

## Research Article

# Confidence Intervals for the Mean Based on Exponential Type Inequalities and Empirical Likelihood

Sandra Vucane,<sup>1</sup> Janis Valeinis,<sup>1</sup> and George Luta<sup>2</sup>

<sup>1</sup> Faculty of Physics and Mathematics, University of Latvia, Zellu Street 8, Riga, LV-1002, Latvia

<sup>2</sup> Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University, Suite 180, Building D, 4000 Reservoir Rd., NW, Washington, DC 20057-1484, USA

Correspondence should be addressed to Sandra Vucane; [sandra.vucane@gmail.com](mailto:sandra.vucane@gmail.com)

Received 17 July 2013; Accepted 19 September 2013

Academic Editors: J. Fellman and O. François

Copyright © 2013 Sandra Vucane et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For independent observations, recently, it has been proposed to construct the confidence intervals for the mean using exponential type inequalities. Although this method requires much weaker assumptions than those required by the classical methods, the resulting intervals are usually too large. Still in special cases, one can find some advantage of using bounded and unbounded Bernstein inequalities. In this paper, we discuss the applicability of this approach for dependent data. Moreover, we propose to use the empirical likelihood method both in the case of independent and dependent observations for inference regarding the mean. The advantage of empirical likelihood is its Bartlett correctability and a rather simple extension to the dependent case. Finally, we provide some simulation results comparing these methods with respect to their empirical coverage accuracy and average interval length. At the end, we apply the above described methods for the serial analysis of a gene expression (SAGE) data example.

## 1. Introduction

Although the classical  $t$ -test is based on the assumption that observations are normally distributed, it is well known that it is a robust test and works well at least for symmetric distributions (see [1]). However, for skewed or heavy tailed distributions, the confidence intervals for the mean based on the inversion of the  $t$ -test may give poor coverage. Using exponential type inequalities (such as Bernstein's inequality), Rosenblum and van der Laan [2] presented a simple approach for constructing confidence intervals for the population mean. Rosenblum and van der Laan [2] dealt with the bounded version of Bernstein's inequality which is applicable only for a few distributions such as, the bounded uniform distribution. To use it for general distributions, we need to use the unbounded version of Bernstein's inequality, which was analyzed by Shilane et al. [3] for the negative binomial distribution. The confidence intervals based on exponential type inequalities have a guaranteed coverage probability under much weaker assumptions than required by the standard methods. Although the obtained confidence

intervals are usually too large, there are situations when they give better coverage accuracy than the classical methods.

In this paper, our goal is to use the empirical likelihood method introduced by Owen [4, 5] for the inference on the mean and compare it with the Rosenblum and van der Laan [2] method. The empirical likelihood method is based on the nonparametric likelihood ratio statistic, which similarly to the parametric case has asymptotically a limiting chi-square distribution. The confidence intervals obtained by the empirical likelihood method have some very appealing characteristics. There are no prespecified parametric assumptions on the distribution of observations and no constraints on the shape of the confidence intervals. Empirical likelihood intervals are Bartlett correctable in most cases. This means that a simple correction for the mean reduces the coverage error from order  $n^{-1}$  to order  $n^{-2}$ , where  $n$  denotes the sample size (see, e.g., [6, 7]). Owen [7] suggested (see Section 2 for more details) that for small sample sizes the  $F$  distribution may provide a better approximation for the limiting distribution of the test statistic. We found in our

TABLE 1: Empirical coverage accuracy for 95% confidence intervals for the mean using different methods for the iid data generated with a marginal negative binomial distribution  $\text{Negbinom}(\mu, \theta)$ , based on 10 000 samples.

	$n$	$t$	EL	$EL_B$	$EL_F$	$EL_{FB}$	$B_{\text{perc}}$	$B_{\text{norm}}$	$B_{\text{basic}}$	$B_{\text{stud}}$	$B_{\text{bn}}$	$B_{\text{ubn}}$
$\theta = 0.1$	20	0.797	0.808	0.821	0.829	0.840	0.797	0.783	0.744	0.926	0.898	0.889
	50	0.848	0.884	0.898	0.897	0.902	0.861	0.837	0.807	0.945	0.961	0.944
	100	0.889	0.917	0.924	0.922	0.928	0.907	0.888	0.856	0.948	0.979	0.966
$\theta = 0.5$	20	0.893	0.890	0.901	0.905	0.918	0.881	0.874	0.853	0.943	0.980	0.965
	50	0.916	0.921	0.928	0.929	0.932	0.916	0.910	0.906	0.945	0.993	0.980
	100	0.937	0.947	0.950	0.950	0.952	0.936	0.929	0.920	0.955	0.998	0.995
$\theta = 1.0$	20	0.918	0.905	0.914	0.920	0.931	0.900	0.902	0.894	0.940	0.985	0.972
	50	0.931	0.934	0.940	0.939	0.944	0.931	0.924	0.919	0.947	0.995	0.985
	100	0.941	0.946	0.948	0.950	0.952	0.936	0.936	0.936	0.955	0.995	0.990

simulation study that this approach can provide a significant improvement (see Tables 1 and 3).

Time series data are often correlated. For this situation, either modified or different statistical procedures of statistical inference are required. There exist many exponential type inequalities for weakly dependent data, characterized by  $\alpha$ -mixing or  $\beta$ -mixing sequences (e.g., see [8–10]). However, it is problematic to use them for practical data problems. First, such inequalities typically contain the  $\alpha$ - or  $\beta$ -mixing coefficients which have to be estimated from the data. Only recently McDonald et al. [11] introduced the estimation procedure for  $\beta$ -mixing coefficients. Second, usually these inequalities are established for bounded distributions. The applicability of exponential type inequalities for the dependent case is therefore limited. The blockwise empirical likelihood method introduced by Kitamura [12] provides a good alternative for the inference regarding the mean. It seems that the Bartlett correction can be applied also in this case.

For our simulation study, we use the negative binomial distribution, which has been analyzed by several authors (e.g., see [3, 13]) and has been widely used for applications involving insurance data. There are several ways how the negative binomial distribution can be parametrized. We will use the parametrization described by Hilbe [13]; that is, for any  $\mu \in \mathbb{R}^+$  and  $\theta \in \mathbb{R}^+$ , the resulting probability mass function for the negative binomial random variable  $X \sim \text{Negbinom}(\mu, \theta)$  is

$$\mathbb{P}(X = k) = \frac{\mu^k}{k!} \frac{\Gamma(\theta + k)}{\Gamma(\theta)(\mu + \theta)^k} \left(1 + \frac{\mu}{\theta}\right)^{-\theta}, \quad k \in \mathbb{Z}^+. \quad (1)$$

We also simulate some autoregressive processes representing data with a marginal negative binomial distribution. For independent data, we compared the empirical coverage accuracy for the confidence intervals based on exponential type inequalities, the  $t$ -test, and several empirical likelihood and bootstrap methods. For dependent data, we used a simulation study to compare the blockwise empirical likelihood method, with and without the Bartlett correction, and we also applied the empirical likelihood method to the serial analysis of gene expression (SAGE) data analyzed by Shilane et al. [3].

This paper is organized as follows. Section 2 deals with the independent case. It contains three subsections which

introduce Bernstein inequalities and the empirical likelihood method while also including a simulation study. In Section 3, weakly dependent data are discussed. Subsequent three subsections introduce the blockwise empirical likelihood method, describe two exponential-type inequalities, provide a simulation study, and analyze the SAGE data example. We finish with the main conclusions presented in Section 4.

## 2. Independent Case

**2.1. Confidence Intervals Using Bernstein's Inequality.** Rosenblum and van der Laan [2] described the method for generating confidence intervals, which bounds the tails of the distribution of the sample mean of  $n$  independent, bounded random variables. This method is based on exponential type inequalities such as Bernstein's inequality: for all  $x > 0$ ,

$$\begin{aligned} & \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n (X_i - \mathbb{E} X_i) \right| > \frac{x}{\sqrt{n}}\right) \\ & \leq 2 \exp\left(-\frac{nx^2}{2(v + W\sqrt{nx}/3)}\right), \end{aligned} \quad (2)$$

where  $W$  and  $v$  are constants defined in Theorem 1.

**Theorem 1** (Rosenblum and van der Laan [2]). *Let  $X_1, \dots, X_n$  be independent and identically distributed (iid) random variables with mean  $\mu$ , and let  $W$  be such that  $\mathbb{P}(|X - \mu| \leq W) = 1$ . Let  $\sigma^{*2}$  be a number satisfying  $\sigma^{*2} \geq \text{Var}(X)$ . Then for the function*

$$\begin{aligned} & q(1 - \alpha, W, n\sigma^{*2}) \\ & \equiv n^{-1} \log\left(\frac{2}{\alpha}\right) \left( Wn^{-1/2} + \sqrt{\frac{(W^2 n)}{9} + \frac{(2n^2 \sigma^{*2})}{\log(2/\alpha)}} \right), \\ & \mathbb{P}\left(\bar{X} - \frac{q(1 - \alpha, W, n\sigma^{*2})}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{q(1 - \alpha, W, n\sigma^{*2})}{\sqrt{n}}\right) \\ & \geq 1 - \alpha. \end{aligned} \quad (3)$$

In particular,

$$\left( \bar{X} - \frac{q(1-\alpha, W, n\sigma^{*2})}{\sqrt{n}}, \bar{X} + \frac{q(1-\alpha, W, n\sigma^{*2})}{\sqrt{n}} \right) \quad (4)$$

is the  $(1-\alpha)100\%$  confidence interval for  $\mu$ .

The accuracy of coverage probabilities depends on the estimation of  $W$  and  $v$ , where  $v = n\sigma^{*2}$ . Rosenblum and van der Laan [2] suggested that a bound  $W$  should be set at least as large as  $\max_i |X_i - \bar{X}|$  and proposed also to use for  $\sigma^{*2}$  an estimator  $s^2 + \text{SE}(s^2)$ , where  $\text{SE}(s^2)$  is the standard error of  $s^2$  (for details, see [2]).

To examine exponential type inequalities for distributions like the negative binomial distribution, we will need a version of Bernstein's inequality that does not require the assumption of boundedness.

**Lemma 2** (Birgé and Massart [14]). *Let  $Z_1, \dots, Z_n$  be independent random variables satisfying the moments conditions*

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} |Z_i^m| \leq \frac{m!}{2} v^2 c^{m-2} \quad (5)$$

for all  $m > 2$ , for some positive constants  $v$  and  $c$ . Then, for any positive  $\varepsilon$  and  $S_n = \sum_{i=1}^n Z_i$ ,

$$\mathbb{P}[S_n - E(S_n) \geq n\varepsilon] \leq \exp(-nx) \leq \exp\left(\frac{-n\varepsilon^2/2}{v^2 + n\varepsilon}\right), \quad (6)$$

where  $x$  is defined by the equation  $\varepsilon = v\sqrt{2x} + cx$ .

Suppose that  $Z_i$ ,  $1 \leq i \leq n$ , are iid negative binomial with parameters  $\mu$  and  $\theta$ , and let  $Y_i = Z_i - \mu$ . In this case, Shilane et al. [3] obtained the confidence interval for the mean of the form  $\bar{X} \pm \varepsilon$ , where  $\varepsilon$  is the solution to the equation

$$\left( \frac{\theta + \mu}{ne + \theta + \mu} \right)^{-\theta} \left( \frac{(\theta + \mu)(ne + \mu)}{\mu(ne + \theta + \mu)} \right)^{-ne-\mu} = \frac{\alpha}{2}. \quad (7)$$

To estimate the parameter  $\theta$ , we can use the method of moments; that is,  $\hat{\theta} = \bar{X}/((s^2/\bar{X}) - 1)$ .

**2.2. Empirical Likelihood for the Mean.** Let  $X_1, \dots, X_n$  be iid observations with the distribution  $F$  and mean  $\mu = \mathbb{E}X_1$ . To obtain confidence intervals for  $\mu$ , we define the profile empirical likelihood function

$$R(\mu) = \sup \left\{ \prod_{i=1}^n np_i \mid p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \mu \right\}. \quad (8)$$

Owen [4, 5] showed that a unique value for the right-hand side of (8) exists when  $\mu$  is inside the convex hull of the data points  $X_1, \dots, X_n$ . This maximization problem can be solved using the Lagrange multiplier method.

**Theorem 3** (Owen [7]). *If  $X_1, \dots, X_n$  are iid random variables with the distribution function  $F_0$ ,  $\mu_0 = \mathbb{E}(X_i)$ , and  $0 < \text{Var}(X_i) < \infty$ , then*

$$-2 \log(R(\mu_0)) \longrightarrow_d \chi_1^2. \quad (9)$$

The empirical likelihood  $(1-\alpha)100\%$  confidence interval for  $\mu$  is of the following form:

$$\{\mu \mid -2 \log(R(\mu)) \leq c\}, \quad (10)$$

where  $c$  is the  $1-\alpha$  quantile of the  $\chi_1^2$  distribution. Owen [7] stated that the proof of Theorem 3 and some simulations suggest that the quantiles  $\chi_1^{2,1-\alpha}$  can be replaced by the quantiles of the  $F$  distribution  $F_{1,n-1}^{1-\alpha}$ . The  $F$  calibration usually gives better results for small sample sizes.

The Bartlett correction replaces the  $1-\alpha$  quantile of the limiting chi-square distribution  $\chi_1^{2,1-\alpha}$  of the test statistic  $-2 \log(R(\mu))$  by  $(1+an^{-1})\chi_1^{2,1-\alpha}$ , where  $a = (1/2)\mu_4 - (1/3)\mu_3^2$  and  $\mu_i$  is  $i$ th moment of  $X_1$  (see [6]). The Bartlett correction improves the asymptotic error rate for all coverage levels  $1-\alpha$ . A Bartlett-corrected empirical likelihood confidence interval has the following form:

$$\{\mu \mid -2 \log(R(\mu)) \leq (1+an^{-1})\chi_1^{2,1-\alpha}\}. \quad (11)$$

**2.3. Simulation Study.** Our simulation study is based on 10,000 independent trials. In every trial, we generated *iid* random variables from the negative binomial distribution Negbinom( $\mu, \theta$ ) with different sample sizes  $n = \{20, 50, 100\}$ ,  $\mu = 5$ , and  $\theta = \{0.1, 0.5, 1.0\}$ . To simulate data, we used the command rnegbin from the package MASS in the R statistical programming language.

In Table 1, the empirical coverage accuracy for nominal 95% confidence intervals for the mean is evaluated for the following methods: the  $t$ -test ( $t$ ), the empirical likelihood method (EL), the empirical likelihood with Bartlett correction ( $EL_B$ ), the empirical likelihood method with the  $F$  distribution calibration ( $EL_F$ ), and the empirical likelihood method with the  $F$  distribution calibration and Bartlett correction ( $EL_{FB}$ ). For comparison, we have also simulated the coverage accuracy of several bootstrap resampling procedures: the percentile ( $B_{\text{perc}}$ ), normal ( $B_{\text{norm}}$ ), basic ( $B_{\text{basic}}$ ), and studentized ( $B_{\text{stud}}$ ) bootstrap resampling methods. The R package boot and command boot.ci are used to construct the confidence intervals for all the above mentioned bootstrap methods. Although only the unbounded version of Bernstein inequalities is valid for the negative binomial distribution, we also apply the bounded version in order to see the possible effects of the violation of this assumption. In Table 2, we present average confidence interval lengths for the same methods as described in Table 1.

For smaller  $\theta$  values, the degree of skewness of the negative binomial distribution is higher. Therefore the coverage accuracy increases along with the increase of the parameter  $\theta$ . From the results from Tables 1 and 2, we can see that the coverage accuracy is higher for the confidence intervals based on the bounded Bernstein's inequality than for those using the unbounded version. However, the resulting confidence intervals have larger average length when the bounded version is used.

Comparing the empirical likelihood and the bootstrap methods, we can see that even the uncorrected EL method works better than the bootstrap methods (the only exception is the studentized bootstrap, which also has one of the widest

TABLE 2: Average confidence interval lengths for the 95% confidence intervals for the mean using different methods for the iid data generated with the marginal negative binomial distribution  $\text{Negbinom}(\mu, \theta)$ , based on 10 000 samples.

	$n$	$t$	EL	$EL_B$	$EL_F$	$EL_{FB}$	$B_{\text{perc}}$	$B_{\text{norm}}$	$B_{\text{basic}}$	$B_{\text{stud}}$	$B_{\text{bn}}$	$B_{\text{ubn}}$
$\theta = 0.1$	20	12.80	12.11	12.91	12.98	13.85	11.11	11.67	11.11	69.33	28.53	26.59
	50	8.17	8.33	8.75	8.56	8.99	7.76	7.89	7.76	15.27	18.07	15.49
	100	5.95	6.17	6.40	6.25	6.49	5.80	5.85	5.80	8.15	12.74	10.53
$\theta = 0.5$	20	6.63	6.21	6.50	6.65	6.96	6.00	6.06	6.00	9.33	13.39	11.20
	50	4.07	4.06	4.18	4.17	4.29	3.93	3.93	3.93	4.66	8.02	6.53
	100	2.90	2.93	2.98	2.97	3.02	2.86	2.85	2.86	3.12	5.48	4.49
$\theta = 1.0$	20	4.93	4.60	4.79	4.93	5.13	4.48	4.50	4.48	5.87	9.53	7.76
	50	3.07	3.04	3.11	3.12	3.19	2.97	2.96	2.97	3.30	5.79	4.70
	100	2.15	2.15	2.98	2.97	3.02	2.12	2.11	2.12	2.23	3.88	3.21

TABLE 3: Empirical coverage accuracy for the 95% confidence intervals for the mean using the blockwise EL method (EL) and the blockwise EL method with the Bartlett correction ( $EL_B$ ) for the data generated from an AR(1) process with  $\phi = 0.5$  and with a marginal negative binomial distribution  $\text{Negbinom}(\mu, \theta)$ , based on 10 000 samples.

$n$	$M = L$	$\theta = 0.1$		$\theta = 0.5$		$\theta = 1.0$	
		EL	$EL_B$	EL	$EL_B$	EL	$EL_B$
50	2	0.630	0.655	0.759	0.772	0.808	0.825
	2	0.671	0.698	0.783	0.791	0.830	0.837
	4	0.674	0.697	0.797	0.808	0.839	0.852
	5	0.658	0.679	0.793	0.802	0.839	0.859
100	2	0.686	0.693	0.805	0.815	0.835	0.839
	4	0.692	0.714	0.826	0.833	0.859	0.867
	5	0.684	0.706	0.833	0.845	0.866	0.877
	2	0.678	0.693	0.832	0.838	0.882	0.886
200	4	0.710	0.719	0.858	0.872	0.912	0.915
	5	0.706	0.712	0.870	0.883	0.920	0.921
	10	0.714	0.730	0.881	0.894	0.919	0.927
	2	0.724	0.735	0.827	0.836	0.871	0.875
500	4	0.746	0.757	0.871	0.877	0.902	0.907
	5	0.751	0.762	0.869	0.872	0.909	0.913
	10	0.749	0.765	0.878	0.887	0.923	0.926
	20	0.749	0.770	0.882	0.897	0.924	0.932

intervals) and also the average lengths of the EL based confidence intervals are only negligibly larger than those for the bootstrap methods.

When the parameter  $\theta = 0.1$  and data become much more asymmetric, the exponential type inequalities give the best coverage. However, as they are exact methods, for greater  $\theta$  values, the empirical coverage becomes much larger than the nominal 95% confidence level. When using the Bartlett correction and the calibration with the  $F$  distribution, the empirical likelihood method provides much better coverage accuracy compared with the  $t$ -test and bootstrap methods.

### 3. Dependent Case

**3.1. Blockwise Empirical Likelihood for the Mean.** In practice, the assumption of independence may be invalid in many situations, and by relaxing this strong assumption, we extend

the scope of our research. As first shown by Kitamura [12], the empirical likelihood method can be applied not only for independent but also for stationary weakly dependent data characterized by some mixing coefficients.

Let  $(X_t)_{t \in \mathbb{Z}}$  be a real-valued strictly stationary process defined on a probability space  $(\Omega, \mathcal{F}, P)$ . For any two  $\sigma$ -fields  $\mathcal{A}$  and  $\mathcal{B} \subset \mathcal{F}$ , define the following coefficient of dependence:

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup |P(A \cap B) - P(A)P(B)|, \quad (12)$$

where the supremum is taken over all  $A \in \mathcal{A}$  and  $B \in \mathcal{B}$ . For  $-\infty \leq J \leq L \leq \infty$ , define  $F_J^L := \sigma(X_k, J \leq k \leq L)$ .  $(X_t)_{t \in \mathbb{Z}}$  is called strongly mixing or  $\alpha$ -mixing (see, e.g., [15]) if  $\alpha(n) := \alpha(F_{-\infty}^0, F_n^\infty) \rightarrow 0$  when  $n \rightarrow \infty$ .

Further, we also define another coefficient for any two  $\sigma$ -fields  $\mathcal{A}$  and  $\mathcal{B} \subset \mathcal{F}$  as follows:

$$\beta(\mathcal{A}, \mathcal{B}) := \sup \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i)P(B_j)|, \quad (13)$$

where the supremum is taken over all  $\{A_1, \dots, A_I\}$  and  $\{B_1, \dots, B_J\}$  from  $\Omega$  such that  $A_i \in \mathcal{A}$  for all  $i$  and  $B_j \in \mathcal{B}$  for all  $j$ . Define  $F_j^L := \sigma(X_k, J \leq k \leq L)$ , when  $-\infty \leq J \leq L \leq \infty$ .  $(X_t)_{t \in \mathbb{Z}}$  is called absolutely regular or  $\beta$ -mixing if  $\beta(n) := \sup_{J \in \mathbb{Z}} \beta(F_{-\infty}^J, F_{J+n}^\infty) \rightarrow 0$  when  $n \rightarrow \infty$ .

Under certain (usually mild) conditions, ARMA, GARCH, nonlinear time series models, and Markov chains are  $\beta$ -mixing processes (see, e.g., [15]). It is important to note that if a process is  $\beta$ -mixing, then it follows that it is also an  $\alpha$ -mixing sequence.

---


$$L(\mu) = \sup \left\{ \prod_{i=1}^Q (Qp_i) : p_1 \geq 0, \dots, p_Q \geq 0, \sum_{i=1}^Q p_i = 1, \sum_{i=1}^Q p_i \sum_{j=1}^M G(X_{(i-1)L+j}, \mu) = 0 \right\}, \quad (15)$$

which generates the blockwise empirical likelihood ratio statistic

$$LR(\mu) = \left( \frac{QM}{n} \right)^{-1} (2L(\mu) - 2L(\bar{\mu})). \quad (16)$$

Under some regularity conditions, Kitamura [12] showed that  $LR(\mu_0)$  also has a  $\chi^2$  limiting distribution.

The coverage error of confidence intervals

$$\mathbb{P}\{LR(\mu_0) \leq z\} = \mathbb{P}\{\chi_1^2 \leq z\} + O(n^{-2/3}) \quad (17)$$

can be improved up to the order  $O(n^{-5/6})$  by using the Bartlett correction; that is,

$$\mathbb{P}\{LR(\mu_0)(1 - n^{-1}a) \leq z\} = \mathbb{P}\{\chi_1^2 \leq z\} + O(n^{-5/6}), \quad (18)$$

where  $a = (1/2)k_4 - (1/3)k_3^2$  and  $k_i$  denotes the  $i$ th moment of  $B_i$ . This rate is slower than the rate obtained for the independent data due to the nonparametric treatment of dependence.

**3.2. Exponential Type Inequalities.** In this section, we introduce several exponential type inequalities for  $\alpha$ - and  $\beta$ -mixing sequences and discuss their applicability in practice.

**Theorem 4** (Doukhan [17]). *Let  $(X_t)_{t \in \mathbb{Z}}$  be a zero-mean real-valued process such that  $\exists \sigma^2 \in \mathbb{R}_+$ , for all  $n, m \in \mathbb{N}$  :  $(1/m)$*

Further, we will use the notation and the results from two papers: Zhang et al. [16] and Kitamura [12]. Let  $G(X, \mu)$  be a real-valued function such that  $\mathbb{E}\{G(X, \mu)\} = 0$ , where  $\mu = E(X_t)$ . Let  $M$  and  $L$  be the integers depending on  $n$ ,  $M \rightarrow \infty$ ,  $M = o(n^{1/2})$ , and  $L = O(M)$  as  $n \rightarrow \infty$ . Set  $Q = [(n - M)/L] + 1$ , where  $[x]$  is the integer part of  $x$ . The  $Q$  blocks of  $M$  consecutive observations

$$B_i := (X_{(i-1)L+1}, \dots, X_{(i-1)L+M}), \quad i = 1, \dots, Q \quad (14)$$

will be used for the estimation. Note that  $M$  is the block length and  $L$  is the separation between the block starting points. Furthermore, we will use nonoverlapping blocks with  $L = M$ .

In order to address the dependence among  $X_i$ , Kitamura, [12] proposed to use the profile empirical likelihood method to the blockwise sample  $\{\sum_{j=1}^M G(X_{(i-1)L+j}, \mu)\}_{i=1}^Q$  instead of  $\{G(X_i, \mu)\}_{i=1}^n$ . More specifically, Kitamura [12] defined the blockwise empirical likelihood function for  $\mu$  as

$$(X_n + \dots + X_{n+m})^2 \leq \sigma^2 \text{ and for all } t \in \mathbb{N} : |X_t| \leq M. \text{ Then for each } \varepsilon > 0, \theta = \varepsilon^2/4, \text{ and } q \leq n/(1+\theta),$$

$$\begin{aligned} \mathbb{P}(|S_n| \geq x) &\leq 4 \exp \left( -\frac{(1-\varepsilon)x^2}{2(n\sigma^2 + qMx/3)} \right) \\ &+ 2^n \frac{\beta(|q\theta| - 1)}{q}. \end{aligned} \quad (19)$$

**Theorem 5** (Bosq [18]). *Let  $(X_t)_{t \in \mathbb{N}}$  be a zero-mean real-valued process such that  $\sup_{1 \leq t \leq n} \|X_t\|_\infty \leq b$ . Then, for each integer  $q \in [1, n/2]$  and each  $\varepsilon > 0$ ,*

$$\begin{aligned} \mathbb{P}(|S_n| > n\varepsilon) &\leq 4 \exp \left( -\frac{\varepsilon^2 q}{8b^2} \right) \\ &+ 22 \left( 1 + \frac{4b}{\varepsilon} \right)^{1/2} q \alpha \left( \left[ \frac{n}{2q} \right] \right). \end{aligned} \quad (20)$$

Both inequalities involve mixing coefficients, which in practice need to be estimated. Recently McDonald et al. [11] proposed a method for the  $\beta$ -mixing coefficient that uses the histogram estimator for stationary time series data. To our knowledge, this is the first result in this area that allows us to estimate such general mixing coefficients. The estimator has the following form:

$$\hat{\beta}(a) = \frac{1}{2} \int |\hat{f}_a^2 - \hat{f} \otimes \hat{f}|, \quad (21)$$

where  $\hat{f}$  is the histogram estimator of the joint density of observations and  $\hat{f}_a^2$  is the 2-dimensional histogram estimator of the joint density of the two sets of observations

TABLE 4: Empirical average interval lengths for the 95% confidence intervals for the mean using the blockwise EL method (EL) and the blockwise EL method with the Bartlett correction ( $\text{EL}_B$ ) for the data generated from an AR(1) process with  $\phi = 0.5$  and with a marginal negative binomial distribution  $\text{Negbinom}(\mu, \theta)$ , based on 10 000 samples.

$n$	$M = L$	$\theta = 0.1$		$\theta = 0.5$		$\theta = 1.0$	
		EL	$\text{EL}_B$	EL	$\text{EL}_B$	EL	$\text{EL}_B$
50	2	1.62	1.70	0.85	0.88	0.65	0.68
	2	1.47	1.54	0.68	0.71	0.51	0.52
	4	1.49	1.55	0.73	0.76	0.54	0.56
	5	1.48	1.55	0.73	0.76	0.55	0.57
100	2	1.25	1.31	0.52	0.53	0.38	0.39
	4	1.32	1.37	0.57	0.58	0.41	0.43
	5	1.31	1.36	0.58	0.59	0.42	0.44
	2	0.97	1.02	0.34	0.35	0.25	0.25
500	4	1.04	1.08	0.38	0.39	0.28	0.28
	5	1.03	1.08	0.39	0.40	0.28	0.29
	10	1.06	1.10	0.40	0.41	0.30	0.31
	2	0.84	0.88	0.25	0.26	0.18	0.18
1000	4	0.90	0.93	0.29	0.29	0.20	0.20
	5	0.91	0.95	0.29	0.30	0.20	0.21
	10	0.92	0.95	0.31	0.31	0.21	0.22
	20	0.92	0.95	0.31	0.32	0.22	0.22

separated by  $a$  time points. Typically mixing coefficients decay exponentially fast; therefore, it is possible to get reasonable estimates only for small values of  $a$ .

In a recent paper, Merlevède et al. [10] obtained the exponential type inequality depending on the following condition on a strong mixing coefficient; that is,

$$\alpha(n) \leq \exp(-2cn), \quad (22)$$

where  $c$  is some constant. Here, we present the inequality given in Corollary 12 by Merlevède et al. [10].

**Corollary 6** (Merlevède et al. [10]). *Let  $(X_t)_{t \in \mathbb{N}}$  be a sequence of centred real-valued random variables. Suppose that the sequence satisfies (22) and that there exists a positive  $M$  such that  $\sup_{i \geq 1} \|X_i\|_\infty \leq M$ . For all  $n \geq 2(c \vee 2)$  and  $x \geq 0$ ,*

$$\mathbb{P}(|S_n| \geq x) \leq \exp\left(\frac{x^2}{n(\log n)4CM^2 + 4Mx/(c \wedge 1)}\right), \quad (23)$$

where

$$C = 6.2K + \left(\frac{1}{c} + \frac{8}{c^2}\right) + \frac{2}{c \log 2}, \quad (24)$$

$K = 1 + \sum_{i \geq 1} \alpha_i$ , and  $\alpha_i$  is defined in (22).

Note that the inequality (23) only depends on the value of the constant  $c$ . Using the  $\beta$ -mixing estimator (21) from McDonald et al. [11], we propose to estimate  $c$  by the linear regression estimator of the slope from the equation

$$\log(\hat{\beta}(a)) = c_1 - 2cn, \quad (25)$$

where  $c_1$  is some constant. The inequalities defined in Theorems 4 and 5 can be used for practical applications after the estimation of  $c$ , by exploiting the relation  $\alpha(a) \leq \beta(a)$ .

**3.3. Simulation Study.** We simulate  $X_1, \dots, X_n$  from a stationary AR(1) process  $\{X_t\}_{t \in \mathbb{Z}}$  defined as

$$X_t - \phi X_{t-1} = Z_t, \quad (26)$$

where  $\{Z_t\}_{t \in \mathbb{Z}}$  is an innovation process which is weakly stationary with mean 0 and autocovariance  $\mathbb{E}(Z_t Z_{t+h}) = \sigma_Z^2 < \infty$  if  $h = 0$  and 0 otherwise, and  $|\phi| \leq 1$  is the coefficient of the process. The data are generated with a marginal negative binomial distribution using the R package `gsarima`. We used the command `garsim` and selected the parameter `link = "identity"`. We used different sample sizes  $n = \{50, 100, 200, 500, 1000\}$  and selected different values for the parameter  $\theta = \{0.1, 0.5, 1.0\}$ . The minimal value of  $\mu$  was set equal to 0.5 for the algorithm used by the command `garsim`.

In Table 3, we present the empirical coverage accuracies for the 95% confidence intervals constructed using the blockwise empirical likelihood method with and without the Bartlett correction. We report the corresponding average lengths of the 95% confidence intervals in Table 4. Similar to the independent case the, Bartlett correction provides some improvement in all cases. Furthermore, the results strongly depend on the block length parameter  $M$ .

A general autoregressive process AR(1) has the mixing property if the process  $Z_t$  defined in (26) has a continuous marginal density (see [15]). Thus, it might not be appropriate to use exponential type inequalities in the case of the negative binomial distribution. We performed more simulations using a continuous distribution such as the normal distribution. In this case, the confidence intervals using Bernstein type inequalities typically are too large in comparison with those produced by the blockwise empirical likelihood method.

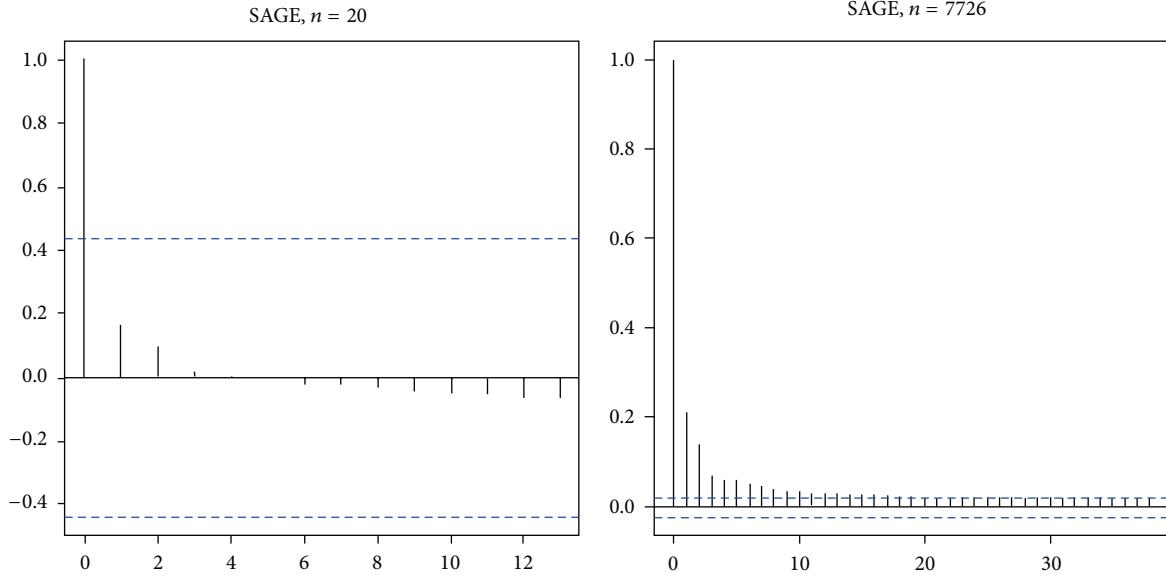
FIGURE 1: Empirical autocorrelations for the SAGE data with different sample sizes,  $n = 20$  and  $n = 7726$ .

TABLE 5: 95% confidence intervals for the mean of the 20 most frequent tag counts for the SAGE data.

Method	Lower limit	Upper limit	Length
Unbounded Bernstein	182.14	430.06	247.92
Bounded Bernstein	-455.02	1067.22	1522.24
Chi square	259.53	356.46	96.93
Gamma	160.82	497.35	336.53
Wald	-38.00	650.20	688.20
$t$ -test	-61.36	673.56	734.92
Bootstrap	111.95	1015.75	903.80
EL	123.17	833.04	709.86

Finally, we analyze the serial analysis of gene expression data (SAGE) from Shilane et al. [3]. This dataset is used in molecular biology to estimate the relative abundance of messenger ribonucleic acid (mRNA) molecules based upon the frequency of corresponding 14 base pair tag sequences that are extracted from a cell. Due to the fact that the cost of sequencing can be quite high, the sample size is often quite small for such data situations.

Shilane et al. [3] picked the 20 most frequent tags from the whole data set and constructed the confidence intervals for the mean of the corresponding counts using several methods: bounded and unbounded Bernstein type inequalities,  $\chi^2$ , Gamma, Wald,  $t$ -test, and bias corrected and accelerated bootstrap methods. Additionally, we report the confidence intervals obtained by the empirical likelihood method in Table 5.

Shilane et al. [3] outlined some strengths and weaknesses of the methods from Table 5. Both the intervals based on the bounded Bernstein inequality and the Wald method include a range of negative numbers as possible values for the mean  $\mu$ , which is unreasonable for the SAGE data. By the contrast,

TABLE 6: 95% confidence intervals for the mean of the 20 most frequent tag counts for different block sizes  $M$  for the SAGE data.

$M = L$	EL			EL <sub>B</sub>		
	Lower	Upper	Length	Lower	Upper	Length
2	2.13	4.60	2.47	2.11	4.74	2.62
4	2.06	4.86	2.80	2.06	4.86	2.80
5	2.04	4.92	2.88	2.04	4.92	2.88
10	1.96	5.16	3.20	1.96	5.16	3.20
15	1.91	5.30	3.39	1.91	5.30	3.39
20	1.87	5.40	3.53	1.87	5.40	3.53
25	1.84	5.49	3.65	1.87	5.40	3.53
30	1.82	5.57	3.75	1.82	5.57	3.75
50	1.74	5.82	4.08	1.74	5.82	4.08

the  $\chi^2$ , Gamma, bootstrap, and empirical likelihood methods give only nonnegative values.

All the methods used in Table 5 require the independence of the data. However, if we look at the whole data set, there are significant correlations (see Figure 1). Still in this case we can use the blockwise empirical likelihood method for statistical inference (see Table 6).

#### 4. Conclusions

We conclude that large confidence intervals are typically obtained when Bernstein type inequalities are used. This method can still provide good results for highly skewed situations when the sample sizes are small. For the independent observations, the empirical likelihood method is a good alternative not only to the classical  $t$ -test statistic but also to the bootstrap methods. The Bartlett correction and the calibration using the  $F$  distribution can provide a significant improvement for small sample sizes. For dependent data, in

principle, it is possible to use Bernstein type inequalities for inference for the mean. However, for correlated time series data, usually bigger sample sizes are necessary. Second, for the negative binomial distribution, these inequalities are not valid for ARMA type processes. A reasonable alternative is to use the blockwise empirical likelihood method.

## Conflict of Interests

The authors of the paper do not have a direct financial relation with any commercial identity mentioned in the paper that might lead to a conflict of interests for any of the authors.

## Acknowledgments

Valeinis acknowledges a partial support of the Project 2009/0223/IDP/1.1.2.0/09/APIA/VIAA/008 of the European Social Fund, and Vucane acknowledges a support by ESF Grant No. 2009/0138/IDP/1.1.2.1.2/09/IPIA/VIAA/004.

## References

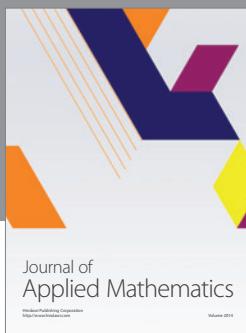
- [1] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, Springer Texts in Statistics, Springer, New York, NY, USA, 2008.
- [2] M. A. Rosenblum and M. J. van der Laan, “Confidence intervals for the population mean tailored to small sample sizes, with applications to survey sampling,” *The International Journal of Biostatistics*, vol. 5, article 4, 2009.
- [3] D. Shilane, S. N. Evans, and A. E. Hubbard, “Confidence intervals for negative binomial random variables of high dispersion,” *International Journal of Biostatistics*, vol. 6, no. 1, article 10, 2010.
- [4] A. B. Owen, “Empirical likelihood ratio confidence intervals for a single functional,” *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.
- [5] A. Owen, “Empirical likelihood ratio confidence regions,” *The Annals of Statistics*, vol. 18, no. 1, pp. 90–120, 1990.
- [6] P. Hall and B. La Scala, “Methodology and algorithms of empirical likelihood,” *International Statistical Review*, vol. 58, pp. 109–127, 1990.
- [7] A. B. Owen, *Empirical Likelihood*, Chapman & Hall/CRC, New York, NY, USA, 2001.
- [8] P. Doukhan, P. Massart, and E. Rio, “Invariance principles for absolutely regular empirical processes,” *Annales de l’Institut Henri Poincaré. Probabilités et Statistiques*, vol. 31, no. 2, pp. 393–427, 1995.
- [9] R. S. Kallabis and M. H. Neumann, “An exponential inequality under weak dependence,” *Bernoulli*, vol. 12, no. 2, pp. 333–350, 2006.
- [10] F. Merlevède, M. Peligrad, and E. Rio, “Bernstein inequality and moderate deviations under strong mixing conditions,” in *High Dimensional Probability V: The Luminy Volume*, vol. 5 of *Inst. Math. Stat. Collect.*, pp. 273–292, Inst. Math. Statist., Beachwood, Ohio, USA, 2009.
- [11] D. J. McDonald, C. R. Shalizi, and M. Schervish, “Estimating beta-mixing coefficients,” in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudík, Eds., JMLR W&CP, 2011.
- [12] Y. Kitamura, “Empirical likelihood methods with weakly dependent processes,” *The Annals of Statistics*, vol. 25, no. 5, pp. 2084–2102, 1997.
- [13] J. M. Hilbe, *Negative Binomial Regression*, Cambridge University Press, Cambridge, UK, 2007.
- [14] L. Birgé and P. Massart, “Minimum contrast estimators on sieves: exponential bounds and rates of convergence,” *Bernoulli*, vol. 4, no. 3, pp. 329–375, 1998.
- [15] R. C. Bradley, *Introduction to Strong Mixing Conditions*, vol. 1, 2, 3, Kendrick Press, Heber City, Utah, USA, 2007.
- [16] R. Zhang, L. Peng, and Y. Qi, “Jackknife-blockwise empirical likelihood methods under dependence,” *Journal of Multivariate Analysis*, vol. 104, pp. 56–72, 2012.
- [17] P. Doukhan, *Mixing: Properties and Examples*, vol. 85 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1994.
- [18] D. Bosq, *Nonparametric Statistics for Stochastic Processes*, vol. 110 of *Lecture Notes in Statistics*, Springer, New York, NY, USA, 1996.



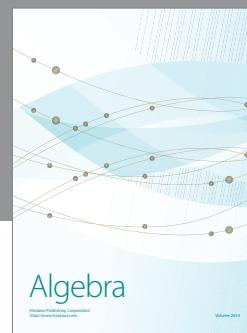
Advances in  
Operations Research



Advances in  
Decision Sciences



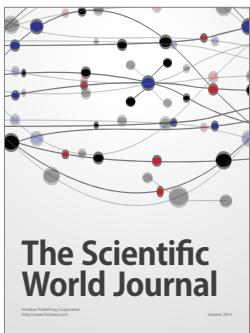
Journal of  
Applied Mathematics



Algebra



Journal of  
Probability and Statistics

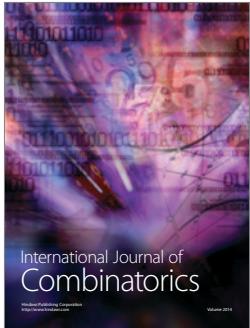


The Scientific  
World Journal

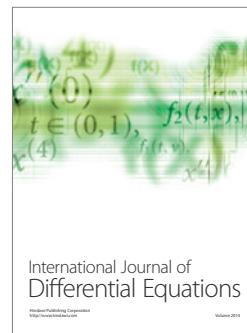


Hindawi

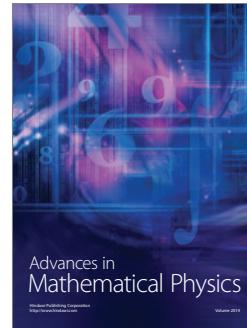
Submit your manuscripts at  
<http://www.hindawi.com>



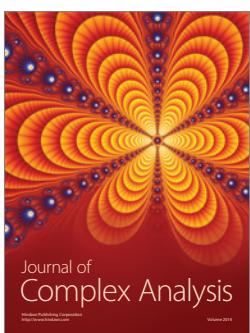
International Journal of  
Combinatorics



International Journal of  
Differential Equations



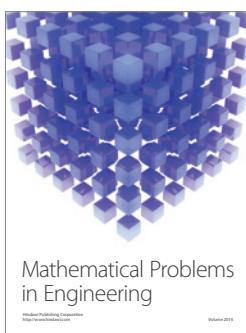
Advances in  
Mathematical Physics



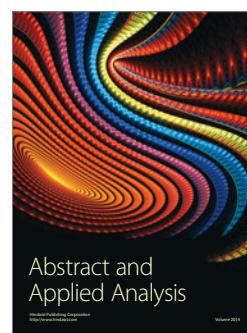
Journal of  
Complex Analysis



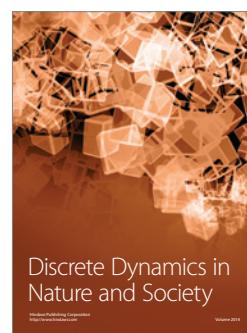
Journal of  
Mathematics



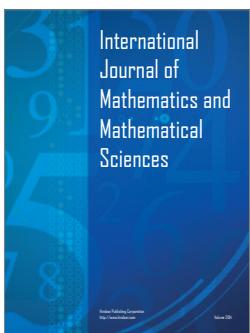
Mathematical Problems  
in Engineering



Abstract and  
Applied Analysis



Discrete Dynamics in  
Nature and Society



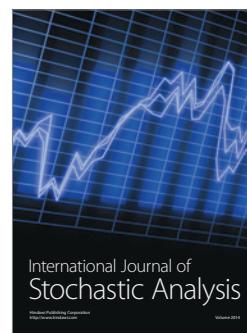
International  
Journal of  
Mathematics and  
Mathematical  
Sciences



Journal of  
Discrete Mathematics



Journal of  
Function Spaces



International Journal of  
Stochastic Analysis



Journal of  
Optimization