Edgeworth expansion for an estimator of the adjustment coefficient

Margarida Brito *
Centro de Matemática & Departamento de Matemática Aplicada
Fac. Ciências, Univ. Porto
Ana Cristina Moreira Freitas
Centro de Matemática & Fac. Economia
Univ. Porto

Abstract

We establish an Edgeworth expansion for an estimator of the adjustment coefficient \( R \), directly related to the geometric-type estimator for general exponential tail coefficients, proposed in Brito and Freitas (2003). Using the first term of the expansion, we construct more accurate confidence bounds for \( R \). The accuracy of the approximation is illustrated using an example from insurance (cf. Schultze and Steinebach (1996)).

Key words and phrases: Adjustment coefficient; Edgeworth expansions; Parameter estimation; Sparre Andersen model; Tail index

JEL Classification: C13, C14
AMS 2000 Classification: 62G05, 62P05

*Corresponding author. Departamento de Matemática Aplicada, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal. Tel: +351-220402202. Fax: +351-220402209. E-mail: mabrito@fc.up.pt.
1 Introduction

Let $Z_1, Z_2, \ldots$ be independent, nonnegative random variables (r.v.’s) with common distribution function (d.f.) $F$ satisfying

$$1 - F(z) = P(Z_1 > z) = r(z)e^{-Rz}, \quad z > 0,$$

where $r$ is a regularly varying function at infinity and $R$ is a positive constant. Denoting by $F^{-1}$ the left continuous inverse of $F$, i.e., $F^{-1}(s) := \inf\{x: F(x) \geq s\}$, (1) is equivalent to

$$F^{-1}(1 - s) = -\frac{1}{R} \log s + \log \tilde{L}(s), \quad 0 < s < 1,$$

where $\tilde{L}$ is a slowly varying function at zero (see e.g. Schultze and Steinebach (1996) and the references therein).

The problem of estimating $R$ or related tail indices has received considerable attention. Several estimators for $R$ have been proposed in the literature (see e.g. Dekkers et al. (1989), Bacro and Brito (1995), Csorgő and Viharos (1998) and references therein). Common applications may be found in a big variety of domains, as for example in economics, insurance, telecommunications, traffic, geology and sociology. We shall consider here the problem of estimating the adjustment coefficient in risk theory.

Based on least squares considerations, Schultze and Steinebach (1996) proposed three estimators, denoted by $\hat{R}_1(k_n)$, $\hat{R}_2(k_n)$ and $\hat{R}_3(k_n)$, for the tail coefficient $R$. Brito and Freitas (2003) have introduced a geometric-type estimator of $R$, $\hat{R}(k_n)$, which is related to the least squares estimators $\hat{R}_1(k_n)$ and $\hat{R}_3(k_n)$ and defined by

$$\hat{R}(k_n) = \frac{\sqrt{i_n(k_n)}}{\sqrt{\frac{1}{k_n} \sum_{i=1}^{k_n} Z_{n-i+1,n}^2 - \left( \frac{1}{k_n} \sum_{i=1}^{k_n} Z_{n-i+1,n} \right)^2}},$$

where

$$i_n(k_n) := \frac{1}{k_n} \sum_{i=1}^{k_n} \log^2(n/i) - \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \log(n/i) \right)^2,$$

$Z_{1,n} \leq Z_{2,n} \leq \ldots \leq Z_{n,n}$ denote the order statistics of the sample $Z_1, Z_2, \ldots, Z_n$ and $k_n$ is a sequence of positive integers satisfying

$$1 \leq k_n < n, \quad \lim_{n \to \infty} k_n = \infty \quad \text{and} \quad \lim_{n \to \infty} k_n/n = 0.$$ 

(2)

An important application in risk theory was given by Schultze and Steinebach (1996). In particular, these authors establish the consistency of $\hat{R}_1(k_n)$ and $\hat{R}_3(k_n)$ for estimating the adjustment coefficient for a subclass of the Sparre
Andersen model. More recently, Brito and Freitas (2006) have shown that the consistency properties of all these estimators still hold in the Sparre Andersen model under standard conditions.

It is also known that these estimators, when properly normalized, are asymptotically normal, with convergence rate of order $k_n^{-1/2}$. For applications, it is important to derive higher-order approximations. In this paper, our main goal is the construction of more accurate confidence bounds for $R$, using asymptotic expansions. We begin by establishing an Edgeworth expansion for a slightly modified version of $\hat{R}(k_n)$, given by

$$\tilde{R}(k_n) = \left( \frac{1}{k_n} \sum_{i=1}^{k_n} Z^2_{n-i+1,n} - \left( \frac{1}{k_n} \sum_{i=1}^{k_n} Z_{n-i+1,n} \right)^2 \right)^{-1/2}.$$  

Based on this result, we construct more accurate interval estimates for $R$.

Another and commonly used estimator for tail coefficients is the Hill estimator. Cheng and Pan (1998) established the one-term Edgeworth expansion for this popular estimator and constructed more accurate interval estimates for the tail index. We recall that the least squares type estimators have the interesting property of being universal asymptotically normal under the model (1), a property not shared by the Hill estimator. Also, they are particularly adequate in situations where the coefficient $R$ is expected to be small (cf. Csörgő and Viharos (1998)). This is the case of the application considered here and constitutes the main motivation for this work.

We also point out that the derived Edgeworth expansion is of independent interest and may be used for other applications.

We present our results in the following section and the corresponding proofs are given in Section 3. The confidence bounds for $R$ are derived in Section 4. In order to illustrate the finite sample behaviour of the confidence intervals for the adjustment coefficient, we consider, in Section 5, a typical example from insurance and present a small-scale simulation study. In this study, we also include the confidence bounds based on the asymptotic normality and on the bootstrap approximation.

2 Main results

2.1 General results

In the sequel, $\xrightarrow{D}$ and $\xrightarrow{=} D$ stand, respectively, for convergence and equality in distribution. In the same way, $\xrightarrow{P}$ denotes convergence in probability.

We begin by stating the asymptotic properties of $\hat{R}(k_n)$, which follow directly from the known corresponding properties for $\hat{R}(k_n)$, using the fact
that \( \tilde{R}(k_n) = \frac{R(k_n)}{\sqrt{i_n(k_n)}} \), with

\[
i_n(k_n) = 1 + O\left( \frac{\log^2 k_n}{k_n} \right),
\]
as \( n \to \infty \) (cf. Lemma 2 of Brito and Freitas (2003)).

In this way, we may conclude that \( \tilde{R}(k_n) \) is a consistent estimator of \( R \).

The general conditions ensuring the asymptotic normality are stated in the theorem below. This result follows directly from Proposition 1 of Brito and Freitas (2003).

**Theorem 1** Assume that \( F \) satisfies (1) and \( k_n \) is a sequence of positive integers such that (2) holds. If we suppose that, as \( n \to \infty \),

\[
k_n^{1/2} \sup_{1/k_n \leq y \leq 1} \left| \log \left( \frac{L(yt_n)}{L(k_n/n)} \right) \right| \to 0
\]

uniformly in \( t \) on compact subsets of \((0, \infty)\), then

\[
\frac{R^2}{\sqrt{8} k_n^{1/2}} \left( \frac{1}{\sqrt{R^2(k_n)}} - \frac{1}{R^2} \right) \overset{D}{\to} N(0, 1).
\]

Using the results established in Brito and Freitas (2003) and Brito and Freitas (2006), it is also easy to conclude that \( \tilde{R}(k_n) \) is universal asymptotically normal under the model (1) and that the tail bootstrap method introduced by Bacro and Brito (1998) works for this estimator.

Consider now the normalized estimator

\[
R_n := \frac{R^2}{\sqrt{8} k_n^{1/2}} \left( \frac{1}{\sqrt{R^2(k_n)}} - \frac{1}{R^2} \right).
\]

We will investigate how close the distribution of \( R_n \) is to the normal distribution. To this end, we will give the asymptotic expansion for the d.f. of the normalized estimator.

Before presenting this result it is convenient to introduce some notation. We assume that \( U_1, U_2, \ldots \) is a sequence of independent uniform \((0, 1)\) r.v.’s. The order statistics of the sample \((U_1, U_2, \ldots, U_n)\) are denoted by \( U_{1,n} \leq U_{2,n} \leq \ldots \leq U_{n,n} \). We denote by \( W_1, W_2, \ldots, W_{k_n} \) the \( k_n \) exceedances of the random level \( Z_{n-k_n,n} \), that is

\[
W_i := Z_{n-k_n+i,n} - Z_{n-k_n,n}, \quad 1 \leq i \leq k_n.
\]

With the above notation,

\[
\tilde{R}(k_n) = \left( \frac{1}{k_n} \sum_{i=1}^{k_n} W_i^2 - \left( \frac{1}{k_n} \sum_{i=1}^{k_n} W_i \right)^2 \right)^{-1/2}.
\]
Since $Z_i \overset{D}{=} F^{-1}(U_i), i \geq 1$, we write, without loss of generality,

$$W_i = F^{-1}(U_{i-k_n+i,n}) - F^{-1}(U_{i-k_n,n}).$$

As in Bacro and Brito (1998) (cf. Theorem 1) we shall make use of the following equivalent representation for $W_i, i = 1, \ldots, k_n$:

$$W_i = -\frac{1}{R} \log Y_i + \log \frac{\tilde{L}(Y_i(1 - U_{i-k_n,n}))}{L(1 - U_{i-k_n,n})}, \quad (3)$$

where

$$Y_i = \frac{1 - U_{i-k_n+i,n}}{1 - U_{i-k_n,n}}, \quad \text{for } i = 1, \ldots, k_n. \quad (4)$$

We recall also that $(Y_i)_{1 \leq i \leq k_n}$ is distributed as the vector of the order statistics of an i.i.d. $k_n$-sample from a uniform $(0, 1)$ distribution. Now, we define

$$V(k_n) := k_n^{1/2} \sup_{Y_{k_n} \leq y \leq 1} \left| \log \frac{\tilde{L}(y(1 - U_{k_n,n}))}{L(k_n)} \right|. \quad (5)$$

Moreover, we write $\Phi$ for the standard normal d.f. and $\phi$ for the corresponding density.

We present now the Edgeworth expansion for the d.f. of $R_n$ as a smooth function of a sample variance of i.i.d. unit exponential r.v.’s.

**Theorem 2** Assume that $F$ satisfies condition (1) and $k_n$ is a sequence of positive integers satisfying (2). Let $j$ be a positive integer and suppose that, for some $\varepsilon > 0$,

$$P \left[ V(k_n) > ak_n^{-j/2-\varepsilon} \right] = o(k_n^{-j/2}), \forall a > 0. \quad (5)$$

Then,

$$P[R_n \leq x] = \Phi(x) + k_n^{-1/2}p_1(x)\phi(x) + \ldots + k_n^{-j/2}p_j(x)\phi(x) + o(k_n^{-j/2}), \quad (6)$$

uniformly in $x$, where the function $p_j$ is a polynomial of degree at most $3j-1$, odd for even $j$ and even for odd $j$, with coefficients depending on moments of the $\text{Exp}(1)$ distribution up to order $2(j + 2)$.

In order to give an example, we will derive the expression of $p_1$ (To compute higher order polynomials see for example Hall (1992), Section 2.4). We use some calculations obtained by Hall (1992), Section 2.6, and the fact, proved in the next section that, under condition (5) the first $j$ terms of the Edgeworth expansion of $R_n$ are exactly the first $j$ terms of the Edgeworth expansion of a smooth function based on the variance of a $k_n$-sample of i.i.d.
r.v.'s with distribution $\text{Exp}(1)$. Let $T$ be a r.v. with distribution $\text{Exp}(1)$. Then,

$$p_1(x) = (\tau^1)^{-1} + (\tau^1)^{-3}(\rho^2 - \lambda^1/6)(x^2 - 1),$$

(7)

where

$$(\tau^1)^2 = E(T - E(T))^4(E(T - E(T))^2)^{-2} - 1 = 8,$$

$$\rho = E(T - E(T))^3(E(T - E(T))^2)^{-3/2} = 2,$$

$$\lambda^1 = E((T - E(T))^2(E(T - E(T))^2)^{-1} - 1)^3 = 240.$$  

Simplifying (7), we then obtain

$$p_1(x) = \frac{\sqrt{2}}{8}(11 - 9x^2).$$

(8)

So, if (5) holds for $j = 1$, then we obtain

$$P[R_n \leq x] = \Phi(x) + k_n^{-\frac{1}{2}} \frac{\sqrt{2}}{8}(11 - 9x^2)\phi(x) + o(k_n^{-1/2}).$$

### 2.2 Application

Some further notation will be needed before we apply the above result to the estimation of the adjustment coefficient. We begin by considering the Sparre Andersen model, where the claims $C_1, C_2, \ldots$ occur at times $T_1, T_1 + T_2, \ldots$, such that $\{C_i\}$ and $\{T_i\}$ are independent sequences of i.i.d. r.v.'s with finite means. Denoting by $C(t)$ the total sum of claims up to time $t$, we have

$$C(t) = \sum_{i=1}^{N(t)} C_i,$$

where $N(t)$ is the number of claims observed up to time $t$. Starting with an initial capital $x$ and compensating the claim process $\{C(t)\}$ by incoming premiums with constant rate $\gamma > 0$ per unit time, the risk reserve of the company is defined by

$$S(t) = x + \gamma t - C(t).$$

The probability of ruin is then given by

$$U(x) = P\left(\inf_{t>0} S(t) < 0\right)$$

$$= P\left(\sup_{n\geq1} \sum_{i=1}^{n} (C_i - \gamma T_i) > x\right).$$
Define now i.i.d. r.v.’s by $D_i := C_i - \gamma T_i$ for $i = 1, 2, \ldots$ and assume that $E(D_1) < 0$. Denote by $S_n$ the associated random walk, with $S_0 = 0$ and $S_n = D_1 + \cdots + D_n$, $n = 1, 2, \ldots$.

The distribution of the r.v. $D_1$ considered here is supposed to have a sufficiently regular tail, such that the following conditions hold:

(H1) There exists $R > 0$ such that $E(e^{RD_1}) = 1$.

(H2) $E(D_1 | e^{RD_1}) < \infty$.

The unique positive solution $R$ is called the adjustment coefficient. It is well-known that, under (H1), the Lundberg inequality gives

$$U(x) \leq e^{-Rx}, \quad \text{for all } x > 0,$$

and that, under both conditions (H1) and (H2), the Cramér-Lundberg approximation holds,

$$U(x) \sim ae^{-Rx}, \quad \text{as } x \to \infty,$$

where $a$ is a positive constant (see e.g. Rolski et al. (1999)). Different approaches have been used for estimating the adjustment coefficient $R$ (see e.g. Pitts et al. (1996) and references therein). Here, we will follow the approach suggested by Csörgő and Steinebach (1991) of estimating $R$ by means of a sequence of auxiliary r.v.’s $\{Z_k\}$, recursively defined as follows:

\[
M_0 = 0, \quad M_n = \max\{M_{n-1} + D_n, 0\} \quad \text{for } n = 1, 2, \ldots, \\
\nu_0 = 0, \quad \nu_k = \min\{n \geq \nu_{k-1} + 1 : M_n = 0\} \quad \text{for } k = 1, 2, \ldots, \\
Z_k = \max_{\nu_{k-1} < j \leq \nu_k} M_j \quad \text{for } k = 1, 2, \ldots.
\]

$L_1, L_2, \ldots$ defines a sequence of i.i.d. r.v.’s. In the context of queueing models the r.v. $L_k$ may be interpreted as the maximum waiting time in the $k$-th busy cycle of a GI/G/1 queueing system. The distribution of $L_1$ was computed by Cohen (1969) in the cases where

(H3) $\{C(t)\}$ is a compound Poisson process or the claims $C_i$ are exponentially distributed.

As a consequence, Csörgő and Steinebach (1991) observed that, in both cases,

$$P(Z_1 > z) = ce^{-Rz}(1 + O(e^{-Az})) \quad \text{as } z \to \infty,$$

with positive constants $c$ and $A$. Hence, under one of the assumptions stated in (H3) condition (1) holds. In this way, the estimators of the exponential tail coefficient in the family (1) may be applied to the estimation of the adjustment coefficient under (H3).
More recently, Brito and Freitas (2006) have shown that the consistency properties of the estimators still hold in the Sparre Andersen model under (H1) and (H2), since under these conditions

$$P(Z_1 > z) = ce^{-Rz}(1 + o(1)) \text{ as } z \to \infty.$$

On the other hand, it is clear that we need to assume stronger conditions than (H1) and (H2) in order to apply the result of Theorem 2 to this context. In the following corollary, we establish the Edgeworth expansion for the family (9), for an adequate choice of the sequence \( k_n \).

**Corollary 1** Assume that (9) holds and \( k_n \) is a sequence of positive integers satisfying (2). Let \( j \) be a positive integer and fix a small \( \varepsilon > 0 \). For sequences \( k_n \to \infty \) such that

$$k_n = o \left( n^{4A/(2(A+R)+(2R+A)+4R_0)} \right),$$

we have

$$P[R_n \leq x] = \Phi(x) + k_n^{-1/2}p_1(x)\phi(x) + \ldots + k_n^{-j/2}p_j(x)\phi(x) + o\left(k_n^{-j/2}\right),$$

uniformly in \( x \), where the function \( p_j \) is a polynomial of degree at most \( 3j - 1 \), odd for even \( j \) and even for odd \( j \) with coefficients depending on moments of the \( \text{Exp}(1) \) distribution up to order \( 2(j + 2) \).

In the case where \( \{C(t)\} \) is a compound Poisson claim process with exponentially distributed claims, the exact d.f. of the r.v.’s \( Z_1, Z_2, \ldots \) is given by

$$F(z) = \frac{1 - ae^{-\frac{(1-a)z}{\beta}}}{1 - a^2 e^{-\frac{(1-a)z}{\beta}}}, z > 0,$$

where \( a := \beta/\alpha \) and \( \alpha := E(\gamma T_1) > E(C_1) =: \beta \) (cf. Cohen (1969)). Moreover,

$$R = \frac{\alpha - \beta}{\alpha \beta}.$$

In this case, \( 1 - F(z) = a(1-a)e^{-Rz}\{1 + O(e^{-Rz})\} \) as \( z \to \infty \), and so the Edgeworth expansion (10) is valid for sequences \( k_n \to \infty \) such that \( k_n = o \left( n^{\frac{4}{\gamma + j + \varepsilon}} \right) \). In particular, the one-term Edgeworth expansion \((j = 1)\) holds for intermediate sequences \( k_n \to \infty \) satisfying \( k_n = o \left( n^{\frac{4}{\gamma + \varepsilon}} \right) \).

### 3 Proofs

**Proof of Theorem 2.**

In order to simplify the presentation of the proof of Theorem 2 we will introduce some notation. Let \( T_1, T_2, \ldots \) and \( T'_1, T'_2, \ldots \) be two sequences of
r.v’s, \( T = (T_1, \ldots, T_n) \) and \( T' = (T'_1, \ldots, T'_n) \). We denote by \( S_n^2(T) \) the sample variance of \( T \) and by \( S_n(T, T') \) the sample covariance between \( T \) and \( T' \), that is,

\[
S_n^2(T) = \frac{1}{n} \sum_{i=1}^n (T_i - \frac{1}{n} \sum_{i=1}^n T_i)^2
\]

and

\[
S_n(T, T') = \frac{1}{n} \sum_{i=1}^n \left( (T_i - \frac{1}{n} \sum_{i=1}^n T_i)(T'_i - \frac{1}{n} \sum_{i=1}^n T'_i) \right).
\]

We define also, for \( i = 1, \ldots, k_n \),

\[
E_i := -\frac{1}{R} \log Y_i, \quad F_i := \log \frac{\tilde{L}(Y_i (1 - U_{n-k_n,n}))}{L(1 - U_{n-k_n,n})},
\]

where \( Y_i \) is defined as in (4), and write \( E = (E_1, \ldots, E_{k_n}) \) and \( F = (F_1, \ldots, F_{k_n}) \).

With these notations and using the representation (3) in the definition of the estimator \( \tilde{R}(k_n) \), we may write

\[
\frac{1}{R^2(k_n)} - \frac{1}{R^2} = S_{k_n}^2(E) + S_{k_n}^2(F) + 2S_{k_n}(E, F) - \frac{1}{R^2}
\]

and thus

\[
R_n = \frac{k_n^{1/2}}{\sqrt{8}} (R^2 S_{k_n}^2(E) - 1) + \frac{R^2}{\sqrt{8}} k_n^{1/2} (S_{k_n}^2(F) + 2S_{k_n}(E, F)).
\]

Note that \( R^2 S_{k_n}^2(E) = S_{k_n}^2(RE) \), which is the sample variance of a unit exponential i.i.d. \( k_n \)-sample.

By Theorem 2.2 of Hall (1992) we have that (see also Hall (1992), Section 2.6), for \( j \geq 1 \),

\[
P[k_n^{1/2}(R^2 S_{k_n}^2(E) - 1)/\sqrt{8} \leq x] = \Phi(x) + k_n^{-1/2} p_1(x) \phi(x) + \ldots + k_n^{-j/2} p_j(x) \phi(x) + o(k_n^{-j/2}), \quad (12)
\]

uniformly in \( x \), where the functions \( p_j \) are polynomials of degree at most \( 3j - 1 \), odd for even \( j \) and even for odd \( j \), with coefficients depending on moments of the \( \text{Exp}(1) \) distribution up to order \( 2(j + 2) \).

Now, we are going to study the term \( \frac{R^2}{\sqrt{8}} k_n^{1/2} (S_{k_n}^2(F) + 2S_{k_n}(E, F)) \).

The well-known inequality

\[
(S_{k_n}(E, F))^2 \leq S_{k_n}^2(E)S_{k_n}^2(F),
\]

implies that

\[
|S_{k_n}(E, F)| \leq (1/R + o_P(1)) S_{k_n}(F).
\]

9
By a routine calculation, it is easily verified that
\[ S_{k_n}^2(F) \leq \sup_{y_n \leq y \leq 1} \left( \log \frac{\tilde{L}(y(1 - U_{n - k_n,n}))}{L(k_n/n)} \right)^2. \]

Thus, if we suppose that (5) holds for some \( \varepsilon > 0 \), then we have that
\[ P \left[ \frac{R^2}{\sqrt{8}} k_n^{1/2} S_{k_n}^2(F) > k_n^{-j/2 - \varepsilon} \right] = o(k_n^{-j/2}) \]  \hspace{1cm} (13)

and
\[ P \left[ \frac{R^2}{\sqrt{8}} k_n^{1/2} 2S_{k_n}(E, F) > k_n^{-j/2 - \varepsilon} \right] = o(k_n^{-j/2}). \] \hspace{1cm} (14)

For simplifying the presentation we write
\[ R_n = P_n + \Delta_1 + \Delta_2, \]
where
\[ P_n = \frac{k_n^{1/2}}{\sqrt{8}} (R^2 S_{k_n}^2(E) - 1), \]
\[ \Delta_1 = \frac{R^2}{\sqrt{8}} k_n^{1/2} S_{k_n}^2(F) \]
and
\[ \Delta_2 = \frac{R^2}{\sqrt{8}} k_n^{1/2} 2S_{k_n}(E, F). \]

Now, using the delta method and relations (13) and (14), we obtain
\[
P[R_n \leq x] = P[P_n + \Delta_1 + \Delta_2 \leq x] \\
\leq P[P_n + \Delta_1 \leq x + k_n^{-j/2 - \varepsilon}] + P[|\Delta_2| > k_n^{-j/2 - \varepsilon}] \\
\leq P[P_n \leq x + 2k_n^{-j/2 - \varepsilon}] + P[|\Delta_1| > k_n^{-j/2 - \varepsilon}] + o(k_n^{-j/2}) \\
= P[P_n \leq x] + o(k_n^{-j/2}).
\]

Similarly,
\[
P[R_n \leq x] \geq P[P_n \leq x] + o(k_n^{-j/2}).
\]

Thus, the results follows from (12). \( \square \)

**Proof of Corollary 1.**

For the family (9), we have that
\[ \tilde{L}(s) = c^{1/R} \{1 + O(s^{A/R})\}, \quad 0 < s < 1. \]
Let $0 < \delta < 1$ such that $|s| < \delta \Rightarrow |\frac{O(s^{A/R})}{s^{A/R}}| < k$, for some $k \in \mathbb{R}^+$. Consider the event $A_n(\delta) = \{1 - U_{n-k,n} < \delta\}$. On $A_n(\delta)$, we have, for $n$ sufficiently large and for $0 < y \leq 1$,

$$
\tilde{L}(y(1 - U_{n-k,n})) \leq 1 + b_1(y(1 - U_{n-k,n}))^{A/R} + b_2 \left(\frac{k_n}{n}\right)^{A/R}, \quad b_1, b_2 \in \mathbb{R}^+.
$$

So,

$$
\left| \log \frac{\tilde{L}(y(1 - U_{n-k,n}))}{\tilde{L}(\frac{k_n}{n})} \right| \leq b_1(y(1 - U_{n-k,n}))^{A/R} + b_2 \left(\frac{k_n}{n}\right)^{A/R},
$$

and consequently,

$$
\sup_{y_{k,n} \leq y \leq 1} \left| \log \frac{\tilde{L}(y(1 - U_{n-k,n}))}{\tilde{L}(\frac{k_n}{n})} \right| \leq b_1(1 - U_{n-k,n})^{A/R} + b_2 \left(\frac{k_n}{n}\right)^{A/R}.
$$

Now, fix a small $\epsilon > 0$. On $A_n(\delta)$, we then have, for $a > 0$ and $n$ sufficiently large,

$$
P \left[ V(k_n) > ak_n^{-j/2-\epsilon} \right] = P \left[ \sup_{y_{k,n} \leq y \leq 1} \left| \log \frac{\tilde{L}(y(1 - U_{n-k,n}))}{\tilde{L}(\frac{k_n}{n})} \right| > ak_n^{-(j+1)/2-\epsilon} \right]
$$

$$
\leq P \left[ b_1(1 - U_{n-k,n})^{A/R} + b_2 \left(\frac{k_n}{n}\right)^{A/R} > ak_n^{-(j+1)/2-\epsilon} \right]
$$

$$
= P \left[ (1 - U_{n-k,n})^{A/R} > \frac{a}{b_1} k_n^{-(j+1)/2-\epsilon} - b_2 \left(\frac{k_n}{n}\right)^{A/R} \right]
$$

$$
= P \left[ U_{k,n+1,n} > \left(\frac{a}{b_1} k_n^{-(j+1)/2-\epsilon} - b_2 \left(\frac{k_n}{n}\right)^{A/R} \right)^{R/A} \right]
$$

$$
\leq P \left[ U_{k,n+1,n} - \frac{k_n + 1}{n + 1} > \left(\frac{a}{b_1} k_n^{-(j+1)/2-\epsilon} - b_2 \left(\frac{k_n}{n}\right)^{A/R} \right)^{R/A} - \frac{k_n + 1}{n + 1} \right]
$$

$$
\leq \frac{E(U_{k,n+1,n} - \frac{k_n + 1}{n + 1})}{\left(\left(\frac{a}{b_1} k_n^{-(j+1)/2-\epsilon} - b_2 \left(\frac{k_n}{n}\right)^{A/R} \right)^{R/A} - \frac{k_n}{n} \right)^2}.
$$

Recalling that $E(U_{k,n}) = \frac{k_n}{n+1}$ and $V(U_{k,n}) = O\left(\frac{k_n}{n^2}\right)$, we conclude that

$$
P \left[ V(k_n) > ak_n^{-j/2-\epsilon} \right] = \frac{O\left(\frac{k_n}{n^2}\right)}{O\left((k_n^{-(j+1)/2-\epsilon})^{2R/A}\right)} = O\left(\frac{k_n^{1+2R/A}((j+1)/2+\epsilon)}{n^2}\right).
$$
Note now that $1 - P[A_n(\delta)] = O\left(\frac{k_n}{n^2}\right)$ since, for $n$ sufficiently large,

$$1 - P[A_n(\delta)] = P[U_{k_n+1,n} - \frac{k_n+1}{n+1} \geq \delta - \frac{k_n+1}{n+1}] \leq \frac{E(U_{k_n+1,n} - E(U_{k_n+1,n}))^2}{(\delta - \frac{k_n}{n})^2}. $$

So,

$$P[V(k_n) > ak_n^{-j/2-\varepsilon}] = O\left(\frac{k_n^{1+2R((j+1)/2+\varepsilon)}n^{1/2}}{n^2}\right).$$

Thus, if $k_n \to \infty$ is such that $k_n^{1+2R((j+1)/2+\varepsilon)}n^{1/2} = o(k_n^{-j/2})$, which is equivalent to $k_n = o\left(n^{2(A+R)/(2R+4A)+4R}\right)$, the condition (5) of Theorem 2 holds and the result follows. □

4 Confidence bounds

Suppose that the d.f. $F$ satisfies (1) and (2) holds.

If we knew the d.f. of $R_n$, we could calculate the corresponding $q$-quantile $x_q$, i.e $P(R_n \leq x_q) = q$. The lower 100$p\%$ confidence intervals for $R$ would be:

$$CI_{R}(k_n,p) = \{R : R_n > x_q\} = \left(\bar{R}(k_n) \left(1 + \sqrt{8x_qk_n^{-1/2}}\right)^{1/2}, +\infty\right)$$

$$= \left(\bar{R}(k_n) \left(1 + \sqrt{2x_qk_n^{-1/2} + o(k_n^{-1/2})}\right), +\infty\right),$$

where $p = 1 - q$.

For estimating $x_q$ we may use, as usual, the asymptotic normality of $R_n$. $P(R_n \leq x_q) = q$ may be written as $\Phi(x_q) + o(1) = q$, and so $x_q = \Phi^{-1}(q) + o(1)$, where $\Phi^{-1}(q)$ denotes the $q$-quantile of the $N(0,1)$. We then may construct the asymptotic confidence bounds

$$NI_{R}(k_n,p) = \left(\bar{R}(k_n) \left(1 + \sqrt{2\Phi^{-1}(q)k_n^{-1/2} + o(k_n^{-1/2})}\right), +\infty\right).$$

Next, we are going to use the asymptotic expansion obtained in Theorem 2 to get more accurate interval estimates for $R$. For this we suppose that condition (5) holds for $j = 1$ and we consider the polynomial derived in (8):

$$p_1(x) = \frac{2\sqrt{2}}{8} (11 - 9x^2).$$

(The following procedure can be extended for higher orders.) In the case $j = 1$, $P(R_n \leq x_q) = q$ may be written as

$$\Phi(x_q) + k_n^{-1/2}p_1(x_q)\phi(x_q) + o(k_n^{-1/2}) = q.$$
Solving this equation for $x_q$, we have

$$x_q = \Phi^{-1}(q) - p_1(\Phi^{-1}(q)) k_n^{-1/2} + o(k_n^{-1/2})$$

(see e.g. Hall (1992), Section 2.5).

Then, we obtain the following confidence bounds

$$EI_{\tilde{R}}(k_n, p) = \left( \tilde{R}(k_n) \left( 1 + \sqrt{2} \Phi^{-1}(q) k_n^{-1/2} - \{(251, 226),(254, 226)} \right) k_n^{-1} + o(k_n^{-1}) \right), +\infty \right).$$

The performance of these bounds will be investigated in Section 5.

We will include in our study the so-called percentile confidence bounds given by:

$$BI_{\tilde{R}}(k_n, p) = \left( \tilde{R}(k_n) \left( 1 + \sqrt{2} \xi_q k_n^{-1/2} + o(k_n^{-1/2}) \right), +\infty \right),$$

where $\xi_q$ is the $q$-quantile of the tail bootstrap d.f. of $R_n$ (see e.g. Brito and Freitas (2006)).

5 Simulation results

In this section we present a small simulation study, in order to examine the finite sample behaviour of the confidence intervals $EI_{\tilde{R}}(k_n, p)$ constructed in Section 4.

For the sake of comparison with previous simulation studies, we consider the family defined by (11) with $(\alpha, \beta) = (24000, 10000)$.

We also include in the study, the confidence bounds based on the asymptotic normality of the geometric-type estimator $\tilde{R}(k_n)$, $NI_{\tilde{R}}(k_n, p)$, and on the tail bootstrap approximation, $BI_{\tilde{R}}(k_n, p)$.

We take here $n = 500$ and $n = 1000$. For a very limited illustration of the influence of the choice of $k_n$ on the coverage accuracy, we present here three values of $k_n$ for each $n$. The empirical coverage rates for 95% and 99% confidence bounds are reported in Tables 1 ($n = 500$) and 2 ($n = 1000$). The confidence bounds were constructed from 1000 samples for each $n$. Each bootstrap bound was computed using 2500 bootstrap samples.

We start by noting that all the bootstrap confidence intervals show undercoverage.

We may observe that the confidence bounds based on the first order Edgeworth expansion are always more accurate than the ones based on the standard normal distribution, which are too conservative.

We may conclude that, in general, for both values of $n$ and $p$, the coverage rates present a better behaviour using the first order Edgeworth expansion.
<table>
<thead>
<tr>
<th>$k_n$</th>
<th>$NI_R(k_n,p)$</th>
<th>$EI_R(k_n,p)$</th>
<th>$BI_R(k_n,p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.993</td>
<td>0.967</td>
<td>0.945</td>
</tr>
<tr>
<td>50</td>
<td>0.976</td>
<td>0.955</td>
<td>0.940</td>
</tr>
<tr>
<td>70</td>
<td>0.973</td>
<td>0.948</td>
<td>0.938</td>
</tr>
</tbody>
</table>

Table 1: Empirical coverage rates of the confidence bounds, $n = 500$, $p = 0.95$.

<table>
<thead>
<tr>
<th>$k_n$</th>
<th>$NI_R(k_n,p)$</th>
<th>$EI_R(k_n,p)$</th>
<th>$BI_R(k_n,p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1.000</td>
<td>0.994</td>
<td>0.983</td>
</tr>
<tr>
<td>50</td>
<td>1.000</td>
<td>0.992</td>
<td>0.981</td>
</tr>
<tr>
<td>70</td>
<td>0.999</td>
<td>0.988</td>
<td>0.980</td>
</tr>
</tbody>
</table>

Table 2: Empirical coverage rates of the confidence bounds, $n = 500$, $p = 0.99$.

<table>
<thead>
<tr>
<th>$k_n$</th>
<th>$NI_R(k_n,p)$</th>
<th>$EI_R(k_n,p)$</th>
<th>$BI_R(k_n,p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.977</td>
<td>0.958</td>
<td>0.945</td>
</tr>
<tr>
<td>80</td>
<td>0.972</td>
<td>0.951</td>
<td>0.942</td>
</tr>
<tr>
<td>110</td>
<td>0.964</td>
<td>0.949</td>
<td>0.940</td>
</tr>
</tbody>
</table>

Table 3: Empirical coverage rates of the confidence bounds, $n = 1000$, $p = 0.95$.

<table>
<thead>
<tr>
<th>$k_n$</th>
<th>$NI_R(k_n,p)$</th>
<th>$EI_R(k_n,p)$</th>
<th>$BI_R(k_n,p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.999</td>
<td>0.994</td>
<td>0.984</td>
</tr>
<tr>
<td>80</td>
<td>0.998</td>
<td>0.990</td>
<td>0.984</td>
</tr>
<tr>
<td>110</td>
<td>0.997</td>
<td>0.988</td>
<td>0.982</td>
</tr>
</tbody>
</table>

Table 4: Empirical coverage rates of the confidence bounds, $n = 1000$, $p = 0.99$.

**Acknowledgements**

Both authors acknowledge support from CMUP, financed by FCT (Portugal) through the programs POCTI and POSI, with national and European Community structural funds.
References


