursively. *Australian Journal of Statistics* 35, 59–67.

Moore, P.G. (1956). The geometric, logarithmic and discrete Pareto forms of series. *Journal of the Institute of Actuaries* 82, 130–136.

Olkin, I., L.J. Gleser and C. Derman (1980). *Probability Models and Applications.* MacMillan, New York.

Parunak, A. (1979). Graphical analysis of ranked counts (of words). *Journal of the American Statistical Association* 74, 25-30.

Seal, H.L. (1952). The maximum likelihood fitting of the discrete Pareto law. *Journal of the Institute of Actuaries* 78, 115-121.

Seal, H.L. (1947). A probability distribution of deaths at age x when policies are counted instead of lives. *Skandinavisk Aktuarietidskrift* 30, 18-43.

Simon, H.A. and C.P. Bonini (1958). The size distribution of business firms. Reprint 20 of the Graduate School of Business Administration, Carnegie Institute of Technology, Pittsburgh, PA.