BERRY–ESSEEN BOUNDS FOR STATISTICS OF WEAKLY DEPENDENT SAMPLES

V. Bentkus[1] F. Götze[1] A. Tikhomirov[1,2]

University of Bielefeld
Syktyvkar University and the Mathematical Department
of Komi RC of the Russian Academy of Sciences

April 1995

Abstract. We prove Berry–Esseen bounds for a general class of asymptotically normal statistics which are functions of N weakly dependent random variables under easily verifiable conditions. In particular, we show the validity of the bound $\mathcal{O}(N^{-1/2}\log^{\delta} N)$ with some $\delta > 0$, for U-statistics, Studentized means, functions of sample means, functionals of empirical distribution functions and linear combinations of order statistics

1. Introduction and results

Let $X_1, X_2, \ldots$ be a sequence of random variables taking values in arbitrary measurable space $(\mathcal{X}, \mathcal{A})$ which is *stationary in the strong sense.* We shall assume that the sequence satisfies an absolute regularity condition with coefficients $\beta(m) \to 0$, as $m \to \infty$,

$$\beta(m) \overset{\text{def}}{=} \sup_{k\geqslant 1} \mathbf{E} \sup\Big\{ \big|\mathbf{P}\{A \,\big|\, \sigma[1,k]\} - \mathbf{P}\{A\}\big| :\ A \in \sigma[k+m, \infty)\Big\},$$

where $\sigma[a,b]$ denotes the σ-algebra generated by the random variables X_l with $l \in [a,b]$.

The aim of this paper is to prove Berry–Esseen bounds for a sufficiently large class of statistics of weakly dependent random variables. Let $t = t_N$ be a real-valued

---

1991 *Mathematics Subject Classification.* Primary 62E20; secondary 60F05.

Key words and phrases. weakly dependent random variables, mixing, absolute regularity, Berry-Esseen bounds, asymptotically normal statistics, Studentized means, U-statistics, functions of sample means, functionals of the empirical distribution functions, linear combinations of order statistics.

[1] Research supported by the SFB 343 in Bielefeld.

[2] Research supported by Russian Foundation for Fundamental Research. Grant N93–011–1454.

function of N variables. We shall consider statistics $T = t(X_1, \dots, X_N)$ which can be represented as

$$T = S + R, \qquad \text{where} \quad S = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} g(X_j), \tag{1.1}$$

with some function $g : \mathcal{X} \to \mathbf{R}$ such that $\mathbf{E}\, g(X_1) = 0$ and some remainder term $R = R_N(X_1, \dots, X_N)$. The function g may depend on N; throughout we allow such dependence without explicitly mentioning it.

Let

$$\sigma_N^2 \stackrel{\text{def}}{=} \mathbf{E}\, S^2 = \mathbf{E}\, g^2(X_1) + 2 \sum_{j=1}^{N-1} (1 - jN^{-1})\, \mathbf{E}\, g(X_1)\, g(X_{1+j})$$

denote the variance of S, and let $\rho_s = \mathbf{E}\, |g(X_1)|^s$.

In the case of independent identically distributed (i.i.d.) random variables X_1, $X_2, \dots$ the linear part S of the statistic T is asymptotically normal as $N \to \infty$,

$$\mathbf{P}\,\{S < \sigma_N x\} \to \Phi(x), \qquad \text{where} \quad \Phi(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\{-\frac{t^2}{2}\} dt,$$

provided that

$$\sigma_N^2 \geq \Sigma^2 > 0, \qquad \text{for all} \quad N, \tag{1.2}$$

and $\sup_N \rho_{2+\delta} < \infty$, for some fixed $\Sigma > 0$ and $\delta > 0$. In the weakly dependent case the additional condition $\sup_m m^{(1+\varepsilon)(1+1/\delta)} \beta(m) < \infty$, for some $\varepsilon > 0$, ensures the same result (see, Ibragimov and Linnik (1971), Ibragimov (1975), Eberlein (1984)). Thus the statistic T will be asymptotically normal provided that in addition $R \to 0$ in probability.

In order to prove Berry-Esseen bounds some stronger conditions are necessary. It is known (Tikhomirov (1980)) that the conditions

$$\sup_N \rho_3 \leq \rho < \infty \tag{1.3}$$

and

$$\beta(m) \leqslant K \exp\{-\beta m\}, \qquad \text{for all} \quad m \geqslant 1, \tag{1.4}$$

with some $K < \infty$ and $\beta > 0$, together imply

$$\sup_x \Big| \mathbf{P}\,\{S < \sigma_N x\} - \Phi(x) \Big| \leqslant A N^{-1/2} \log^2 N$$

with a constant A depending on K, β, Σ and ρ only. We shall extend this estimate for general nonlinear statistics.

Let $\sigma^c[j,k]$ denote the σ-algebra generated by X_l such that $l \notin [j,k]$ and $1 \leq l \leq N$. In the case $k < j$, set $\sigma^c[j,k] = \sigma(X_1, \dots, X_N)$.