Improved Maximum Likelihood Estimation

for the Zeta Distribution

David E. Giles

Department of Economics
University of Victoria

September, 2024

Abstract

We consider the small-sample properties of the maximum likelihood estimator for the scaling parameter of the zeta probability distribution. Specifically, two analytical approaches to reducing the bias of this estimator are investigated – the "corrective" method of Cox and Snell (1968) and the "preventive" approach of Firth (1993). Both methods are effective, but the latter approach is superior and it essentially eliminates the bias of the estimator. At the same time, the percentage mean squared error is also reduced significantly This leads us to recommend Firth's bias correction in this context. The impact of both bias correction procedures is illustrated in two applications with real-life data.

Key Words Zeta distribution; maximum likelihood estimation; bias reduction

Author contact

58 Rock Lake Court, RR1, Bancroft, Ontario K0L 1C0, CANADA; dgiles@uvic.ca ; +1-613-202 9501

(*The author is Professor Emeritus, Department of Economics, University of Victoria, CANADA*.)

1. Introduction

The Riemann zeta distribution for the discrete random variable, X, has the p.m.f.:

$$p(x) = Pr.[X = x] = x^{-s}/\zeta(s) \quad ; \quad x = 1, 2, 3, \dots \dots \dots\ ; s > 1 \tag{1}$$

where the normalizing constant, $\zeta(s) = \sum_{i=1}^{\infty} i^{-s}$, is Riemann's zeta function. The m'th. raw moment of X is given by

$$E[X^m] = (1/\zeta(s)) \sum_{k=1}^{\infty} \left(1/_{k^{s-m}} \right) \ , \tag{2}$$

and the sum in (2) converges only for $s > m + 1$. If this condition is satisfied, clearly

$$E[X^m] = (\zeta(s-m)/\zeta(s)). \tag{3}$$

Given the key role played by the zeta function in the density in (1), it is hardly surprising that the Zeta distribution arises in a variety of contexts in the context of number theory. For example, Lin and Hu (2001) use this distribution to provide alternative proofs of several existing theorems in that field, and Peltzer (2019) uses it to develop new central limit theorems for prime factors in an arithmetic progression, *inter alia*. Recently, Fujita and Yoshida (2023) use the zeta distribution to improve some established results relating to the distributions of the greatest common divisors and least common multiples of positive integers. By using the zeta distribution, rather than the uniform distribution, their results are exact, rather than being valid only asymptotically.

From a statistical perspective, the zeta distribution is of interest in several fields, including insurance and actuarial science, psychology, network theory, and artificial intelligence. Seal (1947, 1952) provides one of the earliest examples in the field of actuarial science, as we discuss in section 4 of this paper. Also, see Doray and Luong (1995) and Doray and Arsenault (2002). In another early contribution, Haight (1966) applied the zeta distribution (and others) to psychological data relating to word association; and more recently the role of the zeta distribution in certain artificial intelligence problems, has been discussed by Özkural (2018). The zeta distribution is central to scale-free networks, and there is a vast literature associated with this. Broido and Clauset (2019) provide an important and critical summary of this literature. Devianto *et al*. (2019) explore the characteristic function of the zeta distribution in some detail in the complex case; and Dai *et al*. (2021a, 2021b) show that this distribution can be represented as a continuous mixture of either negative binomial or Poisson counts.

This paper examines the properties of the maximum likelihood estimator of the scaling parameter, s, in (1), and evaluates the merits of some alternative analytical methods for reducing the bias of this estimator in finite samples. This estimator, and the bias reduction techniques, are discussed in detail in the next section. Section 3 presents the results of a simulation experiment that investigates the effectiveness of the bias reduction strategies, and provides a clear preference for a particular approach. Two illustrative empirical applications involving real data are provided in section 4; and section 5 offers some concluding remarks.

2. Estimation and bias reduction

Maximum likelihood estimation provides the natural estimator for the power (scaling) parameter, s. Method of moments estimation is thwarted by the fact that the m'th. moment of the zeta distribution is defined only for $m < s - 1$, and the latter value is unknown. From (1), the log-likelihood function for s, based on a sample of N independent observations, is

$$l(s|\boldsymbol{x}) = \log\big(L(s|\boldsymbol{x})\big) = -s\sum_{i=1}^{N}\log(x_i) - Nlog\big(\zeta(s)\big), \tag{4}$$

and the first-order condition for its maximization is

$$\frac{\partial l}{\partial s} = -\sum_{i=1}^{n}\log(x_i) - N\left(\frac{\zeta^{(1)}(s)}{\zeta(s)}\right) = 0 \qquad , \tag{5}$$

where $\zeta^{(j)}(s)$ is the j'th. derivative of the Riemann zeta function with respect to s.

The Maximum Likelihood Estimator (MLE), $\tilde{s}$, is obtained by solving (5) numerically for s. Although this estimator is consistent and asymptotically efficient, it will be biased and inefficient in small samples. Our objective is to reduce this bias analytically, without sacrificing small-sample efficiency. To this end we consider two approaches – the "corrective" procedure proposed by Cox and Snell (1968); and the "preventive" procedure suggested by Firth (1993). Both approaches have been used extensively in the associated literature. For example, see Cordeiro and McCullagh (1991), Cordeiro and Klein (1994), Kosmidis and Firth (2009), Giles (2021), Giles *et al.* (2016), and Schwartz *et al.* (2013), among many other authors.

In what follows, we need the following higher derivatives of (4):

$$\frac{\partial^2 l}{\partial s^2} = \frac{N\left[(\zeta^{(1)}(s))^2 - \zeta(s)\zeta^{(2)}(s)\right]}{\big(\zeta(s)\big)^2} \tag{6}$$

$$\frac{\partial^3 l}{\partial s^3} = \frac{N\left[3\zeta(s)\zeta^{(1)}(s)\zeta^{(2)}(s) - 2\left(\zeta^{(1)}(s)\right)^3 - \left(\zeta(s)\right)^2\zeta^{(3)}(s)\right]}{\left(\zeta(s)\right)^3} \quad . \tag{7}$$

Following the notation of Cordeiro and Klein (1994), we denote the joint cumulants of $l(s|\boldsymbol{x})$ as $\kappa_{11} = E\left[\frac{\partial^2 l}{\partial s^2}\right]$ and $\kappa_{111} = E\left[\frac{\partial^3 l}{\partial s^3}\right]$, and let $\kappa_{11}^{(1)} = \frac{\partial \kappa_{11}}{\partial s}$. Note that κ_{11} is given by (6); and $\kappa_{111} = \kappa_{11}^{(1)}$ is given by (7). Fisher's (scalar) information measure is

$$K = -E\left[\frac{\partial^2 l}{\partial s^2}\right] = \frac{N\left[\zeta(s)\zeta^{(2)}(s) - (\zeta^{(1)}(s))^2\right]}{\left(\zeta(s)\right)^2} \quad , \tag{8}$$

and we define

$$A = \kappa_{11}^{(1)} - ½\kappa_{111} = \frac{N\left[3\zeta(s)\zeta^{(1)}(s)\zeta^{(2)}(s) - 2\left(\zeta^{(1)}(s)\right)^3 - \left(\zeta(s)\right)^2\zeta^{(3)}(s)\right]}{2\left(\zeta(s)\right)^3} \quad . \tag{9}$$

Cordeiro and Klein (1994) Show that the first-order bias of $\tilde{s}$, as determined by Cox and Snell, can be written as

$$b = Bias = AK^{-2} = \left(\frac{\zeta(s)}{2N}\right)\frac{\left[3\zeta(s)\zeta^{(1)}(s)\zeta^{(2)}(s) - 2\left(\zeta^{(1)}(s)\right)^3 - (\zeta(s))^2\zeta^{(3)}(s)\right]}{[(\zeta^{(1)}(s))^2 - \zeta(s)\zeta^{(2)}(s)]^2} + O\left(\frac{1}{N^2}\right) . \tag{10}$$

The Cox-Snell bias-corrected MLE for *s* is then defined as

$$\hat{s} = \tilde{s} - \tilde{b} \quad , \tag{11}$$

where $\tilde{b}$ is obtained by replacing *s* by $\tilde{s}$ everywhere in (10).

Firth's "preventive" method for eliminating the first-order bias of the MLE involves solving the modified score equation,

$$\frac{\partial l}{\partial s} - Kb = 0 \quad , \tag{12}$$

for '*s'* yielding the estimator, $\check{s}$. That is, we solve the following non-linear equation for '*s*':

$$-\sum_{i=1}^{n} \log(x_i) - N\left(\frac{\zeta^{(1)}(s)}{\zeta(s)}\right) + \frac{\left[3\zeta(s)\zeta^{(1)}(s)\zeta^{(2)}(s) - 2\left(\zeta^{(1)}(s)\right)^3 - \left(\zeta(s)\right)^2\zeta^{(3)}(s)\right]}{2\zeta(s)[(\zeta^{(1)}(s))^2 - \zeta(s)\zeta^{(2)}(s)]} = 0 \quad . \tag{13}$$

Because the zeta distribution is a member of the (discrete) linear exponential family, and 's' is its canonical parameter, Firth's preventive method amounts to maximizing the modified log-likelihood function,

$$l^* = l(s|\boldsymbol{x}) + ½\log|K| \ , \tag{14}$$

where K is Fisher's information measure, and $|K|^{½}$ is Jeffreys' invariant prior for s.

In our case,

$$K \propto \frac{\left[\zeta(s)\zeta^{(2)}(s) - (\zeta^{(1)}(s))^2\right]}{\left(\zeta(s)\right)^2} \tag{15}$$

and so

$$l^* = -s\sum_{i=1}^{n}\log(x_i) - (N+1)log\left(\zeta(s)\right) + ½log\left[\zeta(s)\zeta^{(2)}(s) - (\zeta^{(1)}(s))^2\right] + \ const.$$

The Firth estimator, $\breve{s}$, is then obtained by solving

$$\frac{\partial l^*}{\partial s} = -\sum_{i=1}^{n}\log(x_i) - (N+1)\left(\frac{\zeta^{(1)}(s)}{\zeta(s)}\right) + \frac{\left[\zeta(s)\zeta^{(3)}(s) - \zeta^{(1)}(s)\zeta^{(2)}(s)\right]}{2\left[\zeta(s)\zeta^{(2)}(s) - \left(\zeta^{(1)}(s)\right)^2\right]} = 0 \ . \tag{16}$$

Some trivial algebraic manipulations show that the expressions in equations (13) and (16) are identical, for all s, N, and sample values. So, either equation can be solved (numerically) to obtain Firth's bias-reduced estimator, $\breve{s}$.

3. A simulation experiment

3.1 Experimental design

We have conducted a Monte Carlo simulation experiment to evaluate the performances of the Cox-Snell and Firth bias-reduction procedures in the case of the zeta distribution. Various sample sizes (N) have been considered. Recalling that the condition, $s > 1$, must be satisfied we considered values of this parameter from 1.25 to 4.00. This covers the range of estimated values usually encountered in empirical applications using this distribution. The number of Monte Carlo replications is NREP = 100,000, and this ensured stable simulation results.

The zeta-distributed random variates were generated using the 'UnivRNG' package in R (Demirtas *et al*., 2021). This was found to be computationally faster than the 'VGAM' package for this purpose, and the sample moments matched the theoretical moments (where they exist) very accurately. The MLE, $\tilde{s}$, was computed using the zm.ll command in the 'tolerance' package in R (Young, 2020). The other estimators require the calculation of the various derivatives of the Riemann Zeta function. For fast computation, the 'zeta' command in the 'VGAM' package in R (Yee and Moler, 2022) was used to obtain the first two derivatives of this function. Their values were successfully checked for accuracy using the following result from Choudury (1995), equation (20), for the *j*'th. derivative of the Zeta function:

$$\zeta^{(j)}(s) = (-1)^j \frac{j!}{(s-1)^{j+1}} + \sum_{k=0}^{m-1} (-1)^k \frac{\gamma_{j+k}}{k!}(s-1)^k + R_m(j,s). \tag{17}$$

The γ constants in (14) are related to the Stieltjes constants by $A_j = (-1)^j \left(\frac{\gamma_j}{j!}\right)$, and the values of γ_0 to γ_{100} are presented in Table 5 of Choudhary (1995). In practice, a value of *m* is chosen to control the value of the remainder term, *R*, in (17).

The VGAM package does not provide the third or higher derivatives of the zeta function. So, to compute the required third derivative we used equation (14), with $m = 19$. From Choudhury (1995, p. 486), it can be shown that with this choice of *m* the remainder term when $s = 2$ (for example) satisfies $|R_{19}(3,2)| < \left|\frac{\gamma_{19}}{15!}\right| = 3.85 \times 10^{-16}$. Similar results apply for other values of *s*, thereby ensuring the accuracy of our computations.

We used Brent's (1973) algorithm, coded as the 'uniroot' function in the base R package, to obtain numerical solutions for (5) and (13) (or (16)), and hence the values of $\tilde{s}$ and $\check{s}$. While our primary interest is in bias reduction, achieving this at the cost of increased variance in the sampling distributions is not desirable. So, we have simulated both percentage biases and percentage mean squared errors (MSEs) for the original MLE of *s*, as well as the two bias-reduced alternative estimators.

If s_i^* denotes one of the estimators of *s* for the *i*'th. replication of the experiment ($i = 1, 2, \ldots,$ *NREP*), then the simulated percentage bias and percentage MSE of s^* are computed as:

$$\%Bias(s^*) = 100\left[\frac{1}{NREP}\left(\sum_{i=1}^{NREP} s_i^*\right) - s\right]/s$$

and

$$\%MSE(s^*) = 100\left[\frac{1}{NREP}\sum_{i=1}^{NREP}(s_i^* - s)^2\right]/s^2.$$

3.2 Simulation results

The R code for the simulations and the applications in section 4 can be downloaded from https://github.com/DaveGiles1949/r-code. The simulation results are presented in Table 1. In that table, we see that the smallest sample size considered was reduced as s increased. This was necessary to ensure that the mean and variance of all of the simulated zeta variates were defined for every replication.

The results in Table 1 show that the percentage bias of the MLE of the scaling parameter, s, decreases as the true value of that parameter increases. The percentage MSE of $\tilde{s}$ also decreases as s increases, for $s \leq 2.0$, but then increases with the value of s. in the same manner. For any value of s and sample size, N, the Cox-Snell estimator ($\hat{s}$) exhibits less bias than the original MLE. Moreover, this comes without any increase in MSE, and for smaller samples there is a reduction in *both* bias and MSE when the Cox-Snell correction is used. Firth's preventive approach is even more successful in this context. Indeed, we see that $\breve{s}$ has negligible bias, even for very small samples. We conjecture that in fact Firth's modified estimator is *exactly unbiased* for the zeta distribution, and the associated values in Table 1 simply reflect sampling error. However, we have been unable to formally prove this conjecture. Firth's procedure also results in a dramatic reduction in MSE, compared with both the original MLE of s and its Cox-Snell counterpart. All three estimators are weakly consistent, so (for any value of s), both the percentage bias and percentage MSE decrease as N increases.

The computational costs associated with implementing Firth's preventive approach and the Cox-Snell corrective approach to bias reduction are essentially the same. So, our results suggest that the first of these procedures is substantially superior to the second one, when it comes to estimating the parameter of the zeta distribution.

4. Applications

4.1 Insurance policies

First, we present some empirical results based on the data used in early work by Seal (1947, 1952) that involved the zeta distribution. In fact, the second of these papers was the first to present the MLE for the scaling parameter, s, for this distribution. The data in question relate to a sample of the number of 'duplicate' insurance policies held by males with certain British life insurance offices, and are reported in Table 1 of Seal (1952).

These data cover different age groups, such as 15-19 years or 50-54 years, *etc*. Following Seal's approach, we measure the sample observations at the centres of these intervals, such 17.5 years or 52.5 years, *etc*., and then obtain the fitted zeta distribution for the number of policies in each case. The overall sample comprises N = 1,999 observations, with the number of policies ranging from 1 to 16. The sub-samples for the 12 age-groups have sizes ranging from N = 38 to N = 361.

Our maximum likelihood results, including the Cox-Snell and Firth estimates of s, appear in Table 2. The values of the basic MLE, $\tilde{s}$, match those in the first column of results in Seal's Table 3 very closely. The small discrepancies are presumably due to the approximations that Seal used to solve equation (5). These included using a table of approximate values for $\frac{\zeta^{(1)}(s)}{\zeta(s)}$, reported by Walther (1926), and a rather crude interpolation procedure. (See Seal, 1952, p.117.) As might be expected, we see that each of the bias corrections reduces the numerical value of the maximum likelihood estimates slightly. However, as can be seen in Table 3, the impact of bias reduction on the predicted number of insurance policies is very modest.

4.2 The IMDB network

The Internet Movie Database (IMDB, https://www.imdb.com/) is the primary online database providing extensive information about movies, television series, *etc*. Among other things, it facilitates the construction of the network of connectedness between actors. We have used the 'igraph' package in R (Csárdi *et al*., 2024) to analyze the 'actor2' network data file from the Koblenz Network Collection (http://konect.cc/networks/actor2).

The graph for the actors' network for the full IMDB has 896,302 vertices, or nodes. The nodes represent the actors and movies; and a branch (or edge) denotes that an actor played in a movie. The number of branches associated with a node is termed its "degree", and the degrees for the vertices in the full network range from 1 to 1,590. The median degree is 2.0, and the mean of these degrees is 8.440.

The IMDB has been considered by many researchers to be a scale-free network. For example, see Barabási and Albert (1999), Gallos *et al.* (2013). In this case the degrees of the vertices will follow a zeta distribution. Accordingly, we obtain MLE's for the scaling parameter, 's', of $\tilde{s}$ = 1.6000057250, $\hat{s}$ = 1.6000049887, and $\check{s} = 1.5999910446$, with an asymptotic standard error of 0.0006501551 that applies for each of these estimates. Sub-samples of the network data, beginning with the smallest number of degrees, have also been analyzed. The results appear in Table 4. There, we see that the estimates for the scaling parameter lie between 2 and 3, and correcting for the bias in the MLE for this parameter once again has only a modest effect numerically. Moreover, this effect is not statistically significant (on the basis of the reported a.s.e.).

5. Conclusions

The zeta distribution has important connections in pure mathematics, and it arises in a wide range of statistical applications. Its single parameter plays a crucial role in the context of power laws, and so the properties of any estimator of this parameter are of special interest. The maximum likelihood estimator is a natural choice in view of its optimal large-sample properties. However, as if often the case, this estimator can have less desirable properties in small samples.

We have explored the relative bias and relative mean squared error of this maximum likelihood estimator in an extensive simulation study. In addition, two analytical techniques for reducing the order of magnitude of this bias have been investigates. While the widely used "corrective" approach of Cox and Snell (1968) performs very well, the "preventive" approach of Firth (1993) is superior. Indeed, our simulation results lead one to conjecture that Firth's bias correction produces an unbiased estimator in this case. This conjecture warrants further formal investigation. Importantly, the reduction in bias comes with a simultaneous reduction in percentage mean squared error.

Two applications with actual data illustrate the fact that the bias corrections need not necessarily alter the broad conclusions that are reached when the zeta distribution is employed.

Table 1: Percentage Biases and Mean Squared Errors

N	*s*	%Bias($\tilde{s}$)	%Bias($\hat{s}$)	%Bias($\breve{s}$)	%MSE($\tilde{s}$)	%MSE($\hat{s}$)	%MSE($\breve{s}$)
	1.25						
10		30.79	25.09	-0.04	10.46	7.01	0.52
25		20.17	18.44	-0.03	4.33	3.64	0.18
50		15.42	14.67	-0.03	2.49	2.26	0.08
100		12.14	11.80	-0.02	1.52	1.44	0.04
200		9.76	9.61	-0.01	0.97	0.94	0.02
300		8.65	8.55	-0.01	0.76	0.75	0.01
400		7.97	7.90	-0.01	0.65	0.63	0.01
500		7.49	7.43	-0.01	0.57	0.56	0.01
750		6.70	6.66	-0.01	0.45	0.45	0.01
1000		6.20	6.18	-0.01	0.39	0.39	0.00
	1.50						
25		14.41	12.22	-0.04	2.72	2.06	0.50
50		10.29	9.31	-0.03	1.33	1.12	0.24
100		7.59	7.13	-0.01	0.70	0.63	0.12
200		5.69	5.48	-0.00	0.38	0.36	0.06
300		4.84	4.70	-0.00	0.27	0.26	0.04
400		4.33	4.23	-0.00	0.22	0.21	0.03
500		3.98	3.90	-0.00	0.18	0.17	0.02
750		3.41	3.36	-0.00	0.13	0.13	0.02
1000		3.07	3.03	-0.00	0.11	0.10	0.01
	1.75						
25		11.43	8.73	-0.04	2.39	1.70	0.88
50		7.67	6.45	-0.04	1.05	0.84	0.42
100		5.33	4.76	-0.02	0.49	0.43	0.20
200		3.77	3.50	-0.01	0.24	0.22	0.10
300		3.09	2.91	-0.01	0.16	0.15	0.07
400		2.70	2.57	-0.01	0.12	0.11	0.05
500		2.44	2.33	-0.00	0.10	0.09	0.04
750		2.02	1.95	-0.00	0.07	0.06	0.03
1000		1.77	1.72	-0.00	0.05	0.05	0.02
	2.00						
50		6.24	4.76	-0.04	1.05	0.84	0.60
100		4.11	3.41	-0.02	0.47	0.41	0.29
200		2.76	2.42	-0.01	0.22	0.20	0.14
300		2.19	1.97	-0.01	0.14	0.13	0.09
400		1.88	1.71	-0.01	0.11	0.10	0.07
500		1.66	1.53	-0.00	0.08	0.08	0.06
750		1.34	1.25	-0.00	0.06	0.05	0.04
1000		1.15	1.09	-0.00	0.04	0.04	0.03

Table 1 (*continued*): Percentage Biases and Mean Squared Errors

N	*s*	%Bias($\tilde{s}$)	%Bias($\hat{s}$)	%Bias($\check{s}$)	%MSE($\tilde{s}$)	%MSE($\hat{s}$)	%MSE($\check{s}$)
	2.25						
50		5.44	3.65	-0.06	1.17	0.92	0.78
100		3.41	2.57	-0.03	0.52	0.45	0.38
200		2.19	1.79	-0.01	0.24	0.22	0.19
300		1.70	1.43	-0.01	0.15	0.14	0.12
400		1.42	1.23	-0.01	0.11	0.11	0.09
500		1.24	1.09	-0.01	0.09	0.09	0.07
750		0.97	0.87	-0.01	0.06	0.06	0.05
1000		0.83	0.75	-0.00	0.04	0.04	0.04
	2.50						
50		5.03	2.90	-0.05	1.38	1.08	0.98
100		3.01	2.02	-0.03	0.59	0.51	0.47
200		1.86	1.38	-0.01	0.27	0.25	0.23
300		1.40	1.09	-0.01	0.18	0.16	0.15
400		1.16	0.93	-0.01	0.13	0.12	0.12
500		1.00	0.81	-0.01	0.10	0.10	0.09
750		0.77	0.64	-0.01	0.07	0.07	0.06
1000		0.64	0.55	-0.00	0.05	0.05	0.05
	2.75						
50		4.90	2.36	-0.04	1.65	1.26	1.20
100		2.82	1.65	-0.02	0.70	0.61	0.58
200		1.67	1.11	-0.01	0.32	0.29	0.28
300		1.24	0.87	-0.01	0.20	0.19	0.19
400		1.00	0.73	-0.01	0.15	0.14	0.14
500		0.86	0.64	-0.01	0.12	0.11	0.11
750		0.65	0.50	-0.01	0.08	0.08	0.07
1000		0.53	0.42	-0.00	0.06	0.06	0.05
	3.00						
100		2.77	1.37	-0.02	0.82	0.70	0.68
200		1.57	0.92	-0.01	0.37	0.34	0.33
300		1.14	0.71	-0.01	0.24	0.22	0.22
400		0.92	0.59	-0.01	0.17	0.17	0.16
500		0.78	0.52	-0.00	0.14	0.13	0.13
750		0.58	0.41	-0.00	0.09	0.09	0.09
1000		0.46	0.33	-0.00	0.07	0.07	0.07

Table 1: (*continued*): Percentage Biases and Mean Squared Errors

N	***s***	**%Bias($\tilde{s}$)**	**%Bias($\hat{s}$)**	**%Bias($\breve{s}$)**	**%MSE($\tilde{s}$)**	**%MSE($\hat{s}$)**	**%MSE($\breve{s}$)**
	3.25						
100		2.75	1.15	-0.01	0.96	0.81	0.80
200		1.52	0.76	-0.02	0.43	0.39	0.38
300		1.09	0.59	-0.01	0.27	0.26	0.25
400		0.86	0.49	-0.02	0.20	0.19	0.19
500		0.72	0.43	-0.01	0.16	0.15	0.15
750		0.51	0.33	-0.00	0.10	0.10	0.10
1000		0.43	0.28	-0.00	0.08	0.08	0.07
	3.50						
200		1.54	0.65	-0.02	0.49	0.45	0.44
300		1.09	0.51	-0.01	0.32	0.30	0.29
400		0.86	0.49	-0.02	0.20	0.19	0.19
500		0.72	0.43	-0.01	0.16	0.15	0.15
750		0.53	0.33	-0.00	0.10	0.10	0.10
1000		0.41	0.24	-0.00	0.09	0.09	0.09
	3.75						
200		1.59	0.55	-0.02	0.57	0.51	0.51
300		1.10	0.44	-0.01	0.36	0.34	0.34
400		0.87	0.37	-0.01	0.27	0.25	0.25
500		0.72	0.32	-0.00	0.21	0.20	0.20
1000		0.41	0.21	-0.00	0.10	0.10	0.10
	4.00						
200		1.70	0.48	-0.03	0.67	0.59	0.58
300		1.17	0.38	-0.01	0.42	0.39	0.38
400		0.90	0.32	-0.01	0.31	0.29	0.29
500		0.74	0.28	-0.00	0.24	0.23	0.23
750		0.53	0.22	-0.00	0.16	0.15	0.15
1000		0.41	0.19	-0.00	0.12	0.11	0.11
	4.25						
300		1.26	0.35	-0.01	0.48	0.44	0.44
400		0.96	0.28	-0.01	0.35	0.33	0.32
500		0.78	0.25	-0.00	0.28	0.26	0.26
750		0.55	0.19	0.00	0.18	0.17	0.17
1000		0.47	0.16	0.00	0.13	0.13	0.13
	4.50						
300		1.38	0.30	-0.00	0.56	0.50	0.49
400		1.04	0.25	-0.01	0.40	0.37	0.37
500		0.85	0.22	0.00	0.31	0.29	0.29
750		0.58	0.17	-0.00	0.20	0.19	0.19
1000		0.45	0.14	0.00	0.15	0.14	0.14

Table 2: Maximum Likelihood Estimates*

Age	N	$\tilde{s}$	$\hat{s}$	$\check{s}$
17.5	39	4.2039 (0.7477)	3.9335	3.9547
22.5	61	4.7375 (0.7664)	4.4738	4.4943
27.5	111	3.8970 (0.176))	3.8231	3.8248
32.5	275	3.4593 (0.1913)	3.4387	3.4259
37.5	328	3.3143 (0.1610)	3.2991	3.2734
42.5	361	3.1295 (0.1376)	3.1178	3.0921
47.5	282	2.8260 (0.1252)	2.8148	2.7656
52.5	239	3.0844 (0.1614)	3.0681	3.0377
57.5	140	2.7460 (0.1716)	2.7252	2.7171
62.5	80	3.5002 (0.3629)	3.4269	3.4287
67.5	45	2.6701 (0.2866)	2.6101	2.6112
72.5	38	2.3970 (0.2520)	2.3438	2.3445
All	1,999	3.1616 (0.0596)	3.1594	3.1592

*Note: Asymptotic standard errors appear in parentheses, and they apply to each of the three associated MLE's.

Table 3: Actual and Predicted Policy Numbers, All Age Groups*

x = No. of Policies	**Actual**	$\tilde{P}$	$\hat{P}$	$\breve{P}$
1	1,695	1,704.29	1,703.77	1,703.70
2	207	190.46	190.69	190.72
3	46	52.85	52.96	52.98
4	22	21.28	21.34	21.35
5	9	10.51	10.55	10.55
6	8	5.91	5.93	5.93
7	4	3.63	3.64	3.64
8	3	2.38	2.39	2.39
9	1	1.64	1.65	1.65
10	1	1.17	1.18	1.18
11	2	0.87	0.87	0.87
12	0	0.66	0.66	0.66
13	1	0.51	0.52	0.52
14	0	0.41	0.41	0.41
15	0	0.33	0.33	0.33
16	0	0.27	0.27	0.27
Total:	1,999	1,997.17	1,997.16	1,997.15

*Note: The predictions based on the MLE, Cox-Snell estimator, and Firth estimator of s are denoted $\tilde{P}, \hat{P}$, and $\breve{P}$ respectively.

Table 4: Summary of IMDB analysis for various sample sizes

***N*:**		**50**	**100**	**250**	**500**	**896,302**
Degrees						
	Min.	1	1	1	1	1
	Max.	16	16	47	56	1,590
	Median	1	2	1	1	2
	Mean	2.000	2.083	2.041	2.028	8.440
$\tilde{s}$		2.310	2.091	2.324	2.280	1.600
$\hat{s}$		2.273	2.076	2.316	2.276	1.600
$\check{s}$		2.237	2.067	2.294	2.265	1.600
(a.s.e.)		(0.204)	(0.120)	(0.093)	(0.063)	(6.5×10^{-4})

References

Apostol, T. M., 1985. Formulas for higher derivatives of the Riemann Zeta function. *Mathematics of Computation*, 4, 223-232.

Barabási, A-L. and R. Albert, 1999. Emergence of scaling in random networks. *Science*, 286, 509-512.

Brent, R., 1973. *Algorithms for Minimization Without Derivatives*. Prentice-Hall, Englewood Cliffs, NJ.

Broido, A. D. and A. Clauset, 2019. Scale-free networks are rare. *Nature Communications*, 10:1017.

Choudhury, B. K., 1995. The Riemann Zeta-function and its derivatives. *Royal Society Proceedings: Mathematical and Physical Sciences*, 450, 477-499.

Clauset, A., C. R. Shalizi, and M. E. J. Newman, 2009. Power-law distributions in empirical data. *SIAM Review*, 51, 661-703.

Cordeiro, G. M. and R. Klein, 1994. Bias correction in ARMA models. *Statistics and Probability Letters*, 19, 169–176.

Cordeiro, G. M. and P. McCullagh, 1991. Bias correction in generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53, 629-643.

Cox, D. R. and E. J. Snell, 1968. A General definition of residuals. *Journal of the Royal Statistical Society. Series B*, 30, 248–275.

Csárdi, G., T. Nepusz, V. Traag, S. Horvát, F. Zanini, D. Noom, K. Müller, M. Salmon, and M. Antonov, 2024. Package 'igraph': Network analysis and visualization. https://r.igraph.org/.

Dai, J., Z. Huang, M. R. Powers, and J. Xu, J., 2021a. Heavy-tailed loss frequencies from mixtures of negative binomial and Poisson counts. Working paper, Tsinghua University School of Economics and Management.

Dai, J., Z. Huang, M. R. Powers, and J. Xu, 2021b. Characterizing the zeta distribution via continuous mixtures. *arXiv*:2008.06200v4.

Demirtas, H., A. Allozi, and R. Gao, 2021. Package 'UnivRNG': Univariate pseudo-random number generation. https://cran.r-project.org/web/packages/UnivRNG/UnivRNG.pdf.

Devianto, D., H. Yozzo, and Maiyastri, 2019. Characterization of Riemann zeta distribution. *Journal of Physics: Conference Series*, 1317.

Doray, L. G. and Arsenault, M., 2002. Estimators of the regression parameters of the Zeta distribution. *Insurance: Mathematics and Economics*, 30, 439-450.

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

Firth, D., 1993. Bias reduction of maximum likelihood estimates. *Biometrika*, 80, 27-38.

Fujita, T. and N. Yoshida, 2023. On further application of the zeta distribution to number theory. *Research in Number Theory*, 9:81.

Gabaix, X. and R. Ibragimov, 2011. A simple way to improve the estimation of tail exponents. *Journal of Business and Economic Statistics*, 29, 24-39.

Gallos, L. K., F. Q. Potiguar, J. S. Andrade Jr., and H. A. Makse, 2013. IMDB network revisited: Unveiling fractal and modular properties from a typical small-world network. *PLOS One*, 8(6): e66443.

Giles, D. E., 2021. Improved maximum likelihood estimation for the Weibull distribution under length-biased sampling. *Journal of Quantitative Economics*, 19, 59-77.

Giles, D. E., H. Feng, and R. T. Godwin, 2016. Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. *Communications in Statistics – Theory and Methods*, 45, 2465-2483.

Haight, F. A., 1966, Some statistical problems in connection with word association data. *Journal of Mathematical Psychology*, 3, 217-233.

Kosmidis, I. and D. Firth, 2009. Bias reduction in exponential family nonlinear models. *Biometrika,* 96, 793-804.

Lin, D. G. and C-Y. Hu, 2001. The Riemann Zeta distribution. *Bernoulli,* 7, 817-828.

Özkural, E., 2018. Zeta distribution and transfer learning problem. arXiv.1806.08908v1 [cs.AI].

Peltzer, A., 2019. The Riemann zeta distribution. Ph.D. Dissertation, Department of Mathematics, University of California, Irvine.

Schwartz, J., R. T. Godwin, and D. E. Giles, 2013. Improved maximum likelihood estimation of the shape parameter in the Nakagami distribution. *Journal of Statistical Computation and Simulation*, 83, 434-445.

Seal, H. L., 1947. A probability distribution of deaths at age x when policies are counted instead of lives. *Scandinavian Actuarial Journal*, 1947, 18-43.

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

Walther, A., 1926. Anschauliches zur Riemannschen Zetafunktion. *Acta Mathematica*, XLVIII, 393-400.

Yee, T. and C. Moler, 2022. Package 'VGAM': Vector generalized linear and additive models. https://cran.r-project.org/web/packages/VGAM/VGAM.pdf

Young, D. S., 2020. Package 'tolerance': Statistical tolerance intervals and regions.

https://cran.r-project.org/web/packages/tolerance/tolerance.pdf