

MODERATE DEVIATIONS IN SUBSAMPLING DISTRIBUTION ESTIMATION

PATRICE BERTAIL, ANTHONY GAMST, AND DIMITRIS N. POLITIS

(Communicated by Stanley Sawyer)

ABSTRACT. In Politis and Romano (1994) the subsampling methodology was put forth for approximating the sampling distribution (and the corresponding quantiles) of general statistics from i.i.d. and stationary data. In this note, we address the question of how well the subsampling distribution approximates the tail of the target distribution. In the regular setting of the sample mean of an m -dependent sequence we show a moderate deviation property of the subsampling distribution.

1. INTRODUCTION

Let $\underline{X}_n = (X_1, \dots, X_n)$ be an observed stretch of a (strictly) stationary m -dependent sequence of real random variables $\{X_t, t \in \mathbb{Z}\}$ such that $E|X_t|^3 < \infty$. Recall that the m -dependence condition means essentially that the set of random variables (X_1, \dots, X_k) is independent of the set of random variables $(X_{k+m+1}, \dots, X_{2k+m})$ for any $k \in \mathbb{N}$. The special case of independent, identically distributed (i.i.d.) data is covered by 0-dependence.

It is assumed that the mean $\theta = EX_t$ is unknown and estimated by the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ of the observed stretch $\underline{X}_n = (X_1, \dots, X_n)$. For convenience, however, we will assume here that $\sum_{k=-m}^m \text{Cov}(X_0, X_k) = 1$, so that the following Central Limit Theorem (CLT) holds:

$$(1.1) \quad K_n(x, P) \equiv P\{\sqrt{n}(\bar{X}_n - \theta) \leq x\} \longrightarrow \Phi(x), \quad \text{as } n \rightarrow \infty.$$

As a matter of fact, the CLT (1.1) for stationary m -dependent sequences holds under the sole assumption that the sequence $\{X_t, t \in \mathbb{Z}\}$ has finite second moments; see, for example, [5].

Let b be an integer depending on n such that $1 < b < n$, and such that $b \rightarrow \infty$ as $n \rightarrow \infty$, with $b/n \rightarrow 0$. Since b should be viewed as a function of n , a better notation might be to write $b(n)$ or b_n ; nevertheless, for the sake of notational simplicity, the dependence of b on n will not be explicitly denoted.

Also let $\bar{X}_{b,i}$ be the sample mean of the block $(X_i, X_{i+1}, \dots, X_{i+b-1})$, i.e., $\bar{X}_{b,i} = b^{-1} \sum_{k=i}^{i+b-1} X_k$. The subsampling distribution of the sample mean is defined (see

Received by the editors April 6, 1998 and, in revised form, April 30, 1999.
1991 *Mathematics Subject Classification*. Primary 60F05; Secondary 60F10.
Key words and phrases. Central limit theorem, large deviations, resampling.

[6]) as

$$(1.2) \quad K_b(x, \underline{X}_n) \equiv q^{-1} \sum_{i=1}^q 1\{\sqrt{b}(\bar{X}_{b,i} - \bar{X}_n) \leq x\},$$

where $q = n - b + 1$, and $1\{\cdot\}$ is the indicator function of the event $\{\cdot\}$.

The CLT (1.1) together with the design requirements that $b \rightarrow \infty$ as $n \rightarrow \infty$ with $b/n \rightarrow 0$ are sufficient (see [6]) to ensure that the subsampling distribution is consistent, i.e. that

$$(1.3) \quad \sup_x |K_b(x, \underline{X}_n) - \Phi(x)| = o_P(1)$$

and

$$(1.4) \quad \sup_x |K_b(x, \underline{X}_n) - K_n(x, P)| = o_P(1)$$

as n tends to infinity.

The question of large and/or moderate deviations in a convergence such as (1.3) or (1.4) has to do with investigating the rate of convergence as x itself is allowed to increase with n . In particular, the moderate deviation rate of convergence in equation (1.4) is especially interesting in a statistical application where the objective is to approximate the quantiles of the (generally unknown) distribution $K_n(x, P)$ by the corresponding quantiles of $K_b(x, \underline{X}_n)$. Typically, the quantiles of interest are 95% or 99%, a fact which immediately points to the (right) tail of the estimated distribution $K_n(x, P)$; cf. [6] for more details. The rate of convergence in equation (1.4) is the subject of the next section.

2. MODERATE DEVIATIONS IN SUBSAMPLING DISTRIBUTION ESTIMATION

To study moderate deviations we will consider the estimator $K_b(x, \underline{X}_n)$ as x increases with n at an appropriate rate; thus, we consider any nondecreasing sequence $x_n > 0$. Our first result concerns the unbiased version of the subsampling distribution given by

$$\tilde{K}_b(x_n, \underline{X}_n) \equiv q^{-1} \sum_{i=1}^q 1\{\sqrt{b}(\bar{X}_{b,i} - \theta) \leq x_n\}.$$

Lemma 2.1. *Assume that $\{X_t, t \in \mathbb{Z}\}$ is a stationary, m -dependent sequence of real random variables with finite second moments. Let $n \rightarrow \infty$ and let $x_n > 0$ be any nondecreasing sequence; then*

$$(2.1) \quad \tilde{K}_b(x_n, \underline{X}_n) = K_b(x_n, P) + \sqrt{1 - K_b(x_n, P)} O_P\left(\sqrt{\frac{b}{n}}\right),$$

where $\bar{X}_b = b^{-1} \sum_{i=1}^b X_i = \bar{X}_{b,1}$.

Proof. Consider the quantity

$$\tilde{K}_b^{(j)}(x_n, \underline{X}_n) \equiv \frac{b+m}{n} \sum_{i=0}^{[n/(b+m)]-1} 1\{\sqrt{b}(\bar{X}_{b,i(b+m)+j} - \theta) \leq x_n\},$$

for $j = 1, \dots, (b+m)$, where $[\cdot]$ denotes the integer part. Note that due to m -dependence, $\tilde{K}_b^{(j)}(x_n, \underline{X}_n)$ is an average of $[n/(b+m)]$ independent and identical distributed Bernoulli random variables with mean $P\{\sqrt{b}(\bar{X}_b - \theta) \leq x_n\} \equiv K_b(x_n, P)$.

It follows that

$$\begin{aligned} \text{Var}(\tilde{K}_b^{(j)}(x_n, \underline{X}_n)) &= \frac{b+m}{n} K_b(x_n, P)(1 - K_b(x_n, P)) \\ &= O\left(\frac{b+m}{n}(1 - K_b(x_n, P))\right). \end{aligned}$$

Now note that

$$\tilde{K}_b(x_n, \underline{X}_n) = \frac{1}{b+m} \sum_{j=1}^{b+m} \tilde{K}_b^{(j)}(x_n, \underline{X}_n).$$

But

$$\begin{aligned} & \left| \text{Cov}\left(\tilde{K}_b^{(j)}(x_n, \underline{X}_n), \tilde{K}_b^{(k)}(x_n, \underline{X}_n)\right) \right| \\ & \leq \sqrt{\text{Var}\left(\tilde{K}_b^{(j)}(x_n, \underline{X}_n)\right) \text{Var}\left(\tilde{K}_b^{(k)}(x_n, \underline{X}_n)\right)} \\ & = \text{Var}\left(\tilde{K}_b^{(j)}(x_n, \underline{X}_n)\right) = O\left(\frac{b+m}{n}(1 - K_b(x_n, P))\right), \end{aligned}$$

where we used the Cauchy-Schwarz inequality together with the fact that

$$\text{Var}\left(\tilde{K}_b^{(j)}(x_n, \underline{X}_n)\right) = \text{Var}\left(\tilde{K}_b^{(k)}(x_n, \underline{X}_n)\right).$$

Now since

$$\text{Var}\left(\tilde{K}_b(x_n, \underline{X}_n)\right) = (b+m)^{-2} \sum_{j=1}^{b+m} \sum_{k=1}^{b+m} \text{Cov}\left(\tilde{K}_b^{(j)}(x_n, \underline{X}_n), \tilde{K}_b^{(k)}(x_n, \underline{X}_n)\right),$$

it follows that $\text{Var}\left(\tilde{K}_b(x_n, \underline{X}_n)\right) = O\left(\frac{b+m}{n}(1 - K_b(x_n, P))\right)$ as well. Since

$$E\tilde{K}_b(x_n, \underline{X}_n) = K_b(x_n, P),$$

the lemma is proven. □

Remark 1. It is interesting to note that since the random variables

$$Y_{n,i} = \sqrt{b}(\bar{X}_{b,i} - \theta), \quad \text{with } b = o(n),$$

form a triangular array of strong mixing random variables, the quantity $\tilde{K}_b(x_n, \underline{X}_n)$ can be viewed as an empirical process indexed by the class \mathbf{F} of indicator functions $1\{\cdot < x\}$ with x ranging over the set of real numbers \mathbb{R} . Hence, $K_b(x_n, \underline{X}_n)$ is an empirical process with an estimated parameter. The class \mathbf{F} of indicator functions is polynomial with bracketing entropy $N(\delta, \mathbf{F}) \leq \delta^{-1}$ (see, for example, [7], [8]), and this allows us to apply the empirical process results for strong mixing data, surveyed for example in [1], to the empirical process $\tilde{K}_b(x_n, \underline{X}_n)$ indexed by \mathbf{F} . If the sequence of random variables $Y_{n,i}$ is strong mixing with mixing coefficient $\alpha(k) = O(k^{-A})$ with $A > 1/2$, which clearly obtains in the m -dependent setting studied here, then $\sup_x \sqrt{q} |\tilde{K}_b(x, \underline{X}_n) - K_b(x, P)|$, with $q = n - b + 1$, converges in distribution to the supremum of a Brownian bridge on \mathbb{R} with covariance function $\rho(x, y) = K(x, P)(1 - K(y, P))$; take $Q = 2$ and $\gamma = 2$ in Theorem 2.2 of [1]. It is easy to see that such a result will hold for any statistic having a non-degenerate asymptotic distribution and not only for the sample mean (see [6]).

Remark 2. Note that Lemma 2.1 holds even in the case of fixed (nonincreasing) b . However, in order to estimate the limiting distribution $K(x, P)$ via subsampling, b must tend to infinity as well. In addition, in order to obtain a moderate deviations consistency result, we must impose some limitation on the growth of x_n .

Theorem 2.2. *Assume that $\{X_t, t \in \mathbb{Z}\}$ is a stationary, m -dependent sequence of real random variables such that $E|X_t|^3 < \infty$. Let $b \rightarrow \infty$ as $n \rightarrow \infty$ with $b = o(n)$, and let $x_n > 0$ be any nondecreasing sequence such that $x_n = O(b^{1/6})$. Then,*

$$(2.2) \quad K_b(x_n, \underline{X}_n) = \tilde{K}_b(x_n, \underline{X}_n) + O_P \left(\sqrt{\frac{b}{n}} \left(x_n + \sqrt{\ln \frac{n}{b}} \right) (\varphi(x_n))^{1-\delta} \right),$$

for any fixed $\delta \in (0, 1)$.

Proof. We have $\tilde{K}_b(x_n, \underline{X}_n) - K_b(x_n, \underline{X}_n) = q^{-1} \sum_{i=1}^q Y_i$, where

$$Y_i = 1\{\sqrt{b}(\bar{X}_{b,i} - \theta) \leq x_n\} - 1\{\sqrt{b}(\bar{X}_{b,i} - \bar{X}_n) \leq x_n\}$$

and $1\{\cdot\}$ is the indicator function of the event $\{\cdot\}$. Therefore, for any $\eta > 0$ (perhaps depending on n) we have

$$\begin{aligned} |K_b(x_n, \underline{X}_n) - \tilde{K}_b(x_n, \underline{X}_n)| &= q^{-1} \left| \sum_{i=1}^q Y_i \right| \\ &\leq q^{-1} \left| \sum_{i=1}^q Y_i 1\left\{|\bar{X}_n - \theta| > \frac{\eta}{\sqrt{n}}\right\} \right| + q^{-1} \left| \sum_{i=1}^q Y_i 1\left\{|\bar{X}_n - \theta| \leq \frac{\eta}{\sqrt{n}}\right\} \right| \\ &\leq 1\left\{|\bar{X}_n - \theta| > \frac{\eta}{\sqrt{n}}\right\} + q^{-1} \left| \sum_{i=1}^q 1\left\{x_n - \eta\sqrt{\frac{b}{n}} < \sqrt{b}(\bar{X}_{b,i} - \theta) \leq x_n\right\} \right| \\ &\quad + q^{-1} \left| \sum_{i=1}^q 1\left\{x_n < \sqrt{b}(\bar{X}_{b,i} - \theta) \leq x_n + \eta\sqrt{\frac{b}{n}}\right\} \right| \end{aligned}$$

where the last inequality comes from splitting the event $\{Y_i \neq 0\}$ on the value of $(\bar{X}_n - \theta)$. Taking expectations makes it clear that

$$\begin{aligned} E|K_b(x_n, \underline{X}_n) - \tilde{K}_b(x_n, \underline{X}_n)| &\leq P\left\{|\bar{X}_n - \theta| > \frac{\eta}{\sqrt{n}}\right\} + q^{-1} \sum_{i=1}^q P\left\{x_n - \eta\sqrt{\frac{b}{n}} < \sqrt{b}(\bar{X}_{b,i} - \theta) \leq x_n\right\} \\ &\quad + q^{-1} \sum_{i=1}^q P\left\{x_n < \sqrt{b}(\bar{X}_{b,i} - \theta) \leq x_n + \eta\sqrt{\frac{b}{n}}\right\} \\ &= P_I + P_{II} + P_{III}. \end{aligned}$$

Additionally, define

$$z_n = x_n - \eta\sqrt{\frac{b}{n}}.$$

Now from results of Heinrich [2], [3] we have for $x_n = o(b^{1/2})$ that

$$(2.3) \quad \frac{1 - P\{\sqrt{b}(\bar{X}_b - \theta) \leq x_n\}}{1 - \Phi(x_n)} = e_b(x_n)$$

as $b \rightarrow \infty$ where $\gamma = E(X_1 - \theta)^3 / (\text{Var}(X_1))^{3/2}$ is the skewness of X_1 , Φ is the distribution function of a standard normal random variable, and

$$e_b(x_n) = \exp\left(\frac{x_n^3 \gamma}{6\sqrt{b}} + o\left(\frac{x_n^3}{\sqrt{b}}\right)\right) \left(1 + O\left(\frac{1+x_n}{\sqrt{b}}\right)\right),$$

as $b \rightarrow \infty$. Thus we have, uniformly in i ,

$$\frac{1 - P\{\sqrt{b}(\bar{X}_{b,i} - \theta) \leq x_n\}}{1 - \Phi(x_n)} = e_b(x_n)$$

and we can use (2.3) to bound P_{II} and P_{III} .

Now let

$$\eta = x_n + \sqrt{\ln \frac{n}{b}}.$$

Clearly,

$$\begin{aligned} P_{II} &= (1 - \Phi(z_n))e_b(z_n) - (1 - \Phi(x_n))e_b(x_n) \\ &\leq (\Phi(x_n) - \Phi(z_n))e_b(x_n)(1 + o(1)) \\ &= \sqrt{\frac{b}{n}}\eta\varphi(z_n)e_b(x_n)(1 + o(1)) \\ (2.4) \quad &= O\left(\sqrt{\frac{b}{n}}\eta(\varphi(x_n))^{1-\delta}\right) \end{aligned}$$

where the second to last equality comes from a Taylor expansion about x_n , and the last equality from the fact that, with our particular choice for η ,

$$\varphi(z_n)\varphi(x_n)^{\delta-1} = (2\pi)^{\delta/2} \exp(-\delta x_n^2/2 + x_n\eta\sqrt{\frac{b}{n}} - b\eta^2/2n) = O(1).$$

Note also that we have used the facts

$$e_b(x_n) \leq e_b(z_n)(1 + o(1)), \text{ and } e_b(x_n) = O(1),$$

since $x_n = O(b^{1/6})$ by assumption. The P_{III} term is handled similarly.

Using (2.3) again, with n replacing b , we get a bound for the P_I -term as well:

$$\begin{aligned} P_I &\leq (1 - \Phi(\eta))e_n(\eta) \\ (2.5) \quad &= O\left(\frac{\varphi(\eta)}{\eta}e_n(\eta)\right). \end{aligned}$$

It follows that, under the stated assumptions, $e_n(\eta) = O(1)$, and

$$P_I = O\left(\sqrt{\frac{b}{n}}\frac{\varphi(x_n)}{x_n}\right) = O(P_{II}).$$

The above, together with (2.4) and Markov's inequality, gives

$$K_b(x_n, \underline{X}_n) = \tilde{K}_b(x_n, \underline{X}_n) + O_P\left(\sqrt{\frac{b}{n}}\left(x_n + \sqrt{\ln \frac{n}{b}}\right)(\varphi(x_n))^{1-\delta}\right)$$

and this proves the theorem. □

Remark 3. Note that we could have even taken $\delta = 0$ in Theorem 2.2 at the expense of having to add the condition $x_n = o((n/b)^{1/4})$. Nevertheless, this extra condition would only be required if b were equal to or larger than $O(n^{3/5})$, since otherwise $x_n = o((n/b)^{1/4})$ would be implied by the already imposed condition $x_n = O(b^{1/6})$.

The following theorem summarizes our findings regarding moderate deviations in the subsampling distribution of $K_b(x_n, \underline{X}_n)$.

Theorem 2.3. *Under the same conditions as Theorem 2.2 we have*

$$\begin{aligned}
 (2.6) \quad K_b(x_n, \underline{X}_n) &= K_n(x_n, P) + O\left(\frac{\varphi(x_n)}{x_n} \left(e^{x_n^3 \gamma / (6\sqrt{b})} - 1\right)\right) \\
 &+ O_P\left(\sqrt{\frac{b}{n} \frac{\varphi(x_n)}{x_n}}\right) + O_P\left(\sqrt{\frac{b}{n}} \left(x_n + \sqrt{\ln \frac{n}{b}}\right) (\varphi(x_n))^{1-\delta}\right)
 \end{aligned}$$

for any fixed $\delta \in (0, 1)$, where γ is the skewness of X_1 .

Proof. By Lemma 2.1 we have that

$$\begin{aligned}
 \tilde{K}_b(x_n, \underline{X}_n) - K_b(x_n, P) &= \sqrt{1 - K_b(x_n, P)} O_P\left(\sqrt{\frac{b}{n}}\right) \\
 &= \sqrt{(1 - \Phi(x_n))O(1)} O_P\left(\sqrt{\frac{b}{n}}\right) \\
 &= O_P\left(\sqrt{\frac{b}{n}} \frac{1}{\sqrt{x_n}} \exp\left(-\frac{x_n^2}{4}\right)\right)
 \end{aligned}$$

where the second equality comes from the results of Heinrich [2], [3] and the fact that $x_n = O(b^{1/6})$, and the third equality uses the fact that $1 - \Phi(x_n) = O(\varphi(x_n)/x_n)$. Now, with $x_n = O(b^{1/6})$ and $b = o(n)$, the results of Heinrich also imply that

$$\begin{aligned}
 (2.7) \quad \frac{1 - K_n(x_n, P)}{1 - \Phi(x_n)} &= \frac{1 - P\{\sqrt{n}(\bar{X}_n - \theta) \leq x_n\}}{1 - \Phi(x_n)} = e_n(x_n) \\
 &= \exp\left(\frac{x_n^3 \gamma}{6\sqrt{n}} + o\left(\frac{x_n^3}{\sqrt{n}}\right)\right) \left(1 + O\left(\frac{1 + x_n}{\sqrt{n}}\right)\right) = O(1).
 \end{aligned}$$

Finally, combining equation (2.7) with the results of Theorem 2.2 and Lemma 2.1 gives

$$\begin{aligned}
 K_b(x_n, \underline{X}_n) - K_n(x_n, P) &= K_b(x_n, \underline{X}_n) - \tilde{K}_b(x_n, \underline{X}_n) \\
 &\quad + \tilde{K}_b(x_n, \underline{X}_n) - K_b(x_n, P) \\
 &\quad + K_b(x_n, P) - K_n(x_n, P) \\
 &= O_P\left(\sqrt{\frac{b}{n}} \left(x_n + \sqrt{\ln \frac{n}{b}}\right) \exp\left(-\frac{(1 - \delta)x_n^2}{2}\right)\right) \\
 &\quad + O_P\left(\sqrt{\frac{b}{n}} \frac{1}{\sqrt{x_n}} \exp\left(-\frac{x_n^2}{4}\right)\right) \\
 &\quad + (1 - \Phi(x_n))(e_b(x_n) - e_n(x_n));
 \end{aligned}$$

note that $e_b(x_n) - e_n(x_n) = O(e_b(x_n) - 1)$, and the proof is complete. □

Finally, note that our previous results are valid as stated *even* if x_n is a constant. Note that the $O\left(\frac{1}{\sqrt{b}}\right)$ term in the corollary is nothing other than the Berry-Esseen rate in the approximation of $\Phi(x)$ by $K_b(x, P)$.

Corollary 2.4. *Under the same conditions as Theorem 2.2 but with $x_n = x$ (a constant), we have*

$$(2.8) \quad K_b(x, \underline{X}_n) = K_n(x, P) + O\left(\frac{1}{\sqrt{b}}\right) + O_P\left(\sqrt{\frac{b}{n} \ln \frac{n}{b}}\right).$$

Nevertheless, if we assume that $x_n \rightarrow \infty$ even at a slow, logarithmic rate, i.e., we have a *bona fide* moderate deviations set-up, then Theorem 2.3 has the following useful corollary that can be compared with an analogous result of Jing [4] on the moderate deviations property of the block-bootstrap distribution estimator. The corollary follows by letting $\delta = 1/4$ (say) in Theorem 2.3.

Corollary 2.5. *Under the same conditions as Theorem 2.2 but in addition assuming that $x_n \geq \sqrt{2 \ln(n/b)}$, we have*

$$(2.9) \quad K_b(x_n, \underline{X}_n) = K_n(x_n, P) + O_P\left(\sqrt{\frac{b}{n} \frac{\varphi(x_n)}{x_n}}\right).$$

To give an example, equation (2.9) holds true if we choose $x_n = n^\beta$, for some $\beta \in (0, 1/6)$, with the concurrent choice $b = n^\zeta$, for $\zeta \in [6\beta, 1)$.

REFERENCES

- [1] Andrews, D. and Pollard, D. (1994), An introduction to functional central limit theorems for dependent stochastic processes, *Int. Stat. Rev.*, 62, 119-132.
- [2] Heinrich, L. (1982), A method for the derivation of limit theorems for sums of m -dependent random variables, *Z. Wahrsch. verw. Geb.*, 60, 501-515. MR **83j**:60026
- [3] Heinrich, L. (1984), Non-uniform estimates and asymptotic expansions of the remainder in the central limit theorem for m -dependent random variables, *Math. Nachr.* 115, 7-20. MR **85k**:60033
- [4] Jing, B.Y. (1997), On the relative performance of the block bootstrap for dependent data, *Comm. Statist. — Theory and Methods*, vol. 26, 1313–1328. MR **98j**:62034
- [5] Liu, R. and Singh, K. (1992), Moving blocks jackknife and bootstrap capture weak dependence, In *Exploring the limits of bootstrap*, edited by R. LePage and L. Billard, John Wiley, New York, pp. 225-248. MR **94f**:62079
- [6] Politis, D.N. and Romano, J.P. (1994), ‘Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions’, *Ann. Statist.*, vol. 22, No. 4, 2031-2050. MR **96b**:62078
- [7] Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer Verlag, New York. MR **86i**:60074
- [8] Pollard, D. (1991), Bracketing methods in statistics and econometrics, in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press, Cambridge. MR **93m**:62056

INRA-CORELA, 65, BD. DE BRANDEBOURG, 34205 IVRY-SEINE, FRANCE

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT SAN DIEGO, LA JOLLA, CALIFORNIA 92093-0112

E-mail address: `acgamst@osiris.ucsd.edu`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT SAN DIEGO, LA JOLLA, CALIFORNIA 92093-0112

E-mail address: `politis@euclid.ucsd.edu`