Improved Maximum Likelihood Estimation

for the Zeta Distribution

David E. Giles

Department of Economics

University of Victoria

Revised, January, 2026

Abstract

We consider the small-sample properties of the maximum likelihood estimator for the shape 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 reduced significantly. Some illustrative results also show that Firth's bias correction is superior to bootstrap or jackknife bias correction in the context of the zeta distribution. The impact of the Cox-Snell and Firth bias correction procedures is illustrated in three applications with real-life data.

Key Words Zeta distribution; maximum likelihood estimation; bias reduction

Author contact dgiles@uvic.ca ; +64-22-155 3406

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

Disclosure Statement: *The author reports there are no competing interests to declare*.

Acknowledgement: I am grateful to the two referees whose helpful and insightful comments on earlier versions of this paper led to some major improvements in its presentation.

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 zeta distribution is "heavy-tailed", and the m'th. raw moment of X is given by

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

where 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 p.m.f. 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, forensics, 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 roles of the zeta distribution in certain artificial intelligence problems, and in the analysis of forensic data, have been discussed by Özkural (2018), and Coulson *et al*. (2011), respectively. 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 shape 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 analytic bias reduction techniques, are discussed in detail in the next section. Section 3 presents the results of a Monte Carlo simulation experiment that investigates the effectiveness of the bias reduction strategies, and provides a clear preference for a particular approach. This preference is justified further in Appendices 1 and 2. Three illustrative empirical applications involving real data are discussed in section 4; and section 5 offers some concluding remarks.

2. Estimation issues

2.1 Alternative estimators

Various estimators of the shape parameter, s, in (1) can be considered. 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. Maximum likelihood estimation provides a natural estimator for the shape parameter, but it is not without shortcomings.

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. This common weakness of the MLE is augmented in the case of the zeta distribution because of the latter's "heavy-tail" characteristic. This feature of the distribution implies that extreme sample values or outliers can arise with relatively high probability. This impacts the likelihood function and the properties of the MLE, which has led to the use of various modified and potentially more robust estimators for heavy-tail distributions.

It is important to note that these alternative inferential methods also affect the sampling properties of the estimators, such as small-sample bias and efficiency. Some recent examples of these include "regularized" or shrinkage estimation (as used by Beirlant *et al.*, 2019); and penalized likelihood, as applied to the heavy-tail three-parameter Weibull distribution by da Silva *et al*. (2025). Bayesian methods are also appropriate, as is illustrated for the zeta, and other, distributions by Carcassi and Szymanik (2022)

2.2 Analytic bias reduction

Notwithstanding the relevance of these other estimators, this paper deals with the standard MLE for the zeta distribution's shape parameter, with the emphasis on bias reduction, without sacrificing small-sample efficiency. We measure the latter in terms of Mean Squared Error (MSE), recognizing that bias reduction may increase an estimator's variance, and the effect on MSE is often ambiguous.

To this end we focus primarily on two *analytic* approaches – the "corrective" procedure proposed by Cox and Snell (1968); and the "preventive" procedure suggested by Firth (1993). Both of these approaches have been used extensively in the context of continuous distributions. 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), and the many other studies cited therein. This paper adds to the more limited literature associated with bias-reduced MLEs for *discrete* distributions. Examples of the latter include the contributions of Saha and Paul (2005), Chen and Giles (2011), Giles and Feng (2011), Schwartz and Giles (2016), Rasekhi and Hamedani (2020), Pagui *et al.* (2022), and Giles (2024).

Other modifications of the MLE that can be used for the purposes of bias reduction include the jackknife, the (somewhat related) analytic approach of Schucany *et al*. (1971), and the bootstrap. These are discussed and illustrated briefly in Appendix 2, where it becomes clear why they are not given more attention in this paper.

In what follows in the development of the Cox-Snell and Firth estimators, we need the following higher derivatives of (4):

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

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

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

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

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, $|K|^{½}$ is Jeffreys' invariant prior for s. In our case, $-\sum_{i=1}^{n}\log(x_i)$ is a sufficient statistic,

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

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 estimator, $\breve{s}$. Like the Cox-Snell estimator, $\breve{s}$ is unbiased to $O(N^{-2})$. As we discuss in Appendix 1, the structure of (16) as a function of *s* precludes this estimator from being *exactly* unbiased in finite samples. Finally, from (14), it follows that $\breve{s}$ is also the Bayes estimator under a zero-one loss function, which relates to the discussion in section 2.1.

3. A simulation experiment

3.1 Computational details

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 4.25. This covers the range of estimated values usually encountered in empirical applications using this distribution. For example, see Clauset *et al.* (2009). 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 $\check{s}$. The R code written for the simulation experiments and for the applications in section 4 can be downloaded from https://github.com/DaveGiles1949/r-code. It should be noted that the R package, 'mle.tools' (Mazuchelli, 2017) can also be used to compute the Cox-Snell bias-adjusted estimator. Indeed, Mazuchelli *et al*. (2017) illustrate this for 31 distributions, and replicate the results from a number of studies. However, this package uses the R function, 'D', to compute the required derivatives of the likelihood function numerically, and it does not provide analytic an expression for the bias of the MLE in the manner of this paper. Further, 'mle.tools' does not compute Firth's estimator. In addition, Stosic and Cordeiro (2009) provide both 'Mathematica' and 'Maple' code to compute the Cox-Snell estimator for 22 two-parameter distributions.

3.2 Simulation results

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 MSEs for the original MLE of *s*, as well as the two bias-reduced alternative estimators. Ideally, simultaneous reductions in both bias and MSE might be achieved. This can arise even if the variance of the bias-adjusted estimator exceeds that of the MLE, provided that the increase in variance doe not exceed the square of the bias reduction.

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

$$\%Bias(s^*) = 100\left[\frac{1}{NREP}\left(\sum_{i=1}^{NREP} s_i^*\right) - s\right]/s \tag{18}$$

and

$$\%MSE(s^*) = 100\left[\frac{1}{NREP}\sum_{i=1}^{NREP}(s_i^* - s)^2\right]/s^2. \tag{19}$$

The following discussion is based on the full set of simulation results that can be downloaded from https://github.com/DaveGiles1949/Results. There, simulations based on values of the shape parameter in the range [1.25 (0.25) 4.50] are reported. These results are summarized in Figures 1 and 2, and 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 and Figure 1 show that the (positive) percentage bias of the MLE of the shape parameter, s, decreases as the true value of that parameter increases. Moreover, we see in that table (and in Figure 2) that 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 in Table 1 that $\check{s}$ has negligible bias, even for very small samples (which is why this percentage bias is omitted from Figure 1). However, as we discuss in Appendix 1, Firth's modified estimator is not *exactly* unbiased for the zeta distribution. Firth's procedure also results in a dramatic reduction in MSE, compared with both the original MLE of s and its Cox-Snell counterpart in small samples. Of course, all three estimators are weakly consistent, so (for any value of s), both the percentage bias and percentage MSE (and the variance) decrease as N increases, and converge across the estimators.

From (18) and (19), $\%Var.(s^*) = \%MSE(s^*) - \frac{[\%Bias(s^*)]^2}{100}$, for an arbitrary estimator, s^*. Figure 3 illustrates the effect of bias reduction on the small-sample percentage variances of the Cox-Snell and Firth estimators. When s = 1.25, $\%Var.(\check{s}) < \%Var.(\hat{s}) < \%Var.(\tilde{s})$, for all N. For larger value of s the reductions in percentage variance are essentially the same for both the Firth and Cox-Snell estimators, relative to that of the original MLE. Importantly, we see that the previously discussed MSE improvements are associated with reductions in *both* the biases and variances of $\hat{s}$ and $\check{s}$.

The computational costs associated with implementing Firth's preventive approach and the Cox-Snell corrective approach to bias reduction are very similar. So, our overall 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. Further support for this is provided in Appendix 1.

Table 1: Illustrative Monte Carlo experiment results*

(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
	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
	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
	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

***Note:** $\tilde{s}, \hat{s}$, and $\breve{s}$ denote the original MLE, the Cox-Snell adjusted estimator, and Firth's modified estimator, respectively.

Figure 1: Percentage biases of original and bias-adjusted MLEs

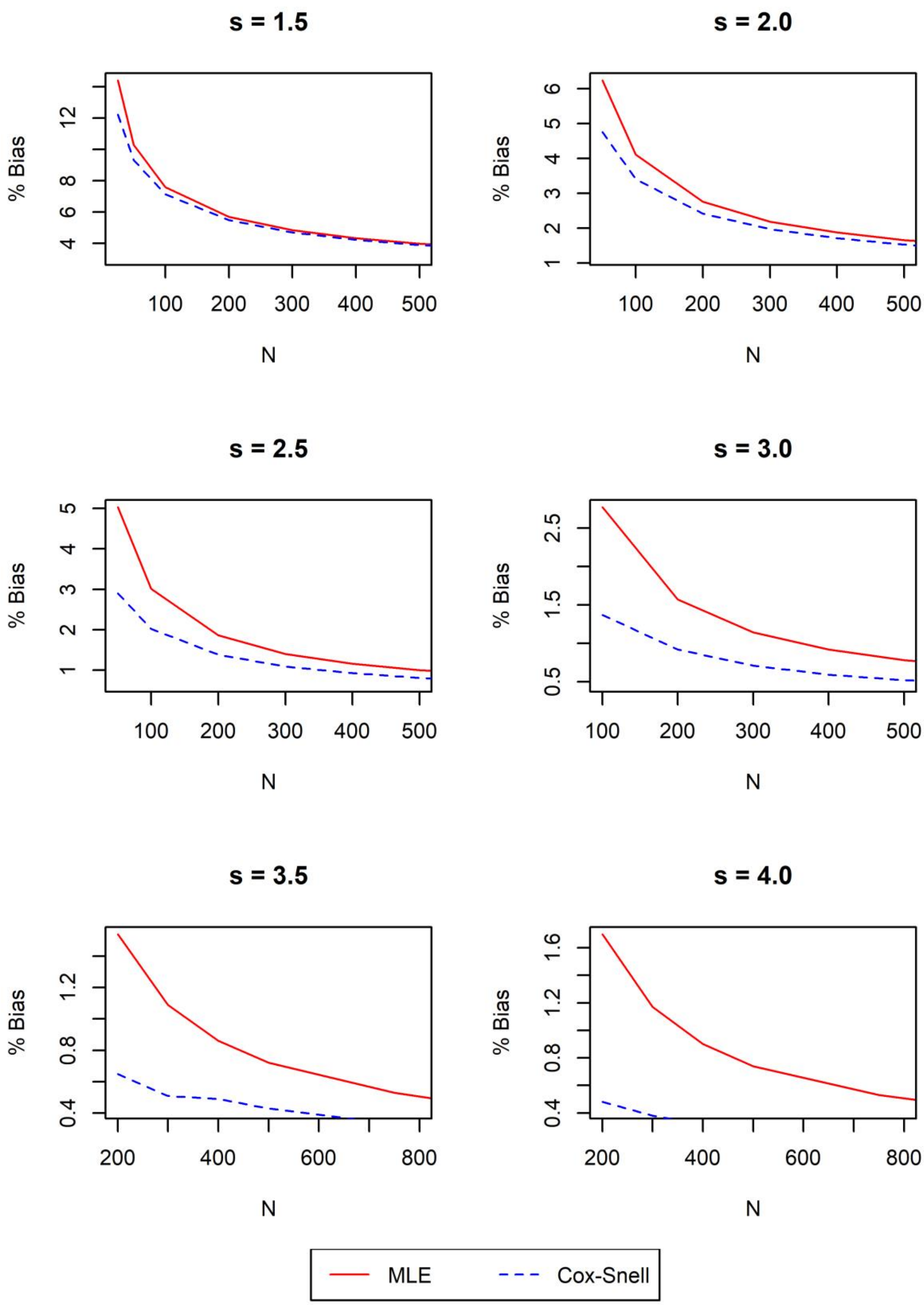


Figure 2: Percentage mean squared errors of original and bias-adjusted MLEs

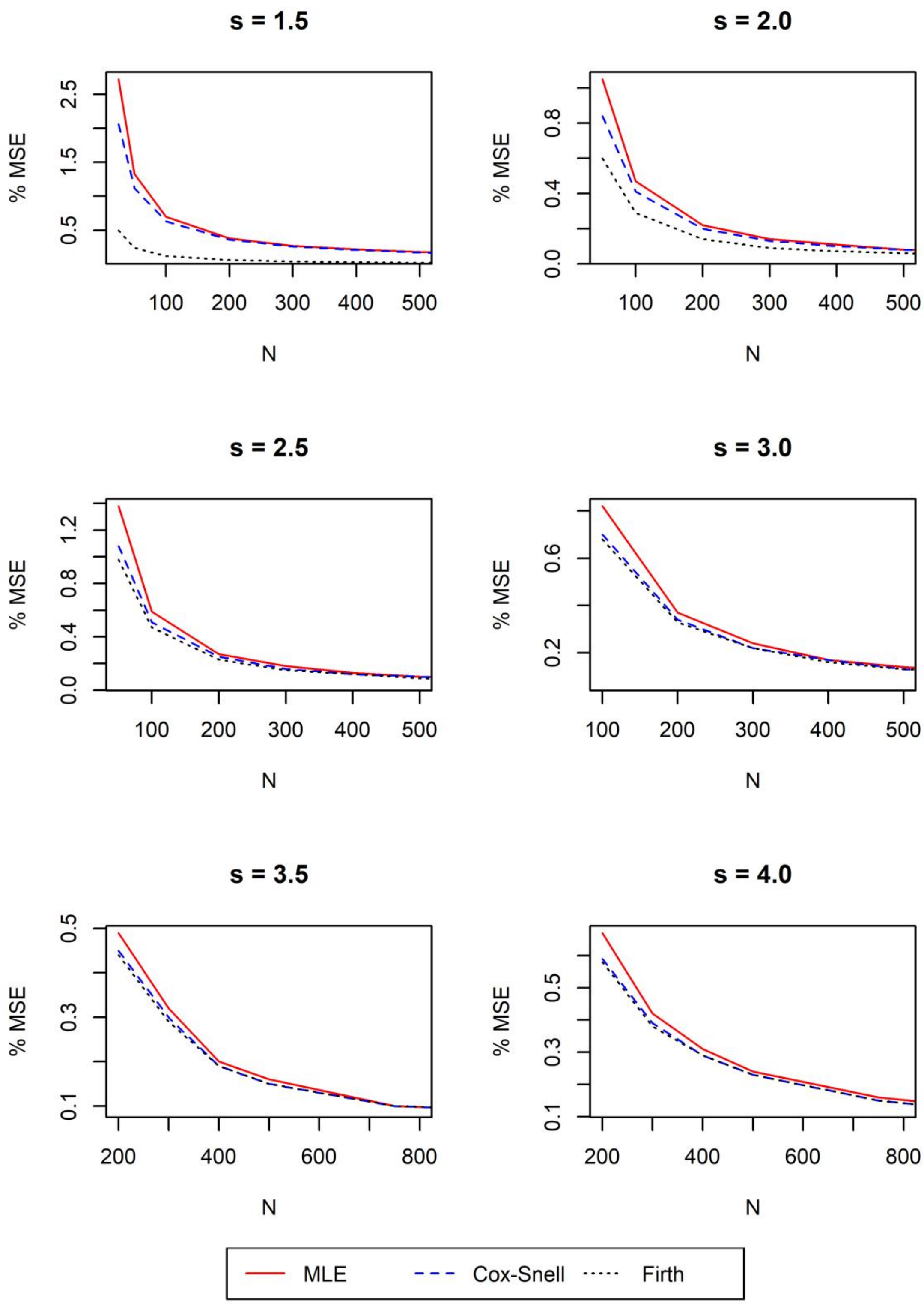
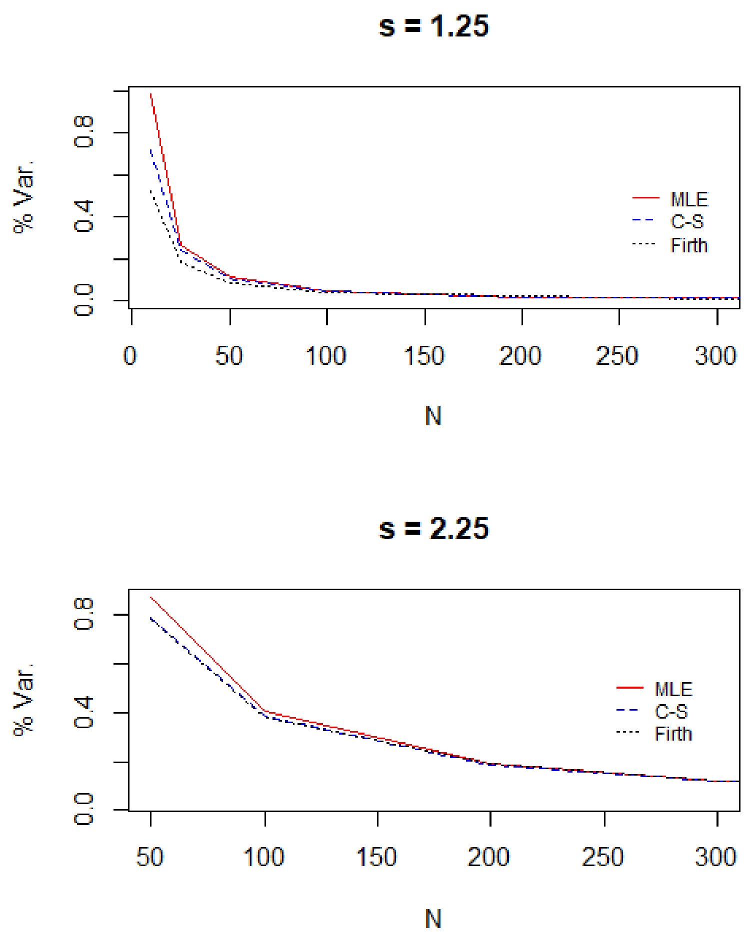


Figure 3: Percentage variances of original and bias-adjusted MLEs



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 very first to present the MLE for the shape 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 are of greater interest in the present context, having 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 expected, each of the bias corrections reduces the numerical value of the maximum likelihood estimates, to an extent depending on the sample size.

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 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. Our interest lies with (relatively) small samples, so we have analyzed sub-samples of the network data, beginning with the smallest number of degrees. Each of these sub-samples has minimum, maximum, and median values of 1, 16, and 2 degrees respectively. The resulting MLEs for the shape parameter of the zeta distribution results appear in Table 3. There, we see that the estimates for the shape parameter lie between 2 and 3, and correcting for the bias in the MLE for this parameter once again has a modest effect numerically. Comparing the base MLEs with their Firth bias-adjusted counterparts there are reductions in the values of the point estimates of 7.9%, 3.2%, and 0.8% for N = 25, 50, and 100 respectively. However, these reductions are not statistically significant (on the basis of the reported standard errors).

Table 2: Maximum likelihood estimates based on Seal's data*

Age (years)	***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

***Note:** Asymptotic standard errors appear in parentheses, and they apply to each of the three associated MLE's. (See Firth, 1993, p.36.)

Table 3: Maximum likelihood estimates based on the IMDB data*

N	**Mean degrees**	$\tilde{s}$	$\hat{s}$	$\check{s}$	**a.s.e.**
25	2.000	2.433	2.358	2.241	(0.323)
50	2.000	2.310	2.273	2.237	(0.204)
100	2.083	2.091	2.076	2.067	(0.120)

***Note:** Asymptotic standard errors (a.s.e.) appear in parentheses, and they apply to each of the three associated MLEs. (See Firth, 1993, p.36.)

4.3 Forensic data

Several recent studies in forensic science have focussed on the number of glass fragments found on the shoes and clothing of surveyed parties. For example, see Lau *et al.* (1997), Pettered *et al.* (1998), Ross and Nguyen (1998), Roux *et al.* (2005), Coulson *et al.* (2011)., and Jackson *et al.* (2013). This led Curran (2025) to create the 'fitPS' package in R to facilitate the use of the zeta distribution to analyze these integer data.

In Table 4 we report the MLEs and their bias-adjusted counterparts for the zeta distribution's scale parameter for four such data-sets. These samples, from Ross and Nguyen (1998) and Roux *et al*. (2001) have sizes that are pertinent to the present study. Comparing the Firth point estimates of the scale parameter with the basic MLE results, we see that in the examples with the largest (and smallest) sample sizes (N = 86, and N = 21), Firth's estimator results in decreases in the values of the point estimates of 6.0% and 3.2% respectively, although these reductions are not statistically significant.

Table 4: Maximum likelihood estimates based on forensic data*

Study	N	$\tilde{s}$	$\hat{s}$	$\check{s}$	**a.s.e.**
Ross *et al.*	86	5.682	5.298	5.341	(0.967)
Roux *et al.* (Fig. 3)	47	2.075	2.046	2.046	(0.168)
Roux *et al.* (Fig. 4)	21	2.105	2.037	2.038	(0.259)
Roux *et al.* (Fig. 5)	46	2.383	2.339	2.340	(0.226)

*Note: Asymptotic standard errors (a.s.e.) appear in parentheses, and they apply to each of the three associated MLEs. (See Firth, 1993, p.36.)

5 Conclusions

The zeta distribution has important connections in pure mathematics, and it arises in applications in a wide range of fields. 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. In, particular it exhibits positive bias.

We have explored the relative bias and relative mean squared error of this maximum likelihood estimator of the distribution's scale parameter in an extensive simulation study. In addition, two analytical techniques for reducing the order of magnitude of this bias have been investigated. While the widely used "corrective" approach of Cox and Snell (1968) performs well, the "preventive" approach of Firth (1993) is clearly superior. Indeed, the latter estimator has negligible bias even in very small samples, and the reduction in bias comes with a simultaneous reduction in percentage mean squared error, relative to the MLE itself. The superiority of Firth's procedure relative to the Cox-Snell correction in this context is in part due to the zeta distribution being a member of the linear exponential family, with s being its canonical parameter. It also reflects the negative effect that the shape of the (unobserved) bias function of the MLE has on the Cox-Snell correction, as is discussed in Appendix 1. Three applications with actual data relating to different disciplines illustrate the fact that the bias corrections that we have considered need not necessarily alter the broad conclusions that are reached when the zeta distribution is employed.

Appendix 1: Analysis of the bias of Firth's estimator

The results in Table 1 of the text indicate that Firth's modified estimator of the scale parameter of the zeta distribution has negligible percentage bias for values of s in the range [1.25, 4.25], even when N = 25 or N = 10. This is an important result, given the values of the scale parameter that are deemed to be relevant in empirical applications using this distribution. For instance, as noted already, Clauset *et al*. (2009) focus on values of s in the range 1.7 to 3.7, and the applications in the present study yield estimates which are essentially within the range 2.0 to 5.0. It is tempting to infer that Firth's modified estimator is *exactly* unbiased, and that the departures from zero for the biases reported in Table 1 are due simply to simulation "noise". However, this is not the case, as we illustrate in this Appendix.

Firth (1993, p.28) notes that if the score function associated with the original likelihood function is linear in the unknown parameter, then the MLE will be strictly unbiased. Extending this argument to the score function for Firth's modified likelihood, and for the case of the zeta distribution, as in equation (16), it follows that $\hat{s}$ will be strictly unbiased if the function,

$$f(s) = \left(\frac{\zeta^{(1)}(s)}{\zeta(s)}\right) + \frac{\left[\zeta(s)\zeta^{(3)}(s) - \zeta^{(1)}(s)\zeta^{(2)}(s)\right]}{2\left[\zeta(s)\zeta^{(2)}(s) - \left(\zeta^{(1)}(s)\right)^2\right]}$$

is linear in s. Recall that $\zeta(s)$ is the Riemann zeta function, defined for $s > 1$, and $\zeta^{(n)}(s)$ is its n'th. derivative.

Figure A.1 shows separate plots $f(s)$ for values of $s \in (1.0, 4.25]$, $s \in (1.25, 5.0]$, and $s \in [3.5, 5.0]$. The *non-linearity* of $f(s)$ is clear, although for some ranges of the shape parameter it is approximately satisfied. For example, see Figure A.1(c) where the fitted ordinary least squares (OLS) line (with standard errors in parentheses) is:

$$\widehat{f(s)} = \underset{(0.0027)}{-0.8177} + \underset{(0.0006)}{0.0782}\, s \qquad ; \quad n = 300 \; ; \quad \bar{R}^2 = 0.9806$$

In addition, Table A.1 presents Monte Carlo simulation results that extend Table 1 to cases where $s \in [1.04, 1.06]$, and $N \leq 50$. Again, 100,000 replications were used. There, we see that percentage biases of the order -0.55% to -0.60% are typical for Firth's estimator for s and N values in this range. Although Firth's adjustment of the MLE estimator is extremely effective for the zeta distribution, it does not produce an *exactly unbiased* estimator of the scale parameter under any situation that we have explored.

Figure A.1(a): f(s)

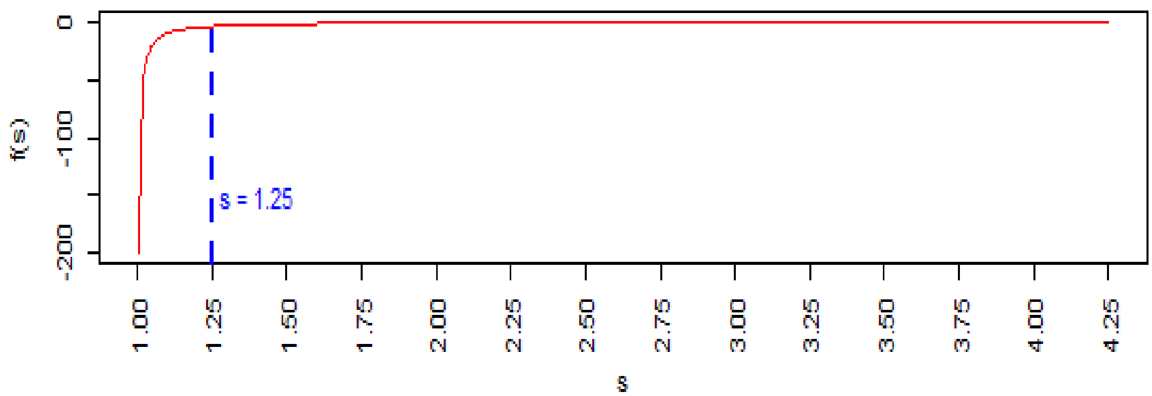


Figure A.1(b): f(s)

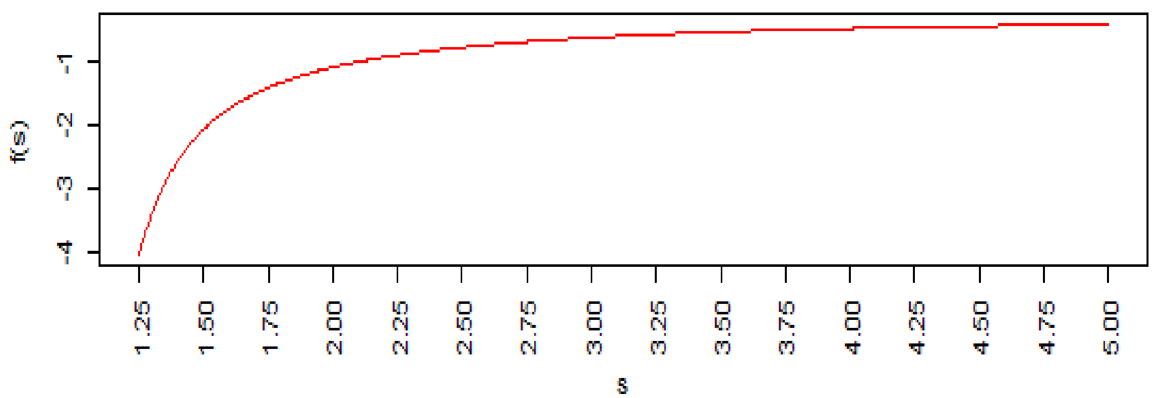


Figure A.1(c): f(s)

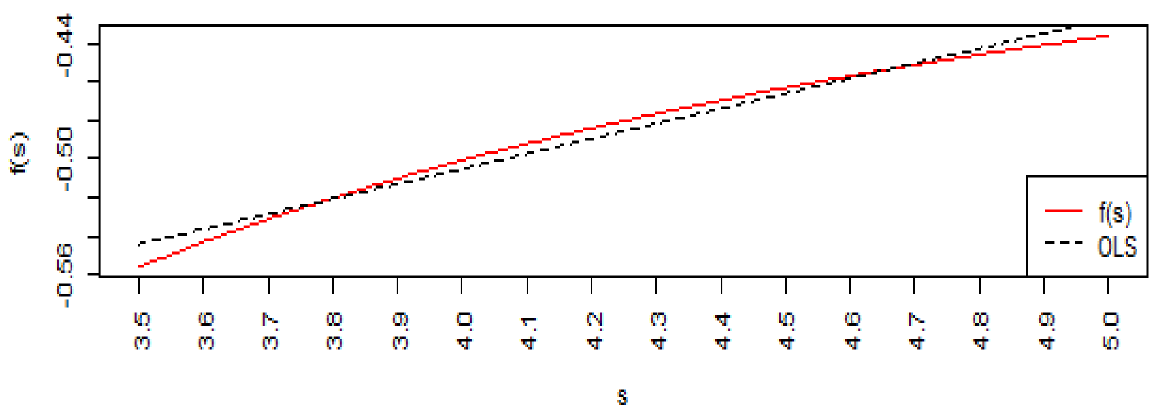


Table A.1: Monte Carlo experiment results for small values of *s* and *N*

(Percentage biases and mean squared errors)*

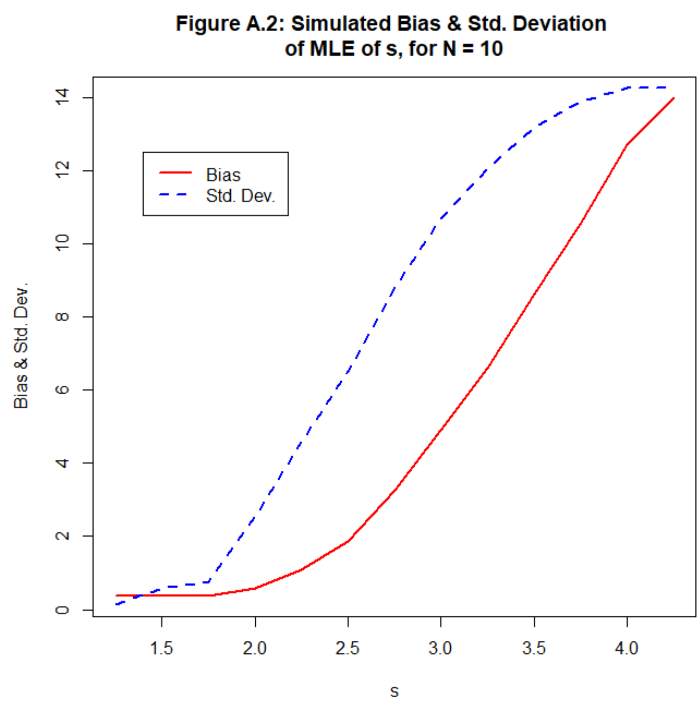
N	*s*	%Bias($\tilde{s}$)	%Bias($\hat{s}$)	%Bias($\breve{s}$)	%MSE($\tilde{s}$)	%MSE($\hat{s}$)	%MSE($\breve{s}$)
	1.040						
10		44.609	39.408	-0.520	19.937	15.559	0.015
25		30.794	29.353	-0.500	9.498	8.630	0.007
50		24.624	24.039	-0.493	6.071	5.786	0.005
	1.045						
10		44.056	38.863	-0.558	19.457	15.140	0.020
25		30.355	28.912	-0.542	9.233	8.376	0.009
50		24.256	23.669	-0.535	5.893	5.611	0.006
	1.050						
10		43.536	38.348	-0.579	19.014	14.751	0.025
25		29.944	28.500	-0.570	8.990	8.143	0.011
50		29.910	23.321	-0.566	5.729	5.449	0.007
	1.055						
10		43.026	37.841	-0.591	18.585	14.375	0.030
25		29.548	28.101	-0.588	8.759	7.922	0.013
50		23.580	22.988	-0.585	5.574	5.297	0.008
	1.060						
10		42.521	37.339	-0.593	18.167	14.008	0.036
25		29.175	27.724	-0.593	6.544	7.715	0.015
50		23.267	22.672	-0.592	5.429	5.155	0.009

***Note:** *N* = sample size; $\tilde{s}, \hat{s}$, and $\breve{s}$ denote the original MLE, the Cox-Snell adjusted estimator, and Firth's modified estimator, respectively.

The high performance of Firth's *preventive* approach to bias reduction in various situations is well documented. For example, see Kosmidis and Firth (2009), Zhang *et al.* (2021), and Stolte *et al.* (2024), among others. As noted in section 2.2, the zeta distribution is a member of the linear exponential family, and '*s*' is its canonical parameter. Such distributions motivated Firth's procedure, so the excellent performance of the latter for the zeta distribution is unsurprising.

It is often overlooked that *corrective* bias reductions, such as that of Cox and Snell, rely heavily on the implicit assumption that the (unobservable) bias function is "flat" with respect to the parameter(s). For example, see MacKinnon and Smith (1998) and Godwin and Giles (2019). Unless this condition is satisfied, corrective bias reduction is likely to be less successful than the preventive approach. More specifically, the former authors (p.229) conclude that even if the bias function is approximately linear (but not “flat”), a *corrective* bias correction may not work well. They also note that if the bias function slopes upward, bias-corrected estimator will have smaller variance than the original estimator, as was shown in Figure 3. Finally, they state that a corrective approach is likely to perform well if the bias is large relative to the variability of the estimator.

Figure A.2 depicts the simulated (total) bias and standard deviation of the MLE of s as a function of that parameter’s values, when N = 10, using 10,000 Monte Carlo replications. Taking into account the conclusions of MacKinnon and Smith, the shape of the bias function and the relative magnitudes of the bias and standard deviation explain why the *corrective* (Cox-Snell) bias adjustment is inferior to Firth’s *preventive* approach in our case.



Appendix 2: Other bias reduction methods

The parametric or non-parametric bootstrap is a common computational method to eliminate an estimator's bias of $O\ (N^{-1})$. Bootstrap bias-adjusted MLEs have been compared with the Cox- Snell and Firth estimators in various simulation studies. For example, see Cribari-Neto and Vasconcellos (2002), Xiao and Giles (2014), and Giles (2023) among others. The bootstrap estimator is unbiased to $O\ (N^{-2})$, but often this comes with a bias-variance trade-off that results in increased mean square error. In addition, the bootstrap has been found to *over-correct* the first-order bias of some MLEs - *e.g.*, see Cribari-Neto and Vasconcellos (2002), and Schwartz *et al*. (2013). In term of small-sample bias and efficiency, Giles *et al*. (2013, 2016) and Xiao and Giles (2014) find that the bootstrap estimator is inferior to the Cox-Snell estimator for the two-parameter Lomax, generalized Pareto and generalized Rayleigh distributions. Giles (2024) reaches the same conclusion when the performance of the bootstrap is compared with that of both the Firth and Cox-Snell estimators for the (discrete) Bell distribution.

Given these findings, and the computational burden associated with bootstrap bias reduction, we did not pursue this option rigorously in the present study. However, Table A.2 reports some illustrative simulation results for the small-sample properties of the (non-parametric) bootstrap bias-corrected MLE, $\tilde{s}_B$, obtained using the 'bootstrap' package in R (Tibshirani *et al*., 2025). The results are based on 1,000 bootstrap samples of size N, and to limit the computational cost this Monte Carlo experiment used just 10,000 replications. As our focus is on small-sample properties, Table A.2 is limited to values of 's' for which results based on $N \leq 200$ appear in Table 1 in the body of the paper. Comparing these results with their counterparts in Table 1 we see that while $\tilde{s}_B$ clearly dominates the Cox-Snell estimator in terms of bias (and slightly in terms of MSE), it is clearly out-performed by Firth's estimator.

The jackknife offers another familiar computational method to eliminate bias to $O\ (N^{-2})$. Some illustrative results for the "leave-one-out" jackknife biased-corrected MLE, $\tilde{s}_J$, are also reported in Table A.2. Again, the 'bootstrap' package in R (Tibshirani *et al*., 2025) was used, and the associated Monte Carlo experiment involved 10,000 replications. We see that $\tilde{s}_J$ dominates $\tilde{s}_B$ (and hence the Cox-Snell estimator) in terms of percentage bias, but the two estimators have similar percentage MSEs unless s is small in value. This reflects the fact that the variance of the former estimator generally exceeds that of the latter estimator in this study. Again, Firth's estimator clearly dominates overall in terms of both bias and efficiency, and in terms of its lower computational cost.

Finally, Schucany *et al.* (1971) illustrate how an exactly unbiased estimator can be constructed from two biased estimators whose biases (in general) are known. They also provide details of the connection between their estimator and the jackknife bias-corrected estimator. In our case we lack estimators whose biases are known, so the methodology of Schucany *et al.* cannot be implemented here. However, we have *estimates* of the biases of three estimator of the scale parameter, s. This provides an opportunity to construct a "pseudo estimator" by replacing the true biases with their estimates. We use this term because the constructed statistic requires knowledge of s itself in order to use the simulated biases from Table 1.

The value of considering this pseudo estimator here is that it provides a cross-check on the precision of the Monte Carlo simulation results reported in section 3. As Firth's estimator has negligible bias, we choose to use the MLE and the Cox-Snell estimator to construct the pseud estimator of s, as follows. Following Schucany *et al.* (1971, p. 524), define

$$R = R(s, N) = Bias(\tilde{s})/Bias(\hat{s}) = [\%Bias(\tilde{s})/\%Bias(\hat{s})]. \tag{A.2.1}$$

The statistic, $t = [\tilde{s} - R\hat{s}]/[1 - R]$ is an *exactly* unbiased estimator of s, but in our case the value of R is unknown. Let R^* be the estimator of R, obtained by replacing the true percentage biases in (A.2.1) by their estimates from Table 1, yielding the "pseudo estimator" of s, namely $\tilde{s}_S = [\tilde{s} - R^*\hat{s}]/[1 - R^*]$. This estimator will be *approximately* unbiased, with the degree of approximation reflecting the performance of the Monte Carlo experiment that generated the results in Table 1. Table A.2 includes illustrative percentage biases and MSEs of $\tilde{s}_S$, based on 100,000 Monte Carlo simulations. The negligible values of the biases suggest that R^* is approximately equal to R, and confirms the accuracy of the results reported in Table 1 and Figures 1 to 3.

Table A.2: Additional Monte Carlo experiment results

(Percentage biases and mean squared errors)*

N	***s***	**%Bias(** $\tilde{s}_B$ **)**	**%Bias(** $\tilde{s}_J$ **)**	**%Bias(** $\tilde{s}_S$ **)**	**%MSE(** $\tilde{s}_B$ **)**	**%MSE(** $\tilde{s}_J$ **)**	**MSE(** $\tilde{s}_S$ **)**
	1.25						
10		20.88	14.76	0.02	5.08	6.20	0.07
25		15.67	11.88	0.02	2.71	1.65	0.05
50		12.50	9.81	-0.01	1.67	1.07	0.03
100		10.06	8.14	0.01	1.06	0.71	0.02
200		8.19	6.80	0.01	0.69	0.48	0.01
	2.25						
50		2.63	1.62	0.02	0.89	0.86	0.64
100		1.84	1.14	0.01	0.40	0.41	0.33
200		1.21	0.75	-0.02	0.20	0.19	0.17
	3.25						
100		0.85	0.32	-0.01	0.76	0.82	0.74
200		0.51	0.20	-0.01	0.38	0.38	0.37

***Note:** *N* = sample size; $\tilde{s}_B$ and $\tilde{s}_J$ denote the bootstrap and jackknife bias-corrected estimators of 's'; and $\tilde{s}_S$ denotes the Schucany *et al*. "pseudo estimator".

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.

Beirlant, J., G. Maribe, and A. Verster, 2019. Using shrinkage estimators to reduce bias and MSE in estimation of heavy tails. *REVSTAT-Statistical Journal*, 17, 91-108

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.

Carcassi, F. and J. Szymanik, 2022. Heavy tails and the shape of modified numerals. *Cognitive Science*, 56(7), e13176.

Chen, Q. and D. E. Giles, 2011. Finite-sample properties of the maximum likelihood estimator for the Poisson regression model with random covariates. *Communications in Statistics – Theory and Methods*, 40, 1000-1014.

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.

Coulson, S. A, J. S. Buckleton, A. B. Gummer, and C. M. Triggs, 2011. Glass on clothing and shoes of members of the general population and people suspected of breaking crimes. *Science and Justice*, 41, 39–48.

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

Cribari-Neto, F. and K. L. P. Vasconcellos, 2002. Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*, 72, 107-118.

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/.

Curran, J., 2025. Package ‘fitPS’: Fit zeta distributions to forensic data.

da Silva, A., F. Quintino, F. Almeida, and D. Aguiar, 2025. Bias reduction of modified maximum likelihood estimates for a three-parameter Weibull distribution. *Entropy*, 27(5), 485.

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., 2024. The performance of the maximum likelihood estimator for the Bell distribution for count data. *Journal of Modern Applied Statistical Methods*, 23 (2), article 3.

Giles, D. E. and H. Feng, 2011. Reducing the bias of the maximum likelihood estimator for the Poisson regression model. *Economics Bulletin*, 31 (4), 2933-2943.

Giles, D. E., H. Feng, and R. T. Godwin, 2013. On the bias of the maximum likelihood estimator for the two- parameter Lomax distribution. *Communications in Statistics – Theory and Methods*, 42, 1934-1950.

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.

Godwin, R. T. and D. E. Giles, 2019. Improved analytic bias correction for maximum likelihood estimators. *Communications in Statistics – Simulation and Computation*, 48, 15-26.

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

Jackson F., P. Maynard, K. Cavanagh-Steer, T. Dusting, and C. Roux, 2013. A survey of glass found on the headwear and head hair of a random population *vs*. people working with glass. *Forensic Science International*, 226, 125–131.

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

Lau L., A. D. Beveridge, B. C. Callowhill, N. Conners, K. Foster, R. J. Groves, K. N. Ohashi, A. M. Sumner, and H. Wong, 1997. The frequency of occurrence of paint and glass on the clothing of high school students. *Canadian Society of Forensic Science Journal*, 30, 233–240.

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

MacKinnon, J. G. and A. A. Smith, Jr., 1998. Approximate bias correction in econometrics. *Journal of Econometrics*, 85, 205-230.

Mazuchelli, J., 2017. Package 'mle.tools': Expected/Observed Fisher Information and Bias-Corrected Maximum Likelihood Estimate(s). https://cran.r-project.org/web/packages/mle.tools/index.html

Mazuchelli, J., A. F. Berdusco, and S.Nadarajah, 2017. mle.tools: An R package for maximum likelihood bias correction. *R Journal*, 9, 268-290.

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

Pagui, E. C. K., A. Salvan, and N. Sartori, 2022. Improved estimation in negative binomial regression. *Statistics in Medicine*, 41, 2403-2416.

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

Petterd, C. I., I. McCallum, L. Bradford, K. Brinch, and S. Stewart, 1998. Glass partic les in the clothing of the general population in Canberra — A survey. In *Proceedings of the 14th International Symposium on the Forensic Sciences.*

Rasekhi, M. and G. Hamedani, 2020. Bias-corrected maximum-likelihood estimator for the parameter of the logarithmic series distribution and its characterizations. *Journal of Statistical Research of Iran*, 16, 1-11

Ross. P. and H. Nguyen, 1998. A survey of clothing for the presence of glass fragments. In *Proceedings of the 14th International Symposium on the Forensic Sciences*.

Roux, C., R. Kirk, S. Benson, T. van Haren, and C. I. Petterd, 2001. Glass particles in footwear of members of the public in south-eastern Australia—a survey. *Forensic Science International*, 116, 149-156.

Saha, K. and S. Paul, 2005. Bias-corrected maximum likelihood estimator of the negative binomial dispersion parameter. *Biometrics*, 61, 179-185.

Schucany, W. R., H. L. Gray and D. B. Owen, 1971. On bias reduction in estimation. *Journal of the American Statistical Association*, 66, 524- 533.

Schwartz, J. and D. E. Giles, 2016. Bias-reduced maximum likelihood estimation for the zero-inflated Poisson distribution. *Communications in Statistics – Theory and Methods*, 45, 465-478.

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.

Stolte, M., S. Herbrandt, and U. Ligges, 2024. A comprehensive review of bias reduction methods for logistic regression. *Statistics Surveys*, 18, 139-162.

Stosic, B. D. and G. M. Cordeiro, 2009. Using Maple and Mathematica to derive bias corrections for two parameter distributions. *Journal of Statistical Computation and Simulation*, 79, 751–767.

Tibshirani, R., F. Leisch, and S. Kostyshack, 2025. Package 'bootstrap': Functions for the book *An Introduction to the Bootstrap*. https://gitlab.com/scottkosty/bootstrap

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

Xiao, L.and D. E. Giles, 2014. Bias reduction for the maximum likelihood estimator of the generalized Rayleigh family of distributions. *Communications in Statistics – Theory and Methods*, 43, 1778-1792.

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

Zhang, X., S. Paul, and Y-G. Wang, 2021. Small sample bias correction or bias reduction? *Communications in Statistics – Simulation and Computation*, 50, 1165-1177.