Estimators of the regression parameters of the zeta distribution

Louis G. Doray*, Michel Arsenault

*Département de Mathématiques et de Statistique, Université de Montréal, C.P. 6128,
Succursale Centre-ville, Montreal, Que., Canada H3C 3J7*

Abstract

The zeta distribution with regression parameters has been rarely used in statistics because of the difficulty of estimating the parameters by traditional maximum likelihood. We propose an alternative method for estimating the parameters based on an iteratively reweighted least-squares algorithm. The quadratic distance estimator (QDE) obtained is consistent, asymptotically unbiased and normally distributed; the estimate can also serve as the initial value required by an algorithm to maximize the likelihood function. We illustrate the method with a numerical example from the insurance literature; we compare the values of the estimates obtained by the quadratic distance and maximum likelihood methods and their approximate variance–covariance matrix. Finally, we calculate the bias, variance and the asymptotic efficiency of the QDE compared to the maximum likelihood estimator (MLE) for some values of the parameters.
© 2002 Elsevier Science B.V. All rights reserved.

1. Introduction

The zeta distribution has been used by many authors. While it has found applications in various fields (see Doray and Loung (1995) for references to specific examples in insurance, scientific production, bibliometry and a patronymic study), we could find only one example of a situation where explanatory variables were involved. Seal (1947) used the age of an insured as a covariate to model the distribution of the number of insurance policies per person in the portfolio of an insurance company. However, his method for estimating the parameters was ad hoc, using a graphical procedure, with no indication of the variance of the estimates. Formal estimation of the regression parameters and the properties of the estimators have not been considered. The reason for the rare use is that, even with today's powerful computers, maximum likelihood estimation of the regression parameters takes a lot of time.

In this paper, we propose a new estimator for the regression parameters of the zeta distribution, based on the quadratic distance estimation (QDE) method used by Luong and Garrido (1993) for the (a, b) family, Doray and Luong (1995) for the zeta distribution without covariates and Doray and Luong (1997) for the Good distribution; these estimators can be obtained much more rapidly than the maximum likelihood estimators (MLEs) and possess desirable asymptotic properties.

---

* Corresponding author.
E-mail address: doray@dms.umontreal.ca (L.G. Doray).

The paper is organized as follows. In Section 2, we define the zeta regression model; we derive the equations to solve to calculate the MLE of the regression parameters and we obtain the expression for the observed information matrix (Section 3). In Section 4, we propose the QDE and derive its asymptotic properties. We show that it is asymptotically unbiased and obtain its variance–covariance matrix. We then rework the numerical example considered by Seal with both estimation methods and compare the results obtained (Section 5); we test the fit of the model (Section 6). Finally, we calculate the asymptotic efficiency of the QDE compared to the MLE for the zeta distribution with one covariate, as well as the bias and variance of the QDE with a finite sample.

2. The zeta model with covariates

The probability mass function (pmf) of the one-parameter zeta distribution is equal to

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

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

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

This series converges for all positive values of the parameter ρ. For more on the properties of this distribution, see Johnson et al. (1992) or Doray and Luong (1995). To reflect the heterogeneity which could be present in certain populations, we will modify the basic model where all observations are identically distributed, by introducing a covariate. We also assume as Seal (1947) did, that a linear relation exists between the parameter ρ and this explanatory variable,

$$\rho + 1 = ax + b,$$

where a and b are the parameters of the model to be estimated and x is a known covariate value. The pmf of a random variable N following a zeta$(a, b; x)$ distribution equals

$$P[N = i; x] = \frac{i^{-(ax+b)}}{\zeta(ax+b)}, \quad ax + b > 1, \; i = 1, 2, \dots.$$

Let the sample be $(N_1, x_1), (N_2, x_2), \dots, (N_n, x_n)$, where $N_i \sim$ zeta$(a, b; x_i)$ and x_i is the explanatory variable for subject i. If the explanatory variable is a continuous variate, we will form a finite number of intervals. The n observations of the sample can then be grouped according to the value of x_i into m intervals, or classes, which will be denoted by $I_1, I_2, \dots, I_m$. Let w_j be the number of observations in class I_j, where $\sum_{j=1}^{m} w_j = n$. In what follows, we assume that x is a continuous variate and we denote its central value for the jth interval by x_j^c. Let N be the random variable with explanatory variable x belonging to the jth interval ($N \sim$ zeta$(a, b; x_j^c)$); we denote its pmf by $p_j(i)$,

$$P[N = i; x_j^c] = p_j(i) = \frac{i^{-(ax_j^c+b)}}{\zeta(ax_j^c+b)}, \quad ax_j^c + b > 1, \; i = 1, 2, \dots.$$

For a sample of size n, we thus have w_j identically distributed random variables zeta$(a, b; x_j^c)$ for $j = 1, 2, \dots, m$. In the following sections, we consider two methods to estimate the parameters a and b, maximum likelihood and quadratic distance.

3. Maximum likelihood estimation

The likelihood function for the sample is given by

$$L(a, b) = \prod_{j=1}^{m} \prod_{i=1}^{w_j} p_j(n_{i,j}; a, b),$$

where $n_{1,j}, n_{2,j}, \ldots, n_{w_j,j}$ are the observed values of the sample which are in class j; the log-likelihood $l(a, b) = \sum_{j=1}^{m} \sum_{i=1}^{w_j} \ln[p_j(n_{i,j}; a, b)]$ becomes equal to

$$l(a, b) = -\sum_{j=1}^{m} \sum_{i=1}^{w_j} (ax_j^c + b) \ln(n_{i,j}) - \sum_{j=1}^{m} w_j \ln[\zeta(ax_j^c + b)].$$

Taking the partial derivatives of $l(a, b)$ with respect to a and b, we obtain

$$\frac{\partial l(a, b)}{\partial a} = -\sum_{j=1}^{m} x_j^c \sum_{i=1}^{w_j} \ln(n_{i,j}) + \sum_{j=1}^{m} w_j \frac{x_j^c \sum_{i=1}^{\infty} i^{-(ax_j^c+b)} \ln(i)}{\sum_{i=1}^{\infty} i^{-(ax_j^c+b)}},$$

$$\frac{\partial l(a, b)}{\partial b} = -\sum_{j=1}^{m} \sum_{i=1}^{w_j} \ln(n_{i,j}) + \sum_{j=1}^{m} w_j \frac{\sum_{i=1}^{\infty} i^{-(ax_j^c+b)} \ln(i)}{\sum_{i=1}^{\infty} i^{-(ax_j^c+b)}}.$$

The MLEs $(\hat{a}, \hat{b})$ are the values (a, b) which satisfy the following system of equations:

$$\sum_{j=1}^{m} x_j^c \sum_{i=1}^{w_j} \ln(n_{i,j}) = \sum_{j=1}^{m} w_j \frac{x_j^c \sum_{i=1}^{\infty} i^{-(ax_j^c+b)} \ln(i)}{\sum_{i=1}^{\infty} i^{-(ax_j^c+b)}}, \quad \sum_{j=1}^{m} \sum_{i=1}^{w_j} \ln(n_{i,j}) = \sum_{j=1}^{m} w_j \frac{\sum_{i=1}^{\infty} i^{-(ax_j^c+b)} \ln(i)}{\sum_{i=1}^{\infty} i^{-(ax_j^c+b)}}.$$

(1)

The asymptotic variance–covariance matrix of vector $(\hat{a}, \hat{b})$ is the inverse of the observed information matrix $I(a, b)$ evaluated at $(\hat{a}, \hat{b})$, where

$$I(a, b) = -\begin{pmatrix} \frac{\partial^2}{\partial a^2} l(a, b) & \frac{\partial^2}{\partial a \partial b} l(a, b) \\ \frac{\partial^2}{\partial a \partial b} l(a, b) & \frac{\partial^2}{\partial b^2} l(a, b) \end{pmatrix}.$$

We obtain

$$\text{Var}(\hat{a}, \hat{b}) = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1}$$

with

$$c_{11} = \sum_{j=1}^{m} w_j \frac{[(x_j^c)^2 C_j(a, b)][A_j(a, b)] - [x_j^c B_j(a, b)]^2}{[A_j(a, b)]^2}, \quad c_{22} = \sum_{j=1}^{m} w_j \frac{[C_j(a, b)][A_j(a, b)] - [B_j(a, b)]^2}{[A_j(a, b)]^2},$$

$$c_{12} = c_{21} = \sum_{j=1}^{m} w_j \frac{[x_j^c C_j(a, b)][A_j(a, b)] - x_j^c[B_j(a, b)]^2}{[A_j(a, b)]^2},$$

and where

$$A_j(a,b) = \sum_{i=1}^{\infty} i^{-(ax_j^c+b)}, \qquad B_j(a,b) = -\sum_{i=1}^{\infty} i^{-(ax_j^c+b)} \ln(i), \qquad C_j(a,b) = \sum_{i=1}^{\infty} i^{-(ax_j^c+b)} \ln^2(i).$$

Solving numerically system of equations (1) for a and b is very long if successful. With MATHEMATICA, we could solve it, but it usually took more than one hour. Another problem is that, in general, initial values for the parameters a and b must be known to start the maximization algorithm. In the following section, we propose a quadratic distance method to estimate the vector of parameters (a, b). The estimates are obtained in a few seconds with MATHEMATICA and they can be used as initial values for the maximum likelihood method.

4. Quadratic distance estimation

As seen in Section 2, the sample of size n can be divided into m samples of size w_j, $j = 1, \ldots, m$, where $N_{1,j}, \ldots, N_{w_j,j}$ are i.i.d. zeta$(a, b; x_j^c)$. The observed frequencies in the jth class will be denoted by $f_{1,j}, f_{2,j}, \ldots, f_{k_j+1,j}$. For each class, it can be verified that

$$\ln \frac{p_j(i+1)}{p_j(i)} = (ax_j^c + b) \ln \frac{i}{i+1}, \quad j = 1, \ldots, m, \ i = 1, 2, \ldots, \tag{2}$$

i.e. a linear relation in a and b is obtained. To estimate $p_j(i)$, we will use the MLE $f_{i,j}/w_j$. Following Doray and Luong (1995), we will consider the following linear regression model for all the observations:

$$\ln \frac{f_{i+1,j}}{f_{i,j}} = (ax_j^c + b) \ln \frac{i}{i+1} + \varepsilon_{i,j}, \quad j = 1, \ldots, m, \ i = 1, \ldots, r_j, \tag{3}$$

where $f_{1,j}, f_{2,j}, \ldots, f_{r_j+1,j}$ are different from 0 for $j = 1, \ldots, m$ and $\varepsilon_{i,j}$ is a random error. This model can be rewritten in matrix notation as the linear model $Y = X\beta + \varepsilon$, where the vectors Y, β and ε are equal to

$$Y = \left[\ln \frac{f_{2,1}}{f_{1,1}}, \ \ln \frac{f_{3,1}}{f_{2,1}}, \ldots, \ \ln \frac{f_{r_1+1,1}}{f_{r_1,1}}, \ \ln \frac{f_{2,2}}{f_{1,2}}, \ldots, \ \ln \frac{f_{r_m+1,m}}{f_{r_m,m}} \right]',$$

$$\beta = [a, b]', \qquad \varepsilon = [\varepsilon_{1,1}, \varepsilon_{2,1}, \ldots, \varepsilon_{r_1,1}, \varepsilon_{1,2}, \ldots, \varepsilon_{r_m,m}]',$$

and matrix X to

$$X = \begin{pmatrix} x_1^c \ln \dfrac{1}{2} & x_1^c \ln \dfrac{2}{3} & \cdots & x_1^c \ln \dfrac{r_1}{r_1+1} & x_2^c \ln \dfrac{1}{2} & \cdots & x_m^c \ln \dfrac{r_m}{r_m+1} \\ \ln \dfrac{1}{2} & \ln \dfrac{2}{3} & \cdots & \ln \dfrac{r_1}{r_1+1} & \ln \dfrac{1}{2} & \cdots & \ln \dfrac{r_m}{r_m+1} \end{pmatrix}'.$$

The following two propositions will help us in obtaining the asymptotic distribution of the QDE. They are proved in Appendix A.

Proposition 1. *Asymptotically, vector ε has a mean equal to vector 0.*

Proposition 2. *The asymptotic variance–covariance matrix of vector ε is equal to*

$$\Sigma = \begin{pmatrix} \Sigma_1 & Q_{12} & \cdots & & Q_{1m} \\ Q_{21} & \Sigma_2 & & & \vdots \\ \vdots & & \ddots & Q_{m-1,m} \\ Q_{m1} & \cdots & Q_{m,m-1} & \Sigma_m \end{pmatrix}, \tag{4}$$

where Q_{ij} is an $r_i \times r_j$ matrix of zeros and

$$\Sigma_j = \frac{1}{w_j} \begin{pmatrix} \dfrac{p_j(1)+p_j(2)}{p_j(1)p_j(2)} & -\dfrac{1}{p_j(2)} & 0 & 0 & \cdots & 0 \\[2ex] -\dfrac{1}{p_j(2)} & \dfrac{p_j(2)+p_j(3)}{p_j(2)p_j(3)} & -\dfrac{1}{p_j(3)} & 0 & \cdots & 0 \\[2ex] 0 & -\dfrac{1}{p_j(3)} & \dfrac{p_j(3)+p_j(4)}{p_j(3)p_j(4)} & -\dfrac{1}{p_j(4)} & \ddots & \vdots \\[2ex] 0 & 0 & \ddots & \ddots & \ddots & 0 \\[2ex] \vdots & \vdots & \ddots & -\dfrac{1}{p_j(r_j-1)} & \dfrac{p_j(r_j-1)+p_j(r_j)}{p_j(r_j-1)p_j(r_j)} & -\dfrac{1}{p_j(r_j)} \\[2ex] 0 & 0 & \cdots & 0 & -\dfrac{1}{p_j(r_j)} & \dfrac{p_j(r_j)+p_j(r_j+1)}{p_j(r_j)p_j(r_j+1)} \end{pmatrix}$$

is the asymptotic variance–covariance matrix of the errors for the jth class. Matrix Σ_j is a square $r_j \times r_j$ matrix and Σ is of dimension $\sum_{j=1}^{m} r_j \times \sum_{j=1}^{m} r_j$.

The QDE $(\tilde{a}, \tilde{b})$ is the vector (a, b) which minimizes $[Y - X\beta]' \Sigma^{-1} [Y - X\beta]$. We obtain

$$\tilde{\beta} = (\tilde{a}, \tilde{b}) = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y. \tag{5}$$

$\tilde{\beta}$ is a consistent estimator of (a, b); asymptotically, it follows a bivariate normal distribution with mean (a, b) and variance–covariance matrix equal to

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

Since matrix Σ is a function of $p_j(i)$ and therefore of the unknown parameters a and b, an iteratively reweighted least-squares method will be needed to obtain $(\tilde{a}, \tilde{b})$. A consistent estimator for β is first obtained by replacing Σ by the identity matrix I in (5),

$$\tilde{\beta}_0 = (\tilde{a}_0, \tilde{b}_0) = (X'X)^{-1}X'Y.$$

This value of $(\tilde{a}_0, \tilde{b}_0)$ is used to obtain a consistent estimator of the variance–covariance matrix of ε, denoted by $\Sigma_{(\tilde{a}_0, \tilde{b}_0)}$, from (4). A new consistent estimator for (a, b) is obtained from

$$(\tilde{a}_1, \tilde{b}_1) = (X'\Sigma_{(\tilde{a}_0, \tilde{b}_0)}^{-1}X)^{-1}X'\Sigma_{(\tilde{a}_0, \tilde{b}_0)}^{-1}Y.$$

This iterative process is continued until convergence is attained.

5. Numerical example

Table 1 from Seal (1947) presents the number of persons having i insurance contracts in the portfolio of an insurance company, grouped into 10-year intervals from 15 to 75 years of age. For x_j^c, we will use the midpoint of the six intervals $(20, 30, \ldots, 70)$. The information in Table 1 can be read as follows: among $f_1 = 1695$ people with only one insurance policy $(i = 1)$, $f_{1,1} = 94$ are aged between 15 and 25, $f_{1,2} = 342$ between 25 and 35,..., $f_{1,6} = 59$ are aged between 65 and 75. We see that $\bar{N}_j$, the average number of policies per person in class j increases regularly from 1.06 for people aged 15–25 to 1.51 for those aged 65–75.

For each class, we estimated separately the parameter ρ_j with the maximum likelihood and quadratic distance methods and obtained the values $\hat{\rho}_j$ and $\tilde{\rho}_j$, $j = 1, \ldots, 6$ in the last two lines of Table 1. The two methods give

Table 1
Observed frequencies by age group for Seal's data

i	f_i	I_j	(15, 25)	(25, 35)	(35, 45)	(45, 55)	(55, 65)	(65, 75)
1	1695		94	342	590	433	177	59
2	207		6	34	66	59	30	12
3	46		–	5	16	11	6	8
4	22		–	3	8	7	2	2
5	9		–	0	0	3	4	2
6	8		–	2	3	3	0	–
7	4		–	–	2	2	0	–
8	3		–	–	1	1	1	–
9	1		–	–	1	0	0	–
10	1		–	–	0	1	0	–
11	2		–	–	2	0	0	–
12	0		–	–	–	0	0	–
13	1		–	–	–	1	0	–
14	0		–	–	–	–	0	–
15	0		–	–	–	–	0	–
16	0		–	–	–	–	0	–
17	0		–	–	–	–	0	–
18	1		–	–	–	–	1	–
$n = 2000$		w_j	100	386	689	521	221	83
		x_j^c	20	30	40	50	60	70
$\bar{N} = 1.28$		$\bar{N}_j$	1.06	1.16	1.27	1.32	1.40	1.51
$r = 10$		r_j	1	3	3	7	4	4
$\hat{\rho} = 2.15$		$\hat{\rho}_j$	3.50	2.55	2.20	2.02	1.88	1.53
$\tilde{\rho} = 2.14$		$\tilde{\rho}_j$	2.97	2.51	2.19	2.01	1.83	1.15

very close results except for classes 1 and 6; however, these two classes are the ones with the smallest number of observations ($w_1 = 100$, $w_6 = 83$).

We see that the value of ρ_j decreases with age, suggesting we could use the age as a covariate in the zeta model. We will assume the zeta regression model with

$$\rho + 1 = ax_j^c + b.$$

Proceeding as explained in Section 4 for the linear regression model $Y = X\beta + \varepsilon$, where

$$Y = [\ln \tfrac{6}{94},\ \ln \tfrac{34}{342},\ \ln \tfrac{5}{34}, \ldots,\ \ln \tfrac{2}{8},\ \ln \tfrac{2}{2}],$$

and

$$X = \begin{pmatrix} 20 \ln \tfrac{1}{2} & 30 \ln \tfrac{1}{2} & 30 \ln \tfrac{2}{3} & \ldots & 70 \ln \tfrac{3}{4} & 70 \ln \tfrac{4}{5} \\ \ln \tfrac{1}{2} & \ln \tfrac{1}{2} & \ln \tfrac{2}{3} & \ldots & \ln \tfrac{3}{4} & \ln \tfrac{4}{5} \end{pmatrix}',$$

we obtained with the iteratively reweighted least-squares method the values of $(\tilde{a}, \tilde{b})$ in the first half of Table 2 after only four iterations.

With MATHEMATICA, this result is obtained in a few seconds. To calculate $(\tilde{a}, \tilde{b})$, we need to truncate 15 observations, but this represents only 0.75% of the whole sample. The estimated asymptotic variance–covariance of vector $(\tilde{a}, \tilde{b})$ in Table 2 is calculated with $(X'\Sigma^{-1}_{(\tilde{a},\tilde{b})} X)^{-1}$. To evaluate the matrix Σ^{-1} of dimension $\sum_{j=1}^{6} r_j \times \sum_{j=1}^{6} r_j$, we used $r_j = 250$ for $j = 1, \ldots, 6$.

Table 2
Estimated values of the regression parameters

	a	b	Var. cov. matrix	
QDE	−0.0278562	4.36554	$\begin{pmatrix} 0.0000213361 & -0.00102895 \\ -0.00102895 & 0.0528243 \end{pmatrix}$	
MLE	−0.0250794	4.28831	$\begin{pmatrix} 0.0000231356 & -0.0011019 \\ -0.0011019 & 0.0559271 \end{pmatrix}$	
Seal	−0.032	4.38		

Table 3
QDEs with different covariate values

Covariate	$\tilde{a}$	$\tilde{b}$
$x_j^c - 5$	−0.278562	4.22626
x_j^c	−0.278562	4.36554
$x_j^c + 5$	−0.278562	4.50482

Since $\tilde{a}$ and $\tilde{b}$ are asymptotically unbiased and normally distributed, we can easily obtain an approximate 95% confidence interval for a, $[-0.03691, -0.01880]$ and for b, $[3.915, 4.816]$. The estimated correlation coefficient between $\tilde{a}$ and $\tilde{b}$ is equal to -0.9692.

To calculate the MLE $(\hat{a}, \hat{b})$, we used MATHEMATICA to solve the system of equations (1). As initial values for the algorithm, knowing the QDE $(\tilde{a}, \tilde{b})$, we used the interval $[-0.02, -0.03]$ for a and $[4,5]$ for b. After more than 1 hour of calculation, we obtained the values in the second half of Table 2 for $(\hat{a}, \hat{b})$ and their asymptotic variance–covariance matrix. A 95% confidence interval for a is given by $[-0.03451, -0.01565]$ and for b by $[3.825, 4.752]$.

Using a graphical procedure, Seal had obtained the estimates of (a, b) in the last line of Table 2, but no variance was given for those estimates. With MATHEMATICA, the QDE was obtained in 3 seconds, while it took 1380 times more to obtain the MLE, even when we used the QDE as initial value to solve the system of equations.

To measure the sensitivity of the QD method to the grouping of a continuous covariate into classes, we recalculated the QDEs of a and b by changing the covariate value used for class I_j. In Table 2, the central value of age interval $I_j, x_j^c = \{20, 30, \ldots, 70\}$, was used for the covariate. Table 3 gives the values of the QDEs $\tilde{a}$ and $\tilde{b}$ when we use the lower bound of the age interval $I_j, x_j^c - 5 = \{15, 25, \ldots, 65\}$, and the upper bound of $I_j, x_j^c + 5 = \{25, 35, \ldots, 75\}$, instead of x_j^c, in the first line of regression matrix X.

As can be seen, $\tilde{a}$ is unaffected by a change in x_j^c, while $\tilde{b}$ changes linearly with x_j^c; increasing x_j^c by 5 changes the estimate of $\tilde{b}$ by $-5\tilde{a}$.

6. Model testing

To test whether the fit of model (2) to a data set is appropriate, an exploratory graph should first be done. If the model is adequate, the points $(\ln(i/(i + 1)), \ln(f_{i+1,j}/f_{i,j}))$ will be close to a straight line of slope $(ax_j^c + b)$ passing through the origin. Fig. 1 presents those six straight lines for $x_j^c = \{20, 30, \ldots, 70\}$ with $\tilde{a}$ and $\tilde{b}$, together with the data points. It must be remembered, however, that with the QD method, a weighted least-squares regression is performed, which puts more weight on points with more observations and more probability, explaining the better fit on the left of the graph than on the right. After this plot, the parameters can be estimated and formal χ^2 goodness-of-fit tests can be performed.

Table 4 presents the observed (f_i) and expected (f_t) frequencies under the zeta regression models with the estimators of (a, b) obtained by the quadratic distance method. The chi-square statistic for class j is denoted by χ_j^2.
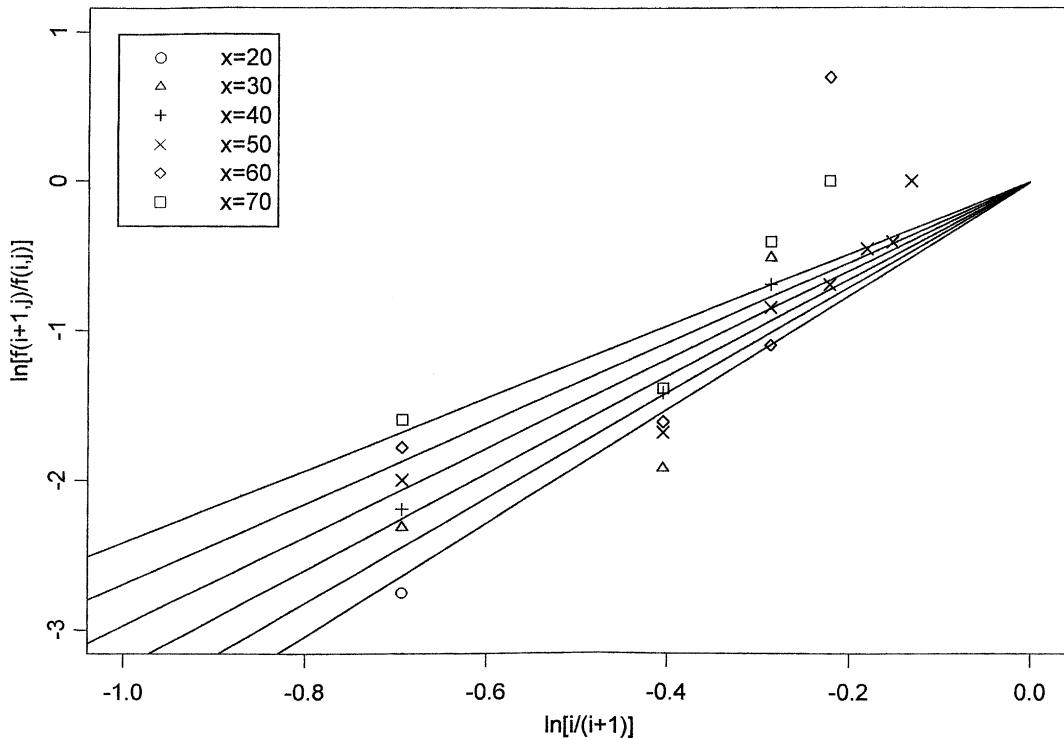
Fig. 1. Adequacy of the zeta regression model.

For the whole sample, the χ^2-statistic equals $\sum_{j=1}^{6} \chi_j^2 = 14.77$ with 19 degrees of freedom. Since $\chi_{19, 0.95}^2 = 30.1$, the zeta model with age as a covariate is not rejected for Seal's data. Using the maximum likelihood method to estimate (a, b), we reach the same conclusion with an observed value of $\chi^2 = 15.12$.

The statistic $[Y - X\tilde{\beta}]' \Sigma^{-1}(\tilde{\beta})[Y - X\tilde{\beta}]$ can also be used to test the goodness-of-fit of the model to a data set. Under the composite hypothesis $H_0 : \beta \in$ the full parameter space, it has an asymptotic χ^2 distribution with number of degrees of freedom which is equal to $\dim(Y) - \dim(\beta)$. The proof is analogous to the one provided in Luong and Doray (1996) for the case without covariates. In our example, the above statistic has a value $\chi^2 = 12.4381$, smaller than the critical value $\chi_{20, 0.95}^2 = 31.4$, indicating again a good fit of the zeta model with age as a covariate.

To test the simple null hypothesis $H_0 : a = 0$, two tests are possible, a test based on the asymptotic normality of estimators $\hat{a}$ or $\tilde{a}$ and the likelihood ratio test. Calculating the statistics $|\hat{a}/\hat{\sigma}_{\hat{a}}|$ and $|\tilde{a}/\hat{\sigma}_{\tilde{a}}|$, we obtain, respectively, 5.21 and 6.03, larger than the critical value 1.96 at the significance level $\alpha = 0.05$, leading to a rejection of hypothesis H_0.

For the other test, under H_0, the likelihood ratio test statistic defined as $\Lambda = -2 \ln \{L[0, \hat{b}(0)]/L[\hat{a}, \hat{b}]\}$ has an asymptotic χ^2 distribution with one degree of freedom, where $\hat{b}(0)$ is the MLE of b when $a = 0$, and L is the value of the likelihood function.

When $a = 0$, the equation $\rho + 1 = ax + b$ becomes $\rho + 1 = b$, and so $\hat{b}(0) = \hat{\rho} + 1$ and we obtain $\hat{b}(0) = 3.15161$, from Table 1.

From the likelihood functions

$$L[0, \hat{b}(0)] = \prod_{j=1}^{6} \prod_{i=1}^{w_j} \frac{n_{i,j}^{-\hat{b}(0)}}{\zeta(\hat{b}(0))}, \qquad L[\hat{a}, \hat{b}] = \prod_{j=1}^{6} \prod_{i=1}^{w_j} \frac{n_{i,j}^{-(\hat{a}x_j^c + \hat{b})}}{\zeta(\hat{a}x_j^c + \hat{b})},$$

Table 4
Observed and expected frequencies with the QDE

i	I_j											
	(15, 25)		(25, 35)		(35, 45)		(45, 55)		(55, 65)		(65, 75)	
	f_i	f_t	f_i	f_t	f_i	f_t	f_i	f_t	f_i	f_t	f_i	f_t
1	94	91.2	342	343.6	590	594.5	433	431.5	177	173.2	59	60.3
2	6	6.5	34	29.7	66	62.4	59	55.0	30	26.8	12	11.3
3	–	1.4	5	7.1	16	16.7	11	16.5	6	9.0	8	4.2
4	–	0.5	3	2.6	8	6.6	7	7.0	2	4.1	2	2.1
5	–	0.2	0	1.2	0	3.2	3	3.6	4	2.3	2	1.2
6	–	0.1	2	0.6	3	1.8	3	2.1	0	1.4	–	0.8
7	–	0.1	–	0.4	2	1.1	2	1.3	0	0.9	–	0.5
8	–	–	–	0.2	1	0.7	1	0.9	1	0.6	–	0.4
9	–	–	–	0.1	1	0.5	0	0.6	0	0.5	–	0.3
10	–	–	–	0.1	0	0.3	1	0.5	0	0.4	–	0.2
11	–	–	–	0.1	2	0.2	0	0.3	0	0.3	–	0.2
12	–	–	–	0.1	–	0.2	0	0.3	0	0.2	–	0.1
13	–	–	–	–	–	0.1	1	0.2	0	0.2	–	0.1
14	–	–	–	–	–	0.1	–	0.2	0	0.1	–	0.1
15	–	–	–	–	–	0.1	–	0.1	0	0.1	–	0.1
16	–	–	–	–	–	0.1	–	0.1	0	0.1	–	0.1
17	–	–	–	–	–	0.1	–	0.1	0	0.1	–	0.1
18	–	–	–	–	–	–	–	0.1	1	0.1	–	0.1
≥ 19	–	–	–	–	–	–	–	0.1	–	0.1	–	–
w_j	100		386		689		521		221		83	
χ_j^2	0.99		1.29		2.50		2.16		3.85		3.98	

we obtain $\Lambda = 26.2$; at the significance level $\alpha = 0.05$, $\chi_{1,\,0.95}^2 = 3.841$, and so there is strong evidence in the data against $H_0 : a = 0$. The linear regression model with age as a covariate is appropriate for the data.

7. Asymptotic efficiency of QDE

In this section, we calculate the bias, variance of the QDE and its asymptotic efficiency compared to the MLE, when a covariate is present in the zeta distribution. For the model without regression parameters, Doray and Luong (1995) calculated the ratio of the asymptotic variances $\mathrm{Var}(\tilde{\rho})/\mathrm{Var}(\hat{\rho})$ for values of the parameter ρ between 1.0 and 5.0. In most practical applications, Johnson et al. (1992) observed that the value of ρ was slightly larger than 1. For these values, the QDE was highly efficient.

First, we calculate the asymptotic efficiency of the QDE compared to the MLE with the explanatory variable used in the example of Section 6, $x \in [20, 70]$. The values of the parameters a and b (first two columns of Table 5) were chosen so that $\rho = ax + b$ is approximately between 1 and 3.5 (last column of Table 5). The explanatory variables used are $x_j^c = 20, 30, 40, 50, 60, 70$ and the number of observations in each class is $w_j = 100, 386, 689, 521, 221, 83$, for a total of $n = 2000$ observations. To calculate Σ, we used $r_j = 250$, $j = 1, 2, \ldots, 6$.

In Table 5, we calculated the ratios $\mathrm{Var}(\tilde{a})/\mathrm{Var}(\hat{a})$, $\mathrm{Var}(\tilde{b})/\mathrm{Var}(\hat{b})$ and $1/\mathrm{eff}(\tilde{\beta})$, where $\mathrm{eff}(\tilde{\beta})$, the asymptotic efficiency of the QDE, is defined as the ratio of the determinants of the asymptotic variance–covariance-matrices,

$$\mathrm{eff}(\tilde{\beta}) = \frac{|\mathrm{Var}(\hat{\beta})|}{|\mathrm{Var}(\tilde{\beta})|} = \frac{|I_{(\beta)}^{-1}|}{|(X'\Sigma^{-1}X)^{-1}|},$$

Table 5
Ratio of asymptotic variances

a	b	$\mathrm{Var}(\tilde{a})/\mathrm{Var}(\hat{a})$	$\mathrm{Var}(\tilde{b})/\mathrm{Var}(\hat{b})$	$1/\mathrm{eff}(\tilde{\beta})$	Range of ρ
−0.01	2.5	1.13155	1.10281	1.2508	[0.8, 1.3]
−0.01	3.0	1.01527	1.01190	1.0275	[1.3, 1.8]
−0.01	3.5	1.00162	1.00125	1.0030	[1.8, 2.3]
−0.01	4.0	1.00016	1.00012	1.0002	[2.3, 2.8]
−0.01	4.5	1.00001	1.00001	1.0001	[2.8, 3.3]
−0.02	3.5	1.02944	1.02039	1.0475	[1.1, 2.1]
−0.02	4.0	1.00323	1.00224	1.0049	[1.6, 2.6]
−0.02	4.5	1.00032	1.00022	1.0006	[2.1, 3.1]
−0.03	4.0	1.06613	1.04354	1.1050	[0.9, 2.4]
−0.03	4.5	1.00758	1.00510	1.0112	[1.4, 2.9]
−0.03	5.0	1.00078	1.00053	1.0012	[1.9, 3.4]
−0.04	5.0	1.01836	1.01207	1.0280	[1.2, 3.2]

Table 6
Estimated bias and variance of the QDE $(\tilde{a}, \tilde{b})$

a	b	$\mathrm{Bias}(\tilde{a})$	$\mathrm{Var}(\tilde{a})$	$\mathrm{Bias}(\tilde{b})$	$\mathrm{Var}(\tilde{b})$	Range of ρ
−0.01	2.5	0.000184	0.0000185	0.0107	0.0390	[0.8, 1.3]
−0.01	3.0	0.000228	0.0000289	0.0078	0.0611	[1.3, 1.8]
−0.01	3.5	−0.000109	0.0000518	0.0203	0.1090	[1.8, 2.3]
−0.01	4.0	−0.000338	0.0000882	0.0297	0.1921	[2.3, 2.8]
−0.01	4.5	−0.000340	0.0001700	0.0301	0.3592	[2.8, 3.3]
−0.02	3.0	0.000082	0.0000196	0.0158	0.0435	[0.6, 1.6]
−0.02	3.5	−0.000134	0.0000316	0.0261	0.0725	[1.1, 2.1]
−0.02	4.0	0.000058	0.0000504	0.0075	0.1144	[1.6, 2.6]
−0.02	4.5	−0.000621	0.0000948	0.0511	0.2185	[2.1, 3.1]
−0.03	4.0	0.000448	0.0000301	−0.0033	0.0719	[0.9, 2.4]
−0.03	4.5	0.000443	0.0000545	−0.0055	0.1337	[1.4, 2.9]
−0.03	5.0	−0.000118	0.0000914	0.0241	0.2279	[1.9, 3.4]
−0.04	4.5	0.000185	0.0000339	0.0084	0.0856	[0.7, 2.7]
−0.04	5.0	−0.000252	0.0000582	0.0312	0.1491	[1.2, 3.2]
−0.05	5.0	0.000181	0.0000359	0.0097	0.0944	[0.5, 3.0]

a notion of efficiency introduced by Bhapkar (1972), with $|A|$ denoting the determinant of matrix A. As observed for the zeta distribution without regression parameters, the efficiency of the QDE increases as the value of the parameter ρ increases (i.e. as a increases for b fixed or b increases with a fixed). In our example, $\tilde{a}$ and $\tilde{b}$ are close to −0.028 and 4.4, respectively, so that for these values, $\mathrm{eff}(\tilde{\theta})$ is at least 0.99.

In Table 6, we calculate the bias and variance of $(\tilde{a}, \tilde{b})$ for some values of a and b, with a sample of size $n = 2000$, $x \in [20, 70]$ and the preceding values of w_j. These calculations were done by generating 1000 samples of size 2000 each; because of reasons explained in Section 5 (more than a 1000 times longer to calculate the MLE than the QDE), it was not possible to estimate the MLE $(\hat{a}, \hat{b})$. The bias and variance are reported only for the QDE $\tilde{a}$ and $\tilde{b}$.

8. Conclusion

We have shown one way to introduce covariates into the zeta or discrete Pareto model. The continuous Pareto distribution has been used in economics to model the income of individuals in certain populations. The discrete version of this distribution could be used to model a number of events proportional to the income. Instead of maximum

likelihood, we have proposed a new way of estimating the parameters of this distribution, based on quadratic distance. This estimator is easier to calculate than the MLE; it is also consistent, has a very high asymptotic efficiency and an asymptotic normal distribution.

Acknowledgements

Appendix A.

Proof of Proposition 1. From (3), $\epsilon_{i,j} = \ln(f_{i,j+1}/f_{i,j}) - (ax_j^c + b)\ln(i/(i+1))$. The random variable $f_{i,j}$ has a binomial $(w_j, p_j(i))$ distribution. As w_j gets large, $f_{i,j}$ converges to a normal distribution, i.e.

$$\text{as } w_j \to \infty, \quad \sqrt{w_j}\left(\frac{f_{i,j}}{w_j} - p_j(i)\right) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in law, and $\sigma^2 = p_j(i)(1 - p_j(i))$.

By the δ-method (see Theorem C1 of Lawless (1982)), since $g(x) = \ln x$ is a differentiable function

$$\text{as } w_j \to \infty, \quad \sqrt{w_j}\left(\ln\frac{f_{i,j}}{w_j} - \ln p_j(i)\right) \xrightarrow{\mathcal{D}} N(0, [g'(p_j(i))]^2\sigma^2).$$

Therefore, as $w_j \to \infty$, $E(\ln(f_{i,j}/w_j) - \ln p_j(i)) = 0$. Similarly, as $w_j \to \infty$, $E(\ln(f_{i+1,j}/w_j) - \ln p_j(i+1)) = 0$. Subtracting, we get, as $w_j \to \infty$, $E(\ln(f_{i+1,j}/f_{i,j}) - \ln(p_j(i)/p_j(i+1))) = 0$, from which we obtain $E(\epsilon_{i,j}) = 0$, as $w_j \to \infty$. $\qquad\square$

Proof of Proposition 2. From the above theorem, we get the asymptotic variance

$$\text{Asvar}\left(\ln\frac{f_{i,j}}{w_j} - \ln p_j(i)\right) = \frac{1}{(p_j(i))^2}\frac{p_j(i)(1 - p_j(i))}{w_j} = \frac{1 - p_j(i)}{w_j p_j(i)}.$$

Since $(f_{i,j}, f_{i+1,j})$ has a trinomial distribution with parameters $(w_j, p_j(i), p_j(i+1))$, the covariance between $f_{i,j}$ and $f_{i+1,j}$ is equal to $-w_j p_j(i)p_j(i+1)$. By Theorem C2 of Lawless (1982), we get the asymptotic covariance

$$\text{Ascov}\left(\ln\frac{f_{i,j}}{w_j}, \ln\frac{f_{i+1,j}}{w_j}\right) = \left(\frac{d\ln x}{dx}\bigg|_{x=p_j(i)}\right)\left(\frac{d\ln y}{dy}\bigg|_{y=p_j(i+1)}\right)\text{Ascov}\left(\frac{f_{i,j}}{w_j}, \frac{f_{i+1,j}}{w_j}\right)$$

$$= \frac{1}{p_j(i)} \times \frac{1}{p_j(i+1)} \times -\frac{p_j(i)p_j(i+1)}{w_j} = -\frac{1}{w_j}.$$

Therefore,

$$\text{Asvar}(\epsilon_{i,j}) = \text{Asvar}\left(\ln\frac{f_{i+1,j}}{f_{i,j}}\right) = \text{Asvar}\left(\ln\frac{f_{i,j}}{w_j}\right) + \text{Asvar}\left(\ln\frac{f_{i+1,j}}{w_j}\right) - 2\,\text{Ascov}\left(\ln\frac{f_{i,j}}{w_j}, \ln\frac{f_{i+1,j}}{w_j}\right)$$

$$= \frac{1 - p_j(i)}{w_j p_j(i)} + \frac{1 - p_j(i+1)}{w_j p_j(i+1)} + \frac{2}{w_j} = \frac{p_j(i) + p_j(i+1)}{w_j p_j(i)p_j(i+1)}.$$

In the jth class, the asymptotic covariance between two successive errors is given by

$$
\begin{aligned}
\mathrm{Ascov}(\epsilon_{i,j}, \epsilon_{i+1,j}) &= \mathrm{Ascov}\left(\ln \frac{f_{i+1,j}}{f_{i,j}},\ \ln \frac{f_{i+2,j}}{f_{i+1,j}} \right) \\
&= \mathrm{Ascov}\left(\ln \frac{f_{i+1,j}}{w_j},\ \ln \frac{f_{i+2,j}}{w_j} \right) - \mathrm{Asvar}\left(\ln \frac{f_{i+1,j}}{w_j} \right) \\
&\quad - \mathrm{Ascov}\left(\ln \frac{f_{i,j}}{w_j},\ \ln \frac{f_{i+2,j}}{w_j} \right) + \mathrm{Ascov}\left(\ln \frac{f_{i,j}}{w_j},\ \ln \frac{f_{i+1,j}}{w_j} \right) \\
&= -\frac{1}{w_j} - \frac{1 - p_j(i+1)}{w_j\, p_j(i+1)} + \frac{1}{w_j} - \frac{1}{w_j} = -\frac{1}{w_j\, p_j(i+1)}.
\end{aligned}
$$

If $|i - k| > 1$, $\mathrm{Ascov}(\epsilon_{i,j}, \epsilon_{k,j}) = 0$ and we obtain Σ_j for the variance–covariance matrix of the errors in the jth class.

For errors in different classes, since $f_{i,j}$ and $f_{k,l}$ $(j \neq l)$ are independent random variables, $\mathrm{Cov}(\epsilon_{i,j}, \epsilon_{k,j}) = 0$ so that the matrix $Q_{i,j}$ equals 0. □

References

Bhapkar, V.P., 1972. On a measure of efficiency of an estimating equation. Sankhya, Series A 34, 467–472.

Doray, L.G., Luong, A., 1995. Quadratic distance estimators for the zeta family. Insurance: Mathematics and Economics 16, 255–260.

Doray, L.G., Luong, A., 1997. Efficient estimators for the Good family. Communication in Statistics: Simulation and Computation 26, 1075–1088.

Johnson, N.L., Kotz, S., Kemp, A.W., 1992. Univariate Discrete Distributions. Wiley, New York.

Lawless, J.F., 1982. Statistical Models and Methods for Lifetime Data. Wiley, New York.

Luong, A., Doray, L.G., 1996. Goodness-of-fit test statistics for the zeta family. Insurance: Mathematics and Economics 19, 45–53.

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

Seal, H.L., 1947. A probability distribution of deaths at age x when policies are counted instead of lives. Scandinavisk Aktuarietidskrift 30, 18–43.