Journal of Modern Applied Statistical Methods
https://jmasm.com

Article

Testing for Size-Biased Sampling in Lifetime Data

David E. Giles

Department of Economics, University of Victoria, Victoria, BC V8W 2Y2, Canada; dgiles@uvic.ca; Tel.: +64-22-155-3406

How To Cite: Giles, D. Testing for Size-Biased Sampling in Lifetime Data. *Journal of Modern Applied Statistical Methods* **2026**, *25*(2), 9. https://doi.org/10.53941/jmasm.2026.100009

Received: 3 March 2025
Revised: 13 October 2025
Accepted: 4 June 2026
Published: 9 July 2026

Abstract: Size-biased sampling is a phenomenon that arises in many disciplines and it is common with "survival", or "lifetime" data. This type of sampling occurs when the probability of a population element being sampled is proportional to some weight function that depends on the observed value of that element. This alters the assumed ("base") distribution of the sample data to a "weighted distribution". In practice, it is often not known if size-biased sampling has been used, so there is a need for formal testing to select between the "base" distribution and the "weighted" distribution. We consider two variants of a particular such test in the context of distributions from the generalized gamma family. A Monte Carlo simulation experiment is used to generate critical values for the tests, and to evaluate their power properties. The tests are illustrated in two empirical applications using real-life data.

Keywords: size-biased sampling; hypothesis testing; generalized gamma distribution; power

1. Introduction

Knowledge of the method by which data have been sampled from a population is of crucial importance in any statistical analysis. Many standard inferential procedures are based on the assumption of simple random sampling, and if a different sampling scheme has been adopted this may have significant implications for the validity and properties of inferences that are drawn from the data. One important "non-standard" sampling scheme that has been studied for the past century is the so-called size-biased sampling. For example, see the seminal contributions of [1–4]. This form of sampling arises when the probability of a population element being selected for the sample is proportional to some predetermined weight function, where the latter depends on the observed value of the data for that element.

For example, if the probability of a certain sample value being recorded is directly proportional to the magnitude of that value, we have so-called "length-biased" sampling. If the probability is proportional to the square of that magnitude we have "area-biased" sampling, and so on. Size-biased sampling occurs in many fields, including epidemiology, ecology, environmental and resource economics, forestry, genome mapping, marketing, meteorology, and in a range of reliability and survival studies. For example, see [5–14], among others.

In any empirical analysis it is important to know if size-biased sampling has occurred, because it affects the formulation of the correct data density function, and hence the likelihood function. If size-biased sampling has been used, but this is either unknown or ignored, the likelihood function is mis-specified and any inferences that are based upon it will be incorrect, even asymptotically. This applies to both frequentist and Bayesian inference. Examples of applications that illustrate the negative impact of failing to account for size-biased sampling include those of [15,16], among others. To ensure valid inferences the appropriate "weighted density" (or "moment density") is required as the foundation for the likelihood function. A detailed discussion of this issue is provided by [17], for example.

One practical problem that arises in this context is that the analyst may not know whether or not size-biased sampling has been used, as it is not part of any pre-specified experimental design. Rather, it is a characteristic of the context of the sampling process. So, in order to ensure valid subsequent inferences, one might consider the application of some test of the hypothesis, H_0: "Unweighted (simple) random sampling has been used", against the simple alternative hypothesis, H_1: "A specific weighted (size-biased) sampling has been used". Several parametric and nonparametric tests of H_0 against H_1 have been suggested in the literature. For example, see [18–20]. In this paper, we focus on the distribution-free test of the above hypotheses proposed by [18] and extended by [19]. The latter authors also consider both likelihood ratio and Wald tests of H_0 against the composite alternative hypothesis, H_A: "Some general size-biased sampling has been used". However, they find the latter tests to be substantially inferior in terms of power to the test against the simple alternative, H_1, and this motivates the present study.

The principal contribution of this paper, and hence its novelty, is the computation of appropriate critical values, and an evaluation of the powers, of the above-mentioned distribution-free test when applied to a selection of distributions previously not studied in this context. These distributions are all ones that are employed widely in the analysis of lifetime (duration) data, and are members of the "generalized gamma" family of distributions. Accordingly, this study substantially broadens the associated results in the literature.

In the next section we outline the statistical framework that is adopted in this paper, including the formal details of two hypothesis tests that are evaluated, and the base distributions that are considered. Section 3 provides the details of an extensive simulation experiment that we have conducted, first to compute critical values for the tests under a wide range of situations, and second to evaluate the powers of the tests. The latter are presented graphically in Section 3, and the critical values are tabulated for various base distributions and sample sizes in the Appendix A. Section 4 presents two empirical applications with real-life data that illustrate the use of the tests; and some concluding comments and suggestions for further study are provided in Section 5.

2. Statistical Framework

Suppose that a positive random variable, X, has a density function $f(x;\phi)$, for some parameter (vector), ϕ. Given the existence of the cth (raw) population moment of X, namely $\mu'_c = E[X^c] < \infty$, the corresponding "weighted" (or "moment") density of X, to allow for size-biased sampling of order c, is defined as follows.

$$f_c(x;\phi) = f(x;\phi)x^c/\mu'_c \tag{1}$$

The existence of μ'_c ensures that (1) is a proper density. In what follows, we will refer to $f(x;\phi)$ as the "unweighted", or "base" density function of X.

The case $c = 1$ corresponds to "length biased" sampling; $c = 2$ corresponds to "area biased", etc. We will be concerned only with length-biased and area-biased sampling in this paper, as these are the most common forms of size-biased sampling that arise in practice. We can re-phrase our testing problem as one involving the simple null hypothesis, H_0: "the relevant density is $f(x;\ \phi)$"; versus the simple alternative hypothesis, H_1: "the relative density is $f_c(x;\phi)$, for a known c". Or, for example, $H_0: c = 0$ vs. $H_1: c = 1$, in the case of potential length-biased sampling.

The two primary test statistics that we consider are based on the quantity

$$\prod_{i=1}^{n}[f_c(x_i;\phi)/f(x_i;\phi)] \tag{2}$$

a sufficiently large value of which will suggest that size-biased sampling has been used. Using the definition of $f_c(x_i;\phi)$ in (1), an equivalent test procedure is to reject H_0 in favour of size-biased sampling if the statistic:

$$\lambda = [\prod_{i=1}^{n} x_i]^{1/n}/(\mu'_c)^{1/c} \tag{3}$$

is sufficiently large. However, the statistic in (3) will usually be unobservable, because typically m'_c will be a function of one or more of the elements of the parameter vector, ϕ. In this paper we compare the merits of two ways of dealing with this issue. First, following Economou and Tzavelas (2013) [19], we consider the statistic:

$$\hat{\lambda}_c = [\prod_{i=1}^{n} x_i]^{1/n}/(\hat{\mu}'_c)^{1/c} \tag{4}$$

where $\hat{\mu}'_c = (1/n)\sum_{i=1}^{n} x_i^c$ is the cth (raw) sample moment. Alternatively, following the approach of Giles (2026) [21] we can define the statistic:

$$\tilde{\lambda}_c = [\prod_{i=1}^{n} x_i]^{1/n}/(\tilde{\mu}'_c)^{1/c} \tag{5}$$

where $\tilde{\mu}'_c$ is the MLE of μ'_c, obtained by replacing the relevant elements of the parameter vector, ϕ, with maximum likelihood estimates based on the (unweighted) density, $f(x;\phi)$. As is discussed below, in some situations the two test statistics will coincide. However, in general they may differ and the relative performances of the tests are of interest.

In [22] there is a proposal for an alternative testing procedure based on Kullback-Leibler (K-L) divergence that can be applied to our H_0 and H_1 hypotheses. However, some simple re-arrangement of the associated test statistic in that paper's Equation (21) shows that it is simply $\hat{\lambda}_1$ in (4), when c = 1, and $\hat{\lambda}_2^2$ when c = 2. Accordingly, some of the results provided in the following sections are relevant for this test, as we discuss later.

Asymptotically (as $n \to \infty$) the test statistics in (4) and (5) will be identical, due to the weak consistency of both method of moments and maximum likelihood estimation. The results in [19,22] establish the asymptotic normality (and analytic expressions for the the asymptotic means and variances) of $\hat{\lambda}_c$ and their K-L statistic. Clearly, from (4) and (5), asymptotically $\tilde{\lambda}_c$ will also follow a (different) normal distribution. However, our primary concern is with finite sample sizes. The two test statistics $\hat{\lambda}_c$ and $\tilde{\lambda}_c$ may differ in value (and in distribution) in finite samples, in which case different critical values will apply. These critical values are obtained by simulation, as is described in the next section. For a sample of size *n,* and a significance level *α* they will be labelled $\hat{k}_c(\alpha, n, \phi_1)$ and $\tilde{k}_c(\alpha, n, \phi_1)$ respectively, where ϕ_1 is a (possibly null) sub-vector of the parameter vector for the distribution of *X*. Accordingly, the powers of the tests based on (4) and (5) need to be evaluated separately for any given situation. The asymptotic efficiency of a maximum likelihood estimator is never less than that of its method of moments counterpart, but we cannot generalize about the relative performances in finite samples. Accordingly, while one might anticipate that the $\tilde{\lambda}_c$ test may out-perform the $\hat{\lambda}_c$ for some range of sample sizes, this cannot be guaranteed. We explore this possibility in the present study, and discuss it in Section 3.2.

All of the above discussion relates to a specific base (unweighted) distribution, $f(x;\phi)$. As this base distribution changes, so do the specific forms and distributions of $\hat{\lambda}_c$ and $\hat{\lambda}_c$; the values of the associated critical values; and the powers of the tests. In this paper we consider a range of base distributions that are special cases of the so-called "generalized gamma distribution" (introduced in [23]), and are commonly used with lifetime (duration) data. The density function for the generalized gamma distribution is:

$$f(x;a,d,p) = \left(\frac{p}{a^d}\right) x^{d-1} e^{-\left(\frac{x}{a}\right)^p} / \Gamma\left(\frac{d}{p}\right);\ x > 0\ ;\ a,d,p > 0 \tag{6}$$

and its *c*th raw moment is:

$$E[X^c] = a^c \Gamma((d+c)/p)/\Gamma(d/p) \tag{7}$$

So, the size-biased (weighted) counterpart to (6) takes the form

$$f_c(x;a,d,p) = \left(\frac{p}{a^{d+c}}\right) x^{d+c-1} e^{-\left(\frac{x}{a}\right)^p} / \Gamma\left(\frac{d+c}{p}\right);\ x > 0\ ;\ a,d,p,c > 0 \tag{8}$$

The density in (8) is a generalized gamma density with parameters *a*, (*d* + *c*), and *p*, implying that the latter density is "form invariant" under biased sampling. This result follows directly from the fact that the density in (6) belongs to the log-exponential family, as established by [17] (pp. 52–53). The authors in [24] (p. 124) provide a further discussion of such "form-invariance". The particular members of the generalized gamma family that we consider are the exponential, gamma, half-normal, Rayleigh, and Weibull distributions. These correspond to the generalized gamma distribution for the special cases of $(d = p = 1, a = 1/\theta)$, $(d = b, p = 1, a = 1/\theta)$, $(d = 1, p = 2, a = \sqrt{2\sigma^2})$, $(d = p = 2, a = \sqrt{2\sigma^2})$, and $(d = p = k, a = \theta)$. Size-biased variants of each of these distributions have been discussed by numerous authors, including [6–11,25–35].

The base density functions and the associated size-biased densities are provided in Table 1, Together with the associated formulae for $\tilde{\lambda}_c$ (c = 1, 2). It should be noted that the length-biased and area-biased counterparts to the exponential distribution with a rate parameter *θ* are gamma distributions with the same rate parameter, and shape parameters $b = 2$ and $b = 3$ respectively. Further, the length-weighted half-normal distribution is a Rayleigh distribution. We also see in that table we see that $\tilde{\lambda}_c = \hat{\lambda}_c$ when c = 1 for the exponential and gamma distributions, and when c = 2 for each of the half-normal and Rayleigh distributions. In all cases, $\hat{\lambda}_1 = [\prod_{i=1}^n x_i]^{1/n}/\bar{x}$, and $\hat{\lambda}_2 = [\prod_{i=1}^n x_i]^{1/n}/\sqrt{\sum_{i=1}^n x_i^2/n}$.

Table 1. Distributions and test statistics.

Distribution	$\tilde{\lambda}_1$	$\tilde{\lambda}_2$
Exponential		
$f(x;\theta)=\theta e^{-\theta x}$ $f_c(x;\theta)=\theta^{c+1}x^c e^{-\theta x}/\Gamma(c+1)$	$\hat{\lambda}_1$	$[\prod_{i=1}^n x_i]^{1/n}/[\bar{x}\sqrt{2}]$
Half-normal		
$f(x;\sigma)=(1/\sigma)\sqrt{(2/\pi)}e^{-x^2/(2\sigma^2)}$ $f_c(x;\sigma)=\left(\frac{1}{\sigma}\right)^{c+1}x^c e^{-\frac{x^2}{2\sigma^2}}/\Gamma(\frac{c+1}{2})$	$[\prod_{i=1}^n x_i]^{1/n}/\sqrt{2/(n\pi)\sum_{1=1}^n x_i^2}$	$\hat{\lambda}_2$
Rayleigh		
$f(x;\sigma)=(x/\sigma^2)e^{-x^2/(2\sigma^2)}$ $f_c(x;\sigma)=\left(\frac{1}{\sigma}\right)^{c+2}2^{-\frac{c}{2}}x^{c+1}e^{-\frac{x^2}{2\sigma^2}}/\Gamma(1+\frac{c}{2})$	$[\prod_{i=1}^n x_i]^{1/n}/\sqrt{\pi/(2n)\sum_{1=1}^n x_i^2}$	$\hat{\lambda}_2$
Gamma		
$f(x;b,\theta)=x^{b-1}e^{-\theta x}\theta^b/\Gamma(b)$ $f_c(x;b,\theta)=x^{b+c-1}e^{-\theta x}\theta^{b+c}/\Gamma(b+c)$	$\hat{\lambda}_1$	$[\prod_{i=1}^n x_i]^{1/n}/\left[\bar{x}\sqrt{1+1/\tilde{b}}\right]$
Weibull		
$f(x;k,\sigma)=\left(\frac{k}{\sigma}\right)(x/\sigma)^{k-1}e^{-(x/\sigma)^k}$ $f_c(x;k,\sigma)=\left(\frac{k}{\sigma}\right)\left(\frac{x}{\sigma}\right)^{k+c-1}e^{-\left(\frac{x}{\sigma}\right)^k}/\Gamma(1+\frac{c}{k})$	$[\prod_{i=1}^n x_i]^{1/n}/[\tilde{\sigma}\Gamma(1+1/\tilde{k})]$	$[\prod_{i=1}^n x_i]^{1/n}/\left[\tilde{\sigma}\sqrt{\Gamma(1+2/\tilde{k})}\right]$

Note: In all cases, $x > 0$; and all of the parameters are positive. $\tilde{b}$, $\tilde{\sigma}$, and $\tilde{k}$, are maximum likelihood estimators obtained numerically from the likelihood function constructed from the associated base (unweighted) density.

3. A Simulation Experiment

3.1. Computational Details

All of the computations in this study were undertaken with the R statistical software, version 4.5.2 [36]. The R code can be downloaded from https://github.com/DaveGiles1949/r-code (accessed on 6 July 2026). A Monte Carlo study was used to generate finite-sample critical values for the various tests, and to explore their power properties. The random variates for the exponential, half-normal, Rayleigh, gamma, and Weibull distributions (and their weighted counterparts) were generated using the 'ggamma' package [37]. The first three of these distributions have a single (scale or rate) parameter, and the MLE's of this parameter can be expressed in closed form in each case. Specifically, these estimators are $\tilde{\theta}=1/\bar{x}$, $\tilde{\sigma}=\sqrt{(1/n)\sum_{i=1}^n x_i^2}$, and $\tilde{\sigma}=,\sqrt{1/(2n)\sum_{i=1}^n x_i^2}$, respectively for the exponential, half-normal, and Rayleigh distributions. In these cases the calculation of $\tilde{\lambda}_c$ and $\hat{\lambda}_c$ can be performed analytically, and 100,000 Monte Carlo simulations were performed to compute the critical values.

For the gamma distribution, it is easily shown that the MLEs of the parameters satisfy $\tilde{b}=\tilde{\theta}\bar{x}$, and so we see in Table 1 that $\tilde{\lambda}_c=\hat{\lambda}_c$ for this distribution when *c* = 1. However, the calculations of $\tilde{\lambda}_c$ for the gamma distribution when *c* = 2, and for the Weibull distribution for *c* = 1 and 2 require parameter MLEs that are obtained by solving the likelihood equations numerically. This was achieved using the 'uniroot' command in the base R 'stats' package. Given the additional computation that is required, the associated critical values in Tables A4 and A5 are based on 50,000 Monte Carlo replications. The power evaluations discussed in Section 3.3 are based on 20,000 Monte Carlo replications.

3.2. Critical Values

The Monte Carlo experiment explores the finite-sample distributions of each of the test statistics for each null distribution, in a variety of situations. The upper percentiles of these distributions provide the critical values needed to apply the tests for size-biased sampling. For these null distributions, the upper percentiles of the simulated distributions of $\hat{\lambda}_c$ and $\tilde{\lambda}_c$ provide the critical values, $\hat{k}_c(\alpha,n,\phi_1)$ and $\tilde{k}_c(\alpha,n,\phi_1)$. In all cases, it is easily shown that these values are independent of the magnitude of the scale or rate parameter, as is also noted by [22] (p. 95). However, as noted in Section 2, they depend upon the chosen significance level, underlying distribution, the sample size, the choice of test statistic, and (in the case of the gamma and Weibull distributions) on the shape parameter. The results appear in Appendix Tables A1–A3 for *c* = 1 and 2, various sample sizes, and significance levels of α = 1%, 5% and 10%.

As noted in Section 2, for our particular null and alternative hypotheses the K-L-based test statistic proposed in Equation (21) of [22] collapses to $\hat{\lambda}_1$, when *c* = 1, and $\hat{\lambda}_2^2$ when *c* = 2. Accordingly, the results for their test when *c* = 1 are already covered in the Appendix Tables. When *c* = 2, squaring the critical values associated with $\hat{\lambda}_2$ reported in the Appendix Tables provides the appropriate critical values for $\hat{\lambda}_2^2$. This is easily verified by simulating some specific results for the Weibull distribution and comparing them with their counterparts in Tables 1 and 2 of Economou and Tzavelas (2014) [22], for uncensored sampling.

Table 2. Applications results.

	Exponential		**Half-Normal**		**Rayleigh**	
	(a) Algonquin Park fishing survey (n = 40)					
	f	f_1	f	f_1	f	f_1
AIC	254.149	230.6805	427.1477	248.6187	220.8928	226.8107
$\tilde{\sigma}$	6.6695	5.4456	[0.0834]	[0.0454]	(0.5107)	(0.4169)
	cv ($\tilde{\lambda}_1$)	**cv ($\hat{\lambda}_1$)**	**cv ($\tilde{\lambda}_1$)**	**cv ($\hat{\lambda}_1$)**	**cv ($\tilde{\lambda}_1$)**	**cv ($\hat{\lambda}_1$)**
10%	0.6632	0.6632	0.7565	0.7935	0.8921	0.9191
5%	0.6877	0.6877	0.7774	0.8251	0.9018	0.9357
1%	0.7317	0.7317	0.8148	0.8825	0.9187	0.9653
	$\tilde{\lambda}_1$	$\hat{\lambda}_1$	$\tilde{\lambda}_1$	$\hat{\lambda}_1$	$\tilde{\lambda}_1$	$\hat{\lambda}_1$
	0.9112	0.9112	0.9112	1.0414	0.9112	0.9376
	(b) Tesla share price data (n = 23)					
	f	f_1	f	f_1	f	f_1
AIC	26.7781	23.7553	125.3392	57.6991	41.7567	67.3432
$\tilde{\theta}$	1.5862	3.1724	[0.0689]	[0.0957]	(0.3536)	(0.7072)
	cv ($\tilde{\lambda}_1$)	**cv ($\hat{\lambda}_1$)**	**cv ($\tilde{\lambda}_1$)**	**cv ($\hat{\lambda}_1$)**	**cv ($\tilde{\lambda}_1$)**	**cv ($\hat{\lambda}_1$)**
10%	0.7004	0.7004	0.7868	0.8391	0.9068	0.9438
5%	0.7311	0.7311	0.8121	0.8799	0.9184	0.9645
1%	0.7838	0.7838	0.8553	0.9518	0.9370	0.9993
	$\tilde{\lambda}_1$	$\hat{\lambda}_1$	$\tilde{\lambda}_1$	$\hat{\lambda}_1$	$\tilde{\lambda}_1$	$\hat{\lambda}_1$
	0.7257	0.7257	0.7257	0.6415	0.7257	0.5776

Note: $\tilde{\sigma}$ is the MLE of the scale parameter for the Rayleigh distributions. $\tilde{\theta}$ is the MLE of the rate parameter for the exponential distributions. Analytic asymptotic standard errors and bootstrap standard errors appear in brackets and parentheses, respectively. "cv" denotes the critical values for the two tests.

The gamma and Weibull distributions have both scale and shape parameters. The critical values are again invariant to the value of the former parameter, but they depend on the (unknown) value of the shape parameter. Accordingly, the critical values presented in Appendix Tables A4 and A5 allow for a range of values for this parameter. Shape parameter values of unity are omitted as both the gamma and Weibull distributions collapse to an exponential distribution in this case. In Section 5.1 of [22] it is shown that choosing the critical value for the Weibull distribution by using the MLE for the shape parameter when its value is unknown distorts the sizes of associated tests in small samples. Undoubtedly this will also be the case for the gamma distribution. They propose alternative methods for dealing with the unknown parameter value issue that could be explored in the context of the set of distributions under consideration here, but we leave this for future research.

The accuracy of these simulated critical values can be assessed by comparing our results for the Weibull distribution in Table A5 with those in Table 1 of [19]. The latter results for the $\hat{\lambda}_c$ test, which are for α = 5%, match ours very closely for c = 1 and 2 and various values of the shape parameter (k) and sample sizes, even though in that study they are based on just 5000 Monte Carlo replications. For all of the distributions, the critical values increase with the sample size, *ceteris paribus*. Also, for each distribution the critical values associated with c = 2 are always less than their counterparts for c = 1, for any sample size, significance level, or shape parameter value (if relevant). In the case of the gamma and Weibull distributions, the critical values increase as the shape parameter increases in value, for any sample size significance level, or value of c.

3.3. Power Evaluations

For each base distribution of interest, and various sample sizes, we have used the simulated critical values to determine the powers of the tests by simulating the rejection rates when the data are generated from the corresponding weighted distribution. The latter distributions are also shown in Table 1. These rejection rates are computed as the fraction of 20,000 simulated values of the test statistics for which $\hat{\lambda}_c > \hat{k}_c(\alpha, n, \phi_1)$, or $\tilde{\lambda}_c > \tilde{k}_c(\alpha, n, \phi_1)$. Again, c = 1 or 2, and significance levels of α = 1%, 5%, and 10% are considered. The tests under consideration in this study involve a simple null hypothesis and a simple alternative hypothesis. For this reason, conventional "power curves", in which the power is plotted against the degree of departure from the null hypothesis, do not apply. However, for any of the distributions under consideration, and for a given significance level and value of c, the powers of the $\hat{\lambda}_c$ and $\tilde{\lambda}_c$ tests as a function of the sample size are still of interest. A selection of such functions is plotted in Figures 1–3.

Several basic results emerge from all of these power plots. First, the tests are all "unbiased", at least for the particular parameter values that are considered. That is, in the case of these parameters, $Pr.[\hat{\lambda}_c > \hat{k}_c(\alpha, n, \phi_1)|\ H_1\ True] \geq \alpha$ (and similarly for $\tilde{\lambda}_c$). Moreover, the powers are generally substantial in value, even for quite moderate sample sizes. Second, for each distribution and sample size (and for all shape parameter values, if applicable), the powers of all of the tests are greater when $c = 2$ than when $c = 1$. Third, all of the tests are consistent. That is, their power approaches unity as $n \to \infty$. (This has been verified separately for those cases where the figures are limited to modest values of n to ensure meaningful plots.)

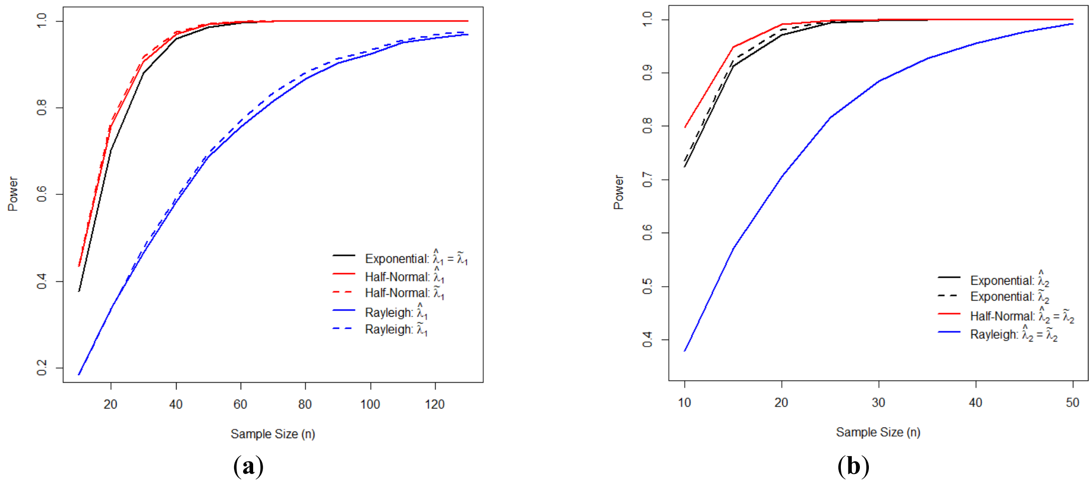


Figure 1. (**a**) Powers against H_1: $c = 1$ ($\alpha = 5\%$); (**b**) Powers against H_1: $c = 2$ ($\alpha = 5\%$).

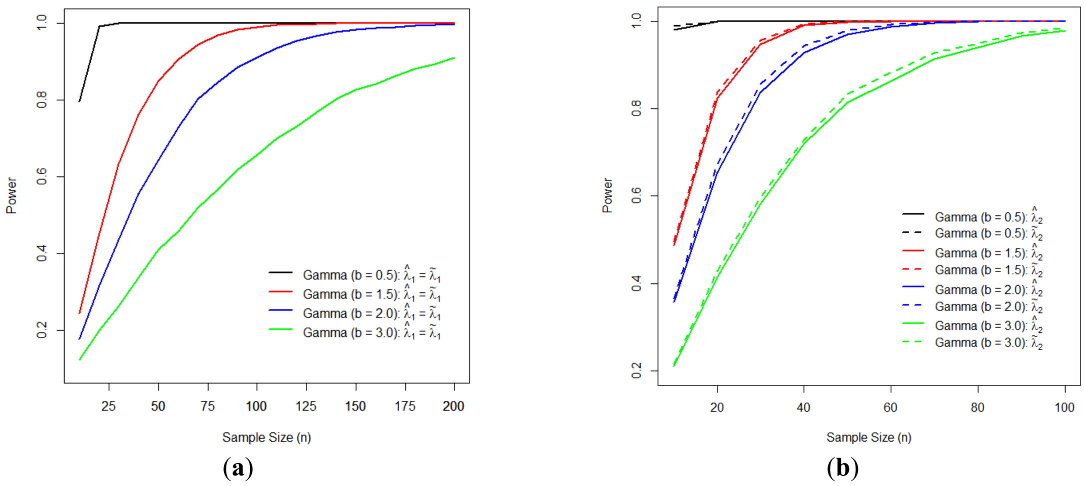


Figure 2. (**a**) Powers against H_1: $c = 1$ ($\alpha = 5\%$); (**b**) Powers against H_1: $c = 2$ ($\alpha = 5\%$).

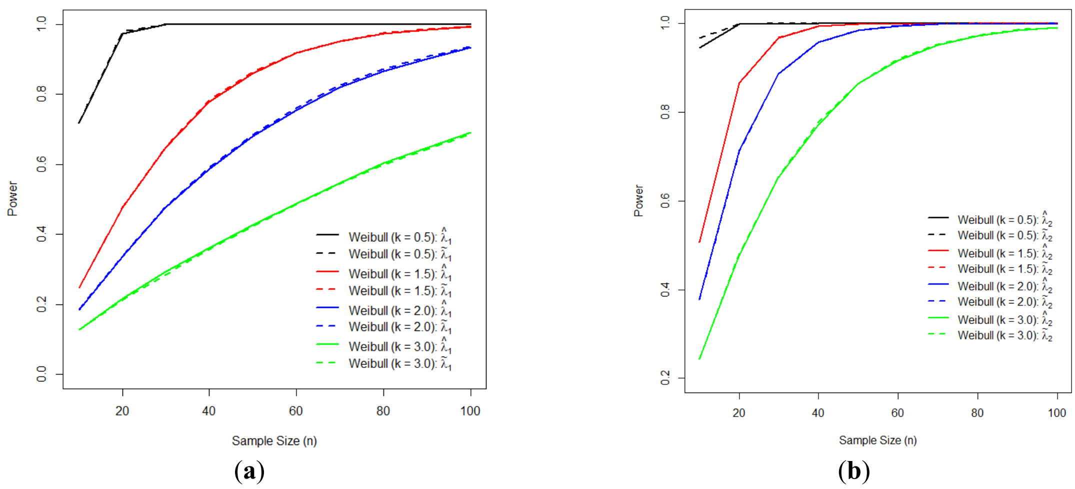


Figure 3. (**a**) Powers against H_1: $c = 1$ ($\alpha = 5\%$); (**b**) Powers against H_1: $c = 2$ ($\alpha = 5\%$).

Further, both Figures 2 and 3 show that the tests' powers decrease (for any n) as the value of the shape parameter increases. For any fixed sample size, this reflects the changes in the first four moments of the distributions of the test statistics under both the null and alternative hypotheses as the shape parameter increases. This is illustrated in Figure 4 for the case where the data follow a gamma distribution when $n = 10$ and $\alpha = 5\%$, and the shape parameter increases from 0.5 to 3.0 in value. For this case, it will be recalled that $\hat{\lambda}_1 = \tilde{\lambda}_1$, and the magnitude of the critical value increases with the shape parameter (as in Table A4). These critical values are marked with vertical lines in Figure 4. The plots in the upper (lower) part of that figure depict the simulated densities of the test statistic under the null (alternative) hypothesis. The reduction in power across the two lower plots as the shape parameter increases reflects both the increase in the 5% critical value as well as the changes in the mean, skewness, and kurtosis of the test statistic's distribution from 0.7, −0.43 and 2.93 to 0.89, −0.74 and 3.71 respectively.

All of the above results relating to the powers of the tests are consistent with the findings of [19] for the $\hat{\lambda}_c$ test in the context of the Weibull distribution. Moreover, the results apply precisely to their K-L test for the other distributions when $c = 1$, and to their test for $c = 2$ when the associated critical values are squared.

One of our primary interests lies in the comparison of the relative powers of the $\hat{\lambda}_c$ and $\tilde{\lambda}_c$ tests. In the cases where these test statistics differ for the exponential, half-normal, and Rayleigh distributions, we see in Figure 1a,b that there is a slight gain in power by using the $\tilde{\lambda}_c$ test with moderate sample sizes. The two test statistics also differ for the case of area-biased sampling for the gamma distribution. Figure 2b shows that for this case there is also a small gain in power by using the $\tilde{\lambda}_2$ test in small samples. In the case of the Weibull distribution we see in Figure 3a,b that the two versions of the tests have essentially the same power, for any n, when testing for either length-biased or area-biased sampling. As noted in Section 2, the values of the $\hat{\lambda}_c$ and $\tilde{\lambda}_c$ test statistics converge as $n \to \infty$, and so the associated powers also converge (ultimately to unity).

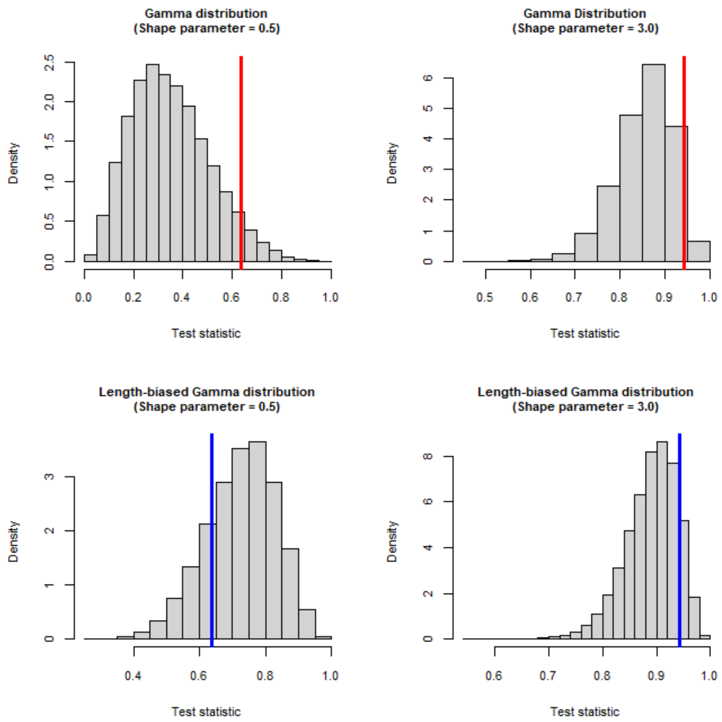


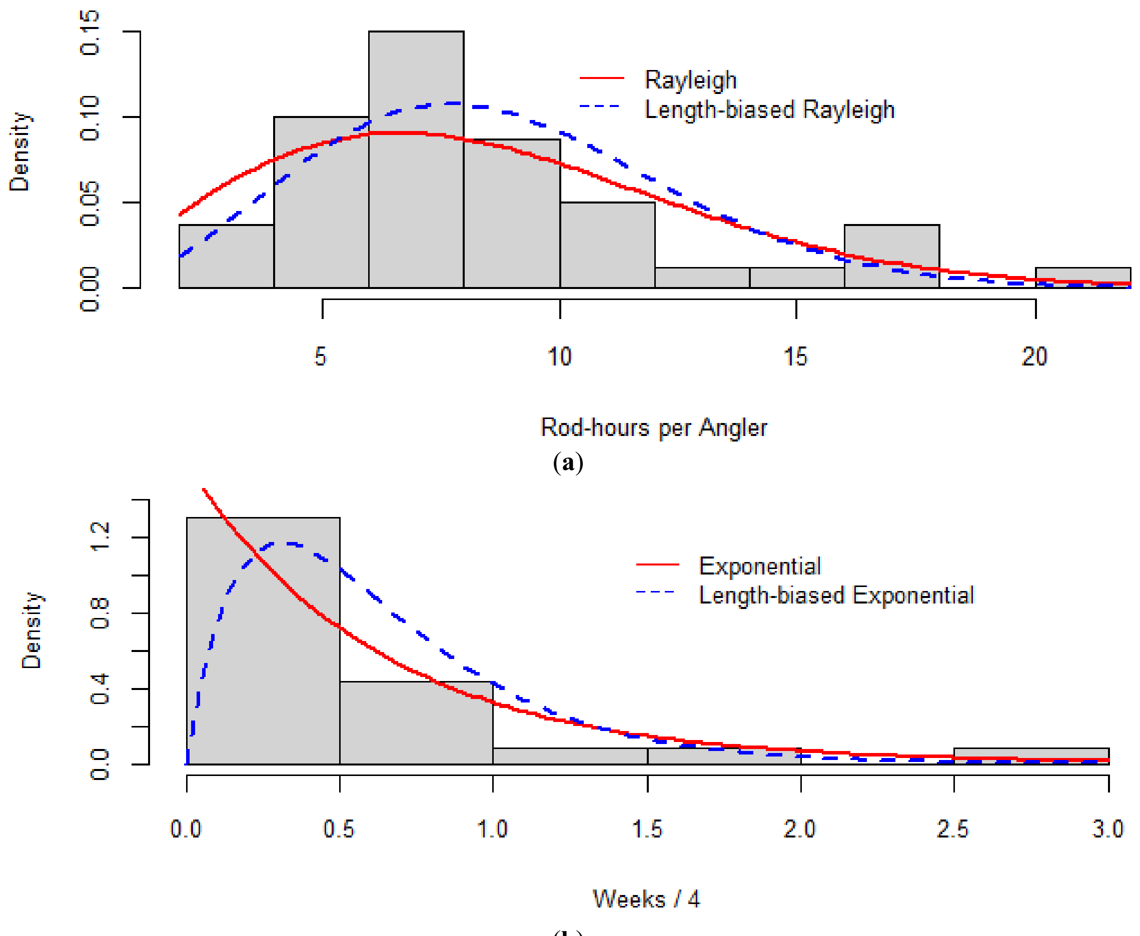
Figure 4. Test statistic densities ($n = 100$).

4. Empirical Applications

To illustrate the use of the tests that are under consideration we provide two empirical applications. The data-sets that are used are those studied by [30] in the context of the Weibull distribution and (possibly) length-biased data. To avoid the complications associated with distributions involving unobserved nuisance parameters, we focus simply on the exponential, half-normal, and Rayleigh distributions. Given the particular sample sizes involved, the critical values for the tests have been simulated as in Section 3, and are reported in Table 2 with the other associated results. Only a very brief discussion of the two data-sets is provided here. Full details can be obtained by downloading the free-access discussion paper version of [30] given in the References section.

The first application involves conservation survey data obtained by the Ontario Ministry of Natural Resources from volunteering fishing parties in Canada's Algonquin Park in 2010. Specifically, we analyze the duration data for rod-hours per angler, for the "Top 40" waterbodies (among the 1500 lakes and 1200 km of streams) in the survey. A greater number of rod-hours may reflect more devotion to the sport, and more concern for information to assist in conservation, thus leading to length-biased sampling. In Table 2, "f" denotes a base distribution, and "f_1" denotes the corresponding length-biased distribution. For each distribution, the values of $\tilde{\lambda}_1$ and $\hat{\lambda}_1$ reject the "no bias" hypothesis in favour of length-biased sampling, at the 5% level. The Akaike's Information Criterion (AIC) values in Table 2a support the Rayleigh distribution over the other two, so the associated MLE results are reported for this case. Given the small sample size, non-parametric bootstrap standard errors based on 10,000 bootstrap samples, are included. The fitted base and length-biased Raleigh densities are compared with the actual data in Figure 5a.

The second application involves data for the rapidly rising price of Tesla shares (TSLA) during the period 31 December 2018 to 30 October 2020, when there were 23 "spells" in which prices rose week-over-week for one or more successive weeks. (The duration values for the 23 spells are divided by 4 to make them non-integer.) The frequency of observing survival data can in itself generate length-biased sampling, as is noted [38], p. 1192. In particular, length-biased sampling arises here if only weekly price observations are available, even though the data are actually generated at a much high frequency. In this application the AIC values in Table 2b favor the exponential distribution, and the associated MLE results are presented. For this selected distribution both of the $\tilde{\lambda}_1$ and $\hat{\lambda}_1$ tests reject the "no bias" hypothesis in favour of length-biased sampling, at the 10% level, and this is consistent with the appearance of Figure 5b.



(**a**)

(**b**)

Figure 5. (**a**): Algonquin Park fishing survey; (**b**): Tesla share price data.

5. Concluding Remarks

In this paper we have explored the properties of three variants of a test for the presence of size-biased sampling, where the degree of the potential weighting function is known, and the base distributions are well-known members of the generalized gamma family. These distributions are widely used in the context of lifetime data, where either length-biased or area-biased sampling is common. Two empirical applications demonstrate the implementation of the tests for length-biased sampling.

Critical values for the test(s) are reported for each of the distributions, and these are then used to evaluate the powers of the tests. The variants of the test differ with respect to the estimation of the population moment in the denominator of the test statistic. In one case the corresponding sample moment is used; while in another case the maximum likelihood estimator of the population moment is employed. The third variant of the test is shown to be just a special case of the other two variants, for the situations under study in this paper. In some situation these two approaches also coincide, and in all cases they are asymptotically equivalent for very large samples. In all of the situations considered, the tests are shown to have very good power, even for moderately sized samples, and the variant based on maximum likelihood estimation has a slight advantage over the method of moments version of the test. These results generalize and confirm those reported for just the latter version of the test for the Weibull distribution by [19].

These encouraging results suggest that the tests in question should be evaluated for additional base distributions. In addition, the present study is limited to uncensored sample data. Censored data for the Weibull distribution are considered by [19], and further research could assess the implications of right-censoring for the other distributions under study here for each version of the test for size-biased sampling. Moreover, extending of the upper and lower bounds for the critical values to allow for data outliers, that are introduced in Section 3.3 of [19], to distributions other than the Weibull, would also provide useful future research. Importantly, it should be recalled that some of the results reported in this study are dependent on the unknown values of certain distributional parameters. This also suggests the need for further research. For example, there may be potential for exploring the development of interpolation techniques to express the critical values for the tests as a function of the nuisance parameter(s). An example of this is provided by [39] in the case of goodness-of-fit tests for the log-normal distribution. Finally, while this paper deals specifically with size-biased sampling, the basic results could easily be extended to apply to other weighted sampling situation. This remains a task for future research.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data used in the empirical applications in this paper can be accessed at https://github.com/DaveGiles1949/Data/tree/main/Size-biased%20sampling%20Applications%20Data.

Acknowledgments

I am grateful to two anonymous referees for their extremely helpful comments and suggestions relating to an earlier version of this paper.

Conflicts of Interest

The author declares no conflict of interest.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

Appendix A

Critical Values

Table A1. Critical values—Exponential distribution.

Sig. Level (α):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
10	0.7813	0.6556	0.8202	0.7076	0.8825	0.7972
		(0.5524)		(0.5799)		(0.6241)
25	0.6935	0.5460	0.7225	0.5814	0.7738	0.6445
		(0.4904)		(0.5109)		(0.5472)
50	0.6518	0.4970	0.6740	0.5216	0.7134	0.5668
		(0.4609)		(0.4766)		(0.5044)
100	0.6237	0.4651	0.6393	0.4821	0.6692	0.5149
		(0.4410)		(0.4521)		(0.4732)
150	0.6113	0.4511	0.6246	0.4654	0.6488	0.4918
		(0.4322)		(0.4426)		(0.4588)
200	0.6046	0.4435	0.6159	0.4558	0.6367	0.4791
		(0.4276)		(0.4355)		(0.4502)
250	0.5997	0.4382	0.6100	0.4491	0.6288	0.4697
		(0.4241)		(0.4313)		(0.4446)
300	0.5964	0.4345	0.6058	0.4445	0.6230	0.4635
		(0.4217)		(0.4284)		(0.4405)
500	0.5881	0.4253	0.5954	0.4331	0.6091	0.4483
		(0.4158)		(0.4210)		(0.4307)
1000	0.5801	0.4168	0.5852	0.4222	0.5952	0.4327
		(0.4102)		(0.4138)		(0.4208)
10,000	0.5673	0.4031	0.5689	0.4048	0.5720	0.4080
		(0.4011)		(0.4023)		(0.4045)

Note: For each sample size, the critical values in the first row are $\hat{k}_c(\alpha, n, \phi_1)$. If $\hat{\lambda}_c$ differs from $\tilde{\lambda}_c$, then the corresponding critical values for the latter, namely, $\tilde{k}_c(\alpha, n, \phi_1)$ appear in parentheses in the second row. In this table, ϕ_1 is a null vector as there are no “nuisance parameters”.

Table A2. Critical values—Half-normal distribution.

Sig. Level (α):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
10	0.8515	0.7542	0.8815	0.7975	0.9258	0.8655
		(0.9453)		(0.9996)		(1.0847)
25	0.7818	0.6637	0.8064	0.6945	0.8482	0.7494
		(0.8319)		(0.8704)		(0.9392)
50	0.7468	0.6216	0.7660	0.6444	0.8004	0.6863
		(0.7790)		(0.8076)		(0.8601)
100	0.7215	0.5923	0.7356	0.6089	0.7619	0.6392
		(0.7424)		(0.7631)		(0.8011)
150	0.7104	0.5803	0.7222	0.5933	0.7436	0.6180
		(0.7273)		(0.7437)		(0.7746)
200	0.7042	0.5730	0.7143	0.5846	0.7332	0.6061
		(0.7182)		(0.7326)		(0.7596)
250	0.6999	0.5685	0.7094	0.5788	0.7266	0.5987
		(0.7125)		(0.7254)		(0.7504)
300	0.6967	0.5650	0.7053	0.5744	0.7213	0.5926
		(0.7081)		(0.7199)		(0.7428)
500	0.6893	0.5568	0.6959	0.5641	0.7083	0.5776
		(0.6978)		(0.7070)		(0.7239)
1000	0.6817	0.5487	0.6865	0.5539	0.6955	0.5638
		(0.6877)		(0.6942)		(0.7066)
10,000	0.6696	0.5357	0.6711	0.5374	0.6740	0.5404
		(0.6714)		(0.6735)		(0.6773)

Note: For each sample size, the critical values in the first row are $\hat{k}_c(\alpha, n, \phi_1)$. If $\hat{\lambda}_c$ differs from $\tilde{\lambda}_c$, then the corresponding critical values for the latter, namely, $\tilde{k}_c(\alpha, n, \phi_1)$ appear in parentheses in the second row. In this table, ϕ_1 is a null vector as there are no “nuisance parameters”.

Table A3. Critical values—Rayleigh distribution.

Sig. Level (α)	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
10	0.9365	0.8839	0.9497	0.9065	0.9692	0.9406
		(0.9973)		(1.0229)		(1.0613)
25	0.9043	0.8329	0.9158	0.8506	0.9343	0.8802
		(0.9398)		(0.9598)		(0.9932)
50	0.8869	0.8071	0.8961	0.8205	0.9118	0.8444
		(0.9107)		(0.9258)		(0.9528)
100	0.8750	0.7897	0.8819	0.7998	0.8942	0.8180
		(0.8911)		(0.9024)		(0.9230)
150	0.8694	0.7820	0.8753	0.7903	0.8860	0.8056
		(0.8824)		(0.8918)		(0.9090)
200	0.8662	0.7775	0.8713	0.7847	0.8806	0.7981
		(0.8773)		(0.8854)		(9006)
250	0.8640	0.7744	0.8686	0.7810	0.8773	0.7933
		(0.8738)		(0.8812)		(0.8951)
300	0.8623	0.7721	0.8668	0.7783	0.8747	0.7895
		(0.8713)		(0.8782)		(0.8908)
500	0.8584	0.7668	0.8618	0.7714	0.8680	0.7802
		(0.8652)		(0.8705)		(0.8803)
1000	0.8546	0.7616	0.8571	0.7650	0.8618	0.7715
		(0.8594)		(0.8632)		(0.8705)
10,000	0.8484	0.7532	0.8492	0.7543	0.8507	0.7563
		(0.8499)		(0.8511)		(0.8534)

Note: For each sample size, the critical values in the first row are $\hat{k}_c(\alpha, n, \phi_1)$. If $\hat{\lambda}_c$ differs from $\tilde{\lambda}_c$, then the corresponding critical values for the latter, namely, $\tilde{k}_c(\alpha, n, \phi_1)$ appear in parentheses in the second row. In this table, ϕ_1 is a null vector as there are no "nuisance parameters".

Table A4. Critical values—Gamma distribution.

Sig. Level (α):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***b* = 0.5**			
10	0.5695	0.4228	0.6359	0.4921	0.7488	0.6166
		(0.4061)		(0.4738)		(0.6028)
25	0.4402	0.2967	0.4835	0.3348	0.5630	0.4103
		(0.2855)		(0.3230)		(0.3971)
50	0.3857	0.2487	0.4143	0.2722	0.4676	0.3178
		(0.2418)		(0.2649)		(0.3093)
100	0.3509	0.2185	0.3700	0.2340	0.4074	0.2646
		(0.2143)		(0.2290)		(0.2587)
150	0.3366	0.2062	0.3522	0.2182	0.3832	0.2434
		(0.2027)		(0.2146)		(0.2386)
200	0.3285	0.1992	0.3419	0.2101	0.3681	0.2312
		(0.1967)		(0.2067)		(0.2273)
250	0.3230	0.1950	0.3350	0.2039	0.2222	0.3587
		(0.1926)		(0.2015)		(0.2193)
300	0.3190	0.1917	0.3300	0.2006	0.3513	0.2172
		(0.1897)		(0.1978)		(0.2143)
500	0.3100	0.1847	0.3183	0.1912	0.3339	0.2035
		(0.1831)		(0.1893)		(0.2011)
1000	0.3010	0.1776	0.3068	0.1820	0.3174	0.1904
		(0.1767)		(0.1808)		(0.1886)
10,000	0.2870	0.1668	0.28888	0.1682	0.2922	0.1707
		(0.1665)		(0.1678)		(0.1702)
			***b* = 1.5**			
10	0.8552	0.7552	0.8822	0.7957	0.9232	0.8588
		(0.7500)		(0.7910)		(0.8573)
25	0.7924	0.6674	0.8136	0.6966	0.8511	0.7487
		(0.6616)		(0.6911)		(0.7450)
50	0.7608	0.6244	0.7772	0.6459	0.8067	0.6858
		(0.6200)		(0.6412)		(0.6800)
100	0.7401	0.5970	0.7520	0.6119	0.7735	0.6397
		(0.5940)		(0.6084)		(0.6359)
150	0.7304	0.5847	0.7404	0.5975	0.7591	0.6219
		(0.5822)		(0.5944)		(0.6180)
200	0.7250	0.5777	0.7337	0.5891	0.7499	0.6096
		(0.5756)		(0.5864)		(0.6065)
250	0.7213	0.5731	0.7292	0.5827	0.7435	0.6012

Table A4. *Cont.*

			***b* = 1.5**			
		(0.5710)		(0.5806)		(0.5979)
300	0.7188	0.5695	0.7258	0.5785	0.7395	0.5960
		(0.5677)		(0.5762)		(0.5933)
500	0.7123	0.5614	0.7180	0.5684	0.7287	0.5815
		(0.5600)		(0.5667)		(0.5794)
1000	0.7062	0.5538	0.7103	0.5589	0.7180	0.5685
		(0.5528)		(0.5576)		(0.5670)
10,000	0.6961	0.5413	0.6973	0.5428	0.6998	0.5458
		(0.5410)		(0.5424)		(0.5453)
Sig. Level (*α*):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***b* = 2.0**			
10	0.8917	0.8100	0.9124	0.8427	0.9443	0.8962
		(0.8064)		(0.8393)		(0.8937)
25	0.8435	0.7368	0.8600	0.7610	0.8883	0.8056
		(0.7328)		(0.7577)		(0.8021)
50	0.8192	0.7018	0.8319	0.7194	0.8543	0.0.7516
		(0.6983)		(0.7162)		(0.7485)
100	0.8022	0.6773	0.8118	0.6904	0.8287	0.7140
		(0.6750)		(0.6877)		(0.7113)
150	0.7947	0.6668	0.8026	0.6779	0.8173	0.6989
		(0.6647)		(0.6758)		(0.6958)
200	0.7902	0.6606	0.7971	0.6702	0.8099	0.6876
		(0.6589)		(0.6683)		(0.6856)
250	0.7872	0.6565	0.7935	0.6652	0.8051	0.6808
		(0.6549)		(0.6633)		(0.6787)
300	0.7852	0.6535	0.7911	0.6615	0.8015	0.6763
		(0.6520)		(0.6600)		(0.6740)
500	0.7800	0.6461	0.7845	0.6523	0.7928	0.6635
		(0.6451)		(0.6509)		(0.6617)
1000	0.7750	0.6394	0.7782	0.6439	0.7842	0.6520
		(0.6386)		(0.6428)		(0.6506)
10,000	0.7669	0.6281	0.7679	0.6295	0.7698	0.6322
		(0.6279)		(0.6293)		(0.6318)
			***b* = 2.5**			
10	0.9138	0.8454	0.9309	0.8739	0.9560	0.9167
		(0.8429)		(0.8717)		(0.9159)
25	0.8744	0.7824	0.8882	0.8036	0.9110	0.8406
		(0.7795)		(0.8011)		(0.8382)
50	0.8543	0.7516	0.8650	0.7671	0.8832	0.7959
		(0.7489)		(0/7649)		(0/7941)
100	0.8404	0.7303	0.8484	0.7421	0.8624	0.7632
		(0.7286)		(0/7404)		(0.7612)
150	0.8341	0.7212	0.8409	0.7311	0.8531	0.7492
		(0.7196)		(0/7295)		(0.7475)
200	0.8307	0.7160	0.8365	0.7244	0.8468	0.7401
		(0.7147)		(0.7232)		(0.7384)
250	0.8282	0.7125	0.8334	0.7200	0.8430	0.7343
		(0.7113)		(0.7186)		(0.7327)
300	0.8264	0.7096	0.8311	0.7167	0.8398	0.7296
		(0.7085)		(0.7153)		(0.7279)
500	0.8222	0.7034	0.8260	0.7090	0.8332	0.7193
		(0.7026)		(0.7080)		(0.7181)
1000	0.8180	0.6975	0.8208	0.7014	0.8258	0.7089
		(0.6969)		(0.7007)		(0.7079)
10,000	0.8112	0.6874	0.8121	0.6887	0.8137	0.6910
		(0.6873)		(0.6885)		(0.6907)
Sig. Level (*α*):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***b* = 3.0**			
10	0.9285	0.8699	0.9427	0.8942	0.9638	0.9313
		(0.8679)		(0.8926)		(0.9306)
25	0.8954	0.8148	0.9069	0.8334	0.9266	0.8665
		(0.8125)		(0.8317)		(0.8646)
50	0.8782	0.7874	0.8871	0.8018	0.9028	0.8272
		(0.7857)		(0.7999)		(0.8258)
100	0.8664	0.7690	0.8731	0.7791	0.8853	0.7981
		(0.7674)		(0.7775)		(0.7962)
150	0.8611	0.7608	0.8667	0.7693	0.8772	0.7853
		(0.7596)		(0.7678)		(0.7839)

Table A4. *Cont.*

Sig. Level (*α*):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***b* = 3.0**			
200	0.8580	0.7557	0.8630	0.7634	0.8721	0.7777
		(0.7548)		(0.7622)		(0.7760)
250	0.8560	0.7527	0.8605	0.7595	0.8684	0.7719
		(0.7518)		(0.7585)		(0.7705)
300	0.8545	0.7504	0.8586	0.7568	0.8660	0.7680
		(0.7495)		(0.7556)		(0.7669)
500	0.8509	0.7449	0.8542	0.7498	0.8600	0.7589
		(0.7443)		(0.7491)		(0.7577)
1000	0.8473	0.7393	0.8496	0.7429	0.8540	0.7495
		(0.7388)		(0.7423)		(0.7489)
10,000	0.8414	0.7304	0.8422	0.7316	0.8436	0.7336
		(0.7303)		(0.7314)		(0.7333)
			***b* = 3.5**			
10	0.9391	0.8870	0.9513	0.9085	0.9695	0.9413
		(0.8857)		(0.9073)		(0.9407)
25	0.9103	0.8387	0.9202	0.8555	0.9368	0.8837
		(0.8369)		(0.8538)		(0.8822)
50	0.8955	0.8147	0.9032	0.8270	0.9170	0.9490
		(0.8131)		(0.8253)		(0.8481)
100	0.8852	0.7978	0.8909	0.8087	0.9016	0.8238
		(0.7965)		(0.8054)		(0.8227)
150	0.8806	0.7902	0.8855	0.7982	0.8944	0.8123
		(0.7894)		(0.7970)		(0.8113)
200	0.8779	0.7859	0.8822	0.7930	0.8900	0.8051
		(0.7851)		(0.7920)		(0.8042)
250	0.8762	0.7832	0.8800	0.7893	0.8872	0.8009
		(0.7824)		(0.7884)		(0.8000)
300	0.8748	0.7810	0.8784	0.7866	0.8851	0.7974
		(0.7803)		(0.7858)		(0.7962)
500	0.8717	0.7761	0.8745	0.7805	0.8797	0.7886
		(0.7756)		(0.7800)		(0.7881)
1000	0.8685	0.7712	0.8705	0.7744	0.8744	0.7801
		(0.7708)		(0.7739)		(0.7798)
10,000	0.8634	0.7630	0.8640	0.7640	0.8653	0.7660
		(0.7629)		(0.7639)		(0.7657)
Sig. Level (*α*):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***b* = 4.0**			
10	0.9465	0.9002	0.9574	0.9191	0.9731	0.9482
		(0.8990)		(0.9183)		(0.9478)
25	0.9215	0.8568	0.9303	0.8719	0.9449	0.8977
		(0.8556)		(0.8706)		(0.8969)
50	0.9084	0.8348	0.9152	0.8464	0.9276	0.8670
		(0.8338)		(0.8449)		(0.8659)
100	0.8992	0.8200	0.9043	0.8287	0.9136	0.8433
		(0.8191)		(0.8273)		(0.8425)
150	0.8952	0.8132	0.8996	0.8203	0.9076	0.8335
		(0.8125)		(0.8196)		(0.8325)
200	0.8928	0.8095	0.8967	0.8157	0.9036	0.8268
		(0.8088)		(0.8150)		(0.8263)
250	0.8913	0.8069	0.8947	0.8126	0.9011	0.8226
		(0.8063)		(0.8119)		(0.8220)
300	0.8901	0.8049	0.8933	0.8101	0.8991	0.8196
		(0.8044)		(0.8094)		(0.8188)
500	0.8874	0.8005	0.8898	0.8046	0.8943	0.8119
		(0.8000)		(0.8040)		(0.8112)
1000	0.8846	0.7960	0.8863	0.7988	0.8897	0.8040
		(0.7957)		(0.7984)		(0.8036)
10,000	0.8800	0.7886	0.8806	0.7895	0.8817	0.7912
		(0.7885)		(0.7894)		(0.7910)

Note: For each sample size, the critical values in the first row are $\hat{k}_c(\alpha, n, \phi_1)$. If $\hat{\lambda}_c$ differs from $\tilde{\lambda}_c$, then the corresponding critical values for the latter, namely, $\tilde{k}_c(\alpha, n, \phi_1)$ appear in parentheses in the second row. In this table, $\phi_1 = b$.

Table A5. Critical values—Weibull distribution.

Sig. Level (α):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***k* = 0.5**			
10	0.4315	0.2901	0.5038	0.3546	0.6348	0.4895
	(0.4318)	(0.2695)	(0.5047)	(0.3386)	(0.6354)	(0.4801)
25	0.2976	0.1738	0.3378	0.2049	0.4180	0.2718
	(0.2971)	(0.1565)	(0.3374)	(0.1878)	(0.4191)	(0.2574)
50	0.2459	0.1315	0.2708	0.1502	0.3223	0.1895
	(0.2454)	(0.1193)	(0.2705)	(0.1372)	(0.3233)	(0.1765)
100	0.2153	0.1080	0.2317	0.1196	0.2642	0.1431
	(0.2150)	(0.0990)	(0.2315)	(0.1099)	(0.2643)	(0.1329)
150	0.2029	0.0985	0.2158	0.1076	0.2412	0.1259
	(0.2027)	(0.0913)	(0.2159)	(0.0997)	(0.2416)	(0.1167)
200	0.1962	0.0934	0.2074	0.1011	0.2282	0.1157
	(0.1961)	(0.0871)	(0.2069)	(0.0940)	(0.2286)	(0.1079)
250	0.1917	0.0903	0.2016	0.0968	0.2203	0.1100
	(0.1915)	(0.0843)	(0.2015)	(0.0905)	(0.2206)	(0.1028)
300	0.1883	0.0878	0.1975	0.0938	0.2146	0.1055
	(0.1882)	(0.0823)	(0.1973)	(0.0879)	(0.2144)	(0.0989)
500	0.1809	0.0820	0.1876	0.0866	0.2001	0.0954
	(0.1809)	(0.0778)	(0.1874)	(0.0818)	(0.2002)	(0.0896)
1000	0.1735	0.0764	0.1781	0.0796	0.1865	0.0854
	(0.1736)	(0.0735)	(0.1781)	(0.0762)	(0.1868)	(0.0813)
10,000	0.1625	0.0681	0.1639	0.0691	0.1664	0.0709
	(0.1625)	(0.0671)	(0.1639)	(0.0679)	(0.1664)	(0.0694)
			***k* = 1.5**			
10	0.8918	0.8121	0.9136	0.8462	0.9442	0.8971
	(0.8891)	(0.8122)	(0.9110)	(0.8458)	(0.9423)	(0.8956)
25	0.8405	0.7368	0.8586	0.7621	0.8878	0.8055
	(0.8387)	(0.7374)	(0.8564)	(0.7624)	(0.8858)	(0.8057)
50	0.8144	0.7009	0.8283	0.7194	0.8533	0.7538
	(0.8130)	(0.7015)	(0.8268)	(0.7200)	(0.8512)	(0.7540)
100	0.7961	0.6770	0.8064	0.6904	0.8250	0.7149
	(0.7952)	(0.6773)	(0.8053)	(0.6907)	(0.8233)	(0.7152)
150	0.7882	0.6666	0.7968	0.6777	0.8123	0.6984
	(0.7875)	(0.6670)	(0.7959)	(0.6781)	(0.8111)	(0.6989)
200	0.7836	0.6607	0.7911	0.6702	0.8048	0.6881
	(0.7829)	(0.6611)	(0.7902)	(0.6706)	(0.8037)	(0.6885)
250	0.7803	0.6566	0.7872	0.6655	0.7995	0.6813
	(0.7797)	(0.6570)	(0.7864)	(0.6659)	(0.7986)	(0.6819)
300	0.7780	0.6536	0.7843	0.6618	0.7957	0.6766
	(0.7774)	(0.6539)	(0.7837)	(0.6621)	(0.7949)	(0.6770)
500	0.7725	0.6468	0.7774	0.6529	0.7865	0.6644
	(0.7721)	(0.6471)	(0.7769)	(0.6531)	(0.7857)	(0.6647)
1000	0.7670	0.6398	0.7705	0.6441	0.7771	0.6522
	(0.7667)	(0.6400)	(0.7701)	(0.6443)	(0.7766)	(0.6527)
10,000	0.7580	0.6287	0.7591	0.6302	0.7613	0.6327
	(0.7579)	(0.6288)	(0.7590)	(0.6302)	(0.7611)	(0.6328)
Sig. Level (α):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***k* = 2.0**			
10	0.9365	0.8842	0.9497	0.9067	0.9680	0.9389
	(0.9343)	(0.8831)	(0.9481)	(0.9053)	(0.9678)	(0.9373)
25	0.9041	0.8325	0.9156	0.8505	0.9340	0.8801
	(0.9024)	(0.8321)	(0.9142)	(0.8497)	(0.9329)	(0.8791)
50	0.8869	0.8069	0.8962	0.8206	0.9121	0.8449
	(0.8857)	(0.8068)	(0.8949)	(0.8201)	(0.9110)	(0.8438)
100	0.8747	0.7894	0.8815	0.7992	0.8938	0.8172
	(0.8738)	(0.7893)	(0.8806)	(0.7990)	(0.8928)	(0.8167)
150	0.8692	0.7817	0.8751	0.7900	0.8853	0.8053
	(0.8686)	(0.7817)	(0.8742)	(0.7899)	(0.8845)	(0.8048)
200	0.8661	0.7774	0.8713	0.7846	0.8805	0.7977
	(0.8654)	(0.7774)	(0.8706)	(0.7844)	(0.8796)	(0.7975)
250	0.8638	0.7744	0.8685	0.7809	0.8768	0.7929
	(0.8633)	(0.7743)	(0.8679)	(0.7808)	(0.8761)	(0.7926)
300	0.8623	0.7720	0.8666	0.7781	0.8744	0.7893
	(0.817)	(0.7720)	(0.8660)	(0.7781)	(0.8736)	(0.7891)
500	0.8585	0.7669	0.8618	0.7715	0.8682	0.7801
	(0.8581)	(0.7669)	(0.8614)	(0.7714)	(0.8675)	(0.7800)
1000	0.8546	0.7616	0.8571	0.7649	0.8616	0.7712
	(0.8543)	(0.7616)	(0.8567)	(0.7649)	(0.8611)	(0.7711)
10,000	0.8484	0.7532	0.8492	0.7542	0.8507	0.7563
	(0.8483)	(0.7532)	(0.8491)	(0.7542)	(0.8505)	(0.7563)

Table A5. *Cont.*

			***k* = 2.5**			
10	0.9585	0.9222	0.9673	0.9379	0.9794	0.9598
	(0.9573)	(0.9206)	(0.9667)	(0.9363)	(0.9802)	(0.9585)
25	0.9364	0.8852	0.9444	0.8982	0.9570	0.9193
	(0.9354)	(0.8844)	(0.9437)	(0.8971)	(0.0571)	(0.9180)
50	0.9244	0.8663	0.9309	0.8764	0.9419	0.8944
	(0.9236)	(0.8657)	(0.9302)	(0.8764)	(0.9416)	(0.8931)
100	0.9158	0.8532	0.9206	0.8606	0.9292	0.8740
	(0.9152)	(0.8528)	(0.9200)	(0.8600)	(0.9286)	(0.8732)
150	0.9119	0.8475	0.9160	0.8537	0.9233	0.8648
	(0.9114)	(0.8471)	(0.9156)	(0.8533)	(0.9228)	(0.8644)
200	0.9096	0.8441	0.9133	0.8495	0.9199	0.8595
	(0.9092)	(0.8438)	(09129)	(0.8492)	(0.9195)	(0.8589)
250	0.9080	0.8418	0.9113	0.8467	0.9173	0.8557
	(0.9076)	(0.8415)	(0.9110)	(0.8464)	(0.9168)	(0.8553)
300	0.9068	0.8400	0.9099	0.8447	0.9155	0.8530
	(0.9065)	(0.8398)	(0.9096)	(0.8444)	(0.9151)	(0.8525)
500	0.9041	0.8361	0.9065	0.8396	0.9111	0.8462
	(0.9039)	(0.8359)	(0.9063)	(0.8394)	(0.9107)	(0.8458)
1000	0.9014	0.8320	0.9031	0.8346	0.9064	0.8394
	(0.9012)	(0.8319)	(0.9029)	(0.8344)	(0.9061)	(0.8391)
10,000	0.8968	0.8255	0.8974	0.8264	0.8985	0.8279
	(0.8967)	(0.8255)	(0.8973)	(0.8263)	(0.8984)	(0.8279)
Sig. Level (α):	**10%**		**5%**		**1%**	
n	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**	***c* = 1**	***c* = 2**
			***k* = 3.0**			
10	0.9708	0.9444	0.9771	0.9558	0.9856	0.9717
	(0.9703)	(0.9429)	(0.9772)	(0.9545)	(0.9873)	(0.9707)
25	0.9549	0.9168	0.9607	0.9266	0.9698	0.9422
	(0.9545)	(0.9158)	(0.9606)	(0.9254)	(0.9708)	(0.9412)
50	0.9461	0.9024	0.9508	0.9101	0.9588	0.9237
	(0.9458)	(0.9016)	(0.9507)	(0.9093)	(0.9593)	(0.9226)
100	0.9397	0.8922	0.9432	0.8980	0.9496	0.9083
	(0.9394)	(0.8917)	(0.9431)	(0.8974)	(0.9496)	(0.9074)
150	0.9368	0.8878	0.9399	0.8926	0.9453	0.9013
	(0.9366)	(0.8874)	(0.9397)	(0.8921)	(0.9452)	(0.9006)
200	0.9350	0.8852	0.9379	0.8895	0.9427	0.8971
	(0.9349)	(0.8848)	(0.9377)	(0.8890)	(0.9427)	(0.8965)
250	0.9339	0.8834	0.9363	0.8873	0.9408	0.8942
	(0.9337)	(0.8830)	(0.9362)	(0.8868)	(0.9407)	(0.8937)
300	0.9330	0.8820	0.9353	0.8856	0.9395	0.8920
	(0.9328)	(0.8817)	(0.9352)	(0.8853)	(0.9394)	(0.8916)
500	0.9310	0.8789	0.9328	0.8817	0.9362	0.8869
	(0.9309)	(0.8787)	(0.9327)	(0.8814)	(0.9360)	(0.8864)
1000	0.9289	0.8758	0.9302	0.8778	0.9326	0.8815
	(0.9288)	(0.8756)	(0.9301)	(0.8776)	(0.9326)	(0.8812)
10,000	0.9254	0.8706	0.9259	0.8713	0.9267	0.8725
	(0.9254)	(0.8706)	(0.9258)	(0.8712)	(0.9267)	(0.8724)
			***k* = 3.5**			
10	0.9784	0.9583	0.9831	0.9671	0.9894	0.9790
	(0.9784)	(09571)	(0.9837)	(0.9660)	(0.9916)	(0.9768)
25	0.9664	0.9371	0.9708	0.9447	0.9776	0.9567
	(0.9664)	(0.9361)	(0.9712)	(0.9438)	(0.9792)	(0.9559)
50	0.9597	0.9258	0.9633	0.9319	0.9694	0.9425
	(0.9597)	(0.9250)	(0.9636)	(0.9311)	(0.9704)	(0.9416)
100	0.9548	0.9178	0.9575	0.9223	0.9623	0.9304
	(0.9548)	(0.9172)	(0.9577)	(0.9217)	(0.9628)	(0.9296)
150	0.9525	0.9142	0.9549	0.9180	0.9590	0.9248
	(0.9525)	(0.9138)	(0.9550)	(09175)	(0.9594)	(0.9242)
200	0.9512	0.9121	0.9533	0.9156	0.9570	0.9216
	(0.9512)	(0.9117)	(0.9534)	(0.9150)	(0.9574)	(0.9210)
250	0.9502	0.9106	0.9522	0.9138	0.9556	0.9192
	(0.9503)	(0.9103)	(0.9523)	(0.9133)	(0.9559)	(0.9187)
300	0.9496	0.9096	0.9514	0.9124	0.9546	0.9175
	(0.9496)	(0.9093)	(0.9514)	(0.9121)	(0.9547)	(0.9171)
500	0.9480	0.9071	0.9494	0.9093	0.9520	0.9135
	(0.9480)	(0.9069)	(0.9495)	(0.9090)	(0.9521)	(0.9131)
1000	0.9464	0.9046	0.9474	0.9062	0.9493	0.9092
	(0.9464)	(0.9044)	(0.9474)	(0.9059)	(0.9494)	(0.9089)
10,000	0.9437	0.9004	0.9440	0.9010	0.9447	0.9020
	(0.9437)	(0.9004)	(0.9440)	(0.9009)	(0.9447)	(0.9019)

Table A5. *Cont.*

Sig. Level (α):	10%		5%		1%	
n	*c* = 1	*c* = 2	*c* = 1	*c* = 2	*c* = 1	*c* = 2
			k = 4.0			
10	0.9834	0.9680	0.9870	0.9751	0.9919	0.9845
	(0.9838)	(0.9669)	(0.9881)	(0.9743)	(0.9945)	(0.9845)
25	0.9740	0.9508	0.9774	0.9569	0.9827	0.9664
	(0.9744)	(0.9500)	(0.9783)	(0.9561)	(0.9848)	(0.9659)
50	0.9687	0.9418	0.9716	0.9467	0.9764	0.9550
	(0.9691)	(0.9411)	(0.9723)	(0.9460)	(0.9777)	(0.9544)
100	0.9648	0.9352	0.9670	0.9389	0.9708	0.9454
	(0.9651)	(0.9348)	(0.9675)	(0.9384)	(0.9717)	(0.9449)
150	0.9631	0.9323	0.9653	0.9355	0.9689	0.9409
	(0.9632)	(0.9320)	(0.9653)	(0.9350)	(0.9689)	(0.9405)
200	0.9620	0.9306	0.9637	0.9335	0.9666	0.9383
	(0.9622)	(0.9303)	(0.9640)	(0.9330)	(0.9672)	(0.9379)
250	0.9612	0.9294	0.9628	0.9320	0.9655	0.9364
	(0.9614)	(0.9291)	(0.9630)	(0.9316)	(0.9659)	(0.9360)
300	0.9607	0.9286	0.9621	0.9309	0.9647	0.9351
	(0.9608)	(0.9283)	(0.9624)	(0.9306)	(0.9650)	(0.9347)
500	0.9595	0.9266	0.9606	0.9284	0.9627	0.9318
	(0.9596)	(0.9263)	(0.9608)	(0.9281)	(0.9629)	(0.9314)
1000	0.9582	0.9245	0.9590	0.9258	0.9605	0.9282
	(0.9582)	(0.9243)	(0.9591)	(0.9256)	(0.9607)	(0.9280)
10,000	0.9560	0.9211	0.9563	0.9215	0.9568	0.9223
	(0.9560)	(0.9210)	(0.9563)	(0.9214)	(0.9569)	(0.9222)

Note: For each sample size, the critical values in the first row are $\hat{k}_c(\alpha, n, \phi_1)$. If $\hat{\lambda}_c$ differs from $\tilde{\lambda}_c$, then the corresponding critical values for the latter, namely, $\tilde{k}_c(\alpha, n, \phi_1)$ appear in parentheses in the second row. In this table, $\phi_1 = k$.

References

1. Cox, D.R. Some Sampling Problems in Technology. In *New Developments in Survey Sampling*; Johnson, N.L., Smith, H., Eds.; Wiley: New York, NY, USA, 1969.
2. Fisher, R.A. The Effect of Methods of Ascertainment upon the Estimation of Frequencies. *Ann. Eugen.* **1934**, *6*, 13–25.
3. Rao, C.R. On Discrete Distributions Arising out of Methods of Ascertainment. *Sankhyā Indian J. Stat. A* **1965**, *27*, 311–324.
4. Wicksell, S.D. The Corpuscle Problem: A Mathematical Study of a Biometric Problem. *Biometrika* **1925**, *17*, 84–99.
5. Beghriche, A.; Zeghdoudi, H. A Size Biased Gamma Lindley Distribution. *Thailand Stat.* **2019**, *17*, 179–189.
6. Das, K.K.; Roy, T.D. On Some Length-Biased Weighted Weibull Distributions. *Adv. Appl. Sci. Res.* **2011**, *2*, 465–475.
7. Das, K.K.; Roy, T.D. Applicability of Length Biased Weighted Generalized Rayleigh Distribution. *Adv. Appl. Sci. Res.* **2011**, *2*, 320–327.
8. Gove, J.H. Some Observations on Fitting Assumed Diameter Distributions to Horizontal Point Sampling Data. *Can. J. For. Res.* **2000**, *30*, 521–533.
9. Gove, J.H. Moment and Maximum Likelihood Estimators for Weibull Distributions under Length- and Area-Biased Sampling. *Environ. Ecol. Stat.* **2003**, *10*, 455–467.
10. Gove, J.H. Estimation and Applications of Size-Biased Distributions in Forestry. In *Modeling Forest Systems*; Amaro, A., Reed, D., Soares, P., Eds.; CABI Publishing: Wallingford, UK, 2003; pp. 201–212.
11. Nowell, C.; Evans, M.A.; McDonald, L. Length-Biased Sampling in Contingent Valuation Studies. *Land Econ.* **1988**, *64*, 367–371.
12. Nowell, C.; Stanley, L.R. Length-Biased Sampling in Mall Intercept Surveys. *J. Mark. Res.* **1991**, *28*, 475–479.
13. Salant, S.W. Search Theory and Duration Data: A Theory of Sorts. *Q. J. Econ.* **1977**, *91*, 39–57.
14. Simon, R. Length Biased Sampling in Etiologic Studies. *Am. J. Epidemiol.* **1980**, *111*, 444–452.
15. Tzavelas, G.; Batsidis, A.; Economou, P. Size Biased Fréchet Distribution: Properties and Statistical Inference. *J. Stat. Theory Appl.* **2024**, *23*, 456–479.
16. Tzavelas, G.; Douli, M.; Economou, P. Model Misspecification Effects for Biased Samples. *Metrika* **2017**, *80*, 171–185.
17. Patil, G.P.; Ord, J.K. On Size-Biased Sampling and Related Form-Invariant Weighted Distributions. *Sankhyā Indian J. Stat. B* **1976**, *38*, 48–61.
18. Akman, O.; Gamage, J.; Jannot, J.; et al. A Simple Test for Detection of Length-Biased Sampling. *J. Biostat.* **2007**, *1*, 189–195.
19. Economou, P.; Tzavelas, G. Sample Tests for Detection of Size-Biased Sampling Mechanism. *Commun. Stat. Theory Methods* **2013**, *42*, 3280–3295.
20. Navarro, J.; Ruiz, J.; del Aguila, Y. How to Detect Biased Samples? *Biom. J.* **2003**, *45*, 91–112.

21. Giles, D.E. Some Implications of Preliminary-Test Estimation in the Context of Size-Biased Sampling. In *Advances in Shrinkage and Penalized Estimation Strategies: Honoring the Contributions of A. K. Md. Ehsanes Saleh*; Arashi, M., Norouzirad, M., Eds.; Springer: Cham, Switzerland, 2026; pp. 131–157.
22. Economou, P.; Tzavelas, G. Kullback–Leibler Divergence Measure Based Tests Concerning the Biasness in a Sample. *Stat. Methodol.* **2014**, *21*, 88–108.
23. Stacy, E.W. A Generalization of the Gamma Distribution. *Ann. Math. Stat.* **1962**, *33*, 1187–1192.
24. Ducey, M.J.; Gove, J.H. Size-Biased Distributions in the Generalized Beta Distribution Family, with Applications to Forestry. *Forestry* **2015**, *88*, 143–151.
25. Ajami, M.; Jahanshahi, S.M.A. Comparison of Various Estimation Methods for Size-Biased Rayleigh Distribution. In Proceedings of the 13th Iranian Statistical Conference, Kerman, Iran, 24–26 August 2016.
26. Bashir, S.; Rasul, M. Record Values from Size-Biased Half Normal Distribution: Properties and Recurrence Relations for the Single and Product Moments. *Am. J. Comput. Commun. Control* **2018**, *5*, 1–6.
27. Bashir, S.; Rasul, M. A New Weighted Rayleigh Distribution: Properties and Applications on Lifetime Time Data. *Open J. Stat.* **2018**, *8*, 640–650.
28. Blumenthal, S. Proportional Sampling in Life Length Studies. *Technometrics* **1967**, *9*, 205–218.
29. Ducey, M.J. Sampling Trees with Probability Nearly Proportional to Biomass. *For. Ecol. Manag.* **2009**, *258*, 2110–2116.
30. Giles, D.E. Improved Maximum Likelihood Estimation for the Weibull Distribution under Length-Biased Sampling. *J. Quant. Econ.* **2021**, *19*, 59–77.
31. Jabeen, S.; Jan, T.R. Information Measures of Size Biased Generalized Gamma Distribution. *Int. J. Sci. Eng. Appl. Sci.* **2015**, *1*, 2395–3470.
32. Mir, K.A.; Ahmed, A.; Reshi, J.A. On Size Biased Exponential Distribution. *J. Mod. Math. Stat.* **2013**, *7*, 21–25.
33. Perveen, Z.; Ahmad, M. Some Properties of Size-Biased Weighted Weibull Distribution. *Int. J. Adv. Appl. Sci.* **2018**, *5*, 92–98.
34. Scheaffer, R.L. Size-Biased Sampling. *Technometrics* **1972**, *14*, 635–644.
35. Tzavelas, G.; Panagiotakos, D. Statistical Inference for the Size-Biased Weibull Distribution. *J. Stat. Comput. Simul.* **2013**, *83*, 1252–1265.
36. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025.
37. Saldanha, M.H.J.; Suzuki, A.K. Package 'ggamma'—The Generalized Gamma Probability Distribution. Available online: https://ftp.fau.de/cran/web/packages/ggamma/ggamma.pdf (accessed on 6 July 2026).
38. Shen, Y.; Ning, J.; Qin, J. Analyzing Length-Biased Data with Semiparametric Transformation and Accelerated Failure Time Models. *J. Am. Stat. Assoc.* **2009**, *104*, 1192–1202.
39. Batsidis, A.; Economou, P.; Tzavelas, G. Tests of Fit for a Lognormal Distribution. *J. Stat. Comput. Simul.* **2015**, *86*, 215–235.