Itô, K. and Nisio, M.
Osaka J. Math.
5 (1968), 35-48

ON THE CONVERGENCE OF SUMS OF INDEPENDENT BANACH SPACE VALUED RANDOM VARIABLES

KIYOSI ITÔ AND MAKIKO NISIO

(Received February 26, 1968)

1. Introduction

The purpose of this paper is to discuss the convergence of sums of independent random variables with values in a separable real Banach space and to apply it to some problems on the convergence of the sample paths of stochastic processes.

For the real random variables, we have a complete classical theory on the convergence of independent sums due to P. Lévy, A. Khinchin and A. Kolmogorov. It can be extended to finite dimensional random variables without any change. In case the variables are infinite dimensional, there are several points which need special consideration. The difficulties come from the fact that bounded subsets of Banach space are not always conditionally compact.

In Section 2 we will discuss some preliminary facts on Borel sets and probability measures in Banach space. In Section 3 we will extend P. Lévy's theorem. In Section 4 we will supplement P. Lévy's equivalent conditions with some other equivalent conditions, in case the random variables are symmetrically distributed. Here the infinite dimensionality will play an important role. The last section is devoted to applications.

2. Preliminary facts

Throughout this paper, E stands for a separable real Banach space and the topology in E is the norm topology, unless stated otherwise. E^* stands for the dual space of E, $\mathcal{B}$ for all Borel subsets of E and $\mathcal{P}$ for all probability measures on $(E, \mathcal{B})$.

The basic probability measure space is denoted by $(\Omega, \mathcal{F}, P)$ and the generic element of Ω by ω. An E-valued random variable X is a map of Ω into E measurable $(\mathcal{F}, \mathcal{B})$. The probability law μ_X of X is a probability measure in $(E, \mathcal{B})$ defined by

$$\mu_X(B) = P(X \in B), \qquad B \in \mathcal{B}.$$

According to Prohorov [5], every $\mu \in \mathcal{P}$ is tight, i.e.

$$\forall \varepsilon>0 \quad \exists K \text{ compact} \subset E \quad \mu(K)>1-\varepsilon .$$

A subset $\mathcal{M}$ of $\mathcal{P}$ is called *uniformly tight* if

$$\forall \varepsilon>0 \quad \exists K \text{ compact} \subset E \quad \forall \mu \in \mathcal{M} \quad \mu(K)>1-\varepsilon .$$

$\mathcal{P}$ is a complete metric space with respect to the Prohorov metric [5]. $\mathcal{M} \subset \mathcal{P}$ is *conditionally compact*, if and only if $\mathcal{M}$ is uniformly tight.

Let $\mathcal{C}$ denote the algebra of all cylinder sets;

$$\{x \in E: (\langle z_1, x\rangle, \langle z_2, x\rangle, \cdots, \langle z_n, x\rangle) \in \Gamma\}$$
$$n=1, 2, \cdots, z_i \in E^*, \Gamma \in \mathcal{B}(R^n)$$

and $\mathcal{B}[\mathcal{C}]$ the σ-algebra generated by $\mathcal{C}$.

Proposition 2.1. $\mathcal{B}[\mathcal{C}]=\mathcal{B}$.

Proof. Let $\{b_n\}$ be a countable dense subset of E and take $\{z_n\} \subset E^*$ such that

$$\|z_n\|=1, \qquad \langle z_n, b_n\rangle=\|b_n\|, \qquad n=1, 2, \cdots .$$

Such z_n exists by Hahn-Banach's extension theorem for each n. Now we shall prove that

$$\{x: \|x\| \leq r\}=\bigcap_n \{x: \langle z_n, x\rangle \leq r\} .$$

Write B_1 and B_2 for the sets on both sides. $B_1 \subset B_2$ is obvious. To prove $B_1^c \subset B_2^c$, take an arbitrary point b in B_1^c. Then we have $\|b\|>r$. As $\{b_n\}$ is dense, we can find b_n such that

$$\|b-b_n\|<\tfrac{1}{2}(\|b\|-r) .$$

Then we get

$$\|b_n\| \geq \|b\|-\|b-b_n\|>\tfrac{1}{2}(\|b\|+r) ,$$
$$|\langle z_n, b\rangle-\|b_n\||=|\langle z_n, b\rangle-\langle z_n, b_n\rangle| \leq \|z_n\| \, \|b-b_n\|<\tfrac{1}{2}(\|b\|-r)$$

and so

$$\langle z_n, b\rangle>\|b_n\|-\tfrac{1}{2}(\|b\|-r)>\tfrac{1}{2}(\|b\|+r)-\tfrac{1}{2}(\|b\|-r)=r .$$

This shows $b \in B_2^c$. Thus $B_1=B_2$ is proved. Since $B_2 \in \mathcal{B}[\mathcal{C}]$, $B_1 \in \mathcal{B}[\mathcal{C}]$. Since $\mathcal{B}[\mathcal{C}]$ is translation invariant, $\{x: \|x-a\|<r\}$ also belongs to $\mathcal{B}[\mathcal{C}]$. Therefore $\mathcal{B} \subset \mathcal{B}[\mathcal{C}]$. Since $\mathcal{B}[\mathcal{C}] \subset \mathcal{B}$ is obvious, we have $\mathcal{B}=\mathcal{B}[\mathcal{C}]$.

The *characteristic functional* of $\mu \in \mathcal{P}$ is defined by

$$C(z: \mu)=\int_E e^{i\langle z, x\rangle} \mu(dx), \qquad z \in E^* .$$

It is clear that

$$C(z: \mu_X)=E[e^{i\langle z, X\rangle}] .$$

$\mu_n \to \mu$ (Prohorov metric) implies $C(z: \mu_n) \to C(z: \mu)$ for every $z \in E^*$. $C(z: \mu)$ is continuous in the norm topology in E^*.

Proposition 2.2.

$$C(z: \mu) = C(z: \nu) \Rightarrow \mu = \nu \,.$$

Proof. Setting $z = \sum_{j=1}^{n} t_j z_j$ and using the one-to-one correspondence between the probability measures and the characteristic functions in R^n, we can easily see that $\mu = \nu$ on $\mathcal{C}$. Since $\mathcal{C}$ is an algebra which generates $\mathcal{B}$ by Proposition 2.1, we have $\mu = \nu$ on $\mathcal{B}$.

Proposition 2.3. *If we have $r > 0$ such that*

$$C(z: \mu) = 1 \qquad \textit{for} \qquad \|z\| < r \,,$$

then μ is concentrated at 0, i.e. $\mu = \delta$.

Proof. Let $\varphi(t) = C(tz: \mu)$, t real, $z \neq 0$. Then $\varphi(t)$ is a characteristic function in R^1 and

$$\varphi(t) = 1 \qquad \text{for} \qquad |t| < \frac{r}{\|z\|} \,.$$

Using the inequality

$$|\varphi(t) - \varphi(s)| \leq \sqrt{2|1 - \varphi(t-s)|}$$

we can get $\varphi(t) = 1$ for every t. Setting $t = 1$ we have

$$C(z: \mu) = 1 = C(z: \delta) \qquad \text{for} \qquad z \neq 0 \,.$$

This is obvious for $z = 0$. Hence $\mu = \delta$ follows by Proposition 2.2.

3. Sums of independent random variables

Let $X_n(\omega)$, $n = 1, 2, \cdots$ be a sequence of independent E-valued random variables and set

$$S_n = \sum_{1}^{n} X_i \,, \qquad \mu_n = \text{the probability law of } S_n \,.$$

Then we have

Theorem 3.1.[(1)] (*Extension of P. Lévy's theorem*). *The following conditions are equivalent.*

(a) *S_n converges a.s.* (*=almost surely*),

(b) *S_n converges in probability,*

(c) *μ_n converges* (*Prohorov metric*).

---

(1) In the course of printing the authors noticed that a more general fact was proved by A. Tortrat [7].

Proof. (a)⇒(b)⇒(c) is obvious. We can prove (b)⇒(a) in the same way as in the real case, using the inequality that can be verified also as in the real case:

$$P(\max_{m<k\leq n}\|X_{m+1}+X_{m+2}+\cdots+X_k\|>2c)$$
$$\leq\frac{P(\|X_{m+1}+\cdots+X_n\|>c)}{1-\max\limits_{m<k\leq n}P(\|X_{k+1}+\cdots+X_n\|\geq c)}$$

To prove (c)⇒(b), let us denote the probability law of $S_n-S_m=\sum_{m+1}^{n}X_i$ by μ_{mn}, $m<n$. As μ_n tends to a probability measure μ on E by the assumption (c), $\{\mu_n\}$ is conditionally compact and so uniformly tight, i.e.

$$\forall\varepsilon>0\quad\exists K\text{ compact}\quad\forall n\quad\mu_n(K)>1-\varepsilon\,.$$

Let K_1 denote the set $\{x-y\colon x, y\in K\}$. K_1 is also compact by the continuity of the map $(x, y)\rightarrow x-y$. As S_n, $S_m\in K$ implies $S_n-S_m\in K_1$, we have

$$\begin{aligned}\mu_{mn}(K_1)&\geq P(S_n\in K,\ S_m\in K)\\&\geq 1-P(S_n\in K^c)-P(S_m\in K^c)\\&=1-\mu_n(K^c)-\mu_m(K^c)\\&>1-2\varepsilon\,.\end{aligned}$$

This shows that $\{\mu_{mn}\colon m<n\}$ is also conditionally compact. We shall now prove (b), i.e.

$$(1)\qquad\forall\varepsilon>0\quad\exists N\quad\forall m<n<N\quad\mu_{mn}(U_\varepsilon)>1-\varepsilon\,,$$

where U_ε denotes the ε-neighbourhood of the origin 0 in E. Suppose to the contrary that

$$(2)\qquad\exists\varepsilon>0\quad\forall N\quad\exists n(N)>m(N)>N\quad\mu_{m(N)n(N)}(U_\varepsilon)\leq 1-\varepsilon\,.$$

As $\{\mu_{mn}\}$ is conditionally compact, we can assume that $\mu_{m(N)n(N)}$ converges to a probability measure ν on $(E, \mathcal{B})$, then

$$(3)\qquad\nu(U_\varepsilon)\leq\varliminf_{N\rightarrow\infty}\mu_{m(N)n(N)}(U_\varepsilon)\leq 1-\varepsilon\,.$$

On the other hand we have

$$\begin{aligned}C(z\colon\mu_{n(N)})&=E[e^{i<z,S_{n(N)}>}]\\&=E[e^{i<z,S_{m(N)}>}]\,E[e^{i<z,S_{n(N)}-S_{m(N)}>}]\\&=C(z\colon\mu_{m(N)})\,C(z\colon\mu_{m(N)n(N)})\end{aligned}$$

by the independence of X_n, $n=1, 2, \cdots$. Letting $N\rightarrow\infty$, we have

$$C(z\colon\mu)=C(z\colon\mu)\,C(z\colon\nu)\,.$$

Since $C(0:\mu)=1$, we have $r>0$ such that

$$C(z:\mu)\neq 0 \qquad \text{for} \qquad \|z\|\leq r\,.$$

Then

$$C(z:\nu)=1 \qquad \text{for} \qquad \|z\|\leq r\,,$$

so that $\nu=\delta$ by Proposition 2.3. This contradicts (3).

Theorem 3.2. *The uniform tightness of* $\{\mu_n\}$ *implies that we have a sequence* $c_n\in E$, $n=1,2,\cdots$ *such that* S_n-c_n *converges a.s.*

Proof. Let $(Y_n,\ n=1,2,\cdots)$ be a copy of $(X_n,\ n=1,2,\cdots)$ independent of this random sequence. Then $X_1, X_2, \cdots, Y_1, Y_2, \cdots$ are independent. Now set

$$T_n=\sum_1^n Y_i\,, \qquad U_n=\sum_1^n (X_i-Y_i)=S_n-T_n$$

$$\nu_n = \text{the probability law of } U_n\,.$$

Then $X_1-Y_1, X_2-Y_2, \cdots$ are independent.

By our assumption we have

$$\forall\varepsilon>0 \quad \exists K=K(\varepsilon) \ \text{compact} \quad \forall n \quad \mu_n(K)>1-\varepsilon\,.$$

Write $K_1=K_1(\varepsilon)$ for the set $\{x-y:\ x,\ y\in K\}$. Then K_1 is also compact and we have

$$\begin{aligned}\nu_n(K_1)&=P(S_n-T_n\in K_1)\\ &\geq P(S_n\in K,\ T_n\in K)\\ &\geq 1-P(S_n\in K^c)-P(T_n\in K^c)\\ &=1-2\mu_n(K^c)>1-2\varepsilon\,.\end{aligned}$$

Therefore $\{\nu_n\}$ is also uniformly tight and so conditionally compact.

Since $X_1, X_2, \cdots Y_1, Y_2 \cdots$ are independent and X_n and Y_n have the same distribution, we have

$$\begin{aligned}C(z:\nu_n)&=E[e^{i<z,U_n>}]\\ &=\prod_{j=1}^n E[e^{i<z,X_j>}]E[e^{-i<z,Y_j>}]\\ &=\prod_{j=1}^n |E(e^{i<z,X_j>})|^2\,.\end{aligned}$$

Since $0\leq|E[e^{i<z,X_n>}]|^2\leq 1$, $\lim_{n\to\infty} C(z:\nu_n)$ exists for every $z\in E^*$.

Now we shall prove that $\{\nu_n\}$ is convergent. Since it is conditionally compact, it is enough to prove that two arbitrary convergent subsequences $\{\nu_n'\}$, $\{\nu_n''\}$ of $\{\nu_n\}$ have the same limit. Since $\lim_{n\to\infty} C(z:\nu_n)$ exists, we have, for

$\nu'=\lim \nu'_n$ and $\nu''=\lim \nu''_n$,

$$C(z:\nu') = \lim_n C(z:\nu'_n) = \lim_n C(z:\nu''_n) = C(z:\nu''),$$

and so $\nu'=\nu''$ by Proposition 2.3.

By Theorem 3.1 the convergence of ν_n implies the a.s. convergence of $U_n \equiv S_n - T_n$. Since $\{S_n\}$ and $\{T_n\}$ are independent, we can use Fubini's theorem to see that for almost every sample sequence $(c_1, c_2, \cdots)$ of $(T_1, T_2, \cdots)$, $S_n - c_n$ converges a.s. This completes the proof.

4. Sum of independent random variables with symmetric distributions

Let $\{X_n\}$ be independent E-valued random variables, S_n denote $\sum_1^n X_i$, $n=1, 2, \cdots$ and μ_n the probability law of S_n, $n=1, 2, \cdots$ as before. In this section we shall impose an additional condition:

(SD) Each X_n is symmetrically distributed.

Theorem 4.1. *The conditions* (*a*), (*b*) *and* (*c*) *in Theorem* 3.1 *and the following conditions are all equivalent.*

(d) *$\{\mu_n\}$ is uniformly tight.*

(e) *There exists an E-valued random variable S such that $\langle z, S_n\rangle \to \langle z, S\rangle$ in probability for every $z \in E^*$.*

(f) *There exists a probability measure μ on E such that*

$$E[e^{i\langle z, S_n\rangle}] \to C(z:\mu)$$

for every $z \in E^*$.

Remark. In the finite dimensional case, (SD) is not necessary for the proof of the equivalence of all conditions except (d). But (f) does not always imply (c) in the infinite dimensional case without (SD). For example, let E be a Hilbert space and $\{e_n\}$ be an orthonormal base. Now set

$$X_1(\omega) \equiv e_1, \qquad X_n(\omega) \equiv e_n - e_{n-1}, \qquad n = 1, 2, \cdots.$$

Then $S_n(\omega) \equiv e_n$ and

$$\langle z, S_n\rangle = \langle z, e_n\rangle \to 0 = \langle z, S\rangle, \qquad S \equiv 0.$$

But S_n does not converge to S.

Proof. (a)$\Leftrightarrow$(b)$\Leftrightarrow$(c)$\Rightarrow$(d) and (a)$\Rightarrow$(e)$\Rightarrow$(f) are both obvious. Therefore it remains only to prove (f)$\Rightarrow$(e)$\Rightarrow$(d)$\Rightarrow$(a).

Suppose that (d) holds. By Theorem 3.2 we have $\{c_n\}$ such that $S_n - c_n$

converges a.s. Since each X_n is symmetrically distributed and $X_1, X_2, \cdots$ are independent, the random sequence $(X_1, X_2, \cdots)$ has the same probability law as $(-X_1, -X_2, \cdots)$. Since $S_n - c_n \equiv \sum_1^n X_i - c_n$ converges a.s., $-S_n - c_n = \sum_1^n (-X_i) - c_n$ also converges a.s. and so does $S_n = [(S_n - c_n) - (-S_n - c_n)]/2$. Thus (d)$\Rightarrow$(a) is proved.

Suppose that (e) holds. Take $z_1, z_2, \cdots z_p \in E^*$ arbitrarily and fix them. Then the sequence of random vectors

$$\sigma_m(\omega) \equiv (\langle z_1, S_m(\omega)\rangle, \cdots, \langle z_p, S_m(\omega)\rangle), \qquad m = 1, 2, \cdots.$$

converges in probability to the random vector

$$\sigma(\omega) \equiv (\langle z_1, S(\omega)\rangle, \cdots, \langle z_p, S(\omega)\rangle).$$

Since σ_n and $\sigma_m - \sigma_n$, $(m > n)$, prove to be independent by the assumption, σ_n and $\sigma - \sigma_n$ are also independent, i.e.

$$P(\sigma_n \in \Gamma_1, \sigma - \sigma_n \in \Gamma_2) = P(\sigma_n \in \Gamma_1) P(\sigma - \sigma_n \in \Gamma_2), \quad \text{for} \quad \Gamma_1, \Gamma_2 \in \mathscr{B}[R^p].$$

Writing this in terms of S_n and $S - S_n$, we have

$$P(S_n \in C_1, S - S_n \in C_2) = P(S_n \in C_1) P(S - S_n \in C_2)$$

for $C_1, C_2 \in \mathcal{C}$. Since $\mathcal{C}$ is an algebra which generates $\mathscr{B}$ by Proposition 2.1, we have

$$P(S_n \in B_1, S - S_n \in B_2) = P(S_n \in B_1) P(S - S_n \in B_2)$$

for $B_1, B_2 \in \mathscr{B}$, namely S_n and $S - S_n$ are independent. Thus we have

$$P(S \in K) = \int_E P(S_n + x \in K) P(S - S_n \in dx)$$

and so we can find $x_0 = x_0(K, n)$ such that

$$P(S_n + x_0 \in K) \geq P(S \in K).$$

Since S_n proves to be symmetrically distributed by our assumption, we have

$$P(-S_n + x_0 \in K) = P(S_n + x_0 \in K) \geq P(S \in K).$$

Writing K_1 for the set $\{(x-y)/2 \colon x, y \in K\}$, we get

$$\begin{aligned} \mu_n(K_1) = P(S_n \in K_1) &\geq P(S_n + x_0 \in K, -S_n + x_0 \in K) \\ &\geq 1 - P(S_n + x_0 \in K^c) - P(-S_n + x_0 \in K^c) \\ &\geq 1 - 2P(S \in K^c). \end{aligned}$$

By taking a compact set $K = K(\varepsilon)$ for $\varepsilon > 0$, we can make the right hand side greater than $1-\varepsilon$. Since K_1 is compact with K, $\{\mu_n\}$ proves to be uniformly tight. This proves (e)⇒(d).

We shall now prove (f)⇒(e).

Take $z\in E^*$ and fix it for the moment. $\langle z, X_n\rangle$, $n=1, 2, \cdots$ are independent real random variables and

$$\langle z, S_n\rangle = \sum_1^n \langle z, X_i\rangle .$$

By our assumption, we have

$$E[e^{it\langle z, S_n\rangle}] = E[e^{i\langle tz, S_n\rangle}] \to C(tz:\ \mu)$$

for every real t. Since the right hand side is a characteristic function of t for the probability measure on R^1 induced from μ by the map $x\to\langle z, x\rangle$, the probability law of $\langle z, S_n\rangle$ converges to this measure. Therefore $\langle z, S_n\rangle$ converges a.s. to a real random variable, say Y_z, by Lévy's theorem. Notice that the exceptional ω-set depends on z. Since a countable sum of null sets also is a null set, we have a P-null ω-set $N=N(z^1, z^2, \cdots)$ such that

$$\langle z^{(k)}, S_n\rangle \to Y_{z^{(k)}} \qquad k = 1, 2, \cdots$$

for every $\omega\in N^c$.

Now we shall compare two systems of real random variables:

$$Y_z(\omega)\,, \qquad \omega\in(\Omega, \mathcal{F}, \mathcal{P})\,, \qquad z\in E^*$$

and

$$\langle z, x\rangle, \qquad x\in(E, \mathcal{B}, \mu)\,, \qquad z\in E^*.$$

They have the same finite jont distributions. To prove this, take $z^{(1)}, z^{(2)}, \cdots, z^{(p)}\in E^*$. Then

$$\begin{aligned} E\left[e^{i\sum_j t_j Y_{z^{(j)}}}\right] &= \lim_{n\to\infty} E\left[e^{i\sum_j t_j\langle z^{(j)}, S_n\rangle}\right] = \lim_{n\to\infty} E\left[e^{i\langle\sum_j t_j z^{(j)}, S_n\rangle}\right] \\ &= C(\sum_j t_j z^{(j)}:\ \mu) = E_\mu\left[e^{i\langle\sum_j t_j z^{(j)}, x\rangle}\right] \\ &= E_\mu\left[e^{i\sum_j t_j\langle z^{(j)}, x\rangle}\right] \end{aligned}$$

by our assumption (f), where E_μ is the expectation sign based on the measure μ.

Let R^∞ be a countable product space $R^1\times R^1\times\cdots$ and $\mathcal{B}(R^\infty)$ the σ-algebra generated by all cylindrical Borel sets in R^∞. $\mathcal{B}(R^\infty)$ is also the σ-algebra of all Borel subsets of R^∞ with respect to the product topology.

Let $z', z'', \cdots$ be any sequence in E^*. Then we have

$$(1) \qquad P[(Y_{z'}, Y_{z''}, \cdots)\in B] = \mu[(\langle z', x\rangle, \langle z'', x\rangle\cdots)\in B]$$

for $B\in\mathcal{B}(R^\infty)$; in fact, if B is a cylindrical Borel set, this identity holds because of the same joint distributions mentioned above and so it proves to hold for every $B\in\mathcal{B}(R^\infty)$ by usual argument.

Now we shall prove the existence of an E-valued random variable $S(\omega)$ with $Y_z=\langle z, S\rangle$ a.s. for each $z\in E^*$, which will complete the proof of (e). To find out such $S(\omega)$, we shall use the facts mentioned above.

Since μ is tight, we can find an increasing sequence of compact sets K_1, K_2, $\cdots\subset E$ such that $\mu(K_n)\to 1$, so that $\mu(K_\infty)=1$ for $K_\infty\equiv\bigcup_n K_n$. K_∞ is clearly a Borel subset of E. Take $z_1, z_2, \cdots\in E^*$ in Proposition 2.1. The map $\theta\colon E\to R^\infty$ denfined by

$$\theta x = (\langle z_1, x\rangle, \langle z_2, x\rangle\cdots)$$

is continuous and one-to-one from E onto θE; in fact, if $\theta x=(0, 0, \cdots)$, then

$$x\in \bigcap_n \{x\colon \langle z_n, x\rangle\leq r\} = \{x\colon ||x||\leq r\}$$

for every $r>0$ (see the proof of Proposition 2.1) and so $x=0$. Therefore θK_n is compact and the restriction $\theta|K_n$ has a continuous inverse map. Since $\theta K_\infty=\bigcup_n \theta K_n$ is a Borel subset of R^∞, the restriction of θ to K_∞ has an inverse which can be extended to a map $\varphi\colon R^\infty\to E$ measurable $(\mathcal{B}(R^\infty), \mathcal{B})$. It is clear that

$$x = \varphi(\theta x) = \varphi(\langle z_1, x\rangle, \langle z_2, x\rangle, \cdots) \quad \text{on} \quad K_\infty . \tag{2}$$

Since

$$\mu((\langle z_1, x\rangle, \langle z_2, x\rangle\cdots)\in\theta K_\infty)\geq\mu(K_\infty) = 1 ,$$

we have

$$P((Y_{z_1}, Y_{z_2}, \cdots)\in\theta K_\infty) = 1 .$$

Now we set

$$S(\omega) = \varphi(Y_{z_1}(\omega), Y_{z_2}(\omega), \cdots) .$$

Take an arbitrary $z\in E^*$. Then

$$\begin{aligned} \mu(\langle z, x\rangle &= \langle z, \varphi(\langle z_1, x\rangle, \langle z_2, x\rangle\cdots)\rangle) \\ &\geq\mu(x = \varphi(\langle z_1, x\rangle, \langle z_2, x\rangle, \cdots)) \\ &\geq\mu(x = \varphi(\langle z_1, x\rangle, \langle z_2, x)\cdots\rangle, x\in K_\infty) \\ &= \mu(K_\infty) = 1 \qquad \text{by} \quad (2) . \end{aligned} \tag{3}$$

Since φ is measurable $(\mathcal{B}(R^\infty),\mathcal{B})$ and $x\to\langle z, x\rangle$ is continuous; the condition on $(\xi_0, \xi_1, \xi_2, \cdots)$: $\xi_0=\langle z, \varphi(\xi_1, \xi_2, \cdots)\rangle$ is given by a Borel subset of R^∞. By (1) and (3) we have

$$P(Y_z = \langle z, \varphi(Y_{z_1}, Y_{z_2}, \cdots)\rangle) = 1$$

and so

$$P(Y_z = \langle z, S \rangle) = 1 ,$$

which completes our proof.

5. Examples.

a. Generalized Wiener expansion of Brownian motion

Let $B(t)$, $0 \leq t \leq 1$, be a Brownian motion with $B(0) \equiv 0$. Then we have an isomorphism between the real Hilbert Space $L^2[0, 1]$ and the real Hilbert Space $\boldsymbol{M}(B)$ spanned by $B(t)$, $0 \leq t \leq 1$ in $L^2(\Omega, \mathcal{F}, P)$ by

$$\varphi \to \int_0^1 \varphi(u) dB(u) .$$

The indicator e_{0t} of the interval $(0, t)$ corresponds to $B(t)$. Let $\{\varphi_n\}_n$ be an orthonormal base in $L^2[0, 1]$. Then the $\{\xi_n\}_n$ that correspond to $\{\varphi_n\}_n$ form an orthogonal base in $\boldsymbol{M}(B)$. $\{\xi_n\}_n$ are independent since $B(t)$ is a Gaussian process. As e_{0t} has an orthogonal expansion

$$e_{0t} = \sum_n b_n(t) \varphi_n ,$$

where

$$b_n(t) = \int_0^1 e_{0t}(u) \varphi_n(u) du = \int_0^t \varphi_n(u) du ,$$

we have an orthogonal expansion

$$B(t) = \sum_n b_n(t) \xi_n(\omega) = \sum_n \xi_n(\omega) \int_0^t \varphi_n(u) du . \tag{1}$$

For each t, this series converges in the mean square and so converges a.s. by Lévy's theorem. We shall make use of Theorem 4.1 ((e)$\Rightarrow$(a)) to prove

Theorem 5.1. *The right hand side of* (1) *converges uniformly in t to $B(t)$ a.s.*

Proof. Let us introduce a sequence of stochastic processes

$$X_n(t, \omega) = \xi_n(\omega) \int_0^t \varphi_n(u) du , \qquad n = 1, 2, \cdots$$

and write the sample paths of $X_n(t, \omega)$, $n=1, 2, \cdots$ and $B(t, \omega)$ as $\boldsymbol{X}_n(\omega)$, $n=1, 2, \cdots$ and $\boldsymbol{B}(\omega)$ respectively. Then these are symmetrically distributed random variables with values in the Banach space $E \equiv C[0, 1]$ of continuous functions on $[0, 1]$. $\boldsymbol{X}_n$, $n=1, 2, \cdots$ are independent. Set

$$\boldsymbol{S}_n(\omega) = \sum_1^n \boldsymbol{X}_i(\omega) , \qquad n = 1, 2, \cdots .$$

$\boldsymbol{S}_n(\omega)$ is the sample path of the process $S_n(t, \omega) = \sum_1^n X_i(t, \omega)$. Our theorem claims that $\boldsymbol{S}_n \to \boldsymbol{B}$ a.s. To prove this it is enough by Theorem to prove that for every $z \in E^*$ i.e. for every signed measure $z(dt)$ on $[0, 1]$, $\langle z, S_n \rangle$ converges in probability to $\langle z, \boldsymbol{B} \rangle$. It is now enough to observe

$$\begin{aligned} &E[|\langle z, \boldsymbol{S}_n\rangle - \langle z, \boldsymbol{B}\rangle|] \\ &\quad = E\left[\left|\int_0^1 z(dt)(S_n(t)-B(t))\right|\right], \qquad |z| = \text{total variation of } z , \\ &\quad \leq \int_0^1 |z|(dt) E[|S_n(t)-B(t)|] \to 0 , \end{aligned}$$

because

$$(E[|S_n(t)-B(t)|])^2 \leq E[|S_n(t)-B(t)|^2] = \sum_{i=n+1}^{\infty} b_i(t)^2$$

and so

$$E[|S_n(t)-B(t)|] \begin{cases} \leq \left(\int_0^1 e_{0t}(u)^2 du\right)^{1/2} \leq \sqrt{t} & \text{for } \forall t \\ \to 0 & \text{as } n \to \infty . \end{cases}$$

b. Definition of Brownian motion

Theorem 5.1 suggests that we can define Brownian motion as follows. Let $\xi_n(\omega)$, $n=1, 2, \cdots$ be an independent sequence of real random variables with the distribution $N(0, 1)$, (whose existence is guaranteed by Kolmogorov's extension theorem) and an orthonormal base φ_n, $n=1, 2, \cdots$ in $L^2[0, 1]$.

Theorem 5.2.

$$(2) \qquad \sum_n \xi_n(\omega) \int_0^t \varphi_n(u) du , \qquad 0 \leq t \leq 1$$

converges uniformly in t a.s. The limit process $S(t)$ is a stochastic process with independent increaments and continuous paths such that $S(t)S(s)$ is $N(0, t-s)$-distributed for $t>s$, i.e. $S(t)$ is a Brownian motion.

REMARK. N. Wiener [6] defined Brownian motion in this way by taking $\varphi_n(u) = \sqrt{2} \sin n\pi t$, $n=1, 2, \cdots$. He proved the a.s. uniform convergence of the grouped sums

$$\sum_{n=0}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \xi_k(\omega) \int_0^t \sqrt{2} \sin k\pi u du ,$$

which was sufficient for his purpose. G. Hunt [4] proved a theorem that ensures the a.s. uniform convergence of the sum

$$\sum_{n=1}^{\infty} \xi_n(\omega) \int_0^t \sqrt{2} \sin n\pi u du .$$

We claim that this holds for a general orthonormal base $\{\varphi_n\}$, (Delporte. [2], Walsh [8]).

Proof. The only difficult part of our theorem is the a.s. uniform convergence. We shall use $\boldsymbol{X}_n(\omega)$, $S_n(t, \omega)$ and $\boldsymbol{S}_n(\omega)$ as before. By Theorem 3.2 it is enough to prove that the probability laws of $\boldsymbol{S}_n$, $n=1, 2, \cdots$ are uniformly tight.

Observing

$$E[|S_n(t)-S_n(s)|^2] = \sum_{i=1}^{n} |b_i(t)-b_i(s)|^2 \leq \int_0^1 |e_{0t}(u)-e_{0s}(u)|^2 du = |t-s| ,$$

we have

$$E[|S_n(t)-S_n(s)|^4] = 3E[|S_n(t)-S_n(s)|^2]^2 \leq 3(t-s)^2$$

because $S_n(t)-S_n(s)$ is Gauss distributed with the mean 0. Using the same technique of diadic expansions as in the proof of Kolmogorov's theorem, we can prove that for $\varepsilon>0$, there exists $\delta=\delta(\varepsilon)$ such that

$$P(\boldsymbol{S}_n \in K) > 1-\varepsilon, \quad n = 1, 2, \cdots$$

where

$$\begin{aligned} K &= K(\delta) = K(\varepsilon) \\ &= \{f \in C[0, 1]: f(0)= 0, \ |f(t)-f(s)| \leq 5|t-s|^{1/8} \quad \text{for} \quad |t-s| \leq \delta(\varepsilon)\} , \end{aligned}$$

It is easy to see that K is an equi-continuous and equibounded family. Therefore K is conditionally compact in $C[0, 1]$ by Ascoli-Arzela's theorem. This completes the proof.

c. Gaussian stationary processes

Let $S(t)$ be a Gaussian stationary process continuous in the square mean such that $E(S(t))=0$. Let us consider the sample path $\boldsymbol{S}$ of $S(t)$ on a bounded time interval $a \leq t \leq b$. By taking a measurable separable version we have $\boldsymbol{S} \in L^p[a, b]$ a.s. for every $p \geq 1$; we shall identify two functions on $[a, b]$ equal to each other a.e. In fact we have

$$\begin{aligned} E[\|\boldsymbol{S}\|_p^p] &= E[\int_a^b |S(t)|^p dt] \\ &= \int_a^b E[|S(t)|^p] dt = E[|S(0)|^p](b-a) \quad < \infty , \end{aligned}$$

noticing that $S(0, \omega)$ is Gauss distributed and so

$$P[||\boldsymbol{S}||_p < \infty] = 1 .$$

Consider the spectral decomposition of $S(t)$:

$$S(t, \omega) = \int_{-\infty}^{\infty} e^{i\lambda t} \Phi(d\lambda, \omega)$$

and set

$$X_n(t, \omega) = \int_{n-1 \leqq |\lambda| < n} e^{i\lambda t} \Phi(d\lambda, \omega) , \quad n = 1, 2, \cdots .$$

Then $X_n(t)$ has a version whose sample path (on $a \leq t \leq b$) is continuous a.s. and so belongs to $L^p[a, b]$ a.s. Now set

$$S_n(t) = \sum_1^n X_k(t) = \int_{|\lambda| < n} e^{i\lambda t} \Phi(d\lambda)$$

Then it is easy to see that

$$E[|S_n(t) - S(t)|^2] = \int_{|\lambda| \geq n} F(d\lambda) \to 0 \quad \text{as } n \to \infty$$

where $F(d\lambda)$ is the spectral measure of the covariance function of $S(t)$.

Theorem 5.3.

$$\int_a^b |S_n(t) - S(t)|^p dt \to 0 \quad \text{a.s. } (p \geq 1) .$$

Proof. Using the same notation for the sample paths we can see that $\boldsymbol{X}_n$, $n=1, 2, \cdots$ are independent random variables with values in $L^p[a, b]$.

$$\begin{aligned} E[||\boldsymbol{S}_n - \boldsymbol{S}||_p^p] &= E\left[\int_a^b |S_n(t) - S(t)|^p dt\right] \\ &= \int_a^b E[|S_n(t) - S(t)|^p] dt = c_p \int_a^b (E[|S_n(t) - S(t)|^2])^{p/2} dt \\ &= c_p(b-a)\left[\int_{|\lambda| \geq n} F(d\lambda)\right]^{p/2} \to 0 \quad \text{as } n \to \infty , \end{aligned}$$

where $c_p = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(\xi^2/2)} |\xi|^p d\xi$. By Theorem 3.1 ((b)$\Rightarrow$(a)) for $E = L^p[a, b]$, this implies $||\boldsymbol{S}_n - \boldsymbol{S}||_p \to 0$ a.s.

Theorem 5.4. *If the sample path of* $\boldsymbol{S}(t, \omega)$ *is continuous a.s., then*

$$\max_{a \leq t \leq b} |S_n(t) - S(t)| \to 0 \quad \text{a.s.}$$

REMARK. Sufficient conditions for the a.s. continuity of the sample path of $S(t)$ in terms of the correlation function of $S(t)$ were given by G. Hunt [4], Belayev [1] and X. Fernique [3].

Proof. Using the same idea as above, we can apply Theorem 4.1 ((e)$\Rightarrow$(a)) for $E=C[0, 1]$ by observing

$$
\begin{aligned}
&E\left[\left|\int_a^b z(dt)\, S_n(t, \omega) - \int_a^b z(dt)\, S(t, \omega)\right|\right] \\
&\leq \int_a^b |z|(dt) E[|S_n(t, \omega) - S(t, \omega)|] \\
&\leq \int_a^b |z|(dt) [E(|S_n(t) - S(t)|^2)]^{1/2} \\
&= \int_a^b |z|(dt) \left[\int F(d\lambda) \right]^{1/2} \to 0 , \quad \text{as } n \to \infty .
\end{aligned}
$$

AARHUS UNIVERSITY
KOBE UNIVERSITY

References

[1] Yu. K. Belyaev: *Continuity and Hölder's conditions for sample function of stationary processes*, Proc. 4th Berkeley Symp. Math. Stat. Prob. II, University of Calif. Press, 1961, 23–33.

[2] J. Delporte: *Fonctions aléatoires presque sûrement continues sur un intervalle fermé*, Ann. Inst. Henri-Poincaré **1** (1964), 111–215.

[3] X. Fernique: *Continuité des processus gaussiens*, C.R. Acad. Paris **258** (1964), 6058–6060.

[4] G.A. Hunt: *Random Fourier transforms*, Trans. Amer. Math. Soc. **71** (1951), 38–69.

[5] Yu. V. Prohorov: *Convergence of random processes and limit theorems in probability theory*, Theor. Probability Appl. **1** (1956), 157–214.

[6] R. Paley and N. Wiener: Fourier transforms in the complex domain, Amer. Math. Soc. Coll. Publ. XIX, 1934.

[7] A. Tortrat: *Lois de probabilité sur un espace topologique complètement régulier et produits infinis à terms indépendants dans un groupe topologique*, Ann. Inst. Henri-Poincaré **1** (1965), 217–237.

[8] J.B. Walsh: *A note on uniform convergence of stochastic processes*, Proc. Amer. Math. Soc. **18** (1967), 129–132.