

Chapter 1

Likelihood-Based Inference and Finite-Sample Corrections: A Brief Overview

Abstract This chapter introduces the likelihood function and estimation by maximum likelihood. Some important properties of maximum likelihood (ML) estimators are outlined. We also briefly present several important concepts that will be used throughout the book. Three asymptotic testing criteria are also introduced. The chapter also motivates the use of Bartlett and Bartlett-type corrections. The underlying idea is to transform the test statistic in such a way that its null distribution is better approximated by the reference χ^2 distribution. We also investigate the use of bias corrections. They are used to reduce systematic errors in the point estimation process. Finally, we motivate the use of a data resampling method: the bootstrap.

Keywords Bartlett correction · Bartlett-type correction · Bias correction · Bootstrap · Likelihood ratio test · Maximum likelihood · Score test · Wald test

1.1 Introduction

Statistics deals with measurement under uncertainty. Its ultimate goal is to perform inference on a population or phenomenon from which data can be sampled. This is achieved by first considering a model that represents the phenomenon of interest. A model is a simplified representation of a more comprehensive reality. A good model must retain the most important features of the phenomenon it represents. A statistical model has a stochastic component—since it represents a phenomenon that occurs in uncertain fashion—and is typically indexed by fixed and (usually) unknown quantities known as parameters. Statistical inference is then performed on such parameters using data previously collected. By performing inference on the model parameters, we make inference on the model and hence on the phenomenon it is supposed to describe.

Inference can be carried out in three different ways, namely (1) point estimation, (2) interval estimation, and (3) hypothesis testing. Several approaches for

parameter point estimation were proposed in the literature, the maximum likelihood (ML) method being the most commonly employed. The maximum likelihood estimator (MLE) enjoys desirable properties and can be used when constructing confidence intervals and regions and also in test statistics.

In what follows, let Y be a random variable whose density function with respect to the Lebesgue or the counting measure on the real line is $f_Y(\cdot; \theta)$ which we shall also write as $f(\cdot; \theta)$. Here, θ is the parameter vector that indexes the distribution. It is usually unknown and takes values in the parameter space Θ . We shall also consider random samples obtained from $f(\cdot; \theta)$, which shall be denoted as Y_1, \dots, Y_n . We wish to perform inference on θ using the n -dimensional sample (Y_1, \dots, Y_n) . An estimator is a quantity that depends on the data and optionally on known quantities that can be used to estimate θ (or a given function of θ).

1.2 Likelihood Inference

One of the most widely used estimation methods is the ML method. Its underlying motivation is simple and intuitive. Let Y_1, \dots, Y_n be an n -dimensional sample and assume that each variate has probability density function (pdf) $f(\cdot; \theta)$, θ being a p -dimensional vector, i.e., $\theta = (\theta_1, \dots, \theta_p)^\top \in \Theta$. Assume that Y_1, \dots, Y_n are independent and identically distributed (i.i.d.). Their observed values are denoted as y_1, \dots, y_n .

The likelihood function is the joint density function $f_{Y_1, \dots, Y_n}(y_1, \dots, y_n; \theta)$ considered as a function of the parameter $\theta = (\theta_1, \dots, \theta_p)^\top$. Since the n variates in our sample are i.i.d., the likelihood function is the product of the n marginal pdfs. We shall denote the likelihood function as $L(\theta; Y_1, \dots, Y_n)$. The MLE is the value of θ in the parameter space Θ which maximizes the likelihood function, if such a value exists. It is noteworthy that the MLE also maximizes the log-likelihood function, $\ell = \ell(\theta; Y_1, \dots, Y_n) = \log L(\theta; Y_1, \dots, Y_n)$. The log-likelihood derivative with respect to θ is known as the score function. It is possible to show that

$$\mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \right) = 0,$$

that is, the score function has mean zero. The Fisher information matrix is

$$K(\theta) = \mathbb{E} \left(\frac{\partial \ell}{\partial \theta} \frac{\partial \ell}{\partial \theta^\top} \right).$$

Under the conditions given below, it can be shown that $K(\theta) = \mathbb{E}(-\partial^2 \ell / \partial \theta \partial \theta^\top)$. Notice that the Fisher information equals the score function variance.

The relevant assumptions for ML inference can be stated as follows:

1. The parameter space Θ is open, and the log-likelihood function assumes a global maximum in Θ .
2. For almost all y , the fourth-order derivatives $\partial^4 f(y; \theta) / \partial \theta_h \partial \theta_j \partial \theta_k \partial \theta_l$ ($h, j, k, l \in \{1, \dots, p\}$) exist and are continuous in an open neighborhood $M \subset \Theta$ that contains the true parameter value.
3. Consider the integral with respect to y of a function that can be represented as a polynomial in one or more variables in f , $\log f$ and their derivatives of any order with respect to θ . The derivative of such an integral with respect to any component of θ can be obtained by differentiating inside the integral.
4. Consider the $p \times p$ matrix $K = K(\theta)$, whose (h, j) element is

$$K_{hj}(\theta) = \mathbb{E} \left(\frac{\partial \ell}{\partial \theta_h} \frac{\partial \ell}{\partial \theta_j} \right),$$

$h, j = 1, \dots, p$. Here, K is positive definite and finite for all $\theta \in M$.

5. There exist functions M_{hjk} such that

$$\left| \frac{\partial^3 \log f(y; \theta)}{\partial \theta_h \partial \theta_j \partial \theta_k} \right| \leq M_{hjk}(y),$$

for all $\theta \in M$ and for all $h, j, k = 1, \dots, p$, and $\mathbb{E}_0[M_{hjk}(Y)] < \infty$, for all $h, j, k = 1, \dots, p$, where \mathbb{E}_0 denotes expectation under the true parameter value.

These regularity conditions hold in most, nearly all applications. They are thus not restrictive.

1.3 Some Properties of Maximum Likelihood Estimators

An important property of the MLE is its invariance. Let $\hat{\theta}$ be the MLE of θ , and let $g(\cdot)$ be a function from \mathbb{R}^p to \mathbb{R}^s (not necessarily one to one). Then, $g(\hat{\theta})$ is the MLE of $g(\theta)$. For instance, suppose that $\hat{\sigma}^2$ is the MLE of a given variance. Then, $\sqrt{\hat{\sigma}^2}$ is the MLE of σ (standard deviation).

The MLE enjoys other important properties, including large sample ones. It is noteworthy that it is consistent for θ , i.e., $\hat{\theta}_n \xrightarrow{P} \theta$, where \xrightarrow{P} denotes convergence in probability and the subscript n indicates dependence on the sample size. This property means that in large samples, the estimator will be close to the true parameter with high probability.

Another important property is related to the asymptotic distribution of the MLE. It follows that

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N} (0, K^{-1}(\theta)),$$

where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution (or law). Notice that the inverse of Fisher's information equals the MLE asymptotic variance and that the estimator is asymptotically Gaussian. It should also be noted that the MLE is asymptotically efficient; that is, its asymptotic variance achieves the Cramér-Rao lower bound; see Bickel and Doksum (2001, pp. 181–182).

1.4 A Simple Example

Let y_1, \dots, y_n be a random sample from the beta distribution $\mathcal{B}(a, b)$. The density of y_i , for each $i = 1, \dots, n$, is

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1},$$

where $0 < y < 1$, $a > 0$, $b > 0$ and $\Gamma(\cdot)$ is the gamma function. The interest lies in the estimation of $\theta = (a, b)^\top$. The log-likelihood function is

$$\ell(a, b) = n \left\{ (a-1) \log g_1 + (b-1) \log g_2 + \log \left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right) \right\},$$

where the sufficient statistics g_1 and g_2 are the geometric means of the y_i 's and $(1 - y_i)$'s, respectively.

The MLEs \hat{a} and \hat{b} of a and b are the solution to the nonlinear system

$$\begin{aligned} \psi(a) - \psi(a+b) &= \log g_1 \\ \psi(b) - \psi(a+b) &= \log g_2, \end{aligned}$$

where $\psi(\cdot)$ denotes the digamma function, i.e., the first derivative of the log-gamma function. The MLEs of a and b cannot be expressed in closed form. They are obtained by numerically maximizing the log-likelihood function using a nonlinear optimization algorithm.

We shall now compute the estimates of the parameters that index the beta law using the R software (<http://www.R-project.org>). To that end, we shall use the `fitdistr` function of the MASS package.

```
> library(MASS)
> set.seed(16851750) # random number generator seed
> randomsample <- rbeta(100, shape1=0.5, shape2=1.5)
> MLfit = fitdistr(random, dbeta, start=list(shape1=1,
+ shape2=1), lower=c(0,0))
> MLfit
      shape1      shape2
0.50056877  1.50584321
(0.05925167) (0.22864059)
```

In this example, the true parameter values are $a = 0.5$ and $b = 1.5$ and the sample size is $n = 100$. The MLEs are $\hat{a} = 0.501$ and $\hat{b} = 1.506$. The number in parentheses are the standard errors of the estimates, which are obtained as the square roots of the diagonal elements of Fisher's information matrix inverse after the unknown parameters are replaced by their MLEs.

1.5 Likelihood-Based Testing Inference

Let θ denote the p -dimensional parameter vector that indexes the model used to represent the population or phenomenon of interest and partition it as $\theta = (\theta_1^\top, \theta_2^\top)^\top$, $\dim(\theta_1) = q$ and $\dim(\theta_2) = p - q$. Suppose our interest lies in making testing inference on θ_1 , i.e., we wish to test $H_0 : \theta_1 = \theta_1^0$ against $H_1 : \theta_1 \neq \theta_1^0$, where θ_1^0 is a given q -vector. We say θ_1 is the parameter of interest and θ_2 is the nuisance parameter. For instance, we have a random sample from the beta distribution $\mathcal{B}(a, b)$ and wish to test $H_0 : b = 1$ against a two-sided alternative. Here, b is the parameter of interest and a is the nuisance parameter; additionally, $\theta_1^0 = 1$.

Let $\hat{\theta} = (\hat{\theta}_1^\top, \hat{\theta}_2^\top)^\top$ and $\tilde{\theta} = (\theta_1^{0\top}, \hat{\theta}_2^\top)^\top$ denote the unrestricted and restricted MLEs of θ , respectively. The restricted MLE is obtained by imposing $\theta_1 = \theta_1^0$ and maximizing $\ell(\theta)$ over θ_2 . The likelihood ratio (LR) test statistic is

$$LR = 2 \left[\ell(\hat{\theta}) - \ell(\tilde{\theta}) \right]. \quad (1.1)$$

When the null hypothesis is true, little is lost by imposing it when estimating the parameter vector and, as a consequence, the likelihood function evaluated at the unrestricted and restricted MLEs should be approximately the same; that is, the difference between the likelihood functions evaluated at the two point estimates is small. It then follows that LR is small and the null hypothesis is not rejected. Notice that likelihood ratio testing inference requires the estimation of both null and non-null models.

Alternatively, the null hypothesis can be tested using Rao's score test (Rao 1948). The underlying idea is that if the null hypothesis is true, the score function (i.e., the log-likelihood derivative) should be close to zero when evaluated at the restricted MLE, $\tilde{\theta}$. The score test statistic is

$$S_R = s(\tilde{\theta})^\top K^{-1}(\tilde{\theta}) s(\tilde{\theta}), \quad (1.2)$$

where $s(\theta) = \partial \ell(\theta) / \partial \theta$ is the score function. Notice that the score function and Fisher's information matrix are evaluated at the restricted MLE, $\tilde{\theta}$. It then follows that only the null model is estimated.

It is also possible to base our testing inference on a different test statistic, namely the Wald statistic. It is given by

$$W = (\hat{\theta}_1 - \theta_1^0)^\top K^{11}(\hat{\theta})^{-1}(\hat{\theta}_1 - \theta_1^0), \quad (1.3)$$

where $K^{11}(\theta)^{-1}$ is the upper $q \times q$ block of Fisher's information matrix inverse; i.e., it is the upper q -dimensional block of $K^{-1}(\theta)$. Thus, the null hypothesis is not rejected when the distance between $\hat{\theta}_1$ and θ_1^0 is small. Notice that in order to compute the Wald test statistic, one only estimates the non-null (unrestricted) model.

The exact null distributions of the test statistics given in (1.1), (1.2) and (1.3) are usually unknown. Under certain regularity conditions (Bickel and Doksum 2001, Chap. 6; Serfling 1978, Chap. 4), however, it can be established that they converge to χ_q^2 as $n \rightarrow \infty$. It is then possible to base our testing inference on critical values obtained from such a distribution. It follows that, for $T = LR, S_R, W$ and under the null hypothesis, $\Pr(T \leq \chi_{1-\alpha; q}^2) = \alpha + o(1)$, where α is the test nominal significance level and $\chi_{1-\alpha; q}^2$ is the $(1-\alpha)$ th χ_q^2 upper quantile. That is, the difference between the probability that $T \leq \chi_{1-\alpha; q}^2$ (null rejection rate) and α (the nominal significance level) vanishes as $n \rightarrow \infty$.

1.6 Some Remarks on Bartlett and Bartlett-Type Corrections

The likelihood function $L(\theta) = L(\theta; y)$ is the basis for most methods of statistical inference. A natural rule is to base inference on $L(\hat{\theta})/L(\theta) > c$ in order to decide what is the range of 'plausible' values of θ , where θ is assumed to have dimension p and $\hat{\theta}$ is the MLE of θ . Inference based on the likelihood function can also be calibrated with reference to the probability model $f(y; \theta)$, by examining the distribution of $L(\theta)$ as a random function, or more usually, by examining the distribution of various associated quantities. Let $\ell(\theta) = \log[L(\theta)]$ be the log-likelihood function. The asymptotic likelihood theory is based on a version of the central limit theorem for the score function $U(\theta) = \partial \ell(\theta)/\partial \theta$ and also on Fisher's information matrix $K(\theta)$. These quantities were introduced in Sect. 1.2. If $Y = (Y_1, \dots, Y_n)^\top$ has independent components, then $U(\theta)$ is a sum of n independent components, which, under mild regularity conditions, is asymptotically normal. If $\hat{\theta}$ is consistent for θ and $L(\theta)$ has sufficient regularity, the quantities $(\hat{\theta} - \theta)^\top K(\theta)(\hat{\theta} - \theta)$, $U(\theta)^\top K(\theta)^{-1} U(\theta)$ and $2\{\ell(\hat{\theta}) - \ell(\theta)\}$ converge in distribution to χ_p^2 . It is noteworthy, however, that the use of χ_p^2 as an approximation to the true underlying distributions can lead to inaccurate inferences when the sample size is small. The book by Cox and Hinkley (1974) gives a detailed account of likelihood inference and principles of statistical inference. Other good book-length treatments of likelihood inference are Barndorff-Nielsen and Cox (1994), Pawitan (2000), Severini (2000), and Brazzale et al. (2000).

Large sample tests are commonly used in the applied statistics since exact tests are not always available. These tests rely on what is called 'first-order asymptotics'; that is, they employ critical values obtained from a known limiting null distribution. Generally speaking, the main difficulty of testing a null hypothesis using the

LR statistic lies not so much in deriving its closed-form expression—when it has one—but in finding its exact null distribution, or at least a good approximation to it. In a very influential paper, Bartlett (1937) pioneered the correction to the LR statistic in the context of comparing the variances of several populations. For regular problems, Lawley (1956), through a heroic series of calculations, obtained a general formula for the null expected value of LR and demonstrated that all cumulants of the Bartlett-corrected statistic for testing a composite hypothesis agree with those of the reference χ^2 distribution with error of order $n^{-3/2}$.¹ Alternative expressions for the Bartlett corrections were developed by DiCiccio and Stern (1993), McCullagh and Cox (1986), and Skovgaard (2001). In particular, Cordeiro (1983, 1987) was the first to provide matrix expressions for generalized linear models.

Cordeiro and Ferrari (1991) extended the idea of Bartlett corrections to other test statistics, such as the score (S_R) and Wald (W) statistics. In fact, they derived a general formula for Bartlett-type corrections to improve any test statistic that is, under the null hypothesis, asymptotically distributed as χ^2 . The standard Bartlett correction is a special case of their general result. Bartlett and Bartlett-type corrections intend to bring the empirical sizes of asymptotic tests close to the corresponding nominal sizes. In most cases, they do so quite effectively. It is important to bear in mind that these corrections can lead to a loss in power. However, an important result is that the untransformed statistic and its Bartlett-corrected version have the same local power to order n^{-2} . More precisely, let S be a test statistic which is χ^2 distributed under the null hypothesis and let S^* denote the Bartlett-corrected statistic obtained as a transformation of S . Then, under local (Pitman) alternatives, $\Pr(S^* \geq x) = \Pr(S \geq x) + \mathcal{O}(n^{-2})$.

In this book, we shall restrict ourselves to the LR, score test, and Wald test, since they are the most commonly used large sample testing inference. As is well known, these three statistics are asymptotically distributed as χ^2 when the null hypothesis H_0 is true, where q is the number of restrictions under test. However, it is also well known that this first-order approximation may not be accurate in finite samples, thus leading to size distortions. We address the issue of evaluating such approximation and designing more accurate tests. The question ‘Can we do better?’ can be approached from two distinct viewpoints. First, we can obtain a new test statistic whose null distribution is better approximated by the first-order limiting distribution. Second, we can obtain a new distribution which is ‘closer’ to the test statistic exact null distribution. In this monograph, we shall focus on the former approach. Readers interested in the latter approach are referred to Barndorff-Nielsen and Cox (1979, 1989), Reid (1988, 1991), and Hall (1992) and the references therein.

One of the main goals of our monograph is to provide a unified review of the literature on Bartlett and Bartlett-type corrections, i.e., corrections that can be applied to test statistics (not to critical values). An issue of interest is how to define Bartlett-type corrections since it is possible to write the correction in different ways which are equivalent up to a certain order of magnitude. We address this issue by Monte Carlo simulation. We also include discussions on how to obtain the corrections in

¹ Henceforth, ‘to order n^{-k} ’ means that terms of order smaller than n^{-k} are neglected.

regression models, such as generalized linear models, Birnbaum-Saunders nonlinear regression models, and heteroskedastic linear regressions. We use the linear regression framework to address two important issues through simulation: the influence of the covariate values and of the number of nuisance parameters on the first-order asymptotic approximation used in some asymptotic tests.

Bartlett corrections constitute an important topic of research among statisticians. However, they have not yet found their appropriate space and usage in several applied areas of statistics, in which size corrections are almost always based on transformations of critical values obtained from Edgeworth expansions. We hope this book will help narrow this gap. The authors have established general results and explicit expressions for Bartlett and Bartlett corrections in a series of joint publications, as can be seen in their Web pages: <http://www.de.ufpe.br/~gauss> (Gauss M. Cordeiro) and <http://www.de.ufpe.br/~cribri> (Francisco Cribari-Neto). Some applications of Bartlett-type corrections in regression models include score tests for generalized linear models with known dispersion (Cordeiro et al. 1993) and unknown dispersion (Cribari-Neto and Ferrari 1995), exponential family nonlinear models (Ferrari and Cordeiro 1996), and heteroskedastic t regression models (Barroso et al. 2002), among several others. A detailed account of Bartlett and Bartlett-type corrections can be found in Cribari-Neto and Cordeiro (1996).

1.7 Some Remarks on Bias Corrections

We shall also be concerned with point estimation. To that end, we shall review the literature on bias correction. Bias is a systematic error, and there are strategies that can be used to reduce it. The MLEs can be quite biased in small samples. It is thus important to evaluate the n^{-1} biases of these estimators, where n is the sample size, and then define modified estimators that are bias free to this order of approximation. In particular, it is important to derive closed-form expressions for the second-order biases of estimators in some classes of models which can be used in practical applications in order to evaluate the accuracy of these estimators and also to define estimators with smaller biases.

One of our goals in this monograph is to review the literature on bias correction of MLEs. The obvious difficulty is that many MLEs cannot be expressed as explicit functions of the data. Over the last 25 years, there have been many advances with respect to bias calculation of nonlinear MLEs in special distributions and wider classes of regression models such as generalized linear models and heteroskedastic regressions. The computation of higher-order biases is perhaps one of the most important approximations in the theory of estimation by ML in regression models.

There has been considerable interest in finding simple closed-form expressions for second-order biases of MLEs in some classes of regression models. By ‘closed-form’ we mean expressions that do not involve cumulants of log-likelihood derivatives. In fact, the $\mathcal{O}(n^{-1})$ biases of the MLEs have been derived in homoscedastic normal nonlinear models (Cook et al. 1986), generalized log-gamma regression models (Young

and Bakir 1987), generalized linear models (Cordeiro and McCullagh 1991), multiplicative regression models (Cordeiro 1993), ARMA models (Cordeiro and Klein 1994), multivariate nonlinear regression models with normal errors (Cordeiro and Vasconcellos 1997), univariate nonlinear Student's t -regression models (Cordeiro et al. 1998), multivariate Student's t -regression models (Vasconcellos and Cordeiro 2000), heteroskedastic models (Vasconcellos et al. 2000), beta regression models (Ospina et al. 2006), and heteroscedastic normal linear models (Cordeiro 2008). These results were obtained using the general formula given by Cox and Snell (1968). Simulation results on bias corrections can be found in Cordeiro and Cribari-Neto (1993). An appealing alternative approach to computer-intensive bias correction is described by MacKinnon and Smith (1998). For alternative methods, see Cadigan (1994) and Taniguchi and Puri (1995). Second- and third-order bias corrections for one-parameter models were obtained by Ferrari et al. (1996). More recent general results on bias corrections in regression models can be found in Patriota and Lemonte (2009).

1.8 Some Remarks on the Bootstrap

Even though our focus is on analytical corrections to test statistics and estimators, we also cover alternatives that are based on data resampling, more specifically on bootstrap resampling (Efron 1979). The underlying idea is that additional artificial samples can be obtained by sampling from the original sample as if we were sampling from the population. The random drawing mechanism can be of parametric or nonparametric nature. Higher precision can be achieved by using nested bootstrap schemes; see Hall and Martin (1988). For further details on the bootstrap method, see Efron and Tibshirani (1986, 1993), Hall (1992), Shao and Tu (1995) and the references therein. We also refer readers to Young (1994), who also lists the shortcomings of using data resampling. The relationship between Edgeworth expansions and the bootstrap is discussed in generality by Hall (1992). For an econometric example of this relationship, see Rayner (1990). Rocke (1989) suggested the use of a bootstrap Bartlett adjustment for the log-likelihood ratio statistic in the context of seemingly unrelated regressions. As we shall see, his proposal is to use data resampling to estimate the Bartlett correction factor.

In what follows, we shall describe how the bootstrap can be used as an alternative to analytical finite sample corrections to estimators and tests. Empirical researchers can then choose which method is more appropriate to the application at hand. It is important to note that two researchers who use the same data and perform the same analytical correction will arrive at exactly the same result. The same does not hold true, however, for the bootstrap since the final exact result will depend on the number of bootstrap replication, on the random number generator used, and on other factors. The data resampling mechanism used can also be a discrepancy source. The two approaches are thus different in nature, but they aim at the same goal: delivering more accurate and reliable inferences.

References

- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society A*, 160, 268–282.
- Barroso, L. P., Cordeiro, G. M., & Vasconcellos, K. L. P. (2002). Second-order asymptotic for score tests in heteroskedastic t regression models. *Communications in Statistics, Theory and Methods*, 31, 1515–1529.
- Barndorff-Nielsen, O. E., & Cox, D. R. (1979). Edgeworth and saddle-point approximations with statistical applications. *Journal of the Royal Statistical Society B*, 41, 279–312.
- Barndorff-Nielsen, O. E., & Cox, D. R. (1989). *Asymptotic techniques for use in statistics*. London: Chapman and Hall.
- Barndorff-Nielsen, O. E., & Cox, D. R. (1994). *Inference and asymptotics*. London: Chapman and Hall.
- Bickel, P. J., & Doksum, K. A. (2001). *Mathematical statistics: Basic ideas and selected topics* (2nd ed.). Upper Saddle River: Prentice Hall.
- Brazzale, A. R., Davison, A. C., & Reid, N. (2000). *Applied asymptotics*. Cambridge: Cambridge University Press.
- Cadigan, N. G. (1994). Bias approximation for maximum likelihood estimates. *Journal of Statistical Computation and Simulation*, 51, 89–95.
- Cook, D. R., Tsai, C. L., & Wei, B. C. (1986). Bias in nonlinear regression. *Biometrika*, 73, 615–623.
- Cordeiro, G. M. (1983). Improved likelihood ratio statistics for generalized linear models. *Journal of the Royal Statistical Society B*, 45, 404–413.
- Cordeiro, G. M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika*, 74, 265–274.
- Cordeiro, G. M. (1993). Bartlett corrections and bias correction for two heteroscedastic regression models. *Communications in Statistics, Theory and Methods*, 22, 169–188.
- Cordeiro, G. M. (2008). Corrected maximum likelihood estimators in linear heteroscedastic regression models. *Brazilian Review of Econometrics*, 28, 53–67.
- Cordeiro, G. M., & Cribari-Neto, F. (1993). On Bartlett corrections, bias reduction and a new class of transformations. *Brazilian Journal of Probability and Statistics*, 7, 179–200.
- Cordeiro, G. M., & Ferrari, S. L. P. (1991). A modified score statistic having chi-squared distribution to order n^{-1} . *Biometrika*, 78, 573–582.
- Cordeiro, G. M., Ferrari, S. L. P., & Paula, G. A. (1993). Improved score tests for generalized linear models. *Journal of the Royal Statistical Society B*, 55, 661–674.
- Cordeiro, G. M., & Klein, R. (1994). Bias correction in ARMA models. *Statistics and Probability Letters*, 19, 169–176.
- Cordeiro, G. M., & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society B*, 53, 629–643.
- Cordeiro, G. M., & Vasconcellos, K. L. P. (1997). Bias correction for a class of multivariate nonlinear regression models. *Statistics and Probability Letters*, 35, 155–164.
- Cordeiro, G. M., Vasconcellos, K. L. P., & Santos, M. L. F. (1998). On the second-order bias of parameter estimates in nonlinear regression models with student t errors. *Journal of Statistical Computation and Simulation*, 60, 363–378.
- Cribari-Neto, F., & Cordeiro, G. M. (1996). On Bartlett and Bartlett-type corrections. *Econometric Reviews*, 15, 339–367.
- Cribari-Neto, F., & Ferrari, S. L. P. (1995). Second order asymptotics for score tests in generalized linear models. *Biometrika*, 82, 426–432.
- Cox, D. R., & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- Cox, D. R., & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society B*, 30, 248–275.
- DiCiccio, T. J., & Stern, S. E. (1993). On Bartlett adjustments for approximate bayesian inference. *Biometrika*, 80, 731–740.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1–26.

- Efron, B., & Tibshirani, R. J. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54–96.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Ferrari, S. L. P., Botter, D. A., Cordeiro, G. M., & Cribari-Neto, F. (1996). Second and third order bias reduction in one-parameter family models. *Statistics and Probability Letters*, 30, 339–345.
- Ferrari, S. L. P., & Cordeiro, G. M. (1996). Corrected score tests for exponential family nonlinear models. *Statistics and Probability Letters*, 26, 7–12.
- Hall, P. (1992). *The bootstrap and the Edgeworth expansion*. New York: Springer.
- Hall, P., & Martin, M. A. (1988). On bootstrap resampling and iteration. *Biometrika*, 75, 661–671.
- Lawley, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, 71, 233–244.
- MacKinnon, J. G., & Smith, J. A. A. (1998). Approximate bias correction in econometrics. *Journal of Econometrics*, 85, 205–230.
- McCullagh, P., & Cox, D. R. (1986). Invariants and the likelihood ratio statistic. *Annals of Statistics*, 14, 1419–1430.
- Ospina, R., Cribari-Neto, F., & Vasconcellos, K. L. P. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics and Data Analysis*, 51, 960–981. [Errata: vol. 55, p. 2445, 2011].
- Patriota, A. G., & Lemonte, A. J. (2009). Bias correction in a multivariate regression model with general parameterization. *Statistics and Probability Letters*, 79, 1655–1662.
- Pawitan, Y. (2000). *In all likelihood*. Oxford: Oxford University Press.
- Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44, 50–57.
- Rayner, R. K. (1990). Bootstrap tests for generalized least squares regression models. *Economics Letters*, 34, 261–265.
- Rocke, D. M. (1989). Bootstrap Bartlett adjustment in seemingly unrelated regression. *Journal of the American Statistical Association*, 84, 598–601.
- Reid, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science*, 3, 213–238.
- Reid, N. (1991). Statistical theory and modeling: In honour of Sir David Cox, FRS. In D. V. Hinkley, N. Reid, & E. J. Snell (Eds.), *Approximations and asymptotics*. London: Chapman and Hall.
- Serfling, R. J. (1978). *Approximation theorems of mathematical statistics*. New York: Wiley.
- Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. New York: Springer.
- Severini, T. A. (2000). *Likelihood methods in statistics*. Oxford: Oxford University Press.
- Skovgaard, I. (2001). Likelihood asymptotic. *Scandinavian Journal of Statistics*, 28, 3–32.
- Taniguchi, M., & Puri, M. L. (1995). Higher order asymptotic theory for normalizing transformations of maximum likelihood estimators. *Annals of the Institute of Statistical Mathematics*, 47, 581–600.
- Vasconcellos, K. L. P., & Cordeiro, G. M. (2000). Bias corrected estimates in multivariate Student-*t* regression models. *Communications in Statistics, Simulation and Computation*, 29, 797–822.
- Vasconcellos, K. L. P., Cordeiro, G. M., & Barroso, L. P. (2000). Improved estimation for robust econometric regression models. *Brazilian Journal of Probability and Statistics*, 14, 141–157.
- Young, G. A. (1994). Bootstrap: More than a stap in the dark? *Statistical Science*, 9, 382–415.
- Young, D. H., & Bakir, S. T. (1987). Bias correction for a generalized log-gamma regression model. *Technometrics*, 29, 183–191.

An Introduction to Bartlett Correction and Bias
Reduction

Cordeiro, G.M.; Cribari-Neto, F.

2014, XI, 107 p., Softcover

ISBN: 978-3-642-55254-0