Z. Wahrscheinlichkeitstheorie verw. Gebiete
62, 17–37 (1983)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Strong Approximation of Very Weak Bernoulli Processes

Ernst Eberlein

Institut f. Mathem. Stochastik, Universität Freiburg,
Hebelstraße 27, D-7800 Freiburg, Federal Republic of Germany

Summary. Very weak Bernoulli processes with values in a separable metric space are introduced. An estimate for the Prohorov distance in the central limit theorem is obtained. This estimate is used to establish a strong (almost sure) approximation of the partial sums of a very weak Bernoulli process by a Brownian motion where the error term is of the order $O(t^{1/2-\gamma})$. The proofs are based on a new version of the Berkes-Philipp approximation theorem.

1. Introduction and Statement of Results

Functional central limit theorems as well as the functional law of the iterated logarithm and other asymptotic fluctuation results for processes $X=(x_k)_{k\in Z}$ can be derived from a strong (or almost sure) approximation of the partial sums of the process by a Brownian motion with a sufficiently small error term.

In a remarkable paper Kuelbs and Philipp [14] recently established a number of strong approximation results for Banach space valued sequences satisfying various mixing conditions. Their proofs rest on an approximation theorem given first by Berkes and Philipp [2] and in more general form by Philipp [16]. Dehling and Philipp [7] showed that various approximation results stated in [14] for ϕ-mixing sequences continue to hold for absolutely regular sequences. Dehling [4] succeeded to improve the error term in the approximation of the last mentioned sequences. By constructing a counterexample he pointed out the limits of the argumentation used ([4], [5]). For processes with a finite state space absolute regularity is equivalent to being weak Bernoulli.

In this paper we will get an approximation of the order

$$\sum_{\nu\leq t} x_\nu - X(t) \ll t^{1/2-\gamma} \tag{1.1}$$

for very weak Bernoulli processes (Definition 1). Very weak Bernoulli processes with finite state spaces were introduced by Ornstein (see e.g. [15]) in connection with the solution of the isomorphism problem of ergodic theory.

0044-3719/83/0062/0017/$04.20

Our first result (Theorem 1) is a new version of the Berkes-Philipp approximation theorem. It is used in the proof of Theorem 2 to estimate the Prohorov distance in the central limit theorem for very weak Bernoulli processes. Both Theorem 1 as well as Theorem 2 are needed to prove (1.1) which is stated as Theorem 3. The last result finally is a further application of our basic approximation theorem.

Some remarks concerning the notation: $\mathfrak{L}(X)$ denotes distribution of a random variable X defined on a probability space $(\Omega, \mathfrak{A}, P)$. If $A \in \mathfrak{A}$, $P(A)>0$ we denote by $\mathfrak{L}(X|A)$ the conditional distribution of X given A. $\mathfrak{L}((X_1, \ldots, X_n)|A)$ is used if $X=(X_1, \ldots, X_n)$ is a vector. The variance of a real-valued random variable X is denoted by $\mathrm{Var}(X)$. $N(0, \tau^2)$ stands for a normal distribution with mean zero and variance τ^2. Stationarity of a process will always mean strict stationarity. Given a Banach space B, B^* is the topological dual space. For any subset E of a metric space (S, σ) and $\varepsilon>0$, E^ε denotes the open ε-neighborhood of E. $E^\complement$ means complement of a set E. Z are the integers. "$\ll$" and "$O(.)$" are used with the same meaning.

Theorem 1. *Let $\{(S_k, \sigma_k)|k \geqq 1\}$ be a sequence of complete separable metric spaces and let $(X_k)_{k\geqq 1}$ be a sequence of random variables with values in S_k. Let $(\mathfrak{F}'_k)_{k\geqq 1}$ be a sequence of σ-algebras such that X_k is $\mathfrak{F}'_k$-measurable. Suppose that for every $k \geqq 1$ there exist $D_k \in \mathfrak{A}'_k = \bigvee_{j\leqq k} \mathfrak{F}'_j$, $\varepsilon_k \geqq 0$ and $\eta_k \geqq 0$ such that $P(D_k^\complement) \leqq \eta_k$ and if $A \in \mathfrak{A}'_{k-1}$, $A \subset D_{k-1}$, $P(A)>0$ then*

$$\inf_{\lambda \in \mathfrak{P}_k} \int \sigma_k(u,v)\, d\lambda(u,v) \leqq \varepsilon_k^2/2 \tag{1.2}$$

where $\mathfrak{P}_k$ is the class of all Borel probability measures λ on $S_k \times S_k$ with marginals $\mathfrak{L}(X_k|A)$ and $\mathfrak{L}(X_k)$. Let $(G_k)_{k\geqq 1}$ be a sequence of Borel probability measures on S_k and ρ_k, δ_k nonnegative numbers such that

$$\mathfrak{L}(X_k)(E) \leqq G_k(E^{\rho_k}) + \delta_k \tag{1.3}$$

for all Borel sets $E \subset S_k$.

Then without loss of generality there exists a sequence $(Y_k)_{k\geqq 1}$ of independent random variables such that $\mathfrak{L}(Y_k) = G_k$ and

$$P[\sigma_k(X_k, Y_k) \geqq 2\rho_k + \varepsilon_k] \leqq \varepsilon_k + \eta_{k-1} + \delta_k. \tag{1.4}$$

Here and in Theorem 3 the phrase "without loss of generality ..." is to be understood in the sense of Strassen [20]: without changing its distribution we can redefine the sequence on a new probability space on which there exists a sequence $(Y_k)_{k\geqq 1}$ satisfying (1.4) (resp. a Brownian motion $(X(t))_{t\geqq 0}$ satisfying (1.1)).

(1.2) means that the distributions $\mathfrak{L}(X_k|A)$ and $\mathfrak{L}(X_k)$ have distance less than $\varepsilon_k^2/2$ in the Wasserstein metric which is defined in general as follows. Let μ, ν be two Borel probability measures on a complete separable metric space (S, σ) then we define their Wasserstein distance by

$$\rho(\mu, \nu) = \inf_{\lambda \in \mathfrak{P}} \int \sigma(u,v)\, d\lambda(u,v) \tag{1.5}$$

where $\mathfrak{P}$ is the class of Borel probability measures on $S\times S$ with marginals μ and ν. Since S is complete and separable both of the measures μ and ν are tight. This implies that the infimum in (1.5) is actually a minimum.

Recall that the Prohorov distance is defined by

$$\pi(\mu,\nu)=\inf\{\varepsilon>0\,|\,\mu(E)\leqq\nu(E^\varepsilon)+\varepsilon \text{ for all closed } E\subset S\}. \tag{1.6}$$

We will make use of the following relation

$$\pi(\mu,\nu)\leqq(\rho(\mu,\nu))^{1/2}. \tag{1.7}$$

This inequality can easily be derived from the Čebyšev-Markov inequality and the following alternative definition for π

$$\pi(\mu,\nu)=\inf_{\lambda\in\mathfrak{P}}\inf\{\varepsilon>0\,|\,\lambda(\{(u,v)\,|\,\sigma(u,v)>\varepsilon\})\leqq\varepsilon\}$$

which is a consequence of the Strassen-Dudley theorem [9].

We will now define the Wasserstein distance for distributions of processes. For each n we consider a metric σ_n on the product space $S^n=S\times\ldots\times S$

$$\sigma_n(u,v)=n^{-1}\sum_{i=1}^{n}\sigma(u_i,v_i)$$

if $u=(u_1,\ldots,u_n)$, $v=(v_1,\ldots,v_n)\in S^n$. Given two S-valued processes $X=(x_k)_{k\in Z}$ and $Y=(y_k)_{k\in Z}$ we define for $n\geqq 1$, $k\in Z$

$$\bar\rho_n((x_{k+1},\ldots,x_{k+n}),(y_{k+1},\ldots,y_{k+n}))=\inf_{\lambda\in\mathfrak{P}_n}\int\sigma_n\,d\lambda \tag{1.8}$$

where $\mathfrak{P}_n$ is the class of all Borel probability measures on $S^n\times S^n$ with marginals $\mathfrak{L}((x_{k+1},\ldots,x_{k+n}))$ and $\mathfrak{L}((y_{k+1},\ldots,y_{k+n}))$. Since S^n is separable and complete again the infimum in (1.8) is a minimum. If $k=0$ then we write for short $\bar\rho_n(X,Y)$ in (1.8). The processes Y we consider in the following are processes derived from a fixed process X by conditioning X on certain sets A of positive measure, i.e. we consider measures λ with one of the marginals being $\mathfrak{L}((x_{k+1},\ldots,x_{k+n})\,|\,A)$. Again we shall write $X|A$ for short if $k=0$.

In the following $(\mathfrak{F}_k)_{k\in Z}$ will denote a family of sub-σ-algebras, where we will always assume without further mention that $\mathfrak{F}_k$ is countably generated for each k.

Definition 1. Let $X=(x_k)_{k\in Z}$ be a strictly stationary S-valued process defined on a probability space $(\Omega,\mathfrak{A},P)$ and let $(\mathfrak{F}_k)_{k\in Z}$ be a family of sub-σ-algebras of $\mathfrak{A}$ such that x_k is $\mathfrak{F}_k$-measurable. The process is called *very weak Bernoulli (with respect to* $(\mathfrak{F}_k)_{k\in Z}$*)* if for each $\varepsilon>0$ there is an n such that for all integers $m\geqq 0$, $k\in Z$ there exists a set

$$D=D(n,m,k)\in\mathfrak{A}_{k-m}^{k}=\bigvee_{j=k-m}^{k}\mathfrak{F}_j$$

such that $P(D^{\complement})<\varepsilon$ and if $A\in\mathfrak{A}_{k-m}^{k}$, $A\subset D$, $P(A)>0$ then

$$\bar\rho_n((x_{k+1},\ldots,x_{k+n}),(x_{k+1},\ldots,x_{k+n})\,|\,A)<\varepsilon. \tag{1.9}$$

Given a sequence $(\varepsilon(n))_{n\geqq 1}$ of positive numbers decreasing to 0 we say $X=(x_k)_{k\in Z}$ is *very weak Bernoulli* (*with respect to* $(\mathfrak{F}_k)_{k\in Z}$) *at rate* $(\varepsilon(n))_{n\geqq 1}$ if for all integers $m\geqq 0$, $n\geqq 1$, $k\in Z$ there exists a set $D=D(n,m,k)\in\mathfrak{A}^k_{k-m}$ such that $P(D^{\complement})\leqq\varepsilon(n)$ and if $A\in\mathfrak{A}^k_{k-m}$, $A\subset D$, $P(A)>0$ then

$$\bar{\rho}_n((x_{k+1},\ldots,x_{k+n}),(x_{k+1},\ldots,x_{k+n})|A)\leqq\varepsilon(n). \tag{1.10}$$

The process is called *strictly very weak Bernoulli* resp. *strictly very weak Bernoulli at rate* $(\varepsilon(n))_{n\geqq 1}$ if (1.9) resp. (1.10) hold for any set $A\in\mathfrak{A}^k_{k-m}$, $P(A)>0$.

The dependence structure given by this definition differs from the mixing conditions in the sense that the rate $(\varepsilon(n))_{n\geqq 1}$ has implications on the qualitative dependence structure. More precisely, a process which is very weak Bernoulli at rate $\varepsilon(n)=o(n^{-1})$ is a Bernoulli process, i.e. is independent. It is easy to see that this rate implies that the process is strongly mixing. By a simple extension of the argument it was shown recently [6] that this rate actually implies independence. ("Strongly mixing" is used here in the probabilistic sense, not in the terminology of ergodic theory).

If the state space S is a finite set endowed with the discrete metric and the $\mathfrak{F}_k$ are the σ-algebras generated by the x_k, Definition 1 coincides with the one due to Ornstein. In the following two theorems we consider real-valued processes. Thus $\bar{\rho}_n$ will be computed with respect to

$$\sigma_n(u,v)=n^{-1}\sum_{i=1}^{n}|u_i-v_i| \tag{1.11}$$

for $u,v\in\mathbb{R}^n$.

Theorem 2. *Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate*

$$\varepsilon(n)\ll n^{-1} \tag{1.12}$$

Suppose x_0 is centered and bounded with probability 1 *and* $\lim\limits_{n\to\infty}\operatorname{Var}\left(\sum\limits_{k=1}^{n}x_k\right)=\infty$ *then there exists a finite positive value σ^2 such that* $\lim\limits_{n\to\infty}n^{-1}\operatorname{Var}\left(\sum\limits_{k=1}^{n}x_k\right)=\sigma^2$ *and for some $\kappa>0$*

$$\pi\left(\mathfrak{L}\left(n^{-1/2}\sum_{\nu=1}^{n}x_\nu\right),N(0,\sigma^2)\right)\ll n^{-\kappa}. \tag{1.13}$$

The explicit value of the exponent is $\kappa=1/64$.

Theorem 3. *Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate* (1.12). *Suppose x_0 is centered and bounded with probability* 1 *and* $\lim\limits_{n\to\infty}\operatorname{Var}\left(\sum\limits_{k=1}^{n}x_k\right)=\infty$ *then there exists a finite positive value σ^2 such that* $\lim\limits_{n\to\infty}n^{-1}\operatorname{Var}\left(\sum\limits_{k=1}^{n}x_k\right)=\sigma^2$ *and without loss of generality there exists a Brownian motion $(X(t))_{t\geqq 0}$ with variance σ^2 such that*

$$\sum_{\nu\leqq t}x_\nu-X(t)\ll t^{1/2-\gamma} \tag{1.14}$$

for some $\gamma>0$.

The next definition is a straightforward extension of a notion given in [19]. Let X and Y again be S-valued processes. Recall that $\bar{\rho}_n(X, Y)$ denotes the Wasserstein distance of $\mathfrak{L}((x_1, \ldots, x_n))$ and $\mathfrak{L}((y_1, \ldots, y_n))$. Set

$$\bar{\rho}(X, Y) = \sup_{n \geqq 1} \bar{\rho}_n(X, Y). \tag{1.15}$$

Definition 2. Let $X = (x_k)_{k \in Z}$ be a stationary S-valued process. X is called *almost block independent* if given $\varepsilon > 0$ there is an N such that if $n \geqq N$ and Y is the process defined by the two conditions

$$\mathfrak{L}((y_{nk+1}, \ldots, y_{n(k+1)})) = \mathfrak{L}((x_1, \ldots, x_n)) \quad \text{for all } k \in Z, \tag{1.16}$$

$$\text{for each } k,\ (y_{nk+1}, \ldots, y_{n(k+1)}) \text{ is independent of } \{(y_j)_{j \leqq nk}\} \tag{1.17}$$

then

$$\bar{\rho}(X, Y) \leqq \varepsilon.$$

Theorem 4. *Let $X = (x_k)_{k \in Z}$ be a very weak Bernoulli S-valued process and suppose that σ is bounded, then X is almost block independent.*

Let us finally point out another aspect of this paper. The central limit theorem and the strong approximation result established here are of a rather different nature than the aims and statements one is looking towards in the isomorphism theory. But it became obvious that some of the underlying approximation ideas are exactly the same.

2. Proof of Theorem 1

We consider the case of discrete random variables X_k and probability measures $G_k = \mathfrak{L}(X_k)$ first. This means that the assumptions hold for $\rho_k = \delta_k = 0$ $(k \geqq 1)$.

Define $Y_1 = X_1$ and suppose $Y_1, \ldots, Y_{k-1}$ have been constructed. Suppose furthermore Y_j is $\mathfrak{A}'_j$-measurable $(1 \leqq j \leqq k-1)$. Define $D = D(b_1, \ldots, b_{k-1}) = \{Y_1 = b_1, \ldots, Y_{k-1} = b_{k-1}\} \cap D_{k-1}$ for values $b_1, \ldots, b_{k-1}$ in the range of $Y_1, \ldots, Y_{k-1}$ respectively. Then $D \in \mathfrak{A}'_{k-1}$ and $D \subset D_{k-1}$.

As we mentioned already the infimum in (1.2) is a minimum. Therefore if $P(D) > 0$ there is a Borel probability measure λ on $S_k \times S_k$ with marginals $\mathfrak{L}(X_k | D)$, $\mathfrak{L}(X_k)$ and

$$\lambda(\{(u, v) | \sigma_k(u, v) \geqq \varepsilon_k\}) \leqq \varepsilon_k^{-1} \int \sigma_k \, d\lambda \leqq \varepsilon_k. \tag{2.1}$$

Note that for this part of the proof we use only the bound ε_k^2 in (1.2). Let $\{a_i | i \geqq 1\}$ be the range of X_k. Then λ is concentrated on $\{(a_i, a_j) | i, j \geqq 1\}$ and

$$P[X_k = a_i | D] = \sum_{j \geqq 1} \lambda(a_i, a_j)$$

for each $i \geqq 1$. All the $\mathfrak{F}'_k$ can be assumed to be atomless since $(X_k)_{k \geqq 1}$ can be redefined on a richer probability space if necessary. Therefore $\{X_k = a_i\} \cap D \in \mathfrak{A}'_k$ can be partitioned into $\mathfrak{A}'_k$-measurable sets D_{ij} such that $P(D_{ij} | D) = \lambda(a_i, a_j)$. We define $Y_k = a_j$ on D_{ij} $(i, j \geqq 1)$. Then the joint distribution of (X_k, Y_k) on D is λ. By (2.1) this means

$$P(\{\sigma_k(X_k, Y_k) \geqq \varepsilon_k\} \cap D) \leqq P(D)\, \varepsilon_k. \tag{2.2}$$

On sets D with $P(D)=0$ and on $D' = \{Y_1 = b_1, \ldots, Y_{k-1} = b_{k-1}\} \cap D_{k-1}^{\complement}$ we define Y_k such that $\mathfrak{L}(Y_k | D) = \mathfrak{L}(Y_k | D') = \mathfrak{L}(X_k)$. Now Y_k is defined on the whole space and by (2.2)

$$P[\sigma_k(X_k, Y_k) \geqq \varepsilon_k] \leqq \sum_{D \subset D_{k-1}} P(D)\, \varepsilon_k + \sum_{D' \subset D_{k-1}^{\complement}} P(D') \leqq \varepsilon_k + \eta_{k-1}. \tag{2.3}$$

By the very construction Y_k is $\mathfrak{A}'_k$-measurable, $\mathfrak{L}(Y_k) = \mathfrak{L}(X_k)$ and Y_k is independent of $Y_1, \ldots, Y_{k-1}$.

Now we consider the general case: Fix $k \geqq 1$. Given $\delta > 0$ we can construct a partition $(E_i)_{i \geqq 1}$ of the separable space S_k into Borel sets such that $\operatorname{diam}(E_i) < \delta$ $(i \geqq 1)$. We do this for $\delta = \min(\varepsilon_k^2/4, \rho_k/4)$ and choose a point $u_i \in E_i$ $(i \geqq 1)$. The discrete random variable X'_k defined by $X'_k(\omega) = u_i$ for $\omega \in X_k^{-1}(E_i)$ is $\mathfrak{F}'_k$-measurable and satisfies $\sigma_k(X_k, X'_k) < \delta$. Now let be given a set $A \subset D_{k-1}$, $A \in \mathfrak{A}_{k-1}$, $P(A) > 0$ and a Borel probability measure λ on $S_k \times S_k$ with marginals $\mathfrak{L}(X_k | A)$ and $\mathfrak{L}(X_k)$ such that

$$\int \sigma_k(u, v)\, d\lambda(u, v) \leqq \varepsilon_k^2/2. \tag{2.4}$$

We get a probability measure λ' on $S_k \times S_k$ concentrated on $\{(u_i, u_j) | i, j \geqq 1\}$ if we define $\lambda'(u_i, u_j) = \lambda(E_i \times E_j)$. λ' has marginals $\mathfrak{L}(X'_k | A)$ and $\mathfrak{L}(X'_k)$. Furthermore using (2.4)

$$\begin{aligned} \int \sigma_k(u, v)\, d\lambda'(u, v) &= \sum_{i,j} \int_{E_i \times E_j} \sigma_k(u_i, u_j)\, d\lambda(u, v) \\ &\leqq \sum_{i,j} \int_{E_i \times E_j} (\sigma_k(u, v) + \varepsilon_k^2/2)\, d\lambda(u, v) \\ &= \int \sigma_k\, d\lambda + \varepsilon_k^2/2 \leqq \varepsilon_k^2. \end{aligned}$$

Thus we have constructed a sequence $(X'_k)_{k \geqq 1}$ of discrete, $\mathfrak{F}'_k$-measurable random variables satisfying $\sigma_k(X_k, X'_k) < \min(\varepsilon_k^2/4, \rho_k/4)$ and

$$\inf_{\lambda'} \int \sigma_k(u, v)\, d\lambda'(u, v) \leqq \varepsilon_k^2$$

where infimum is taken over all Borel probability measures λ' on $S_k \times S_k$ with marginals $\mathfrak{L}(X'_k | A)$ and $\mathfrak{L}(X'_k)$ for sets A as considered above. By the first part of the proof there is a sequence $(Y'_k)_{k \geqq 1}$ of independent random variables such that $\mathfrak{L}(Y'_k) = \mathfrak{L}(X'_k)$ and

$$P[\sigma_k(X'_k, Y'_k) \geqq \varepsilon_k] \leqq \varepsilon_k + \eta_{k-1}.$$

The proof is finished following exactly Philipp's arguments in [16] p. 176. We just indicate the changes in the quantities ε_k, η_k, ρ_k, δ_k. By (1.3) we get for any Borel set $E \subset S_k$

$$\mathfrak{L}(Y'_k)(E) \leqq G_k(E^{5\rho_k/4}) + \delta_k.$$

Using the Strassen-Dudley theorem one constructs a probability measure λ'' on $S_k \times S_k$ with marginals $\mathfrak{L}(Y_k')$ and G_k such that

$$\lambda''(\{(u,v) | \sigma_k(u,v) \geqq 3\rho_k/2\}) \leqq \delta_k.$$

This implies that there is a sequence $(Y_k)_{k \geqq 1}$ of independent random variables, $\mathfrak{L}(Y_k) = G_k$ and

$$\begin{aligned} &P[\sigma_k(X_k, Y_k) \geqq 2\rho_k + \varepsilon_k] \\ &\leqq P[\sigma_k(X_k, X_k') \geqq \rho_k/4] + P[\sigma_k(X_k', Y_k') \geqq \varepsilon_k] + P[\sigma_k(Y_k', Y_k) \geqq 3\rho_k/2] \\ &\leqq \varepsilon_k + \eta_{k-1} + \delta_k. \end{aligned}$$

3. Inequalities for Moments

In this section if not specified more precisely B will mean a separable Banach space with norm $\|\cdot\|$, $X = (x_k)_{k \in Z}$ a strictly stationary B-valued stochastic process defined on a probability space $(\Omega, \mathfrak{A}, P)$ and $(\mathfrak{F}_k)_{k \in Z}$ a fixed family of countably generated sub-σ-algebras of $\mathfrak{A}$ such that x_k is $\mathfrak{F}_k$-measurable ($k \in Z$). For $m \leqq n$ we shall write $\mathfrak{A}_m^n$ for the σ-algebra generated by $\{\mathfrak{F}_k | m \leqq k \leqq n\}$. Given a set $A \in \mathfrak{A}$, $P(A) > 0$, $P(.|A)$ denotes conditional probability given A.

Besides of the Banach space norm $\|.\|$ various other norms will be used. $\|.\|_*$ denotes the norm in B^* and $\|.\|_r$ stands for L^r-norm, i.e. if $\int \|x\|^r dP < \infty$ for some B-valued random variable x we write

$$\|x\|_r = (\int \|x\|^r dP)^{1/r} \quad (r \geqq 1).$$

$\|.\|_r$ will also be used for the usual L^r-norm, i.e. if $\int |h|^r dP < \infty$ for some real-valued random variable h then

$$\|h\|_r = (\int |h|^r dP)^{1/r} \quad (r \geqq 1).$$

Lemma 3.1. *Let h be a real-valued random variable measurable with respect to $\mathfrak{A}_{-m}^0$ for some $m \geqq 0$ and having finite r-th absolute moment for some $r > 1$. Let g be a uniformly continuous, bounded function g: $B^n \to \mathbb{R}$ with modulus of continuity $c(\tau)$ and bound $|g| \leqq C$. Then, if $X = (x_k)_{k \in Z}$ is very weak Bernoulli at rate $(\varepsilon(n))_{n \geqq 1}$,*

$$\begin{aligned} &|E[h\, g(x_1, \ldots, x_n)] - E[h]\, E[g(x_1, \ldots, x_n)]| \\ &\leqq \|h\|_r (|c(\tau) + 2C\,\varepsilon(n)\,\tau^{-1}|^s + |2C|^s\,\varepsilon(n))^{1/s}. \end{aligned} \tag{3.1}$$

Here s is defined by $r^{-1} + s^{-1} = 1$ and $\tau > 0$ is any real number.

Proof. We may assume that h is of the form $h = \sum_{i=1}^{l} a_i 1_{A_i}$ for a partition of sets $A_i \in \mathfrak{A}_{-m}^0$ satisfying $P(A_i) > 0$. Then using Hölder's inequality

$$\begin{aligned} &|E[h\, g(x_1, \ldots, x_n)] - E[h]\, E[g(x_1, \ldots, x_n)]| \\ &= \left| \sum_{i=1}^{l} a_i \{\int g(x_1, \ldots, x_n)\, dP(.|A_i) - \int g(x_1, \ldots, x_n)\, dP\}\, P(A_i) \right| \\ &\leqq \|h\|_r \left(\sum_{i=1}^{l} |\int g(x_1, \ldots, x_n)\, dP(.|A_i) - \int g(x_1, \ldots, x_n)\, dP|^s\, P(A_i) \right)^{1/s}. \end{aligned}$$

Since D chosen according to Definition 1 is an element of $\mathfrak{A}^0_{-m}$ we can assume – taking intersections if necessary – that $A_i \subset D$ or $A_i \subset D^{\complement}$ for each i.

Consider the case $A_i \subset D$. Denote $\mathfrak{L}((x_1, \ldots, x_n))$ by μ and $\mathfrak{L}((x_1, \ldots, x_n)|A_i)$ by $\bar{\mu}$. Furthermore let λ be a measure such that

$$\bar{\rho}_n(X, X|A_i) = \int \sigma_n \, d\lambda \tag{3.2}$$

then

$$\begin{aligned}&|\int g(x_1, \ldots, x_n)\, dP(.|A_i) - \int g(x_1, \ldots, x_n)\, dP| \\ &= |\int_{B^n} g(v)\, d\bar{\mu}(v) - \int_{B^n} g(u)\, d\mu(u)| \\ &\leqq \int_{B^n \times B^n} |g(u) - g(v)|\, d\lambda(u, v).\end{aligned}$$

Splitting this integral in the two parts where $\{\sigma_n \leqq \tau\}$ and $\{\sigma_n > \tau\}$ we get the bound $c(\tau) + 2C\lambda(\{\sigma_n > \tau\})$. By the Čebyšev-Markov inequality and (1.10)

$$\lambda(\{\sigma_n > \tau\}) \leqq \tau^{-1} \int \sigma_n \, d\lambda \leqq \tau^{-1} \varepsilon(n). \tag{3.3}$$

Therefore

$$\begin{aligned}&|E[h\, g(x_1, \ldots, x_n)] - E[h]\, E[g(x_1, \ldots, x_n)]| \\ &\leqq \|h\|_r \Big(\sum_{A_i \subset D} |c(\tau) + 2C\tau^{-1} \varepsilon(n)|^s P(A_i) + \sum_{A_i \subset D^{\complement}} |2C|^s P(A_i)\Big)^{1/s}.\end{aligned}$$

(3.1) follows since $P(D^{\complement}) \leqq \varepsilon(n)$. □

Considering less general functions g will allow us to deduce sharper bounds. In the limit theorems to be proved g enters always in the form $g\left(\sum_{i=1}^n x_i\right)$. Write

$$S_n = \sum_{i=1}^n x_i.$$

Lemma 3.2. *Let $X = (x_k)_{k \in Z}$ be a stationary B-valued process which is very weak Bernoulli at rate $(\varepsilon(n))_{n \geqq 1}$. Suppose $\|x_0\| \leqq C$ with probability 1. Let h be a real-valued random variable measurable with respect to $\mathfrak{A}^0_{-m}$ for some $m \geqq 0$ and having finite r-th absolute moment for some $r > 1$. Then for $g \in B^*$ and $n \geqq 1$*

$$\begin{aligned}&|E[h\, g(S_n)] - E[h]\, E[g(S_n)]| \\ &\leqq \|h\|_r \, \|g\|_* ((n\, \varepsilon(n))^s + (2nC)^s \varepsilon(n))^{1/s}\end{aligned} \tag{3.4}$$

where s is defined by $r^{-1} + s^{-1} = 1$. If X is strictly very weak Bernoulli at rate $(\varepsilon(n))_{n \geqq 1}$ and x_0 is integrable then for h and g as above

$$|E[h\, g(S_n)] - E[h]\, E[g(S_n)]| \leqq \|h\|_r \, \|g\|_* \, n\, \varepsilon(n). \tag{3.5}$$

Proof. Let $D = D(n, m, 0) \in \mathfrak{A}^0_{-m}$ be a set chosen according to Definition 1 then for any $A \in \mathfrak{A}^0_{-m}$, $A \subset D$, $P(A) > 0$

$$
\begin{aligned}
&|\int g(S_n)\,dP(.\,|A)-\int g(S_n)\,dP| \\
&=\left|\int_{B^n\times B^n}\left(g\left(\sum_{i=1}^{n} u_i\right)-g\left(\sum_{i=1}^{n} v_i\right)\right)d\lambda((u_1,\dots,u_n),(v_1,\dots,v_n))\right| \\
&\leqq \|g\|_* \int_{B^n\times B^n}\sum_{i=1}^{n}\|u_i-v_i\|\,d\lambda((u_1,\dots,u_n),(v_1,\dots,v_n)) \\
&=\|g\|_* \int_{B^n\times B^n} n\,\sigma_n(u,v)\,d\lambda(u,v)\leqq \|g\|_*\, n\,\varepsilon(n).
\end{aligned}
$$

Here again λ is a measure satisfying (3.2). Now we take the same representation for h as in the proof before and apply Hölder's inequality, then

$$
\begin{aligned}
&|E[h\,g(S_n)]-E[h]\,E[g(S_n)]| \\
&\leqq \|h\|_r\left(\sum_{i=1}^{l}|\int g(S_n)\,dP(.\,|A_i)-\int g(S_n)\,dP|^s\,P(A_i)\right)^{1/s}.
\end{aligned}
$$

(3.5) follows immediately since if X is strictly very weak Bernoulli, the set D has probability 1, i.e. the estimate above applies to each of the A_i. In order to derive (3.4) we split the sum according to $A_i\subset D$ or $A_i\subset D^{\complement}$ and use the trivial estimate $|g(S_n)|\leqq\|g\|_*\,n\,C$ in the latter case. Thus we get the bound

$$
\|h\|_r\{\sum_{A_i\subset D}(\|g\|_*\,n\,\varepsilon(n))^s\,P(A_i)+\sum_{A_i\subset D^{\complement}}(2\,\|g\|_*\,n\,C)^s\,P(A_i)\}^{1/s}.
$$

(3.4) follows since $P(D^{\complement})\leqq\varepsilon(n)$. □

Lemma 3.3. *Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate $(\varepsilon(n))_{n\geq 1}$ given by* (1.12). *Suppose x_0 has mean zero and finite variance then for any $N\geqq 1$ and $M=2N$ or $M=2N+1$*

$$
(2(1-c_M))^{1/2}\,\|S_N\|_2\leqq\|S_M\|_2\leqq\|S_N\|_2(2(1+c_M))^{1/2} \tag{3.6}
$$

where

$$
c_{2N}=C_0/\|S_N\|_2,\qquad c_{2N+1}=2\,C_0/\|S_N\|_2+\|x_0\|_2^2/2\,\|S_N\|_2^2
$$

and C_0 is the constant implied by $\ll$ in (1.12).

Proof. Consider the case $M=2N$ first. We use Lemma 3.2 with $r=2$. Then (3.5) yields

$$
\left|E\left[\left(\sum_{i=-N+1}^{0}x_i\right)\left(\sum_{i=1}^{N}x_i\right)\right]\right|\leqq\|S_N\|_2\,C_0. \tag{3.7}
$$

Therefore by stationarity

$$
\begin{aligned}
\|S_{2N}\|_2^2&=2\,\|S_N\|_2^2+2\,E\left[\left(\sum_{i=1}^{N}x_i\right)\left(\sum_{i=N+1}^{2N}x_i\right)\right] \\
&\leqq 2\,\|S_N\|_2^2+2\,\|S_N\|_2\,C_0.
\end{aligned}
$$

This gives the upper inequality. The lower inequality follows in the same way if we use the representation $S_N=S_{2N}-(S_{2N}-S_N)$.

Now let $M=2N+1$. Applying again (3.5) to $E\left[\left(\sum_{i=-N+1}^{0} x_i\right)\left(\sum_{i=1}^{N+1} x_i\right)\right]$ and to $E\left[\left(\sum_{i=-N+1}^{0} x_i\right)x_1\right]$ we get

$$\begin{aligned}\|S_{2N+1}\|_2^2 &= 2\|S_N\|_2^2+\|x_{2N+1}\|_2^2+2E\left[\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=N+1}^{2N+1} x_i\right)\right]\\ &\quad +2E\left[\left(\sum_{i=N+1}^{2N} x_i\right)x_{2N+1}\right]\\ &\leqq 2\|S_N\|_2^2+\|x_0\|_2^2+4\|S_N\|_2 C_0\\ &=\|S_N\|_2^2(2(1+2C_0/\|S_N\|_2+\|x_0\|_2^2/2\|S_N\|_2^2)).\end{aligned}$$

The lower bound in (3.6) can be derived similarly. □

Lemma 3.4. *Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate $(\varepsilon(n))_{n\geqq 1}$ given by* (1.12). *Suppose x_0 has mean zero, finite variance and $\lim_{n\to\infty}\operatorname{Var}(S_n)=\infty$ then*

$$\operatorname{Var}(S_n)=n\,h(n) \tag{3.8}$$

where $h(n)$ is a slowly varying function of the integral variable n.

Proof. We have to show that for every $k\geqq 1$

$$\lim_{n\to\infty}\operatorname{Var}(S_{nk})/\operatorname{Var}(S_n)=k. \tag{3.9}$$

If we define $y_j=\sum_{i=1}^{n} x_{(j-1)n+i}$ for $j=1,\ldots,k$ then by stationarity

$$\operatorname{Var}(S_{nk})=k\operatorname{Var}(S_n)+2\sum_{l=1}^{k-1} E[y_1(y_2+\ldots+y_{l+1})].$$

By (3.5)

$$|E[y_1(y_2+\ldots+y_{l+1})]|\leqq\left\|\sum_{k=1}^{n} x_k\right\|_2 \ln\varepsilon(\ln)\leqq(\operatorname{Var}(S_n))^{1/2}C_0.$$

Therefore

$$\operatorname{Var}(S_{nk})=k\operatorname{Var}(S_n)+o(\operatorname{Var}(S_n)). \quad □$$

Proposition 3.5. *Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate $(\varepsilon(n))_{n\geqq 1}$ given by* (1.12). *Suppose x_0 has mean zero, finite variance and $\lim_{n\to\infty}\operatorname{Var}(S_n)=\infty$ then there exists a finite positive value σ^2 such that*

$$\lim_{n\to\infty} n^{-1}\operatorname{Var}(S_n)=\sigma^2. \tag{3.10}$$

Before proving this Proposition we restate a lemma from Bradley [3] (Lemma 2). Define under the assumptions of the Proposition above

$$g(n)=n^{-1/2}\|S_n\|_2 \tag{3.11}$$

then the following holds

Lemma 3.6. *Given any $\varepsilon>0$ and any positive integer L_0, there are positive integers N and L with $L \geqq L_0$ such that for all l, $L \leqq l \leqq 2L$ one has*

$$(1-\varepsilon)\, g(N) \leqq g(l) \leqq (1+\varepsilon)\, g(N).$$

The proof of this lemma is the same as in [3], since the basic ingredient used there, namely

$$\lim_{N\to\infty} g(mN)/g(N)=1 \tag{3.12}$$

for all $m \geqq 1$ holds under our assumptions by Lemma 3.4.

Again as in [3] we choose $0<A<1$ sufficiently small that if $(a_n)_{n\geqq 1}$ is any sequence of real numbers such that $\sum_{n\geqq 1} |a_n| \leqq A$ then

$$|1-\prod_{n\geqq 1}(1+a_n)| \leqq 2\sum_{n\geqq 1}|a_n| \quad \text{and} \quad |1-\prod_{n\geqq 1}(1+a_n)^{-1}| \leqq 2\sum_{n\geqq 1}|a_n|.$$

Proof of Proposition 3.5. Let $0<\varepsilon<A$ be given and denote by C_0 the constant implied by $\ll$ in (1.12). Let $r\geqq 2$ be a fixed integer and $C>0$ a constant to be determined later. Since $\lim_{n\to\infty} \mathrm{Var}(S_n)=\infty$ we can choose $L_0 \geqq 0$ sufficiently large such that $L_0^{-1}<\varepsilon/4$ and for all $N \geqq L_0$

$$1-2\,C_0/\|S_N\|_2 - \|x_0\|_2^2/2\,\|S_N\|_2^2 > 2^{-1/r}, \tag{3.13}$$

$$\|S_N\|_2 > C \tag{3.14}$$

and

$$\|x_0\|_2^2/2\,\|S_N\|_2 \leqq C_0. \tag{3.15}$$

By Lemma 3.6 there exist positive integers H and L with $L \geqq L_0$ such that for all l, $L \leqq l \leqq 2L$ one has

$$(1-\varepsilon)\, g(H) \leqq g(l) \leqq (1+\varepsilon)\, g(H).$$

Now let m be any integer $m \geqq 2L$ then $2^M L \leqq m \leqq 2^{M+1} L$ for some M. There exist integers $J_0, J_1, \ldots, J_M$ such that $m=J_M$, $L \leqq J_0 \leqq 2L$ and for all $n=0, 1, \ldots, M-1$, $J_{n+1}=2J_n$ or $J_{n+1}=2J_n+1$. By (3.6)

$$g(J_{n+1}) \leqq g(J_n)(1+3\,C_0/\|S_{J_n}\|_2)^{1/2}$$

where we made use of (3.15). Introducing (3.13) in the lower half of (3.6) and making use of (3.14) we see that $\|S_{J_n}\|_2 \geqq 2^{n(r-1)/2r}\, C$. Therefore

$$g(J_{n+1}) \leqq g(J_n)(1+3\,C_0/C\,2^{n(r-1)/2r})^{1/2}$$

and

$$g(m) \leqq g(J_0)\left(\prod_{n=0}^{\infty}(1+3\,C_0/C\,2^{n(r-1)/2r})\right)^{1/2}.$$

Now we choose C such that $\sum_{n=0}^{\infty} 3\,C_0/C\,2^{n(r-1)/2r} < \varepsilon/4$ then

$$g(m) \leqq g(J_0)(1+\varepsilon)^{1/2} \leqq g(H)(1+\varepsilon)^2. \tag{3.16}$$

Using the lower half of (3.6) we get

$$g(J_{n+1}) \geqq g(J_n)(2J_n/J_{n+1})^{1/2}(1-3\,C_0/\|S_{J_n}\|_2)^{1/2}$$
$$\geqq g(J_n)(1-2^{-(n+1)}L_0^{-1})(1-3\,C_0/C\,2^{n(r-1)/2r})^{1/2}$$

or

$$g(m) \geqq g(J_0)\left(\prod_{n=0}^{\infty}(1-2^{-(n+1)}L_0^{-1})(1-3\,C_0/C\,2^{n(r-1)/2r})\right)^{1/2}.$$

By the choice of L_0 and C

$$g(m) \geqq g(J_0)(1-\varepsilon) \geqq g(H)(1-\varepsilon)^2. \tag{3.17}$$

This together with (3.16) shows that $\lim_{m\to\infty} g(m)$ exists, which proves (3.10). □

Proposition 3.7. *Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate $(\varepsilon(n))_{n\geqq 1}$ given by (1.12). Suppose x_0 has mean zero, finite variance and $\lim_{n\to\infty}\operatorname{Var}(S_n)=\infty$ then*

$$m^{-1}\operatorname{Var}(S_m)-\sigma^2 \ll m^{-1/2} \tag{3.18}$$

where $\sigma^2=\lim_{n\to\infty} n^{-1}\operatorname{Var}(S_n)$.

Proof. Since $\lim_{m\to\infty} g(m)$ exists and is positive by (3.10) we can choose constants $0<C_1<C_2$ such that $C_1 \leqq g(m) \leqq C_2$ for all m. Let C_0 be the same constant as before and let L be a positive integer such that $C_0\left(\sum_{n=0}^{\infty} 2^{-n/2}\right)/C_1\,L<A$. Given an integer m we write $2^M L^2 \leqq m \leqq 2^{M+1} L^2$ and for each $n=0,1,\ldots$ we set $J_n=2^n m$. By (3.6)

$$g(J_{n+1}) \leqq g(J_n)(1+C_0/C_1\,L\,2^{(n+M)/2})^{1/2}.$$

Since $\lim_{m\to\infty} g(m)^2=\sigma^2$ this implies

$$g(m)^2 \geqq \sigma^2 \prod_{n=0}^{\infty}(1+C_0/C_1\,L\,2^{(n+M)/2})^{-1}$$

Using the lower half of (3.6) we get

$$g(m)^2 \leqq \sigma^2 \prod_{n=0}^{\infty}(1-C_0/C_1\,L\,2^{(n+M)/2})^{-1}.$$

Both estimates together yield

$$|g(m)^2-\sigma^2| \leqq 2\sigma^2 \sum_{n=0}^{\infty} C_0/C_1\,L\,2^{(n+M)/2} = O(m^{-1/2}).\quad \square$$

Lemma 3.8. *Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate $(\varepsilon(n))_{n\geqq 1}$ given by (1.12). Suppose x_0 is centered and* *bounded with probability* 1 *and* $\lim_{n\to\infty} \operatorname{Var}(S_n)=\infty$ *then for any* $0\leqq\delta<1$

$$E\left[\left|\sum_{k=1}^{n} x_k\right|^{2+\delta}\right] \ll n^{1+\delta/2}. \tag{3.19}$$

Proof. Write for short $a_n=E[|S_n|^{2+\delta}]$ and $\sigma_n=E[|S_n|^2]^{1/2}$ for any $n\geqq 1$. From Proposition 3.5 we know that

$$\sigma_n^2=\sigma^2 n(1+o(1)) \tag{3.20}$$

for some $\sigma^2>0$. Besides of $S_n=\sum_{i=1}^{n} x_i$ we use $\tilde{S}_n=\sum_{i=n+1}^{2n} x_i$ then by stationarity we get as in Doob [8], p. 226

$$\begin{aligned} a_{2n} &\leqq 2a_n+E[S_n^2\,|\tilde{S}_n|^\delta]+E[|S_n|^\delta\,\tilde{S}_n^2] \\ &\quad +2E[|S_n|^{1+\delta}\,|\tilde{S}_n|]+2E[|S_n|\,|\tilde{S}_n|^{1+\delta}]. \end{aligned} \tag{3.21}$$

Let C_0 again be the constant implied by $\ll$ in (1.12). If $A\in\mathfrak{A}^0_{-n+1}$, $P(A)>0$ and λ is a probability measure satisfying (3.2) then

$$\begin{aligned} &|\textstyle\int |S_n|\,dP(.\,|A)-\int |S_n|\,dP| \\ &=\left|\int_{\mathbb{R}^n\times\mathbb{R}^n}\left(\left|\sum_{k=1}^{n} u_k\right|-\left|\sum_{k=1}^{n} v_k\right|\right)d\lambda(u,v)\right| \\ &\leqq \textstyle\int n\,\sigma_n(u,v)\,d\lambda(u,v)\leqq n\,\varepsilon(n)\leqq C_0. \end{aligned} \tag{3.22}$$

We may assume that

$$\left|\sum_{k=-n+1}^{0} x_k\right|=\sum_{i=1}^{l} a_i 1_{A_i} \tag{3.23}$$

for a partition of sets $A_i\in\mathfrak{A}^0_{-n+1}$ satisfying $P(A_i)>0$. Then by stationarity and Hölder's inequality

$$\begin{aligned} &|E[|S_n|^{1+\delta}\,|\tilde{S}_n|]-E[|S_n|^{1+\delta}]\,E[|\tilde{S}_n|]| \\ &=\left|\sum_{i=1}^{l} a_i^{1+\delta}\{\textstyle\int |S_n|\,dP(.\,|A_i)-\int |S_n|\,dP\}\,P(A_i)\right| \\ &\leqq \sigma_n^{1+\delta}\left\{\sum_{i=1}^{l} |\textstyle\int |S_n|\,dP(.\,|A_i)-\int |S_n|\,dP|^{2/(1-\delta)}\,P(A_i)\right\}^{(1-\delta)/2}\leqq \sigma_n^{2+\delta}\,C_0. \end{aligned}$$

In the last inequality we assumed $\sigma_n\geqq 1$ which is true for sufficiently large n. Since $E[|S_n|^{1+\delta}]\leqq\sigma_n^{1+\delta}$ and $E[|\tilde{S}_n|]\leqq\sigma_n$ we get

$$E[|S_n|^{1+\delta}\,|\tilde{S}_n|]\leqq(C_0+1)\,\sigma_n^{2+\delta}. \tag{3.24}$$

Let C be the bound for x_0 then $\left|\sum_{k=1}^{n} x_k\right|\leqq n\,C$ and we can apply the elementary inequality $|y^2-z^2|\leqq 2\,n\,C\,|y-z|$ which holds for $0\leqq y,z\leqq n\,C$. Then similar to (3.22)

$$|\int |S_n|^2 \, dP(.\,|A) - \int |S_n|^2 \, dP|$$
$$\leqq 2nC \int_{\mathbb{R}^n \times \mathbb{R}^n} \left| \left| \sum_{k=1}^{n} u_k \right| - \left| \sum_{k=1}^{n} v_k \right| \right| d\lambda(u,v) \leqq 2nCC_0.$$

We introduce this in the following estimate where (3.23) and Hölder's inequality are used

$$|E[|S_n|^\delta |\tilde{S}_n|^2] - E[|S_n|^\delta] \, E[|\tilde{S}_n|^2]|$$
$$\leqq \sum_{i=1}^{l} a_i^\delta \, |\int |S_n|^2 \, dP(.\,|A_i) - \int |S_n|^2 \, dP| \, P(A_i) \leqq \sigma_n^\delta \, 2nCC_0.$$

By (3.20) the last term is bounded by $\sigma_n^{2+\delta} 4CC_0 \sigma^{-2}$ for large n. Since $E[|S_n|^\delta] \leqq \sigma_n^\delta$ and $E[|\tilde{S}_n|^2] \leqq \sigma_n^2$ we deduce

$$E[|S_n|^\delta |\tilde{S}_n|^2] \leqq (1 + 4CC_0 \sigma^{-2}) \, \sigma_n^{2+\delta}. \tag{3.25}$$

The following inequality $||y|^{1+\delta} - |z|^{1+\delta}| \leqq (1+\delta)(nC)^\delta ||x| - |y||$ which holds for $0 \leqq y,\ z \leqq nC$ implies together with (3.22)

$$|\int |S_n|^{1+\delta} \, dP(.\,|A) - \int |S_n|^{1+\delta} \, dP| \leqq (1+\delta)(nC)^\delta \, C_0.$$

Therefore by Hölder's inequality

$$|E[|S_n| \, |\tilde{S}_n|^{1+\delta}] - E[|S_n|] \, E[|\tilde{S}_n|^{1+\delta}]|$$
$$\leqq \sum_{i=1}^{l} |a_i| \, |\int |S_n|^{1+\delta} \, dP(.\,|A_i) - \int |S_n|^{1+\delta} \, dP| \, P(A_i)$$
$$\leqq \sigma_n (1+\delta)(nC)^\delta \, C_0 \leqq \sigma_n^{1+2\delta} (1+\delta) \, C^\delta \, C_0 \, 2^\delta \sigma^{-2\delta}.$$

In the last line (3.20) is again used. Our assumptions $\lim_{n\to\infty} \sigma_n^2 = \infty$ and $\delta < 1$ imply $\lim_{n\to\infty} \sigma_n^{1-\delta} = \infty$. Therefore the last term above is less than $\sigma_n^{2+\delta}$ for n large enough. We conclude

$$E[|S_n| \, |\tilde{S}_n|^{1+\delta}] \leqq 2\sigma_n^{2+\delta}. \tag{3.26}$$

Now we consider the remaining summand in (3.21).

$$E[S_n^2 \, |\tilde{S}_n|^\delta] \leqq \sigma_n^{2(1-\delta)} (E[|S_n|^2 \, |\tilde{S}_n|])^\delta.$$

Using (3.22) we easily derive

$$|E[S_n^2 \, |\tilde{S}_n|] - E[S_n^2] \, E[|\tilde{S}_n|]| \leqq C_0 \sigma_n^2.$$

Therefore

$$E[S_n^2 \, |\tilde{S}_n|^\delta] \leqq \sigma_n^{2(1-\delta)} ((C_0 \sigma_n^2)^\delta + \sigma_n^{3\delta}) = C_0^\delta \sigma_n^2 + \sigma_n^{2+\delta}.$$

The last term is bounded by $2\sigma_n^{2+\delta}$ for large n. Thus

$$E[|S_n|^2 \, |\tilde{S}_n|^\delta] \leqq 2\sigma_n^{2+\delta}. \tag{3.27}$$

Introducing (3.24)–(3.27) in (3.21) we get for some constant $C'>0$

$$a_{2n} \leqq 2\,a_n + C'\,\sigma_n^{2+\delta}.$$

Now define $b_n = a_n/\sigma_n^{2+\delta}$ then since $\lim_{n\to\infty} \sigma_{2n}^2/\sigma_n^2 = 2$ the last inequality implies that there exist constants λ, $0<\lambda<1$ and $C''>0$ such that for all n

$$b_{2n} \leqq \lambda\, b_n + C''.$$

Hence

$$\sup_r b_{2^r} \leqq b_1 + C''(1-\lambda)^{-1} < \infty. \tag{3.28}$$

The deduction of $a_n \leqq a\,\sigma_n^{2+\delta}$ for some constant $a>0$ from (3.28) is routine (see Doob [8] p. 227 or Ibragimov-Linnik [13] p. 343). Proposition 3.5 has to be used here. This proves the lemma. □

Let us mention that essentially by using (3.5) it is easy to prove the following inequality for higher moments

Remark. Let $X=(x_k)_{k\in Z}$ be a stationary real-valued process which is strictly very weak Bernoulli at rate $(\varepsilon(n))_{n\geqq 1}$ given by (1.12). Suppose x_0 is centered and bounded then for any integer $l\geqq 2$

$$E\left[\left|\sum_{k=1}^{n} x_k\right|^l\right] \ll n^{l-1}.$$

4. Proof of Theorem 2

The existence of σ^2 was shown in Proposition 3.5. Define for every $n\geqq 1$, $N = N(n) = [n^{13/16}]$, $l=l(n)=[n^{3/16}]$ and

$$X_j = \sum_{\nu=(j-1)N+1}^{jN} x_\nu \quad (1\leqq j\leqq l).$$

Furthermore we consider the following sub-σ-algebras

$$\mathfrak{F}'_j = \bigvee_{\nu=(j-1)N+1}^{jN} \mathfrak{F}_\nu \quad (1\leqq j\leqq l)$$

and as in Theorem 1

$$\mathfrak{A}'_k = \bigvee_{j\leqq k} \mathfrak{F}'_j = \bigvee_{\nu=1}^{kN} \mathfrak{F}_\nu \quad (1\leqq k\leqq l).$$

Let k and $A\in\mathfrak{A}'_{k-1}$, $P(A)>0$ be given. Then according to Definition 1

$$\inf_{\lambda\in\mathfrak{P}_N} \int_{\mathbb{R}^N\times\mathbb{R}^N} \sigma_N\, d\lambda \leqq \varepsilon(N)$$

where $\mathfrak{P}_N$ is the class of Borel probability measures on $\mathbb{R}^N \times \mathbb{R}^N$ with marginals $\mathfrak{L}((x_{(k-1)N+1}, \ldots, x_{kN}))$ and $\mathfrak{L}((x_{(k-1)N+1}, \ldots, x_{kN})|A)$. Let $\lambda \in \mathfrak{P}_N$ be a measure such that $\int \sigma_N \, d\lambda \leqq \varepsilon(N)$ and define $T\colon \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \times \mathbb{R}$ by $T((u_1, \ldots, u_N), (v_1, \ldots, v_N)) = \left(n^{-1/2} \sum_{i=1}^{N} u_i,\ n^{-1/2} \sum_{i=1}^{N} v_i\right)$. Then $T(\lambda)$ has marginals $\mathfrak{L}(n^{-1/2} X_k)$ and $\mathfrak{L}(n^{-1/2} X_k|A)$ and

$$\begin{aligned} &\int_{\mathbb{R}\times\mathbb{R}} |u-v| \, dT(\lambda)(u,v) \\ &= \int_{\mathbb{R}^N \times \mathbb{R}^N} n^{-1/2} \left| \sum_{i=1}^{N} (u_i - v_i) \right| d\lambda((u_1, \ldots, u_N), (v_1, \ldots, v_N)) \\ &\leqq n^{-1/2} N \, \varepsilon(N) \leqq C_0 \, n^{-1/2} \end{aligned}$$

where C_0 means the constant implied by $\ll$ in (1.12). Since constants are of no importance in these estimates we assume $C_0 = 1/2$. By Theorem 1 applied to the sequence $(n^{-1/2} X_k)_{1 \leqq k \leqq l}$ there exist independent random variables $n^{-1/2} Y_k$ such that $\mathfrak{L}(n^{-1/2} X_k) = \mathfrak{L}(n^{-1/2} Y_k)$ and

$$P[|n^{-1/2} X_k - n^{-1/2} Y_k| \geqq n^{-1/4}] \leqq n^{-1/4}.$$

Note that we need here only the special case of Theorem 1 where $G_k = \mathfrak{L}(X_k)$ and $\rho_k = \delta_k = \eta_k = 0$. The last inequality implies

$$\pi\left(\mathfrak{L}\left(n^{-1/2} \sum_{k=1}^{l} X_k\right), \mathfrak{L}\left(n^{-1/2} \sum_{k=1}^{l} Y_k\right)\right) \leqq l \, n^{-1/4} \leqq n^{-1/16}. \tag{4.1}$$

From the definition of X_k and Proposition 3.5 we derive

$$\begin{aligned} &E\left[\left| n^{-1/2} \sum_{k=1}^{l} X_k - n^{-1/2} \sum_{\nu=1}^{n} x_\nu \right|^2\right] \\ &= n^{-1} E\left[\left| \sum_{\nu=1}^{n-lN} x_\nu \right|^2\right] \ll n^{-1}(n - lN) \ll n^{-1}(l+N) \ll n^{-3/16} \end{aligned}$$

which immediately implies

$$\pi\left(\mathfrak{L}\left(n^{-1/2} \sum_{k=1}^{l} X_k\right), \mathfrak{L}\left(n^{-1/2} \sum_{\nu=1}^{n} x_\nu\right)\right) \ll n^{-1/16}. \tag{4.2}$$

The following estimation for the Prohorov distance of certain distributions is due to Yurinskii [22]. We restate it in slightly refined form from [4], Prop. 5.1.

Proposition. *Let $X_1, \ldots, X_n$ be independent $\mathbb{R}^d$-valued random variables with mean zero and $E[\|X_i\|^{2+\delta'}] < \infty$. If μ_n denotes the distribution of $n^{-1/2} \sum_{i=1}^{n} X_i$ and ν_n the Gaussian measure with mean zero and the same covariance as μ_n then*

$$\pi(\mu_n, \nu_n) \leqq c \, n^{-\delta'/9} d^{1/3} (\tilde{\rho}_{2+\delta'})^{2/9}$$

where $\tilde{\rho}_{2+\delta'} = n^{-1} \sum_{i=1}^{n} E[\|X_i\|^{2+\delta'}]$ and c is an absolute constant.

The bound actually given in [22] (and [4]) is more complicated. We will apply this Proposition to the random variables $(l/n)^{1/2} Y_k$ $(1 \leqq k \leqq l)$. By Lemma 3.8

$$E[|(l/n)^{1/2} Y_k|^{2+\delta'}] \ll 1$$

for any $0 \leqq \delta' < 1$. If $\tilde{\tau}_N^2$ denotes the variance of $(l/n)^{1/2} Y_1$ we can conclude using $\delta' = 3/4$

$$\pi\left(\mathfrak{L}\left(n^{-1/2} \sum_{k=1}^{l} Y_k\right), N(0, \tilde{\tau}_N^2)\right) \ll n^{-1/64}. \tag{4.3}$$

In the next two steps we use results of Dehling.

Theorem ([4], Th. 7). *Let μ and ν be two Gaussian measures on $\mathbb{R}^d$ with mean zero and covariance operators S and T. Then the following estimation holds*

$$\pi(\mu, \nu) \leqq C \,|||S - T|||^{1/3} d^{1/6} (1 + |\log |||S - T|||^{-1} d|^{1/2})$$

where C is an absolute constant and $|||\,.\,|||$ is defined by

$$|||R||| = \sup_{(e_i)} \sum_{i=1}^{d} |(R\, e_i, e_i)|$$

and the sup *is taken over all orthonormal bases (e_i) for $\mathbb{R}^d$.*

We denote by τ_N^2 the variance of $N^{-1/2} Y_1$. Proposition 3.7 implies $\tau_N^2 - \sigma^2 \ll N^{-1/2}$. So if we replace the estimation in the theorem above by $\pi(\mu, \nu) \leqq C_0 \,|||S - T|||^{1/4} d^{1/4}$ we get

$$\pi(N(0, \tau_N^2), N(0, \sigma^2)) \ll N^{-1/8} \ll n^{-1/10}. \tag{4.4}$$

The last link in our chain of estimates is a direct consequence of

Lemma ([4], Lemma 2.1). *Let X and Y be $\mathbb{R}^d$-valued square-integrable random variables with mean zero. Then we have the following estimation for the Prohorov distance of the corresponding Gaussian measures*

$$\pi(N(0, \operatorname{cov} X), N(0, \operatorname{cov} Y)) \leqq (E[\|X - Y\|^2])^{1/3}.$$

Therefore

$$\begin{aligned}\pi(N(0, \tilde{\tau}_N^2), N(0, \tau_N^2)) &\leqq (E[|(l/n)^{1/2} Y_1 - N^{-1/2} Y_1|^2])^{1/3} \\ &= (n^{-1}(n - (N n l)^{1/2}))^{2/3} \left(E\left[\left|N^{-1/2} \sum_{\nu=1}^{N} x_\nu\right|^2\right]\right)^{1/3}.\end{aligned}$$

By Proposition 3.5 and the definition of N and l we conclude

$$\pi(N(0, \tilde{\tau}_N^2), N(0, \tau_N^2)) \ll (n^{-1}(l + N))^{2/3} \ll n^{-1/8}. \tag{4.5}$$

Theorem 2 follows from (4.1)–(4.5). □

5. Proof of Theorem 3

We define $N_k=[k^{1+\kappa^{-1}}]$ $(k\geqq 1)$ where κ is the exponent from Theorem 2 and set $t_k=\sum_{i\leqq k-1} N_i$ as well as

$$X_k=N_k^{-1/2}\sum_{i=t_k+1}^{t_{k+1}} x_i. \tag{5.1}$$

Since $(x_k)_{k\in Z}$ is strictly stationary Theorem 2 implies

$$\pi(\mathfrak{L}(X_k), N(0,\sigma^2))\leqq C' N_k^{-\kappa}$$

for some constant $C'>0$ which can be assumed to be 1 for our purposes. Therefore $\mathfrak{L}(X_k)$ and $N(0,\sigma^2)$ satisfy assumption (1.3) with $\rho_k=\delta_k=N_k^{-\kappa}$. Furthermore given a set $A\in\mathfrak{A}'_{k-1}=\bigvee_{j=1}^{t_k}\mathfrak{F}_j$, $P(A)>0$ there is a Borel probability measure λ on $\mathbb{R}^{N_k}\times\mathbb{R}^{N_k}$ with marginals $\mathfrak{L}((x_{t_k+1},\dots,x_{t_{k+1}}))$ and $\mathfrak{L}((x_{t_k+1},\dots,x_{t_{k+1}})|A)$ such that

$$\int N_k^{-1}\sum_{i=1}^{N_k}|u_i-v_i|\,d\lambda(u,v)\leqq\varepsilon(N_k).$$

If T is defined on $\mathbb{R}^{N_k}\times\mathbb{R}^{N_k}$ by

$$T((u_1,\dots,u_{N_k}),(v_1,\dots,v_{N_k}))=\left(N_k^{-1/2}\sum_{i=1}^{N_k}u_i,\,N_k^{-1/2}\sum_{i=1}^{N_k}v_i\right)$$

then $T(\lambda)$ has marginals $\mathfrak{L}(X_k)$ and $\mathfrak{L}(X_k|A)$ and

$$\int_{\mathbb{R}\times\mathbb{R}}|u-v|\,dT(\lambda)(u,v)$$

$$\leqq N_k^{-1/2}\int_{\mathbb{R}^{N_k}\times\mathbb{R}^{N_k}}\sum_{i=1}^{N_k}|u_i-v_i|\,d\lambda(u,v)\leqq C_0N_k^{-1/2}.$$

Again we can assume that $C_0=1/2$. This means that we can apply Theorem 1 with $\varepsilon_k=N_k^{-1/4}$ and $\eta_k=0$. Without loss of generality there exists a sequence $(Y_k)_{k\geqq 1}$ of independent $N(0,\sigma^2)$-distributed random variables satisfying

$$P[|X_k-Y_k|\geqq 2N_k^{-\kappa}+N_k^{-1/4}]\leqq N_k^{-1/4}+N_k^{-\kappa}\ll N_k^{-\kappa}\leqq k^{-(1+\kappa)}$$

which implies by the Borel-Cantelli Lemma

$$|X_k-Y_k|\leqq 2N_k^{-\kappa}+N_k^{-1/4}\ll N_k^{-\kappa}\quad\text{a.s.} \tag{5.2}$$

For any Brownian motion $(X(t))_{t\geqq 0}$ with variance σ^2 the random variables $(t_{k+1}-t_k)^{-1/2}(X(t_{k+1})-X(t_k))$ have the same distribution as Y_k. Therefore without loss of generality we can assume that there exists a Brownian motion $(X(t))_{t\geqq 0}$ with variance σ^2 satisfying

$$(t_{k+1}-t_k)^{-1/2}(X(t_{k+1})-X(t_k))=Y_k\quad(k\geqq 1).$$

Now

$$\left|\sum_{i=t_k+1}^{t_{k+1}} x_i-(X(t_{k+1})-X(t_k))\right|=N_k^{1/2}\,|X_k-Y_k|\ll N_k^{1/2-\kappa}\quad \text{a.s.}$$

If we sum over k and make use of the relation

$$k^{2+\kappa^{-1}}\ll t_{k+1}\ll k^{2+\kappa^{-1}}$$

we get

$$\left|\sum_{\nu=1}^{t_{k+1}} x_\nu-X(t_{k+1})\right|$$

$$\ll\sum_{j=1}^{k} N_j^{1/2-\kappa}\leqq k^{(1+\kappa^{-1})(1/2-\kappa)+1}=k^{1/2-\kappa+1/2\kappa}\ll t_{k+1}^{1/2-\kappa/2}\quad \text{a.s.}$$

For any $t>0$ choose k such that $t_k<t\leqq t_{k+1}$ then

$$|\sum_{\nu\leqq t} x_\nu-X(t)|$$

$$\leqq\left|\sum_{\nu=1}^{t_k} x_\nu-X(t_k)\right|+\left|\sum_{\nu=t_k+1}^{t} x_\nu\right|+|X(t)-X(t_k)|.$$

Therefore our proof is finished if we show that there is a constant $\gamma>0$ such that with probability 1

$$\max_{t_k<t\leqq t_{k+1}}\left|\sum_{\nu=t_k+1}^{t} x_\nu\right|\ll t_k^{1/2-\gamma}\tag{5.3}$$

and

$$\sup_{t_k<t\leqq t_{k+1}}|X(t)-X(t_k)|\ll t_k^{1/2-\gamma}.\tag{5.4}$$

With the help of Lemma 3.8 the first of these two statements is shown analogous to the proof of Proposition 2.2 in [14]. (5.4) finally follows directly from Fernique's theorem [11]. □

6. Proof of Theorem 4

Let X be a very weak Bernoulli process. For a fixed integer N and any $k\in Z$ we define $X_k=(x_{(k-1)N+1},\ldots,x_{kN})$ and $\mathfrak{F}'_k=\bigvee_{\nu=(k-1)N+1}^{kN}\mathfrak{F}_\nu$. Then X_k is an S^N-valued $\mathfrak{F}'_k$-measurable random variable. Define σ_N on $S^N\times S^N$ as before. Given an $\varepsilon>0$ we choose N according to Definition 1 such that for any $k\geqq 1$ there exists a set $D_k\in\mathfrak{A}'_k=\bigvee_{j=1}^{k}\mathfrak{F}'_j$ such that $P(D_k^{\complement})<\varepsilon$ and if $A\in\mathfrak{A}'_{k-1}$, $A\subset D_{k-1}$, $P(A)>0$ then

$$\inf_{\lambda\in\mathfrak{P}_N}\int\sigma_N\,d\lambda<\varepsilon$$

where $\mathfrak{P}_n$ is the class of all Borel probability measures on $S^N \times S^N$ with marginals $\mathfrak{L}(X_k)$ and $\mathfrak{L}(X_k|A)$.

By Theorem 1 there exists a sequence $(Y_k)_{k\geqq 1}$ of independent random variables such that $\mathfrak{L}(Y_k)=\mathfrak{L}(X_k)$ and for all $k\geqq 1$

$$P[\sigma_N(X_k, Y_k)\geqq(2\varepsilon)^{1/2}]\leqq(2\varepsilon)^{1/2}+\varepsilon. \tag{6.1}$$

Let Y be the process generated by the Y_k and let C denote the bound for σ then by (1.8) for any $n\geqq 1$

$$\bar{\rho}_{nN}(X, Y)\leqq n^{-1}\sum_{k=1}^{n}\int_{S^N\times S^N}\sigma_N\, d\mathfrak{L}(X_k, Y_k).$$

Splitting this integral in the part where $\sigma_N\geqq(2\varepsilon)^{1/2}$ and its complement $\{\sigma_N<(2\varepsilon)^{1/2}\}$ we get by (6.1)

$$\bar{\rho}_{nN}(X, Y)\leqq(C+1)(2\varepsilon)^{1/2}+C\varepsilon.$$

From this the result follows. □

Acknowledgment. I am grateful to E. Berger for valuable comments.

References

1. Araujo, A., Giné, E.: The central limit theorem for real and Banach valued random variables. New York-Chichester-Brisbane-Toronto: John Wiley & Sons 1980
2. Berkes, I., Philipp, W.: Approximation theorems for independent and weakly dependent random vectors. Ann. Probability **7**, 29–54 (1979)
3. Bradley, R.C.: A sufficient condition for linear growth of variances in a stationary random sequence. Proc. Amer. Math. Soc. **83**, 586–589 (1981)
4. Dehling, H.: Grenzwertsätze für Summen schwach abhängiger vektor-wertiger Zufallsvariablen. Dissertation. Göttingen 1981
5. Dehling, H.: A note on a theorem of Berkes and Philipp. Z. Wahrscheinlichkeitstheorie verw. Gebiete **62**, 39–42 (1983)
6. Dehling, H., Denker, M., Philipp, W.: Versik processes and very weak Bernoulli processes with summable rates are independent. Preprint 1982 (4 pages)
7. Dehling, H., Philipp, W.: Almost sure invariance principles for weakly dependent vector-valued random variables. Ann. Probability **10**, 689–701 (1982)
8. Doob, J.L.: Stochastic processes. New York-London-Sydney: John Wiley & Sons 1953
9. Dudley, R.M.: Distances of probability measures and random variables. Ann. Math. Statistics **39**, 1563–1572 (1968)
10. Dudley, R.M.: Probabilities and metrics. Lecture Notes Series No. 45. Aarhus Universitet 1976
11. Fernique, X.: Integrabilité des vecteurs Gaussiens. C. R. Acad. Sci. Paris Ser. **A. 270**, 1698–1699 (1970)
12. Gray, R.M., Neuhoff, D.L., Shields, P.C.: A generalization of Ornstein's $\bar{d}$ distance with applications to information theory. Ann. Probability **3**, 315–328 (1975)
13. Ibragimov, I.A., Linnik, Yu.V.: Independent and stationary sequences of random variables. Groningen: Wolters-Noordhoff Publishing 1971
14. Kuelbs, J., Philipp, W.: Almost sure invariance principles for partial sums of mixing B-valued random variables. Ann. Probability **8**, 1003–1036 (1980)
15. Ornstein, D.: Ergodic Theory, Randomness, and Dynamical Systems. New Haven and London: Yale University Press 1974

16. Philipp, W.: Almost sure invariance principles for sums of B-valued random variables. In: Probability in Banach spaces II, Lecture Notes in Math. **709**, 171–193. Berlin-Heidelberg-New York: Springer 1979
17. Philipp, W.: Weak and L^p-invariance principles for sums of B-valued random variables. Ann. Probability **8**, 68–82 (1980)
18. Serfling, R.J.: Contributions to central limit theory for dependent variables. Ann. Math. Statistics **39**, 1158–1175 (1968)
19. Shields, P.C.: Almost block independence. Z. Wahrscheinlichkeitstheorie verw. Gebiete **49**, 119–123 (1979)
20. Strassen, V.: An almost sure invariance principle for the law of the iterated logarithm. Z. Wahrscheinlichkeitstheorie verw. Gebiete **3**, 211–226 (1964)
21. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Statistics **36**, 423–439 (1965)
22. Yurinskii, V.V.: On the error of the Gaussian approximation for convolutions. Theory Prob. Applications **22**, 236–247 (1977)

Received September 30, 1981