

## **Bias Reduction for the Maximum Likelihood Estimator of the Parameters of the Generalized Rayleigh Family of Distributions**

**Xiao Ling & David E. Giles**

*Department of Economics, University of Victoria*

**November, 2011**

### **Abstract**

We derive analytic expressions for the biases, to  $O(n^{-1})$ , of the maximum likelihood estimators of the parameters of the generalized Rayleigh distribution family. Using these expressions to bias-correct the estimators is found to be extremely effective in terms of bias reduction, and generally results in a small reduction in relative mean squared error. In general, the analytic bias-corrected estimators are also found to be superior to the alternative of bias-correction *via* the bootstrap.

**Keywords**      Generalized Rayleigh distribution; maximum likelihood; bias; mean squared error; bias correction

**MSC Codes**    62F10; 62G09; 62P20

---

### **Contact author:**

David E. Giles, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: dgiles@uvic.ca; Phone: (250) 721-8540; FAX: (250) 721-6214

## 1. Introduction

In this paper we investigate the finite-sample bias of the maximum likelihood estimator for the parameters in the so-called Generalized Rayleigh family of distributions. This family was proposed by Voda (1976). The density function for the Generalized Rayleigh Distribution (GRD) takes the form:

$$f(x; \theta, k) = \frac{2\theta^{k+1}}{\Gamma(k+1)} x^{2k+1} \exp\{-\theta x^2\} \quad (1)$$

with  $x > 0$ ,  $\theta > 0$ , and  $k \geq 0$ . The GRD family includes several important probability distributions as special cases.

For example, if  $k = 0$  and  $\theta = 1/2\lambda^2$  we obtain the one-parameter Rayleigh distribution with density function

$$f(x; \lambda) = \frac{x}{\lambda^2} \exp\left\{-\frac{x^2}{2\lambda^2}\right\}, \quad x > 0, \quad \lambda > 0. \quad (2)$$

For  $k = 1/2$  and  $\theta = 1/2\lambda^2$  we obtain the one-parameter Maxwell distribution with the density function

$$f(x; \lambda) = \frac{2}{\lambda^3(2\pi)^{1/2}} x^2 \exp\left\{-\frac{x^2}{2\lambda^2}\right\} \quad x > 0, \quad \lambda > 0. \quad (3)$$

For  $k = ((a/2) - 1)$  and  $\theta = 1/2\tau^2$  we obtain the Chi distribution with “ $a$ ” degrees of freedom, whose density function is

$$f(x; \tau, a) = \frac{x^{a-1}}{2^{\frac{a}{2}} a^{-1} \tau^a \Gamma(\frac{a}{2})} \exp\left\{-\frac{x^2}{2\tau^2}\right\}, \quad x > 0, \quad a \in N, \quad \tau > 0 \quad (4)$$

where  $N$  denotes the set of natural numbers.

If we drop the positivity requirement for  $k$  and take  $k = -1/2$  and  $\theta = 1/2\sigma^2$ , we obtain the Half-Normal distribution with the density function

$$f(x; \sigma) = \frac{2}{\sigma(2\pi)^{1/2}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \quad x > 0, \quad \sigma > 0. \quad (5)$$

Clearly, the GRD family is quite broad and lends itself to widespread application. In the case of reliability modeling, for example, the GRD family is more flexible than the widely used Weibull model, as the latter includes only the Rayleigh distribution as a special case, while the GRD also encompasses the Maxwell and Chi distributions. In addition, other members of the GRD family, such as the Half-Normal distribution, are quite widely applied in the social sciences and elsewhere.

More precisely, consider the following examples of areas of application for each major distribution in the Generalized Rayleigh family. The Rayleigh distribution itself has had many applications in life testing (*e.g.*, electrovacuum devices; Polovko, 1968), and in communications engineering (Dyer and Whisenand, 1973). It also arises when wind speed is analyzed in terms of its orthogonal 2-dimensional vector components (Beenstock, 1995); and in the context of wave heights (*e.g.*, Naess, 1985; Battjes and Groenendijk, 2000); and the size of trapped prawns (Yamane, 1998). The Rayleigh distribution has also been used in economic applications. For example, Abernethy (1984) used it to model historical contract cost data; and Meagher *et al.* (2008) exploited the log-concavity of its p.d.f. to characterize the existence of a unique pure strategy equilibrium in the allocation-price duopoly game.

The Half-Normal distribution is an extension of the Normal distribution. Whenever a difference or deviation is measured and the algebraic sign is unknown, disregarded, lost, or otherwise absent, the resulting distribution of these absolute measurements can range in shape from Half-Normal to the normal distribution as the limit. Typical of this are the examples which arise in industrial practice, such as quality control (*e.g.*, Chou and Liu, 1998). Other applications of the Half-Normal distribution arise in the context of stochastic frontier production functions in economics (*e.g.*, Aigner *et al.*, 1977; Battese and Coelli, 1992).

The Chi distribution arises, for example, when a  $k$ -dimensional vector's orthogonal components are independent and each follow a standard normal distribution. The length of the vector will then have a Chi distribution. The

most familiar example is the Maxwell distribution of (normalized) molecular speeds which is a Chi distribution with 3 degrees of freedom. The Maxwell distribution provides a description of how the speed of a mixture of moving particles varies at a particular temperature. Accordingly, it is widely used in physics.

In this paper we consider maximum likelihood (ML) estimation of the parameters of the GRD family. ML estimation is usually motivated by its appealing large-sample asymptotic properties. Specifically, under the usual regularity conditions, it is consistent, asymptotically efficient, and asymptotically normal. However, it is well known that in finite samples the ML estimator need not possess any desirable sampling properties. In particular, the ML estimator is often biased. The determination of such bias can be complicated, as the likelihood equations (first-order conditions) that determine the maximum of the likelihood function are often highly non-linear, and do not possess a closed-form solution. This is the situation with the GRD family.

We focus on the methodology suggested by Cox and Snell (1968) to obtain analytic expressions for the bias, to  $O(n^{-1})$ , of the ML estimators. These expressions are then used to bias-correct the ML estimator, resulting in estimators that are unbiased to  $O(n^{-2})$ . The effectiveness of this bias correction, in terms of both bias reduction and its impact on mean squared error, is compared with that of a parametric bootstrap bias correction. We conclude that the analytic approach to dealing with the bias is computationally efficient and highly effective.

In the next section we summarize the Cox-Snell methodology. Section 3 applies this methodology to the Generalized Rayleigh family of distributions. A Monte Carlo simulation experiment that compares the Cox-Snell bias-adjusted estimators with the bootstrap bias-corrected estimators is discussed in section 4; and an illustrative empirical application using economic data is presented in section 5. Section 6 gives our conclusions.

## 2. Methodology

Let  $l(\theta)$  be the log-likelihood function based on a sample of  $n$  observations, and  $p$ -dimensional parameter vector,  $\theta$ . In what follows, we require that  $l(\theta)$  is regular with respect to all derivatives up to and including the third order. This condition is satisfied for the GRD family. We will require the joint cumulants of the derivatives of  $l(\theta)$  and the derivatives of certain of these cumulants. The joint cumulants of the derivatives of  $l(\theta)$  are denoted as follows:

$$k_{ij} = E \left( \frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right) \quad ; \quad i, j = 1, 2, \dots, p$$

$$k_{ijl} = E \left( \frac{\partial^3 l}{\partial \theta_i \partial \theta_j \partial \theta_l} \right) \quad ; \quad i, j, l = 1, 2, \dots, p \quad (6)$$

$$k_{ij,l} = E \left( \frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \frac{\partial l}{\partial \theta_l} \right) \quad ; \quad i, j, l = 1, 2, \dots, p. \quad (7)$$

In addition, the derivatives of the second-order cumulants are denoted:

$$k_{ij}^{(l)} = \frac{\partial k_{ij}}{\partial \theta_l} \quad ; \quad i, j, l = 1, 2, \dots, p. \quad (8)$$

All of the expressions in (6) and (7) are assumed to be  $O(n)$ . Cox and Snell (1968) extended earlier results of Bartlett (1953a, 1953b), Haldane (1953), Haldane and Smith (1956), Shenton and Wallington (1962), and Shenton and Bowman (1963), to show that for sample data that are independent (but not necessarily identically distributed) the bias of the  $s^{th}$  element of the MLE of  $\theta$  ( $\hat{\theta}$ ) is:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^p \sum_{j=1}^p \sum_{l=1}^p k^{si} k^{sj} \left( \frac{1}{2} k_{ijl} + k_{ij,l} \right) + O(n^{-2}), \quad s = 1, 2, \dots, p, \quad (9)$$

where  $k^{ij}$  is the  $(i, j)^{th}$  element of the inverse of the (expected) information matrix,  $K = \{-k_{ij}\}$ . Cordeiro and Klein (1994) noted that this bias expression also holds if the data are non-independent, provided that all of the  $k$

terms are  $O(n)$ , and showed that it can be re-written as:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^p k^{si} \sum_{j=1}^p \sum_{l=1}^p k^{sj} (k_{ij}^{(l)} - \frac{1}{2} k_{ijl}) k^{jl} + O(n^{-2}), \quad s = 1, 2, \dots, p. \quad (10)$$

The computational advantage of equation (10) is that it does not involve terms of the form defined in (7). Now, let  $a_{ij}^{(l)} = k_{ij}^{(l)} - \frac{1}{2} k_{ijl}$ , for  $i, j, l = 1, 2, \dots, p$ , and define the following matrices:

$$A^{(l)} = \{a_{ij}^{(l)}\}; \quad i, j, l = 1, 2, \dots, p \quad (11)$$

$$A = [A^{(1)} | A^{(2)} | \dots | A^{(p)}]. \quad (12)$$

Cordeiro and Klein (1994) showed that the expression for the  $O(n^{-1})$  bias of  $\hat{\theta}$  can be re-written as:

$$Bias(\hat{\theta}) = K^{-1} A \text{vec}(K^{-1}) + O(n^{-2}). \quad (13)$$

A “bias-corrected” MLE for  $\theta$  can then be obtained as:

$$\tilde{\theta} = \hat{\theta} - \hat{K}^{-1} \hat{A} \text{vec}(\hat{K}^{-1}), \quad (14)$$

where  $\hat{K} = (K)|_{\hat{\theta}}$  and  $\hat{A} = (A)|_{\hat{\theta}}$ . It can be shown that the bias of  $\tilde{\theta}$  will be  $O(n^{-2})$ .

It is crucial to note that (13) and (14) can be evaluated even when the likelihood equations do not admit a closed-form analytic solution, so that the MLE has to be obtained *via* a numerical algorithm. For this reason, this methodology is very useful for the Generalized Rayleigh distribution.

### 3. Bias Adjustment and the Generalized Rayleigh Distribution

Suppose  $X$  is a random variable following the Generalized Rayleigh distribution. From (1), the likelihood function, based on a sample of  $n$  independent observations, is:

$$L = \prod_{i=1}^n f(x_i; \theta, k) = \left( \frac{2\theta^{k+1}}{\Gamma(k+1)} \right) \left( \prod_{i=1}^n x_i \right)^{2k+1} \exp\{-\theta \sum_{i=1}^n x_i^2\}. \quad (15)$$

Then the corresponding log-likelihood function is:

$$l = n \log(2) + n(k+1) \log \theta - n \log \Gamma(k+1) + (2k+1) \sum_{i=1}^n \log x_i - \theta \sum_{i=1}^n x_i^2. \quad (16)$$

It follows immediately that:

$$\begin{aligned} \frac{\partial l}{\partial \theta} &= \frac{n(k+1)}{\theta} - \sum_{i=1}^n x_i^2 ; & \frac{\partial l}{\partial k} &= n \log \theta - n \left( \frac{1}{k} + \Psi(k) \right) + 2 \sum_{i=1}^n \log x_i ; \\ \frac{\partial^2 l}{\partial \theta^2} &= -\frac{n(k+1)}{\theta^2} ; & \frac{\partial^2 l}{\partial k^2} &= -n \left( -\frac{1}{k^2} + \Psi_{(1)}(k) \right) ; \\ \frac{\partial^3 l}{\partial \theta^3} &= \frac{2n(k+1)}{\theta^3} ; & \frac{\partial^3 l}{\partial k^3} &= -n \left( \frac{2}{k^3} + \Psi_{(2)}(k) \right) . \end{aligned} \quad (17)$$

Here,  $\Psi(k)$  is the usual digamma function:

$$\Psi(k) = d \log \Gamma(k) / dk = -\gamma - \sum_{j=1}^{\infty} \xi(j+3) (-k-1)^j,$$

where  $\xi(s) = \sum_{j=1}^{\infty} (n^{-s})$  is the Riemann zeta function, and  $\gamma = 0.57721\dots$  is the Euler-Mascheroni constant. In what follows we also need the trigamma and tetragamma functions, these being  $\Psi_{(i)}(k) = d^i \log \Gamma(k) / dk^i$ ;  $i = 1, 2$ .

Then, we need to determine the cross derivatives of the two parameters in the generalized Rayleigh distribution. These are:

$$\begin{aligned}
\frac{\partial^2 l}{\partial \theta \partial k} &= \frac{n}{\theta} ; & \frac{\partial^2 l}{\partial k \partial \theta} &= \frac{n}{\theta} ; \\
\frac{\partial^3 l}{\partial \theta^2 \partial k} &= -\frac{n}{\theta^2} ; & \frac{\partial^3 l}{\partial k^2 \partial \theta} &= 0 ; \\
\frac{\partial^3 l}{\partial \theta \partial k^2} &= 0 ; & \frac{\partial^3 l}{\partial k \partial \theta^2} &= -\frac{n}{\theta^2} ; \\
\frac{\partial^3 l}{\partial \theta \partial k \partial \theta} &= -\frac{n}{\theta^2} ; & \frac{\partial^3 l}{\partial k \partial \theta \partial k} &= 0 .
\end{aligned} \tag{18}$$

Based on the higher-order derivatives in (17) and (18), we note that there are no observations,  $x$ 's, i.e. we avoid the calculation of expectations. This shows us that the answers for the joint cumulants of the derivatives and the derivatives of cumulants shown in (6) and (8) are the same as the higher-order derivatives in (17) and (18) as follows:

$$\begin{aligned}
k_{11} &= -\frac{n(k+1)}{\theta^2} \\
k_{12} &= k_{21} = \frac{n}{\theta} \\
k_{22} &= -n \left( -\frac{1}{k^2} + \Psi_{(1)}(k) \right) \\
k_{111} &= k_{111}^{(1)} = \frac{2n(k+1)}{\theta^3} \\
k_{112} &= k_{121} = k_{211} = k_{12}^{(1)} = k_{21}^{(1)} = k_{11}^{(2)} = -\frac{n}{\theta^2} \\
k_{122} &= k_{212} = k_{221} = k_{22}^{(1)} = k_{12}^{(2)} = k_{21}^{(2)} = 0 \\
k_{222} &= k_{222}^{(2)} = -n \left( \frac{2}{k^3} + \Psi_{(2)}(k) \right) .
\end{aligned}$$

The information matrix is

$$K = \begin{pmatrix} \frac{n(k+1)}{\theta^2} & -\frac{n}{\theta} \\ -\frac{n}{\theta} & n \left( -\frac{1}{k^2} + \Psi_{(1)}(k) \right) \end{pmatrix},$$

and

$$A = \frac{n}{2} \begin{pmatrix} \frac{2(k+1)}{\theta^3} & -\frac{1}{\theta^2} & -\frac{1}{\theta^2} & 0 \\ -\frac{n}{\theta} & 0 & 0 & -\left( \frac{2}{k^3} + \Psi_{(2)}(k) \right) \end{pmatrix}.$$

So, to  $O(n^{-1})$ ,



$$\text{Bias}(\hat{\theta}) = \frac{\theta\{[2(k+1)(-\frac{1}{k^2} + \Psi_{(1)}(k)) - 3](-\frac{1}{k^2} + \Psi_{(1)}(k)) - (k+1)(\frac{2}{k^3} + \Psi_{(2)}(k))\}}{2n[(k+1)(-1/k^2 + \Psi_{(1)}(k)) - 1]^2}, \quad (19)$$

$$\text{Bias}(\hat{k}) = \frac{(k+1)[-1/k^2 + \Psi_{(1)}(k) - (k+1)(2/k^3 + \Psi_{(2)}(k))] - 2}{2n[(k+1)(-1/k^2 + \Psi_{(1)}(k)) - 1]^2}. \quad (20)$$

Bias-adjusted estimators are then obtained as

$$\tilde{\theta} = \hat{\theta} - \frac{\hat{\theta}\{[2(\hat{k}+1)(-\frac{1}{\hat{k}^2} + \Psi_{(1)}(\hat{k})) - 3](-\frac{1}{\hat{k}^2} + \Psi_{(1)}(\hat{k})) - (\hat{k}+1)(\frac{2}{\hat{k}^3} + \Psi_{(2)}(\hat{k}))\}}{2n[(\hat{k}+1)(-1/\hat{k}^2 + \Psi_{(1)}(\hat{k})) - 1]^2}, \quad (21)$$

$$\tilde{k} = \hat{k} - \frac{(\hat{k}+1)[-1/\hat{k}^2 + \Psi_{(1)}(\hat{k}) - (\hat{k}+1)(2/\hat{k}^3 + \Psi_{(2)}(\hat{k}))] - 2}{2n[(\hat{k}+1)(-1/\hat{k}^2 + \Psi_{(1)}(\hat{k})) - 1]^2}. \quad (22)$$

Recall that the Rayleigh distribution corresponds to the case where  $k = 0$  in (2). Several of the general joint cumulants of the derivatives or derivatives of cumulants that we use would be undefined in this case. Therefore, we need to re-calculate the bias-corrected estimator for the Rayleigh distribution by the methodology introduced by Cox and Snell, as illustrated in Section 2.1. The following results show how we construct the bias-corrected estimator in this special case. The log-likelihood function from (2) is:

$$l(x_i; \lambda) = -2n \log \lambda + \sum_{i=1}^n \log x_i - \frac{\sum_{i=1}^n x_i^2}{2\lambda^2}.$$

Then,

$$\begin{aligned} \frac{\partial l}{\partial \lambda} &= -\frac{2n}{\lambda} + \frac{\sum_{i=1}^n x_i^2}{\lambda^3}; & k_{11} &= -\frac{4n}{\lambda^2} \\ \frac{\partial^2 l}{\partial \lambda^2} &= \frac{2n}{\lambda^2} - \frac{3 \sum_{i=1}^n x_i^2}{\lambda^4}; & k_{111} &= \frac{20n}{\lambda^3} \\ \frac{\partial^3 l}{\partial \lambda^3} &= -\frac{4n}{\lambda^3} + \frac{12 \sum_{i=1}^n x_i^2}{\lambda^5}; & k_{11}^{(1)} &= \frac{\partial k_{11}}{\partial \lambda} = \frac{8n}{\lambda^3} \\ K &= -[k_{11}] = \frac{4n}{\lambda^2}; & K^{-1} &= \frac{\lambda^2}{4n} \\ a_{11} &= k_{11}^{(1)} - 0.5k_{111} = -\frac{2n}{\lambda^3}. \end{aligned}$$

So,

$$\text{Bias}(\hat{\lambda}) = \hat{K}^{-1} \text{Avec}(\hat{K}^{-1}) = \frac{\lambda^2}{4n} \left( -\frac{2n}{\lambda^3} \right) \frac{\lambda^2}{4n} = -\frac{\lambda}{8n},$$

which is unambiguously negative.

#### 4. Simulation Results

The bias expressions in (19) and (20) are valid only to  $O(n^{-1})$ . The actual bias and mean squared error (MSE) of the maximum likelihood and bias-corrected maximum likelihood estimators are now compared in a Monte Carlo experiment. The maximum likelihood estimates were obtained using the Nelder-Mead algorithm in the *maxLik* package (Toomet and Henningsen, 2008) for the R statistical software environment (R, 2011). The R software also includes routines for generating Rayleigh-distributed, Maxwell-distributed, Half-Normal-distributed, and Chi-distributed random variates.

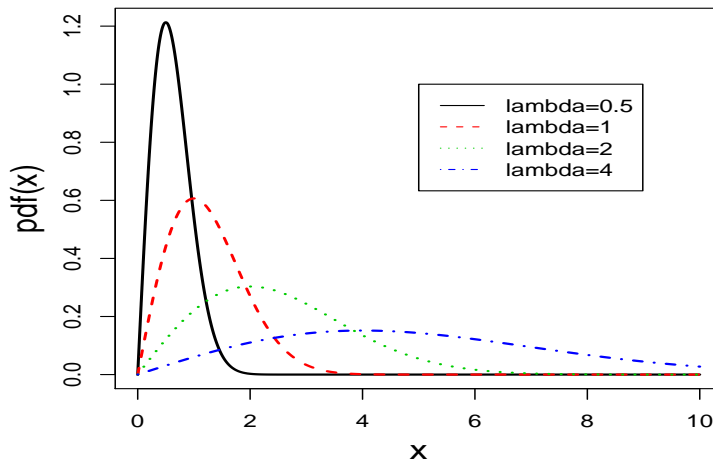


Figure 1: The Rayleigh probability density plot with different parameter values:  $\lambda = 0.5, 1, 2$  and  $4$ .

In addition to  $\hat{\theta}$ ,  $\tilde{\theta}$ ,  $\hat{k}$ , and  $\tilde{k}$ , we have also considered the bootstrap-bias-corrected estimator. In the case of  $\theta$ , for example, this is obtained as  $\check{\theta} = 2\hat{\theta} - (1/N_B) \sum_{j=1}^{N_B} \hat{\theta}_{(j)}$ , where  $\hat{\theta}_{(j)}$  is the MLE of  $\theta$  obtained from the  $j^{th}$  of the  $N_B$  bootstrap samples. Corresponding expressions apply for the estimator of the other parameter.

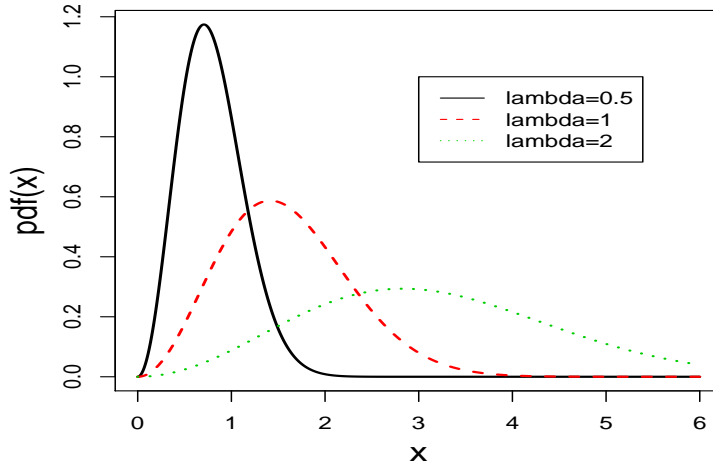


Figure 2: The Maxwell probability density plot with different parameter values:  $\lambda = 0.5, 1$  and  $2$ .

We are interested in the effectiveness of the bias correction when the sample data are generated with different parameter values in a particular distribution. The effect of sample size is also of interest. We have considered sample sizes of 20, 30, 50, 100, and 200 observations. The choice of the parameter values for each of four different distributions in the Generalized Rayleigh family is made by considering the density plots shown in Figures 1 to 4. Table 1 provides the selected parameter values and sample sizes ( $n_{obs}$ ) for the Monte Carlo experiment. We have used Monte Carlo 200,000 replications, and 1,000 re-samples are used in constructing the bootstrap bias corrections.

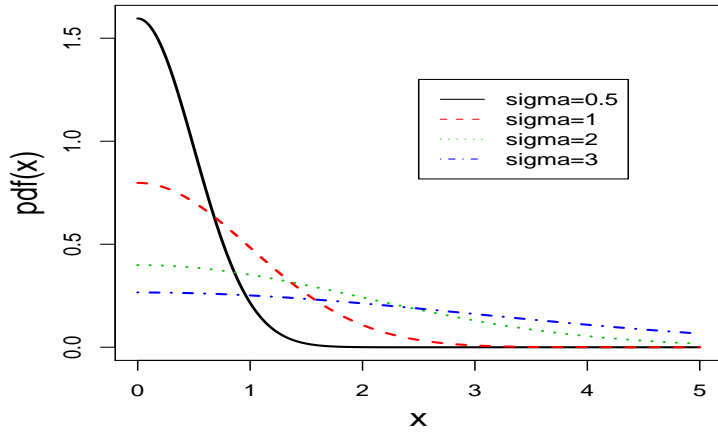


Figure 3: The Half-Normal probability density plot with different parameter values:  $\sigma = 0.5, 1, 2$  and  $3$ .

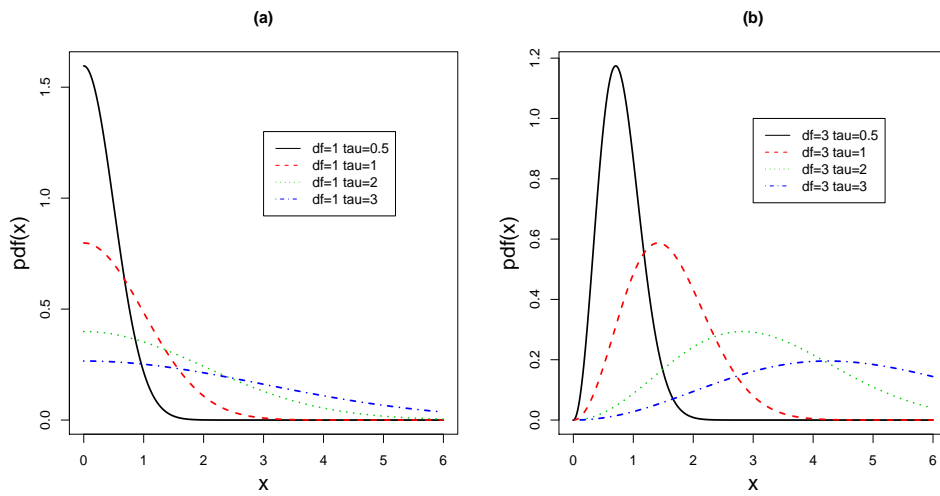


Figure 4: The Chi probability density plot with different parameter values:  $\tau = 0.5, 1, 2$  and  $3$  in both (a) and (b);  $a = 1$  in (a) and  $a = 3$  in (b).

The results associated with the Rayleigh, Maxwell, Half-Normal and Chi distributions appear in Tables 2 to 5 respectively. We begin with the Rayleigh distribution. Recall that the Rayleigh distribution arises when  $k = 0$  in the GRD. Several joint cumulants of the derivatives or derivatives of cumulants would then be undefined. Accordingly, we cannot apply (21) and (22) to compute the bias-corrected estimator,  $\hat{\theta}$ , for the Rayleigh distribution. This case is treated separately as a one-parameter problem, as is explained in Appendix A. *Percentage* biases and MSEs (in square brackets) of the MLE, the analytic bias-adjusted estimator and the bootstrap bias-adjusted estimator of the Rayleigh distribution's parameter are shown in Table 3. The percentage MSEs are defined as  $100 \times (\text{MSE}/k^2)$  or  $100 \times (\text{MSE}/\theta^2)$ .

Table 1: Parameter settings for Monte Carlo experiment

Distribution	sample size	$k$	$\theta$	
Rayleigh	$n_{obs} = 20$	$k = 0$	$\theta = \frac{1}{2\lambda^2}$	$\lambda = 1$
	$n_{obs} = 30$			$\lambda = 1$
	$n_{obs} = 50$			$\lambda = 0.5, 1, 2, \text{ and } 4$
	$n_{obs} = 100$			$\lambda = 1$
	$n_{obs} = 200$			$\lambda = 1$
Maxwell	$n_{obs} = 20$	$k = \frac{1}{2}$	$\theta = \frac{1}{2\lambda^2}$	$\lambda = 1$
	$n_{obs} = 30$			$\lambda = 1$
	$n_{obs} = 50$			$\lambda = 0.5, 1, 2$
	$n_{obs} = 100$			$\lambda = 1$
	$n_{obs} = 200$			$\lambda = 1$
Half-normal	$n_{obs} = 20$	$k = -\frac{1}{2}$	$\theta = \frac{1}{2\sigma^2}$	$\sigma = 1$
	$n_{obs} = 30$			$\sigma = 1$
	$n_{obs} = 50$			$\sigma = 0.5, 1, 2, \text{ and } 3$
	$n_{obs} = 100$			$\sigma = 1$
	$n_{obs} = 200$			$\sigma = 1$
Chi	$n_{obs} = 20$	$k = \frac{a}{2} - 1$ and $a = 1$	$\theta = \frac{1}{2\tau^2}$	$\tau = 1$
	$n_{obs} = 30$	$k = \frac{a}{2} - 1$ and $a = 1$		$\tau = 1$
	$n_{obs} = 50$	$k = \frac{a}{2} - 1$ and $a = 1$ or $3$		$\tau = 0.5, 1, 2$ or $3$
	$n_{obs} = 100$	$k = \frac{a}{2} - 1$ and $a = 1$		$\tau = 1$
	$n_{obs} = 200$	$k = \frac{a}{2} - 1$ and $a = 1$		$\tau = 1$

Table 2: Percentage biases and MSE's of the Rayleigh distribution and the Maxwell distribution

n	Rayleigh distribution			Maxwell distribution				
	$\hat{\theta}$	$\tilde{\theta}$	$\check{\theta}$	$\hat{\theta}$	$\tilde{\theta}$	$\check{\theta}$		
	<b>(a):</b> $k = 0.5, \lambda = 0.5$ or $\theta = 2$							
50	-0.2801 [0.4991]	-0.0308 [0.5008]	0.0312 [0.5762]	0.07116 [33.0095]	-0.4637 [33.0029]	6.9467 [6.3404]	-0.0299 [5.1692]	-0.2062 [5.1912]
	<b>(b):</b> $k = 0.5, \lambda = 1$ or $\theta = 0.5$							
20	-0.6255 [1.2521]	-0.0126 [1.2638]	-0.0173 [1.3237]	0.1335 [97.2173]	-4.9164 [105.0411]	19.3819 [25.4475]	0.0539 [15.6251]	-2.1584 [16.4881]
30	-0.3758 [0.8339]	0.0393 [0.8394]	-0.0342 [0.8407]	0.1363 [59.2491]	-0.2486 [62.5967]	12.2450 [13.0656]	0.0806 [9.3493]	1.0504 [10.6324]
50	-0.2244 [0.4992]	0.0251 [0.5012]	-0.0354 [0.5271]	16.5934 [39.9668]	0.1374 [32.8900]	6.96158 [6.3250]	-0.0164 [5.1539]	0.0199 [5.2091]
100	-0.1292 [0.2480]	-0.0043 [0.2485]	0.0180 [0.2512]	8.0952 [17.2298]	0.0217 [15.5437]	3.4034 [2.7152]	0.0225 [2.4441]	-0.0684 [2.4406]
200	-0.0664 [0.1252]	-0.0039 [0.1254]	0.1247 [0.4976]	4.0092 [8.0070]	0.0333 [7.6173]	1.6897 [1.2543]	0.0253 [1.8889]	-0.0096 [1.2001]
	<b>(c):</b> $k = 0.5, \lambda = 2$ or $\theta = 0.125$							
50	-0.2176 [0.5013]	0.0318 [0.5033]	0.0268 [0.5003]	16.60228 [39.6036]	-0.0519 [32.6314]	6.9521 [6.3073]	-0.0251 [5.1391]	-0.2228 [5.1857]
	<b>(d):</b> $\lambda = 4$ or $\theta = 0.03125$							
50	-0.2668 [0.4978]	-0.0175 [0.4996]	0.0085 [0.4907]					

In Table 2, we see that the percentage biases of the MLEs are much larger than those of both the analytic bias-adjusted estimates and the bootstrapped bias-adjusted estimates,  $\tilde{\theta}$  and  $\check{\theta}$ . All of the percentage biases of the MLEs for the Rayleigh distribution are less than 1 percent. The analytic bias adjustment and the bootstrap bias adjustment are both effective, and generally reduce percentage bias by at least one order of magnitude in all cases. In case (b) in Table 2, the absolute values of percentage biases of MLEs monotonically decline as the sample size increases because of the consistency of the MLEs. On the other hand, the percentage MSEs reported in Table 2 for the different estimators are very close to each other, but for this distribution it seems that bias reduction comes at the expense of some mild increase in variance.

Table 3: Percentage biases and MSE's of the Half-Normal distribution

n	$\hat{k}$	$\tilde{k}$	$\check{k}$	$\hat{\theta}$	$\tilde{\theta}$	$\check{\theta}$
<b>(a): <math>k = -0.5, \sigma = 0.5</math> or <math>\theta = 2</math></b>						
50	-4.6084 [3.5173]	-0.0121 [2.9658]	0.2085 [2.9424]	8.9605 [9.6729]	0.0062 [7.6740]	-0.5042 [7.6818]
<b>(b): <math>k = -0.5, \sigma = 1</math> or <math>\theta = 0.5</math></b>						
20	-12.5300 [12.9837]	0.0692 [8.5404]	1.6159 [8.4679]	24.9904 [41.4694]	-0.1130 [24.2286]	-4.2726 [24.9102]
30	-8.0008 [6.9887]	-0.0249 [5.2710]	0.4314 [5.2178]	15.7146 [20.7703]	0.0270 [14.3319]	-1.2662 [14.0965]
50	-4.5916 [3.5124]	0.0039 [2.9627]	-0.1817 [7.7405]	8.8967 [9.6484]	-0.0527 [7.6661]	-0.1817 [7.7405]
100	-2.2525 [1.5457]	-0.0196 [1.4178]	0.0837 [1.4084]	4.2993 [4.0349]	-0.0185 [3.5847]	-0.1406 [3.5852]
200	-1.1169 [0.7285]	-0.0162 [0.6975]	0.0042 [0.6987]	2.1283 [1.8427]	0.0065 [1.7348]	0.0048 [1.7523]
<b>(c): <math>k = -0.5, \sigma = 2</math> or <math>\theta = 0.125</math></b>						
50	-4.6569 [3.5429]	-0.0579 [2.9847]	0.2085 [2.9421]	9.0187 [9.7361]	0.0616 [7.7241]	-0.5042 [7.6813]
<b>(d): <math>k = -0.5, \sigma = 3</math> or <math>\theta = 1/18</math></b>						
50	-4.6425 [3.5242]	-0.0443 [2.9692]	0.2085 [2.9423]	8.9994 [9.7866]	0.0472 [7.7700]	-0.5041 [7.6817]

Tables 3 to 5 provide corresponding results for the Maxwell, Half-Normal, and Chi distributions. These results are very similar to those for the Rayleigh

Table 4: Percentage biases and MSE's of the Chi distribution

n	$\hat{k}$	$\tilde{k}$	$\bar{k}$	$\hat{\theta}$	$\tilde{\theta}$	$\bar{\theta}$
<b>(a): <math>a = 1</math> or <math>k = -0.5</math>, <math>\lambda = 0.5</math> or <math>\theta = 2</math></b>						
50	-4.5563 [3.4970]	0.0373 [2.9519]	0.1199 [2.9679]	8.9727 [9.7082]	0.0163 [7.7049]	-0.3094 [7.7230]
<b>(a): <math>a = 3</math> or <math>k = 0.5</math>, <math>\lambda = 0.5</math> or <math>\theta = 2</math></b>						
50	16.5948 [39.6932]	-0.0589 [32.7129]	-0.4748 [33.2130]	7.0059 [6.3204]	0.0251 [5.1442]	-0.2162 [5.2102]
<b>(b): <math>a = 1</math> or <math>k = -0.5</math>, <math>\lambda = 1</math> or <math>\theta = 0.5</math></b>						
20	-12.6474 [13.0473]	-0.0326 [8.5651]	1.4878 [8.4769]	25.0599 [41.5974]	-0.0483 [24.2933]	-3.8618 [25.1288]
30	-8.0419 [6.9467]	-0.0631 [5.2307]	0.5362 [5.1906]	15.7766 [20.6236]	0.0829 [14.1990]	-1.4112 [14.2530]
50	-4.6496 [3.5207]	-0.0511 [2.9654]	0.0794 [2.9729]	9.1072 [9.8263]	0.1429 [7.7870]	-0.1768 [7.7815]
100	-2.2866 [1.5475]	-0.0527 [1.4181]	-0.0394 [1.4278]	4.2989 [4.0359]	-0.0181 [3.5858]	0.0189 [3.5942]
200	-1.0849 [0.7280]	0.0155 [0.6977]	0.0161 [0.6989]	2.0831 [1.8415]	-0.0381 [1.7354]	0.0300 [1.7425]
<b>(b): <math>a = 3</math> or <math>k = 0.5</math>, <math>\lambda = 1</math> or <math>\theta = 0.5</math></b>						
50	16.8518 [40.13]	0.1829 [33.0236]	0.3279 [32.9918]	7.0863 [6.3887]	0.1009 [5.1946]	0.1605 [5.1907]
<b>(c): <math>a = 1</math> or <math>k = -0.5</math>, <math>\lambda = 2</math> or <math>\theta = 0.125</math></b>						
50	-4.5785 [3.5103]	0.0163 [2.9620]	0.1019 [2.9615]	9.0437 [9.7865]	0.0821 [7.7620]	-0.2595 [7.8086]
<b>(c): <math>a = 3</math> or <math>k = 0.5</math>, <math>\lambda = 2</math> or <math>\theta = 0.125</math></b>						
50	16.7661 [40.3018]	0.1022 [33.2009]	-0.7921 [32.8446]	7.0316 [6.4147]	0.0495 [5.2243]	-0.2855 [5.1769]
<b>(d): <math>a = 1</math> or <math>k = -0.5</math>, <math>\lambda = 3</math> or <math>\theta = 1/18</math></b>						
50	-4.6888 [3.5385]	-0.0882 [2.9782]	-0.0915 [2.9954]	9.0061 [9.6563]	0.0503 [7.6562]	0.0596 [7.7047]
<b>(d): <math>a = 3</math> or <math>k = -0.5</math>, <math>\lambda = 3</math> or <math>\theta = 1/18</math></b>						
50	16.5731 [39.5854]	-0.0793 [32.6238]	-0.5064 [32.9079]	6.9514 [6.3260]	-0.0258 [5.1557]	-0.2199 [5.1846]



distribution in Table 2, except that the analytic bias correction performs even better, with a reduction of at least two orders of magnitude in terms of percentage bias itself. Any comparison of the two methods of bias adjustment can be made on the basis of percentage bias itself, given the similarities in percentage MSE. On this basis, the overall thrust of the results in Tables 3 to 5 is that the analytical Cox-Snell/Cordeiro-Klein methodology outperforms the use of the bootstrap in almost all of the cases considered, and especially if the sample size is less than  $n = 100$ .

## 5. Illustrative Example

In this section, we use some real data to investigate the practical performance of bias corrected estimates obtained using the method introduced by Cox and Snell. UK consumption of beer, wine and spirits, 1955-1985, in liters *per capita* (Selvanathan and Clements, 1995, p.130) is considered. Using a logisitic regression with Half-Normal errors, we have modelled the expenditure on each beverage as a function of total alcohol expenditure. Expenditures are expressed in current pounds *per capita*. We first estimate the model by the maximum likelihood method, and then apply the Cox-Snell bias correction. The results are shown in Table 6.

We have also calculated the "marginal effects" based on the estimates in Table 5; *i.e.*, the marginal changes in expected probability,

$$\frac{\partial E(y|x)}{\partial x} = \frac{\beta \exp(\hat{\alpha} + \hat{\beta}x)}{(1 + \exp(\hat{\alpha} + \hat{\beta}x))^2},$$

where  $y$  is the conditional budget share for a beverage;  $x$  is total expenditure on alcohol; and  $\alpha$  and  $\beta$  are the intercept and slope parameters from the logistic regression. Table 6 shows the mean value of the marginal effects for each observation for each beverage. The first row shows the mean marginal effects calculated by MLE, while the second row shows those from biased corrected estimates. The sum of the entries in the first row is (essentially) zero, reflecting the fact that expenditure (on alcohol) is fully allocated across

Table 5: Estimates of Logistic Regression Models with Half Normal distribution

	Beer		Wine		Spirits	
	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$
MLE	-5.3399	-0.0005	-6.9225	0.0026	-6.0890	-0.0005
Bias-correct	-5.3266	-0.0003	-6.9093	0.0028	-6.0758	-0.0002

Table 6: Mean Marginal Effects

	Beer	Wine	Spirits
MLE	$-2.4961 \times 10^{-6}$	$3.2660 \times 10^{-6}$	$-1.0300 \times 10^{-6}$
Bias-corrected	$-1.4957 \times 10^{-6}$	$3.6858 \times 10^{-6}$	$-5.4753 \times 10^{-7}$

the three beverage groups noted in Table 5. On the other hand, the second row entries do not sum to zero, because the bias correction does not impose this constraint. In general, the marginal effects in Table 6 are very small, being of the order  $10^{-6}$ . The bias correction does, however, reduce the marginal effect for spirits by an order of magnitude. The parameter estimates in Table 5 are also very small, but in percentage terms, the Cox-Snell bias correction decreases the estimates of  $\beta$  by approximately 50% in absolute value in the case of beer or spirits. These results illustrate that the use of the bias correction can have important implications.

## 6. Conclusions

The two-parameter Generalized Rayleigh distribution provides us with a rich family of specific distributions that have widespread application in many disciplines. Members of this family include the Rayleigh distribution itself, the Half-Normal distribution, the Maxwell distribution, and the Chi distribution. The parameters of these distributions are commonly estimated by the method of maximum likelihood, which in these cases involves solving non-linear first-order conditions. We have applied the analytic bias correction method, introduced by Cox and Snell, to these maximum likelihood estimators, and obtained expressions for the biases to  $O(n^{-1})$ . These results

have then been used to construct the bias-adjusted estimators that are unbiased to order  $O(n^{-2})$ . Using a Monte Carlo simulation experiment, we have investigated the performance of these bias-corrected estimators, in terms of mean squared error and remaining bias, for each of the major members of the Generalized Rayleigh distribution family. The experiment also compared the performance of our bias-corrected estimator with one based on the parametric bootstrap.

In general, our analytic bias correction, which has not previously been implemented or evaluated in this context, is found to be superior to the alternative of bias-correction *via* the bootstrap. Substantial reductions in percentage bias are achieved and, importantly, these gains are usually obtained with only very small increases in relative mean squared error, at least for sample sizes of the magnitude likely to be encountered in practice. Particularly when the computational costs of bootstrapping the bias are taken into account, the use of the Cox-Snell bias correction is strongly recommended when estimating the parameters of the Generalized Rayleigh distribution by maximum likelihood.

### **Acknowledgments**

We are grateful to Ryan Godwin, Farouk Nathoo, and Ken Stewart for their helpful comments on an earlier version of this work.

## References

Abernethy, T. S.(1984). An application of the rayleigh distribution to contract cost data. Masters Thesis, Naval Postgraduate School, Monterey CA.

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production models. *Journal of Econometrics*, 6, 21-37.

Bartlett, M. S. (1953a). Approximate confidence intervals . *Biometrika*, 40, 12-19.

Bartlett, M. S. (1953b). Approximate confidence intervals II. More than one unknown parameter. *Biometrika*, 40, 306-317.

Battese, G. A. and T. J. Coelli (1992). Frontier production functions, technical efficiency and panel data: With applications to paddy farms in India. *Journal of Productivity Analysis*, 3, 153-169.

Battjes, J. A. and H. W. Groenendijk (2000). Wave height distributions on shallow foreshores. *Coastal Engineering*, 40, 161-182.

Beenstock, M. (1995). The stochastic economics of windpower. *Energy Economics*, 17, 27-37.

Chou, C-Y. and H-R. Liu (1998). Properties of the hal-normal distribution and its application to quality control. *Journal of Industrial Technology*, 14, 1-5.

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

Naess, A. (1985). On the distribution of crest to trough wave heights. *Ocean Engineering*, 12, 221-234.

Cox, D. R. and E. J. Snell (1968). A general definition of residuals. *Journal of the Royal Statistical Society, B*, 30, 248-275.

Dyer, D. D. and C. W. Whisenand. (1973). Best linear unbiased estimator of the parameter of the Rayleigh distribution - Part I: Small sample theory for censored order statistics. *IEEE Transactions on Reliability*, 22, 27-34.

Haldane, J. B. S. (1953). The estimation of two parameters from a sample. *Sankhya*, 12, 313-320.

Haldane, J. B. S. and S. M. Smith (1956). The sampling distribution of a maximum likelihood estimate. *Biometrika*, 43, 96-103.

Meagher, K. J., E. G. S. Teo and W. Wang (2008). A duopoly location toolkit: Consumer densities which yield unique spatial duopoly equilibria. *The B.E. Journal of Theoretical Economics* , 8.1, article 14.

Polovko, A. M. (1968). *Fundamentals of Reliability Theory*. Academic Press, New York.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Selvanathan, E. A. and K. W. Clements (1995). *Recent Developments in Applied Demand Analysis*. Springer, Berlin.

Shenton, L. R. and K. Bowman (1963). Higher moments of a maximum-likelihood estimate. *Journal of the Royal Statistical Society, B*, 25, 305-317.

Shenton, L. R. and P. A. Wallington (1962). The bias of moment estimators with an application to the negative binomial distribution. *Biometrika*, 49, 193-240.

Toomet, O. and A. Henningsen (2008). *maxLik: Maximum likelihood estimation*. <http://CRAN.R-project.org>; <http://maxLik.org>

Voda, V. G. (1976). Inferential procedures on a generalized Rayleigh variate (I). *Aplikace Matematiky*, 21, 395-412.

Yamane, T. (1998). Application of the Rayleigh distribution to size selectivity of small prawn pots for the oriental river prawn, *Macrobrachium nipponense*. *Fisheries Research*, 36, 27-33.