MAT320: Probability, Duality, and Infinite Dimensional Geometry

Semon Rezchikov

11/16/2023

Last time we discussed how Fourier series give us "coordinates" on the infinite-dimensional function space $L^2([0,1])$. Specifically, given a function $f \in L^2([0,1])$, we have that $f \in L^1([0,1])$ by Cauchy-Schwarz, so the integrals

$$a_0 = \int_0^1 f(x)dx, a_k = \int_0^1 f(x)\sqrt{2}\sin(2\pi kx), b_k = \int_0^1 \sqrt{2}\cos(2\pi kx)dx, k = 1, 2, \ldots.$$

all make sense. Writing

$$(S_N f)(x) = a_0 + \sum_{k=1}^{N} a_k \sqrt{2}\sin(2\pi kx) + \sum_{k=1}^{n} b_k \sqrt{2}\cos(2\pi kx),$$

you will prove on the problem set assigned today that

$$\|S_N f - f\|_{L^2([0,1])} \to 0 \text{ as } N \to \infty.$$

In particular the Fourier series of f,

$$a_0 + \sum_{k=1}^{\infty} a_k \sqrt{2}\sin(2\pi kx) + \sum_{k=1}^{\infty} b_k \sqrt{2}\cos(2\pi kx)$$

converges almost everywhere, and for any $f \in L^2([0,1])$, we can assign coefficients $\{(a_k), (b_k)\}$ to f uniquely.

1 Duality.

Given a Banach space V, one can define the *space of continuous linear functions on* V, also called the *dual space* of V:

$$V^* = \{\ell : V \to \mathbb{R} : \ell \text{ is linear and continuous .}\}$$

One should restrict to continuous functions because there are too many linear functions otherwise, and the resulting theory is not very interesting. In fact, the existence of bases for a vector space(proven by the axiom of choice) can be used to show that given any two vector spaces V_1, V_2 of the same cardinality, there is a linear map $A : V_1 \to V_2$ which is bijective with linear inverse; so if we define V^* to be the set if linear functions on V then we would have $V_1^* = V_2^*$. However, the point is that we want to think of, say $C^0([0,1])$ and $L^2([0,1])$ as *different*, even though in fact they have the same cardinality. Restricting to continuous linear functions turns out to be very illuminating.

The dual to L^2. Anyway, given $g \in L^2([0,1])$, one can define

$$L_g : L^2([0,1]) \to \mathbb{R}, L_g(f) = \int fg.$$

The latter integral is well defined by Cauchy-Schwarz, since

$$|L_g(f)| \leq \|f\|_2 \|g\|_2$$

In particular there is a constant C such that

$$|L_g(f)| \leq C\|f\|_2$$

i.e., the linear funciton L_g "does not stretch vectors too much". We say that $L_g : L^2([0,1]) \to \mathbb{R})$ is *bounded* (as a linear function). It is easy to show that a bounded linear function (in this sense) on a Banach space is continuous.

It is an important theorem that this assignment

$$g \mapsto L_g, L^2([0,1]) \to L^2([0,1])^*$$

is a bijection: the dual space of $L^2([0,1])$ is just $L^2([0,1])$ itself!

Function evaluation. Let us consider another, rather simple, linear function on a function space: "evaluate the function at a point". This is clearly linear because $(f+g)(x) = f(x)+g(x)$. However, we can't use this formula to define a linear function

$$e_x : L^2([0,1]) \to \mathbb{R}$$

because elements of $L^2([0,1])$ aren't functions – they are *equivalence classes* of functions, and given $f \in L^2([0,1])$ there is always a $g \in L^2([0,1])$ such that $f \sim g$ but $f(x) \neq g(x)$.

However, an element of $C^0([0,1])$ manifestly has pointwise values, and so there is a linear function

$$e_x : C^0([0,1]) \to \mathbb{R}, e_x(f) = f(x),$$

and e_x is clearly bounded and thus continuous.

Now the inclusion $\iota : C^0([0,1]) \subset L^2([0,1])$ is continuous because $\|f\|_{L^2} \leq \|f\|_{C^0}$ for a continuous $f : [0,1] \to \mathbb{R}$. So we have an inclusion

$$L^2([0,1])^* \subset C^0([0,1])^*$$

which is *dual* to ι. We just showed that this inclusion isn't surjective. This is a common pattern: often, when $V \subsetneq W$, we will have that $W^* \subsetneq V^*$ (although when this and isn't true is a theorem in functional analysis).

The dual to C^0. How can we guess what $C^0([0,1])^*$ is? Well, you may know from probability $g \geq 0$ and $\int g(x)dx = 1$,

$$f \mapsto \int_0^1 f(x)g(x)dx$$

is the *expected value* $\mathbb{E}(f(X))$ of the random variable $f(X)$, where the *probability density function* of X is $g(x)$.

We can imagine taking the limit of a sequence of $g(x)$ which become more and more 'concentrated' on a single value, like

$$g_k(x) = \frac{k}{2}\mathbf{1}_{1/2-1/k \le x \le 1/2+1/k}(x).$$

These functions converge pointwise almost everywhere to 0, and they don't converge to anything in L^2. But any $f \in C^0$, the quantities $L_{g_k}(f)$ converge to $f(x)$, and with a little bit of work can show that in fact the linear operators

$$L_{g_k} \in C^0([0,1])^*$$

converge to $e_{1/2}$.

If $f \ge 0$ then we should have that $\mathbb{E}(f(X)) \ge 0$ for any random variable X. It turns out that *random variables $X \in [0,1]$ are exactly the continuous linear functionals $\mathbb{E}_X \in C^0([0,1])^*$ such that*

$$\mathbb{E}_X(f) \ge 0 \text{ if } f(x) \ge 0 \text{ for all } x\text{, and}$$
$$\mathbb{E}_X(1) = 1.$$

In fact, this is a perfectly good *definition* of a random variable – something which lets you assign 'expected values' to continuous functions, in a way that lets you take the limit through the expected value, and satisfies the obvious positivity properties. We will not prove this, but you can reach Chapter 21 of Royden if you are interested.

Remark. In the notes we will be thinking about random variables on $[0,1]$ for a bit, and then will generalize this vastly.

2 Probability via cdfs.

You might have heard that you can describe a random variable via a "cumulative density function", or 'cdf'. This is supposed to be a nondecreasing function

$$c_X : [0,1] \to \mathbb{R}$$

which tells you what is the probability of X being between a and b:

$$\mathbb{P}(a \le X \le b) = c_X(b) - c_X(a).$$

Remark. Here we considering a random variable $X \in [0,1]$ but of course that is not relevant for this part of the discussion; if $X \in \mathbb{R}$ then $c_X : \mathbb{R} \to \mathbb{R}$.

We should now be able to take the expected value of any continuous function f. How should we do that? Well, given a step function $h = \sum_{i=1}^N d_i \mathbf{1}_{a_i,b_i}$ we clearly should have

$$\mathbb{E}_X(h) = \sum_{i=1}^N d_i(c_X(b_i) - c_X(a_i))$$

by linearity. Now we can try to define "Riemann integrals" but with this assignment of an 'integral' to a step function, define 'Riemann integrable' functions with this modified notion, and so forth.

The uniform continuity of any continuous f will make continuous f 'Riemann integrable' in this sense; the corresponding integral, denoted

$$\int f dc_X$$

is called the "Riemann–Stieltjes integral" integral with respect to c_X.

But remember that the Lebesgue integral is almost the Riemann integral – one just takes the completion of the space of Riemann integrable functions with respect to the L^1-norm, and one recovers the Lebesgue measure via

$$E \text{ is measurable if and only if } \mathbf{1}_E \text{ is integrable .}$$

(Actually, we haven't proven this yet – we have only proven one direction and stated the other. It is true and is an easy theorem in the book.)

The same procedure can be done using the Riemann-Stieltjes integral: one takes the completion of the space of functions that are Riemann-Stieltjes integrable with respect to c_X, and then one defines the measure μ_X via

$$\mu_X(A) = \int \nVdash_A dc_X$$

where the latter is defined by approximating $\nVdash_A$ pointwise and in L^1 by Riemann–Stieltjes-integrable functions. It's easy to see that this measure should assign

$$\mu_X([a,b]) = c_X(b) - c_X(a).$$

It turns out that there is a much easier way to construct this measure – you say that the above formula is the measure on intervals. Now the set of intervals $\mathcal{S}$ is a *semi-ring*: if A and B are in $\mathcal{S}$, then $A \cap B \in S$, and also $A \setminus B$ is a finite disjoint union of elements of $\mathcal{S}$. We can more abstractly imagine collection $\mathcal{S}$ if subsets of some Ω with this property. Let $\mathcal{S}'$ be the collection of finite disjoint unions of elements of $\mathcal{S}$.

$$\mu_X : \mathcal{S}' \to \mathbb{R}_{\geq 0}$$

which is a satisfies the finite additivity and countable monotonicity property that measures do.

One now proceeds *exactly as we did when defining the Lebesgue measure*: you define the outer measure

$$\mu_X^*(T) = \inf\{\sum_{i=1}^{\infty} \mu^*(S_i) | S_i \in \mathcal{S}, T \subset \cup_{i=1}^{\infty} S_i\},$$

the notion of a measurable set (now using μ_X^*; so this may be different!) and finally you show that all sets in the smallest σ-algebra containing S', which in this case is the Borel σ-algebra, are measurable in this new sense.

The proof that this works is essentially the same, and this goes under the name of the "Caratheodory Extension theorem". Take a look in Section 17.5 of the book if you are interested in the proof; it is not very difficult. The hardest thing is showing that this measure on 'measurable sets' is the unique extension of μ that agrees with μ; this requires a technical condition, which is X can be written as a countable union of elements of $\mathcal{S}$.

Thus, in summary:

- A random variable has a CDF;
- This CDF defines a notion of expectation value on continuous functions;
- This expectation value corresponds to integration with respect to a measure, which behaves analogously to the PDF associated to the CDF.

Going backwards, given a Borel measure μ on $\mathbb{R}$, one can define its associated cdf via

$$c_\mu(x) = \int_{-\infty}^{x} d\mu$$

where this latter integral is done with respect to the measure μ. Then one can reconstruct μ from c_μ via the procedure detailed above.

Given a nonnegative integrable function f, one can define a measure

$$\mu_f(A) = \int_A f dx;$$

we will write this measure as "$f dx$". Conversely, given a Borel measure μ, we will say that it "has a density" if we can write it as $\mu = f dx$ for an integrable function f.

3 Measure-theoretic Probability.

All this motivates Kolmogorov's perpective on how to think about probability, which is now standard.

Definition. Fix a set Ω and a σ-algebra Σ on Ω. A *probability measure* on (Ω, Σ) is a countably additive measure

$$\mu : \Sigma \to \mathbb{R}_{\geq 0}$$

such that $\mu(\Omega) = 1$.

Now it turns out (you can look at Chapter 18) that the entire theory of Lebesgue integration just abstractly needs a triple (Ω, Σ, μ) where the condition $\mu(\Omega) = 1$ isn't necessary. Theorems like Lebesgue dominated convergence and Fatou's lemma continue to hold; we can talk about $L^1(\mu)$, etc.

Let's define a *random variable* to be a Σ-measurable function $X : \Omega \to \mathbb{R}$, i.e. $X^{-1}(\{x \geq c\}) \in \Sigma$ for all $c \in \mathbb{R}$. Then we can define

$$\mathbb{E}[X] = \int X d\mu$$

whenever X is integrable or nonnegative.

We think of $A \in \Sigma$ as an *event*, and we think of $\mu(A) = \mathbb{P}(A) = \mathbb{E}(\mathbf{1}_A)$.

In general, we should think of Ω as the set of 'possible worlds', Σ as characterizing the *questions* that we can ask about the world (what kinds of 'events' are there?), and μ as defining the probabilities of events.

Now we can talk about independent events: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. We can also talk about expectation values of continuous functions of random variables, since the composition of a continuous function with a measurable function is measurable. We could have *lots* of random variables on the same space Ω, and encode lots of complicated 'nonindependence' relationships between them in μ.

We say that a statement involving random variables with *probability* 1 (with respect to μ, which may be implicit), or *almost always*, if there is a set $A \in \Sigma$ such that $\mu(A) = 1$ and such that for any $x \in A$ the statement is true whenever all random variables in the statement are evaluated at x. We will give an example of this notion in Section 8.

4 Pushforwards of measures

Now you might know about nice continuous random variables, we are supposed to be able to talk about probabilities like $P((X, Y) \in [a, b] \times [c, d])$ for random variables X, Y. So first of all,

$$\mathbb{P}(a \leq X \leq B) = \mathbb{P}(X^{-1}([a, b])$$

makes sense because X is measurable. In fact,

$$E \mapsto \mathbb{P}(X^{-1}(E))$$

is manifestly a probability measure on $(\mathbb{R}, \Sigma_{\mathbb{R}}^{Borel})$, where $\Sigma_{\mathbb{R}}^{Borel}$ is the σ-algebra of Borel sets on $\mathbb{R}$! This measure is called the *pushforward measure*, and is denoted by $X_*\mu$, and this is the generalization of the probability density function of X.

Similarly, the joint probability density function of (X, Y) is replaced with the *pushforward measure* $(X, Y)_*\mu$, where we think of

$$(X, Y) : \Omega \to \mathbb{R}^2$$

as the function sending $x \in \Omega$ to the pair $(X(x), Y(x)) \in \mathbb{R}^2$. The quantity $(X, Y)_*\mu$ is defined in the same way as before: we set

$$(X, Y)_*\mu(A) = \mu((X, Y)^{-1}(A))$$

for all sets $A \in \Sigma_{\mathbb{R}^2}^{Borel}$. Thus $(X, Y)_*\mu$ is a measure on $(\mathbb{R}^2, \Sigma_{\mathbb{R}^2}^{Borel})$, where $\Sigma_{\mathbb{R}^2}^{Borel}$ is the smallest σ-algebra on $\mathbb{R}^2$ containing all the boxes $[a, b] \times [c, d]$.

Notice that, writing $X : \mathbb{R}^2 \to \mathbb{R}$ for the function given by projection to the first coordinate

$$X_*((X, Y)_*\mu)(A) = ((X, Y)_*\mu)(X^{-1}(A)) = (X, Y)_*\mu(A \times \mathbb{R}) = \mu((X, Y)^1(A \times \mathbb{R})) = \mu(X^{-1}(A)) = X_*\mu.$$

This is the measure theoretic analog of the consistency conditions for marginals of probability density functions: if we write $p_X, p_Y, p_{(X,Y)}$ for the pdfs for X, Y, and (X, Y) (when these random variables *have* PDFs), then we have that

$$p_X(x) = \int p_{(X,Y)}(x, y)dy.$$

However, this condition makes sense even when the random variables X or Y do *not* have PDFs.

5 Independence of Random Variables

A pair of random variables $X, Y : \Omega \to \mathbb{R}$ are independent when when all the events defined by these random variables are independent. What this should mean in part is that

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \mathbb{P}(a \leq X \leq b)\mathbb{P}(c \leq Y \leq d)$$

for all $a, b, c, d \in \mathbb{R}$. This is exactly equivalent to the independence condition in measure theoretic probability; by this Caratheodory extension theorem, this means that

$$\mu(X^{-1}(A) \cap Y^{-1}(B)) = \mu(X^{-1}(A))\mu(Y^{-1}(B))$$

for all Borel subsets A, B in $\mathbb{R}$.

More generally, we say that a finite set of random variables $X_1, \ldots, X_n$ are *jointly independent* if

$$\mu(X_1^{-1}(A_1) \cap X_2^{-1}(A) \cap \ldots \cap X_n^{-1}(A_n)) = \prod_{i=1}^{n} \mu(X_i^{-1}(A_i))$$

for all Borel subsets $A_1, \ldots, A_n$ in $\mathbb{R}$.

Given an *infinite set* of random variables $X_1, X_2, \ldots$, we say that they are jointly independent when *all finite subcollections* $X_{i_1}, \ldots, X_{i_n}$ are jointly independent. The fact that there is no 'new independence data' when generalizing to the case of infinitely many random variables is actually a *theorem*: it follows from the continuity of measure that a sequence of random variables is jointly independent in the above sense then we have that

$$\mathbb{P}(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \ldots) = \prod_{i=1}^{\infty} \mathbb{P}(a_i \leq X_i \leq b_i)$$

by taking the limit from the corresponding identities for $X_1, \ldots, X_n$ as $n \to \infty$.

6 Kolmogorov Extension

Now, we really want to be able to consider infinite collections of random variables. One of the nice things about measure theory is that this is not much of a problem. We just need consistency conditions between the finite-dimensional marginals. We won't prove this theorem, but it's basically an application of the Caratheodory extension theorem:

Theorem 1 (Kolmogorov Extension Theorem)**.** Suppose that for every finite subset

$$\vec{a} = \{a_1, \ldots, a_n\} \subset \mathbb{N},$$

there is a probability measure $\mu_{\vec{a}}$ on $\mathbb{R}^n$. Moreover, suppose these probability measures satisfy the following consistency condition: given

$$\vec{b} = \{a_{i_1}, \ldots a_{i_\ell}\} \subset \{a_1, \ldots, a_n\} = \vec{a}$$

suppose that the pushfoward

$$(p^{\vec{b}})_* \mu_{\vec{a}}$$

which is the probability measure on $\mathbb{R}^\ell$ arising by pushforward along the map

$$p^{\vec{b}} : \mathbb{R}^n \to \mathbb{R}^\ell, p^{\vec{b}}(x_1, \ldots, x_n) = (x_{i_1}, \ldots x_{i_\ell}).$$

In other words, this is the condition we have specified the joint probability distributions of all finite subcollections of an infinite collection of random variables. Then there exists a measure on

$$\mathbb{R}^\infty = \mathbb{R} \times \mathbb{R} \times \mathbb{R} \ldots = \{(x_i)_{i=1}^{\infty} | x_i \in \mathbb{R}\}$$

on the σ-algebra $\bar{\Sigma}_{cyl}$, which is the smallest σ-algebra containing all the "cylinder sets"

$$[a_1, b_1] \times \ldots \times [a_n, b_n] \times \mathbb{R} \times \mathbb{R} \times \ldots \subset \mathbb{R}^\infty.$$

In particular, there is a probability measure $\mu_{\mathcal{N}^\infty}$ on $(\mathbb{R}^\infty, \bar{\Sigma}_{cyl})$ such that if we consider the random variables

$$X_i : \mathbb{R}^\infty \to \mathbb{R}, X_i(x_1, x_2, \ldots) = x_i,$$

then these are an infinite collection of independent identically distributed normal random variables; so the joint distribution of $X_{a_i}, \ldots, X_{a_n}$ is always the normal distribution of mean zero and variance 1, i.e.

$$(X_{a_1}, \ldots, X_{a_n})_* \mu_{\mathcal{N}^\infty} = (2\pi)^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) dx_1 \ldots dx_n.$$

7 Infinite-dimensional geometry.

Ok, so Kolmogorov extension lets us construct some kind of "infinite dimensional normal distribution" on $\mathbb{R}^\infty$. That sounds great! Let's try to sample a random vector from this distribution and compute its length.

Well, all the X_i are independent, so expected length of the vector should be

$$\mathbb{E}\left[\sqrt{\sum_{i=1}^{\infty} X_i}\right] \geq \lim_{N\to\infty} \mathbb{E}\left[\sqrt{\sum_{i=1}^{N} X_i}\right] = N$$

for $c > 0$. In other words, the expected length of this vector is infinite! Whatever a typical vector is, it doesn't lie in ℓ_2 with probability 1.

This may feel awkward, because naively speaking one would think that the "pdf" of this distribution

$$\frac{1}{Z}\exp(-\frac{1}{2}\sum_i x_i^2)\text{“}dx\text{”}$$

where Z is some normalization constant, is very natural. Part of what is going on is that an analog of the infinite-dimensional Lebesgue measure "dx" doesn't really exist in any simple way:

Proposition 1. There is no translation invariant Borel measure μ on ℓ^2 which assigns a finite nonzero measure to some open set.

Proof. Given any nonempty open ball $B(x, \epsilon) \subset \ell^2$, we can cover ℓ^2 by a countable union of translates of this ball: choosing a countable dense subset $\{x_i\}_{i=1}^\infty \subset \ell^2$ we have that $\ell^2 = \cup_i B(x_i, \epsilon)$. We can take x_i to be an enumeration of the countable set

$$\{(a_i)_{i=1}^\infty | a_i \in \mathbb{R}, \text{ all but finitely many of the } a_i \text{ are zero.}\} \subset \ell^2.$$

Thus, if $\mu(B(x, \epsilon)) = 0$ for any such ball then μ assigns zero measure to any open subset of ℓ^2 by monotonicity. So we can assume that $\mu(B(x, \epsilon)) > 0$ for all such balls. Suppose that μ assigns a nonzero measure to some open set U. Then it assigns a nonzero measure to every nonempty open ball contained in U; by translation to the origin we have that in particular $\mu(B(0, A)) = C$ for some

A, with $\infty > C > 0$. But there are infinitely many disjoint translated copies of $B(0, A/100)$ side $B(0, A)$; if we write $e_i = (0, 0, \ldots, 0, 1, 0, \ldots)$ with the 1 in the i-th place, then

$$B(0, A) + e_1/(2A), B(0, A) + e_2/(2A), B(0, A) + e_3/(2A), \ldots$$

works. So then by monotonicity and translation invariance of the measure we have that $\mu(0, A/100) = 0$, and thus $\mu(\ell^2) = 0$ by the previous argument.

$\square$

This is why we cannot use pdfs to analyze distributions over functions! There are very effective calculational methods developed in physics which do try to describe such infinite-dimensional distributions using pdfs, but they have some subtleties which are clarified by measure-theoretic constructions.

8 Brownian motion.

However, we have the measure $\mu_{\mathcal{N}^\infty}$ on $\mathbb{R}^\infty$. Lets think of this sequence of independent and identically distributed (iid) normal random variables X_i as instead being indexed by $i \in \mathbb{Z}$ for notational convenience. Let' consider the "function"

$$dB(t) = \sum_{k \in \mathbb{Z}, k \neq 0} X_k e_k$$

where the e_i are the Fourier basis vectors for $L^2([0, 1])$, so

$$e_0(t) = 1, e_k(t) = \sqrt{2}\cos(2\pi kt) \text{ for } k > 0, e_k(t) = \sqrt{2}\sin(2\pi kt) \text{ for } k < 0.$$

This series does not in any sense converge to a function; in fact, the previous computation shows that with probability 1 the L^2 norm of this function is infinite with probability one.

Let's ignore this issue and integrate this series term by term to get the "random function"

$$B(t) = \sum_{i \in \mathbb{Z}, k \neq 0} X_k \frac{e_k}{2\pi k}$$

The first "function" $dB(t)$ is something where all frequencies have the same distribution and are equally variable – it is called "white noise". Its integral is called *Brownian motion*, and turns out to be an actual random function. The contribution of larger frequencies to Brownian motion is damped due to the factor $1/(2\pi k)$ and so one might expect that the random Fourier series has better convergence properties.

This turns out to be true, and is an important theorem. We proved this in class:

Proposition 2. The series for $B(t)$ converges in $L^2([0, 1])$ with probability 1.

Just a little bit more work would show:

Proposition 3. The series for $B(t)$ converges to a continuous function with probability 1, but this function is nowhere differentiable with probability 1.

Sketch of proof of Proposition 2. We have the tail estimate

$$\mathbb{P}[|X_i| > M] = \frac{1}{\sqrt{2\pi}} \int_M^\infty e^{-\frac{x^2}{2}} dx \le e^{-M^2/2}$$

if $M \ge 1$. This can be proven by integration by parts:

$$\frac{1}{\sqrt{2\pi}} \int_M^\infty e^{-\frac{x^2}{2}} dx \le \int_M^\infty \frac{x}{M} e^{-\frac{x^2}{2}} dx = \frac{1}{2M} e^{-M^2/2} \le \frac{1}{2} e^{-M^2/2}.$$

This gives the estimate

$$\mathbb{P}[|X_i| > M] \le e^{-M^2/2}.$$

Therefore for any $\beta > 1$ we have

$$\mathbb{P}[|X_n| > (2\beta \log |n|)^{1/2}] \le \exp -\beta \log |n| = n^{-\beta}.$$

Thus

$$\sum_{n \in \mathbb{Z}, n \neq 0} \mathbb{P}[|X_n| > (2\beta \log |n|)^{1/2}] \le \sum_{n \in \mathbb{Z}, n \ge 0} n^{-\beta} < \infty$$

for any $\beta > 1$.

By Borel-Cantelli, this means that with probability 1 with respect to $\mu_{\mathcal{N}^\infty}$, at most finitely many of the events

$$|X_n| > (2\beta \log |n|)^{1/2}$$

occur. In other words, there exists an event $A \in \bar{\Sigma}_{cyl}$ such that $\mu_{\mathcal{N}^\infty}(A) = 1$, and such that for all $x \in A$ we have that the inequality

$$\mathbb{P}[|X_n(x)| > (2\beta \log |n|)^{1/2}]$$

holds for all but finitely many n. But this means that for all $x \in A$, we have

$$\|B(t)\|_{L^2([0,1]} = \sum_{k \in \mathbb{Z}, k \neq 0} \frac{X_k(x)^2}{(2\pi k)^2} \le D + \sum_{k \in \mathbb{Z}, k \neq 0} \frac{(2\beta \log |k|)^{1/2}}{(2\pi k)^2}$$

for some constant D. But $\log k / k^\epsilon \to 0$ as $k \to \infty$ for any $\epsilon > 0$, so $(\log k)^{1/2} / k^\epsilon \to 0$ as $k \to \infty$ for any $\epsilon > 0$ (by taking sqare root of the previous limit), so

$$\sum_{k \in \mathbb{Z}, k \neq 0} \frac{(2\beta \log k)^{1/2}}{(2\pi k)^2} < \sum_{k \in \mathbb{Z}, k \neq 0} \frac{2\beta |k|^\epsilon}{(2\pi k)^2} = \sum_{k \in \mathbb{Z}, k \neq 0} \frac{2\beta}{(2\pi)^2 |k|^{2-\epsilon}}.$$

Thus sum converges for any $\epsilon < 1$, which proves the proposition.

□