DEMONSTRATIO MATHEMATICA
Vol. XV No 3 1982

Włodzimierz Bryc

ON THE APPROXIMATION THEOREM OF I. BERKES AND W. PHILIPP

1. Introduction

In this paper we obtain a generalization of Berkes and Philipp s theorem [2] to the case when, instead of φ-mixing, one assumes the so called absolute regularity condition (c.f. [4]). Morever, under the condition similiar to φ-mixing our construction gives a stronger thesis which enables us to obtain moment - type inequalities (c.f. Corollary 3.2).

The first draft of this paper was unnecessarily complicated both in proof and in formulation of Theorem 3.1. Afterwards the author learned from W. Philipp and R.Bradley about two results very close to the main result of this paper. First let us mention a paper of Dehling and Philipp [3] where, using the Strassen-Dudley theorem, an approximation theorem in some sense weaker then ours is proved.

The second result is Corollary 4.2.5 on page 94 of Berbee [1]. As it was observed by R. Bradley, our approximation theorem can be obtained from the revised version of Berbee's corollary (our Theorem 3.1 was proved independently, but later then Berbee's corollary 4.2.5). Using a slightly strengthened version of Berbee's corollary 4.2.5 [1] as a lemma for discrete random variables one makes the proof of our main theorem more transparent, thus we will choose this way (suggested by R. Bradley) of presentation of the proof.

Finally, let us mention that in the [illegible] we obtained a complete solution of the approximation problem considered by Berkes and Philipp [2] (c.f. Remark 3.[illegible] below).

2. Lemmas

In the sequel let (S,d) be a Polish space (i.e. separable metric complete one) and consider a sequence of S-valued random variables $(X_n)_{n\in N}$ defined on a probability space $(\Omega,\mathfrak{M}, P)$ and measurable with respect to a Borel σ-field $\mathfrak{B}_S$ in S.

The numbers φ_n (and respectively r_n) defined below are of some interest in study of the φ-mixing (and resp. absolutely regular) sequences of r.v.'s. (c.f. [4]). Let $\varphi_1 = r_1 = 0$ and for $n > 1$ let

$$\varphi_n = \operatorname{ess\,sup} \sup_{A\in\mathfrak{B}^{n-1}} \Big|P\big((X_1,\dots,X_{n-1}) \in A\big|X_n\big) - P\big((X_1,\dots,X_{n-1}) \in A\big)\Big|, \tag{1}$$

$$r_n = E\Big\{\sup_{A\in\mathfrak{B}_S} \big|P(X_n \in A|X_1,\dots,X_{n-1}) - P(X_n \in A)\big|\Big\}. \tag{2}$$

We will also denote by P_X a measure induced on $(S, \mathfrak{B}_S)$ by the random variable X.

Lemma 2.1. For every integer n we have

$$r_n = \frac{1}{2}\operatorname{var}(P_{X_1\dots X_n} - P_{X_1\dots X_{n-1}} \otimes P_{X_n}).$$

For the proof of the lemma see [4] or [1].

Recall that $(\Omega,\mathfrak{M}, P)$ is said to be an atomless probability space iff for every $A\in\mathfrak{M}$ with $P(A) > 0$ there exist $B\in\mathfrak{M}$ such that $0 < P(B) < P(A)$. We will need the following lemma (for the proof see [2]).

Lemma 2.2. If μ is a probability measure on $(S, \mathfrak{B}_S)$ and $(\Omega,\mathfrak{M}, P)$ is an atomless probability space, then there exist a random variable $X : \Omega \to S$ such that for every $A \in \mathfrak{B}_S$

$$P(X \in A) = \mu(A).$$

The next lemma is a revised and slightly strengthened discrete version of Corollary 4.2.5 [1]. Our proof is based on a constructive approach, however a result can be obtained also by Berbee's method combined with a trick which gives (8) and (9) below (this was communicated to me by R. Bradley).

L e m m a 2.3. Let X, Y be discrete S-valued random variables defined on a probability space $(\Omega, \mathcal{M}, P)$. Let

$$\varphi = \text{ess sup} \sup_{A \in \mathcal{B}_S} |P(X \in A|Y) - P(X \in A)|,$$

$$r = E\left\{\sup_{A \in \mathcal{B}_S} |P(Y \in A|X) - P(Y \in A)|\right\}.$$

Then there exist a random variable Z defined on $\Omega \times \langle 0;1\rangle$ such that Z is a function of X, Y, t only ($t \in \langle 0;1\rangle$), $Z \simeq Y$, Z, X are independent random variables and

(3) $$P'(Y \neq Z) = r$$

(4) $P'(d(Y;Z) > \varepsilon) \leq 2\varphi P(d(Y;s) > \frac{\varepsilon}{2})$ for every s S and $\varepsilon > 0$,

where $P' = P \otimes \lambda$, λ is a Lebesque measure on $\langle 0;1\rangle$.

P r o o f . Let x (respectively y) range over all values of X (resp. Y). For brevity we will use the following notation:

$$p_x = P(X = x)$$

$$q_y = P(Y = y)$$

$$p_{x,y} = P(X = x, Y = y)$$

$$e_{x,y} = p_{x,y} - p_x q_y.$$

Note that

$$(5) \qquad \sum_x e_{x,y} = \sum_y e_{x,y} = 0 .$$

Let (x,y) be fixed and denote by $\{B_t\} = \{B_t(x,y)\}$ a disjoint partition of $\langle 0;1\rangle$ with the following properties:

$B_t \neq \emptyset$ only for a finite set of indexes t,

If $e_{x,y} < 0$ then $B_y = \langle 0;1\rangle$ and $B_t = \emptyset$ for $t \neq y$,

If $e_{x,y} \geqslant 0$ then

$$\lambda(B_y) = \frac{p_x q_y}{p_{x,y}}$$

and for $t \neq y$

$$\lambda(B_t) = -\frac{e^+_{x,y}}{p_{x,y}} \frac{e^-_{x,t}}{\sum_z e^+_{x,z}},$$

where $a^+ = \max\{a;0\}$, $a^- = \min\{a;0\}$, λ denotes a Lebesque measure on $\langle 0;1\rangle$.

Note that since $\sum_t \lambda(B_t) = 1$, the partition with required properties exists.

We define the random variable $Z(\omega, t)$ as a function of X,Y,t only by the following identity:

$Z(\omega, t) = z$ if and only if $X(\omega) = x$, $Y(\omega) = y$ and $t \in B_z(x,y)$.

Then from the properties of the partition $\{B_t\}$ it follows that

$$(6) \qquad P(X=x,\ Y=y,\ Z=z) = \begin{cases} -\dfrac{e^+_{x,y}\, e^-_{x,z}}{\sum_t e^+_{x,t}} & \text{if } z \neq y \\ p_x q_y & \text{if } z = y \text{ and } e_{x,y} \geqslant 0 \\ p_{x,z} & \text{if } z = y \text{ and } e_{x,y} < 0 . \end{cases}$$

We will prove that $Y \simeq Z$. Clearly we have

$$P(Z=z) = \sum_{x,y} P(X=x,\ Y=y,\ Z=z) = \sum_{x} p_{x,z}\,\lambda(B_z) -$$

$$- \sum_{x,y} \frac{e^{+}_{x,y}\ e^{-}_{x,z}}{\sum_{t} e^{+}_{x,t}}$$

and

$$P(Y=z) = \sum_{x,y} P(X=x,\ Y=z,\ Z=y) = \sum_{x} p_{x,z}\,\lambda(B_z) -$$

$$- \sum_{x,y} \frac{e^{+}_{x,z}\ e^{-}_{x,y}}{\sum_{t} e^{+}_{x,t}}\,.$$

Thus it suffices to show that

$$(7) \qquad \sum_{x,y} e^{+}_{x,y}\ e^{-}_{x,z} = \sum_{x,y} e^{-}_{x,y}\ e^{+}_{x,z}.$$

However because of (5) we can reduce (7) to the equality

$$- \sum_{x} e^{-}_{x,y} = \sum_{x} e^{+}_{x,z}$$

which is true once more by (5). Thus $Y \simeq Z$.

For the proof of independence of random variablex X and Z notice that

$$P(X=x,\ Z=z) = \sum_y P(X=x,\ Y=y,\ Z=z) =$$

$$= \begin{cases} p_x q_z & \text{if} \quad e_{x,z} \geq 0 \\ p_{x,z} - e_{x,z} & \text{if} \quad e_{x,z} < 0. \end{cases}$$

Thus $P(X=x,\ Z=z) = P(X=x)\ P(Y=z)$ and independence follows from $Z \cong Y$ (which was proved above).

It remains to prove (3) and (4). From (6) it follows that:

(8) $$P(X=x,\ Y=y,\ Y\neq Z) = e^+_{x,y}$$

and

(9) $$P(X=x,\ Z=z,\ Y\neq Z) = e^-_{x,z}.$$

Using (8) and Lemma 2.1 we obtain (3):

$$P(Y\neq Z) = \sum_{x,y} e^+_{x,y} = r.$$

For the proof of (4) fix $s \in S$ and $\varepsilon > 0$. Then

$$P(d(Y;Z) > \varepsilon) \leqslant P(Y\neq Z,\ d(Y;s) > \varepsilon/2) + P(Y\neq Z,\ d(Z;s) > \varepsilon/2).$$

By (8)

$$P(Y\neq Z, d(Y;s) > \varepsilon/2) = \sum_{\{y:d(y;s)>\varepsilon/2\}} \sum_x P(X=x, Y=y, Y\neq Z) =$$

$$= \sum_{\{y=d(y;s)>\varepsilon/2\}} \sum_x e^+_{x,y} \leqslant \varphi\, P(d(Y;s) > \varepsilon/2).$$

Similarly by (9)

$$P(Y\neq Z,\ d(Z;s) > \varepsilon/2) \leqslant \varphi P(d(Y;s) > \varepsilon/2).$$

This ends the proof of (4) and of the lemma.

As an immediate consequence we obtain (up to multiplicative constant) a well known Ibragimov's lemma:

C o r o l l a r y 2.4. If X, Y are real random variables such that $E|X|^p < \infty$, $E|Y|^q < \infty$ $(1/p + 1/q = 1,\ q \neq \infty)$ then

$$|EXY - EX\,EY| \leqslant 2^{\frac{1}{q}+1} \varphi^{\frac{1}{q}} \|X\|_p \|Y\|_q,$$

where φ is defined as in Lemma 2.3.

P r o o f . By a standard approximation argument it suffices to prove the inequality only for discrete random variables. Let Z be a random variable defined in Lemma 2.3. Then

$$|E(XY) - EX\,EY| = |E\,X(Y-Z)| \leqslant \|X\|_p \|Y-Z\|_q.$$

Corollary follows now from (4) and the identity:

$$E|Y-Z|^q = \int_0^\infty q\ t^{q-1}\ P(|Y-Z| > t)dt.$$

3. Theorem

The main result of this paper is the following generalization of the approximation theorem of Berkes and Philipp [2] Theorem 2.

T h e o r e m 3.1. Let $(X_n)_{n\in N}$ be a sequence of S-valued r.v.'s. Then we can redefine (X_n) onto a richer probability space $(\Omega, \mathcal{M}, P)$ on which there exist a sequence (Y_n) of independent random variables such that

(i) $X_n \simeq Y_n$ $(n=1,2,\ldots)$

(ii) $P(X_n \neq Y_n) = r_n$ $(n=1,2,\ldots)$

(iii) $P(d(X_n;Y_n) > \varepsilon) \leq 2\varphi_n P(d(X_n;s) > \varepsilon/2)$ $(n \in N,$ $s \in S,$ $\varepsilon > 0)$,

where r_n and φ_n are defined by formulas (1) and (2) respectively. Morever for every $n > 1$, Y_n and $(X_1,\ldots,X_{n-1})$ are independent r.v.'s.

P r o o f . First suppose that (X_n) is a sequence of discrete random variables. In this case we redefine (X_n) on the probability space $(\Omega, \mathfrak{M}, P)$ which supports a sequence θ_n of independent uniformly $\langle 0,1\rangle$-distributed random variables independent of (X_n).

Let $Y_1 = X_1$ and suppose that $Y_1,\ldots,Y_{n-1}$ are defined. Then we apply Lemma 2.3 to random variables

$$X := (X_1,\ldots,X_{n-1})$$

$$Y := X_n$$

and we define $Y_n := Z(X,Y,\theta_n)$.

It is easy to see that constructed in such a way random variables Y_n have all properties stated in the theorem.

In general case let $X^{(k)} = (X_n^{(k)})_{n\in N}$ be a sequence of discrete random variables such that

$$X^{(k)} \longrightarrow (X_n)_{n\in N} \text{ in distribution } (k \to \infty).$$

For instance one can define $X_n^{(k)} = a_i$ iff $i = \inf\{j: d(X_n;a_j) < k^{-1}\}$ where $(a_k)_{k\in N}$ is a dense set in S. By the discrete part of the proof one can find (after passing to a suitable probability space) a sequence $Y^{(k)}$ of independent random variables such that (a), (b) and (c) are satisfied. Let μ_k be a measure on $S^N \times S^N$ generated by the random variable $(X^{(k)},Y^{(k)})$. Then μ_k is a relatively compact family of probability measures. Indeed let $\varepsilon > 0$ and denote by K_n a compact set in S such that $P(X_n^{(k)} \in K_n) > 1-\varepsilon/2^n$. Then $K = \bigtimes_n K_n \times \bigtimes_n K_n$ is a compact set such that $\mu_k(K) > 1-\varepsilon$ $(k=1,2,\dots)$. Thus by Prochorov's theorem (μ_k) is relatively compact. Passing to a subsequence we may assume that $\lim_{k\to\infty} \mu_k = \mu$. By Lemma 2.2 on every atomless probability space $(\Omega, \mathfrak{M}, P)$ we can find a random variable $((X_n),(Y_n))$ with distribution μ. Clearly Y_n are independent random variables. It remains to prove (b) and (c), since (a) is obvious.

Let $\varepsilon > 0$, $s \in S$ and $n \in N$ be fixed and let δ be any point of continuity of the distribution of real random variables $d(X_n;Y_n) - \varepsilon$ and $2d(X_n;s) - \varepsilon$.

Then

$$P(d(X_n;Y_n) > \varepsilon + \delta) = \lim_{k\to\infty} P(d(X_n^{(k)}, Y_n^{(k)}) > \varepsilon + \delta)$$

thus from the discrete part of the proof (because δ is arbitrarily small) we obtain (b) and (c).

C o r o l l a r y 3.2. Let $f: R_+ \to R_+$ be a continuous increasing function such that $f(0) = 0$ and for some $s \in S$, $n \in N$, $E\, f(2d(X_n;s)) < \infty$. Then (X_n) can be redefined on a richer probability space on which there exists a sequence (Y_n) of independent random variables such that $Y_n \simeq X_n$ and $E\, f(d(X_n;Y_n)) \leq 2\varphi_n\, E\, f(2d(X_n;s))$.

P r o o f . This corollary follows from Theorem 3.1 and from the identity:

For every real random variable $T \geq 0$, $E\, f(T) = \int_0^\infty P(T \geq t)\, df(t)$.

R e m a r k 3.3. The result obtained in Theorem 3.1(b) is best possible. Indeed, suppose that Y_n and $(X_1,\dots,X_{n-1})$ are independent, $X_n \simeq Y_n$. Then by Lemma 2.1

$$r_n = \sup_A |P((X_1,\dots,X_n) \in A) - P((X_1,\dots,X_{n-1},Y_n) \in A)| =$$

$$= \sup_A |P((X_1,\dots,X_n) \in A,\ (X_1,\dots,X_{n-1},Y_n) \notin A) \leq P(X_n \neq Y_n).$$

Acknowledgement. As it was mentioned in the introduction, the presentation of this paper owes much to W. Philipp and R. Bradley. I would like to thank them for useful suggestions.

I would like also to thank S. Kwapień for pointing out an error in the first draft of the paper.

REFERENCES

[1] H.P.C. Berbee: Random walks with stationary increments and renewal theory. Math. Centre Tracts 112. Amsterdam 1979.

[2] I. Berkes, W. Philipp: Approximation theorems for independent and weakly dependent random vectors, Ann. Probability 7 (1979) 29-54.

[3] H. Dehling, W. Philipp: Almost sure invariance principles for weakly dependent vector-valued random variables. Ann. Probability 10 (1982) 689-701.

[4] I.A. Ibragimov, Ju.A. Rozanow: Gaussian Random Process. New York 1978.

INSTITUTE OF MATHEMATICS, TECHNICAL UNIVERSITY OF WARSAW
Received April 9, 1981.