Statistics and Probability Letters 98 (2015) 144–150

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

The difference of symmetric quantiles under long range dependence

G. Tarr *, N.C. Weber, S. Müller

School of Mathematics and Statistics, University of Sydney, Sydney 2006, Australia

A R T I C L E   I N F O

Article history:
Received 2 March 2014
Received in revised form 22 December 2014
Accepted 24 December 2014
Available online 5 January 2015

MSC:
62G30
62G20
60F05

Keywords:
Interquartile range
Long range dependence
Pairwise means
U-statistics

A B S T R A C T

This paper investigates two robust estimators of the scale parameter given data from a stationary, long range dependent Gaussian process. In particular the limiting distributions of the interquartile range and related τ-quantile range statistics are established. In contrast to single quantiles, the limiting distribution of the difference of two symmetric quantiles is determined by the level of dependence in the underlying process. It is shown that there is no loss of asymptotic efficiency for the τ-quantile range relative to the standard deviation under extreme long range dependence which is consistent with results found previously for other estimators of scale.

1. Introduction

Accurate estimation of scale plays a critical role in the analysis of time dependent data. It is well known that the classical standard deviation is not robust to outliers, where even a single outlier can have a severe effect, see, for example Tsay (1988), Chan (1995) and Maronna et al. (2006, Chapter 8).

Long range dependent (LRD) time series form an important class of dependent observations. Applications involving LRD processes are found in a diverse range of settings, such as the analysis of network traffic, heartbeat fluctuations, wind turbine output, climate and financial data, to name just a few. For some recent examples we refer to Park et al. (2011), Ercan et al. (2013) and Beran et al. (2013).

The distinction between short and long range dependence is due to the behaviour of the autocovariance decay rate. In particular, let $\{X_i\}_{i\geq 0}$ be a stationary time series and let $\gamma(h)$ be its autocovariance function at lag h and define the scale parameter, $\sigma = \sqrt{\gamma(0)}$. The sequence $\{X_i\}_{i\geq 0}$ is said to be of long memory if its autocovariances are not summable, $\sum_{h\geq 0} |\gamma(h)| = \infty$. LRD processes are characterised by spurious trends, self-similarity and an autocovariance function that exhibits a slow hyperbolic decay, $\gamma(h) \sim h^{-D}$ for $D \in (0, 1)$.

Formally, let $\{X_i\}_{i\geq 0}$ be the stationary, mean-zero, linear Gaussian process, $X_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$, where $a_0 = 1$, $\sum_{j=0}^{\infty} a_j^2 < \infty$ and the innovations, $\{\varepsilon_k\}_{k\in\mathbb{Z}}$, are i.i.d. mean-zero Gaussian random variables with variance $\mathrm{var}(\varepsilon_1) = \sigma_\varepsilon^2 < \infty$. Assume, without loss of generality, $\mathbb{E}(X_0^2) = 1$. The resulting sequence is long range dependent if $a_j \sim j^{d-1} L_a(j)$ for $d \in (0, 1/2)$ where L_a is a slowly varying function at infinity. In which case $\gamma(h) = \mathbb{E}(X_0 X_h) \sim h^{-D} L(h)$, where $D = 1 - 2d$ is the

* Corresponding author.
E-mail address: garth.tarr@sydney.edu.au (G. Tarr).

http://dx.doi.org/10.1016/j.spl.2014.12.022
0167-7152/

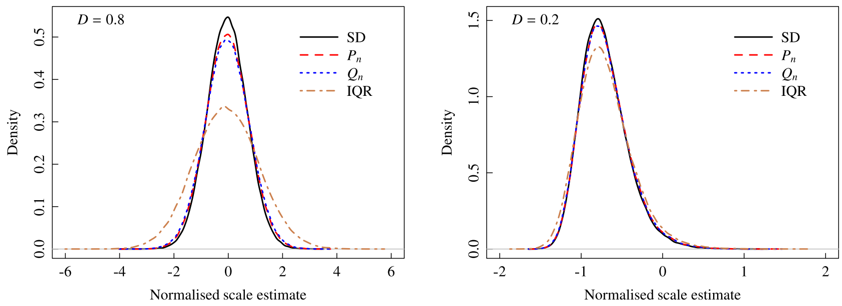

Fig. 1. Empirical densities of normalised scale estimates for ARFIMA(0, 0.1, 0) models, $\sqrt{n}(S-\sigma)$, (left) and ARFIMA(0, 0.4, 0) models, $n^{0.2}(S-\sigma)$, (right) over 100,000 replications of samples of size $n = 1000$.

autocovariance decay rate, $D \in (0, 1)$, and $L(h) = \sigma_\varepsilon^2 L_a^2(h)k(D)$, where $k(D) = \text{Beta}(D, d) = \Gamma(D)\Gamma(d)/\Gamma(D + d)$, see for example Beran et al. (2013, Lemma 2.1). The notation $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.

This paper presents a new limit result for the τ-quantile range, an important special case being the interquartile range. Let $F_n(x) = n^{-1}\sum_{i=1}^n \mathbb{I}\{X_i \le x\}$ be the empirical distribution function with corresponding quantile function, $F_n^{-1}(p) = \inf\{r : F_n(r) \ge p\}$ for $p \in (0, 1)$. Define the empirical τ-quantile range as, $\theta_n = d_\tau[F_n^{-1}(1-\tau) - F_n^{-1}(\tau)]$, for $\tau \in (0, 1/2)$, where $d_\tau = 1/[2\Phi^{-1}(1-\tau)]$ is a correction factor to ensure consistency for the standard deviation when the observations follow a Gaussian distribution. Note that when $\tau = 1/4$, we recover the interquartile range.

When $D \in (1/2, 1)$ we establish a normal limit result for θ_n. In contrast, when $D \in (0, 1/2)$, the limiting distribution is no longer normal. This result is somewhat surprising given that individually the quantiles of a LRD process have normal limit distributions for all $D \in (0, 1)$ and as such, it may be reasonable to expect that the τ-quantile range under LRD follows a normal distribution for all $D \in (0, 1)$. To explain this discrepancy for $D \in (0, 1/2)$ we show that the second order terms in the Bahadur representation of sample quantiles play an important role in the limiting distribution of differences of symmetric quantiles.

We also consider the estimator P_n, the τ-quantile range of the pairwise means (Tarr et al., 2012). Given a set of n observations, $X_1, \ldots, X_n$, the set of $n(n-1)/2$ pairwise means is $\{(X_i + X_j)/2,\ 1 \le i < j \le n\}$. Let $G(r) = \text{P}(X_1 + X_2 \le 2r)$ be the cumulative distribution function of the pairwise means with corresponding empirical distribution function,

$$G_n(r) = \frac{2}{n(n-1)}\sum_{i<j}\mathbb{I}\{X_i + X_j \le 2r\}, \quad \text{for } r \in \mathbb{R}.$$

For $p \in (0, 1)$, define the corresponding U-quantile as $G^{-1}(p) := \inf\{r : G(r) \ge p\}$ and hence the pth sample quantile is $G_n^{-1}(p) := \inf\{r : G_n(r) \ge p\}$. The general form of the scale estimator P_n is,

$$P_n = c_\tau\left[G_n^{-1}(1-\tau) - G_n^{-1}(\tau)\right], \quad \text{for } \tau \in (0, 1/2), \tag{1}$$

where c_τ is a correction factor to make P_n consistent for the standard deviation under the Gaussian distribution. Tarr et al. (2012) highlight the robustness and desirable efficiency properties of the P_n estimator when the data are independent and follow a Gaussian distribution. The estimator P_n, and variants thereof, have also been used in the context of precision matrix estimation under cellwise contamination, see Tarr et al. (submitted for publication).

When $D \in (1/2, 1)$ we obtain a normal limit result for P_n and we motivate the need for a Bahadur representation for U-quantiles under long range dependence to establish a non-normal limit for the $D \in (0, 1/2)$ case. Again, this difference in behaviour is surprising given that Lévy-Leduc et al. (2011a) show that the Hodges–Lehmann estimator, the median of the pairwise means, has a normal limit distribution for all $D \in (0, 1)$ and it is simple to show that the same limit distribution applies to other quantiles of the pairwise means.

The results can be seen graphically in Fig. 1 where the data generating process is an ARFIMA(0, d, 0) model. On the left panel $d = 0.1$ so $D \in (1/2, 1)$ and the limiting distributions for the various scale estimators are normal. On the right panel $d = 0.4$ so $D \in (0, 1/2)$ and the limiting distributions are non-normal.

Our findings are in line with what has been found for other robust scale estimators. The Shamos–Bickel scale estimator is proportional to the median of the interpoint distances, $\hat{\sigma}_{\text{SB}} = a\,\text{median}\{|x_i - x_j|; i < j\}$, where a is a correction factor to ensure consistency for the standard deviation at the normal (Shamos, 1976; Bickel and Lehmann, 1979). A related estimator Q_n, proposed by Rousseeuw and Croux (1993), is asymptotically proportional to the first quartile of the interpoint distances, $Q_n = b\,\{|x_i - x_j|; i < j\}_{(k)}$, where b is a correction factor and $k \approx \binom{n}{2}/4$ for large n. Lévy-Leduc et al. (2011a,b) show that both robust scale estimators have a normal limit distribution when $D \in (1/2, 1)$ and the same non-normal limit distribution as the standard deviation when $D \in (0, 1/2)$ with no loss of asymptotic efficiency.

It is important to note that the results found for the Shamos–Bickel scale estimator and Q_n mentioned above are based on the analysis of a single U-quantile. In Sections 2 and 3 we present theorems that give the limiting distribution of differences of two quantiles or U-quantiles based on LRD processes. In order to apply existing LRD limit results we establish various