Lecture 5. Random Objects in Hilbert Spaces

Functional Data Analysis

Jun Song

Department of Statistics
Korea University

Table of contents

Random Objects

What is a probability space?

A probability space is a triplet $(\mathcal{S}, \mathfrak{S}, P)$:

- $\mathcal{S}$ is the sample space,
- $\mathfrak{S}$ is a σ-algebra of subsets of $\mathcal{S}$,
- P is a probability measure over $\mathfrak{S}$.

What is a probability space?

A probability space is a triplet $(\mathcal{S}, \mathfrak{S}, P)$:

- Intuitively, $\mathfrak{S}$ are of the sets to which we can compute probabilities.
- A σ-algebra is a collection of sets that contains $\emptyset$, $\mathcal{S}$, is closed under complementation ($A \in \mathfrak{S}$ then $A^C \in \mathfrak{S}$) and is closed under countable unions. ($A_i \in \mathfrak{S}$ then $\cup_{i=1}^{\infty} A_i \in \mathfrak{S}$)

A probability measure is a mapping $P : \mathfrak{S} \to [0, 1]$ which satisfies

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

for any countable collection of elements $A_i \in \mathfrak{S}$ which are mutually exclusive, $A_i \cap A_j = \emptyset$ for $i \neq j$.

What is a random variable?

- A **measurable space** is the doublet $(\mathcal{S}, \mathfrak{S})$, i.e., without the measure.
- When $\mathcal{S} = \mathbb{R}$, we usually take $\mathfrak{S} = \mathfrak{B}$, the *Borel σ-algebra*, which is the smallest σ-algebra containing all of the open sets.
- A random variable X, is a mapping from $\mathcal{S}$ to $\mathbb{R}$, which satisfies $X^{-1}(A) \in \mathfrak{S}$ for all $A \in \mathfrak{S}$. Such function is said to be *measurable*
- A random vector, like a random variable, is a measurable mapping from the sample space to $\mathbb{R}^d$.

What is a stochastic process?

- A **stochastic process** is a collection of random variables $\{X_i : i \in \mathcal{I}\}$ where $\mathcal{I}$ is some index set, e.g. $\{1, 2, \ldots, n\}$ or $[0, 1]$.
- If the set, $\mathcal{I}$, is finite, then this is a fairly clear definition as it is just a random vector.
- When $\mathcal{I}$ is infinite/uncountable, it is not so clear, and in fact, quite technical to define fully.

Simpler approach: Random functions

- Instead of piecing together random variables, like with the theory of stochastic processes, it is a bit easier (and more natural from an FDA point of view), to rephrase things.
- In particular, let $\mathcal{B}$ be a Banach space (metric space also works), then one has a topology over $\mathcal{S}$ (i.e. one can define open sets).
- With open sets, one can define a Borel σ-algebra over $\mathcal{B}$.
- **A random function** is then a mapping $X : \mathcal{S} \to \mathcal{B}$ which satisfies $X^{-1}(B) \in \mathfrak{S}$ for any $B \in \mathfrak{S}$. Nearly the same as before!

Modes of convergence

As in scalar/multivariate probability theory, we can define four modes of convergence for sequences of probability distributions and random variables:

- convergence in distribution,
- convergence in probability,
- convergence in mean,
- convergence almost surely (i.e. with probability one).

Convergence in distibution

Let $\{\mu_n : n = 1, 2, \dots\}$ be a sequence of probability measures over a metric space $\mathcal{S}$ (equipped with the Borel σ-algebra $\mathfrak{B}$). We say μ_n **converges in distribution** to μ if

1. $$\int_{\mathcal{S}} f(x)\ d\mu_n(x) \to \int_{\mathcal{S}} f(x)\ d\mu(x),$$

   for all bounded continuous functions $f : \mathcal{S} \to \mathbb{R}$.

2. Equivalently, one can define it as

   $$\mu_n(A) \to \mu(A),$$

   for every μ-continuity set A, which means $\mu(\partial A) = 0$ where ∂A is the boundary of A.

   This definition also works for a sequence of random functions by letting μ_n denote the probability distribution of X_n and μ of X, in this case, we say $X_n \xrightarrow{\mathcal{D}} X$.

Other modes

Let $\{X, X_1, X_2, \dots\}$ be a sequence of random variables. Let $d(\cdot, \cdot)$ be the metric over $\mathcal{S}$,

$$
\begin{aligned}
&d(X_n, X) \xrightarrow{P} 0, && \text{(convergence in probability)} \\
&\mathrm{E}[d(X_n, X)] \to 0, && \text{(convergence in mean)} \\
&d(X_n, X) \xrightarrow{\text{as}} 0 && \text{(convergence almost surely).}
\end{aligned}
$$

Convergence Theorem 1

Theorem

If $X_n \xrightarrow{\mathcal{D}} X$ *and* $d(X_n, Y_n) \to 0$, *then* $Y_n \xrightarrow{\mathcal{D}} X$.

Convergence Theorem 2

Let $\{X_n(u) : n = 1, 2, \ldots, \quad u \in \mathbb{R}^+\}$ be a collection of random elements of $\mathcal{S}$ indexed by n and u. Let $\{X_n\}$ be another sequence of random variables.

Theorem

Assume $X_n(u) \xrightarrow{\mathcal{D}} X(u)$ *as* $n \to \infty$ *and* $X(u) \xrightarrow{\mathcal{D}} X$ *as* $u \to \infty$. *If*

$$\lim_{u \to \infty} \limsup_{n \to \infty} P(d(X_n(u), X(u)) > \varepsilon) = 0,$$

then $X_n \xrightarrow{\mathcal{D}} X$.

Expectation

How to define expectation?

Again we turn to the inner product. Let X be a random element of $\mathcal{H}$. We say that $\mu \in \mathcal{H}$ is the mean of X if it satisfies

$$\mathrm{E}\langle X, x\rangle = \langle \mu, x\rangle \qquad \text{for all } x \in \mathcal{H}.$$

Such a μ need not exist for any random X (as in the scalar case).

- We say that X is *strongly integrable* if $\mathrm{E}\,\|X\| < \infty$.
- We say that X is *weakly integrable* if $\mathrm{E}\,|\langle X, x\rangle| < \infty$ for all $x \in \mathcal{H}$, and if there exists a $\mu \in \mathcal{H}$ such that $\mathrm{E}\langle X, x\rangle = \langle \mu, x\rangle$.

As the name implies strongly integrable implies weakly, but the reverse is not true. In general, one usually works with strong forms of integrability.

Example - $\mathbb{R}^d$

Let $X \in \mathbb{R}^d$ be a random vector. Then we define $\mu \in \mathbb{R}^d$ as

$$\mathrm{E}\langle X, x\rangle = \mathrm{E}\sum_{i=1}^{d} X_i x_i = \sum_{i=1}^{d} \mathrm{E}[X_i] x_i = \sum_{i=1}^{d} \mu_i x_i = \langle \mu, x\rangle.$$

So the mean of X, μ, is just a vector whose coordinates are the coordinate-wise means of X.

Example - $L^2[0,1]$

Let $X \in L^2[0,1]$ be a random function. Then we define $\mu \in L^2[0,1]$ as

$$\begin{aligned} \mathrm{E}\langle X, x\rangle &= \mathrm{E}\int_0^1 X(t)x(t)\ dt \\ &= \int_0^1 \mathrm{E}[X(t)]x(t)\ dt = \int_0^1 \mu(t)x(t)\ dt = \langle \mu, x\rangle. \end{aligned}$$

So the mean of X, μ, is just a function whose coordinates are the coordinate-wise means of X.

Two properties of expection

Below are two potentially useful properties of expectation.

- $\|\mathrm{E}\, X\| \leq \mathrm{E}\, \|X\|$ (contractive property)
- If L is a bounded linear operator, then $\mathrm{E}\, L(X) = L(\mathrm{E}\, X)$.

Proof 1

The first claim follows from the definition of the expected value and the CS inequality:

$$\|\mu\|^2 = \langle \mu, \mu \rangle = \mathrm{E}\langle X, \mu \rangle \leq \mathrm{E}\,\|X\|\|\mu\|.$$

Proof 2

The second property follows from the use of adjoints:

$$\mathrm{E}\langle L(X), x\rangle = \mathrm{E}\langle X, L^*(x)\rangle = \langle \mu, L^*(x)\rangle = \langle L(\mu), x\rangle.$$

So, by definition

$$\mathrm{E}\, L(X) = L(\mu).$$

Covariance

Covariances

- In general Hilbert spaces, covariances are viewed as operators (not as functions or matrices).
- One could also view them as bilinear forms (though this isn't typical).
- We say that C is the covariance operator of X if

$$C(x) = \mathrm{E}[\langle X - \mathrm{E}\, X, x\rangle (X - \mathrm{E}\, X)] \qquad \text{for all } x \in \mathcal{H}.$$

  Such an operator is guaranteed to exist if $\mathrm{E}\,\|X\|^2 < \infty$.

Example - $\mathbb{R}^d$

Let $X \in \mathbb{R}^d$ be a random vector. The requirement that $\mathrm{E}\,\|X\|^2 < \infty$ is given by

$$\mathrm{E}\,\|X\|^2 = \mathrm{E}\sum_{i=1}^{d} X_i^2 = \sum_{i=1}^{d} \mathrm{E}\,X_i^2.$$

So, the covariance operator will exist as long as each coordinate has a finite variance. We characterize the covariance as

$$\begin{aligned}
\langle C(x), y\rangle &= \mathrm{E}[\langle X - \mathrm{E}\,X, x\rangle\langle X - \mathrm{E}\,X, y\rangle] \\
&= \mathrm{E}\left[\sum_i \sum_j (X_i - \mu_i)x_i(X_j - \mu_j)y_j\right] \\
&= \sum_i \sum_j \mathrm{Cov}(X_i, X_j)x_i y_j = x^\top \Sigma y.
\end{aligned}$$

So, the covariance operator equivalent to the covariance matrix, Σ, and is given by

$$C(x) = \Sigma x.$$

Example - $L^2[0,1]$

Let $X \in L^2[0,1]$ be a random function. The requirement that $\mathrm{E}\,\|X\|^2 < \infty$ is given by

$$\mathrm{E}\,\|X\|^2 = \mathrm{E}\int_0^1 X(t)^2 = \int_0^1 \mathrm{E}\,X(t)^2.$$

So, the covariance operator will exist as long as the variance function is integrable. We characterize the covariance as

$$\begin{aligned}
\langle C(x), y\rangle &= \mathrm{E}[\langle X - \mathrm{E}\,X, x\rangle\langle X - \mathrm{E}\,X, y\rangle] \\
&= \mathrm{E}\left[\int\int (X(t) - \mu(t))x(t)(X(s) - \mu(s))y(s)\; dt\; ds\right] \\
&= \int\int \mathrm{Cov}(X(t), X(s))x(t)y(s)\; dt\; ds.
\end{aligned}$$

So, the covariance operator equivalent to the covariance function, $c(t,s)$.

Tensor notation

The notation for $C(x)$ is a bit cumbersome, but it can be simplified using Tensor notation. In particular

$$C := \mathrm{E}[(X - \mu) \otimes (X - \mu)].$$

Covariance operators

Any covariance operator, C, satisfies the following.

- C is symmetric.
- C is nonnegative-definite.
- C is Hilbert-Schmidt.
- C is also called *nuclear* or a *trace class operator* in that it satisfies

$$\sum \lambda_i < \infty,$$

  where λ_i are its eigenvalues.

In fact, any operator that satisfies the above is the covariance operator of some X.

Gaussian Processes

A random function $X \in \mathcal{H}$ is said to be Gaussian if $\langle X, y \rangle$ is normally distributed (with finite variance) for any $y \in \mathcal{H}$. If X is Gaussian then it satisfies the following:

- X is strongly integrable and so $\mathrm{E}\, X$ exists,
- $\mathrm{E}\, \|X\|^2 < \infty$ and so it has a covariance operator, C,
- in fact, $\mathrm{E}\, \|X\|^p < \infty$ for any postive p.

Notice that this means that

$$\langle X, x \rangle \sim \mathcal{N}\left(\langle \mu, x \rangle, \langle C(x), x \rangle\right).$$

Central Limit Theorem

Characteristic functions

The characteristic function of a random function X is defined as

$$\psi(x) = \mathrm{E} \exp\{i\langle X, x\rangle\}.$$

This function exists for any random function in $\mathcal{H}$ and it uniquely characterizes distributions in the sense that X and Y have the same distribution if and only if they have the same characteristic functions. If X is Gaussian with mean μ and covariance C, then

$$\psi(x) = \exp\left\{ i\langle \mu, x\rangle - \frac{1}{2}\langle C(x), x\rangle \right\}.$$

An alternative (but equivalent) definition of a Gaussian process is a random function X whose characteristic function is given as above.

Central limit theorem

We now have enough tools/background to give the CLT for Hilbert spaces.

Theorem

Suppose $\{Y_n : n = 1, 2, \dots\}$ *is an iid sequence of random functions in* $\mathcal{H}$ *which are square integrable,* $\mathrm{E}\,\|Y_n\|^2 < \infty$. *Let* μ *denote their mean and* C *their covariance. Then we have*

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} (Y_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C).$$

Proof - CLT

We will use the approximation result from last class: if $X_N(u) \xrightarrow{\mathcal{D}} X(u)$, $X(u) \xrightarrow{\mathcal{D}} X$ and for any $\varepsilon > 0$ $\lim_{u\to\infty} \limsup_{N\to\infty} P(\|X_N - X_N(u)\| > \varepsilon) = 0$ then $X_N \xrightarrow{\mathcal{D}} X$.

Step 1: Karhunen-Loéve expansion

Let us use the eigenfunctions, v_i, of C, as a basis for expanding the Y_n:

$$Y_n - \mu = \sum_{j=1}^{\infty} \xi_{nj} v_j, \tag{1}$$

where $\xi_{nj} = \langle Y_n - \mu, v_j \rangle$. Notice that

$$\mathrm{E}[\xi_{nj}\xi_{nk}] = \mathrm{E}[\langle Y_n - \mu, v_j \rangle \langle Y_n - \mu, v_k \rangle] = \langle C(v_j), v_k \rangle = \lambda_j 1_{j=k}.$$

So the ξ_{nj} are uncorrelated and have variance λ_j. Expansion (1) is called the *Karhunen-Loéve* expansion.

Step 2: Set up approximations

We are going to use the KL expansion to form our approximation. Let $u \in \{1, 2, \dots\}$ then define the following terms

1. $X_N = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (Y_n - \mu)$
2. $X_N(u) = \sum_{j=1}^{u} \langle X_N, v_j \rangle v_j = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \sum_{j=1}^{u} \langle Y_n - \mu, v_j \rangle v_j$
3. $X \sim \mathcal{N}(0, C)$.

Step 3: $X_n(u) \xrightarrow{\mathcal{D}} X(u)$

This part is fairly straightforward, by the multivariate CLT we have

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \begin{pmatrix} \langle Y_n - \mu, v_1 \rangle \\ \vdots \\ \langle Y_n - \mu, v_u \rangle \end{pmatrix} \xrightarrow{\mathcal{D}} \mathbf{Z}_u \sim \mathcal{N}(0, \mathbf{\Lambda}_u),$$

where Z_u is multivariate normal and $\mathbf{\Lambda}_u$ is a diagonal matrix of $\lambda_1, \ldots, \lambda_u$. This implies that

$$X_N(u) \xrightarrow{\mathcal{D}} X(u) = \sum_{j=1}^{u} Z_j v_j.$$

Note that this follows from the *continuous mapping theorem*.

Step 4: $X(u) \xrightarrow{\mathcal{D}} X$

Now we "ease up" on u allowing it to go to infinity. So, the question becomes, is $\sum_{j=1}^{\infty} Z_j v_j$ an element of $\mathcal{H}$? Is it still a Gaussian process? The answer is yes to both. First notice that, by Parseval's identity

$$\mathrm{E}\,\|X\|^2 = \sum_{j=1}^{\infty} \mathrm{E}\,Z_j^2 = \sum_{j=1}^{\infty} \lambda_j < \infty.$$

So indeed X is in $\mathcal{H}$ and is square integrable. Second, it is clearly Gaussian since

$$\langle X, x \rangle = \sum_{j=1}^{\infty} Z_j \langle x, v_j \rangle,$$

is a linear combination of normals.

Step 5: Probability bound

We now need a probability bound for $\|X_n(u) - X_n\|$. Notice that the difference can be written as

$$X_N(u) - X_N = \sum_{j=u+1}^{\infty} \langle X_N, v_j \rangle v_j.$$

So by Parseval's, we have that

$$\|X_N(u) - X_N\|^2 = \sum_{j=u+1}^{\infty} \langle X_N, v_j \rangle^2.$$

How can find a bound for $P(\|X_N(u) - X_N\| > \varepsilon)$?

Step 5 (cont)

We can use the Markov inequality

$$P(\|X_N(u) - X_N\| > \varepsilon) = P(\|X_N(u) - X_N\|^2 > \varepsilon^2) \leq \frac{\mathrm{E}\,\|X_N(u) - X_N\|^2}{\varepsilon^2}.$$

It is a fairly straightforward calculation to show that

$$\mathrm{E}\,\|X_N(u) - X_N\|^2 = \sum_{j=u+1}^{\infty} \lambda_j.$$

This bound doesn't depend on N so

$$\lim_{u\to\infty} \limsup_{N\to\infty} P(\|X_N - X_n(u)\| > \varepsilon) \leq \lim_{u\to\infty} \sum_{j=u+1}^{\infty} \lambda_j = 0.$$

Why is the last equality true? This completes the proof $\square$.

Alternative proof

There is another proof of the CLT which is shorter but relies on deeper probability theory. In multivariate statistics, a

$$\langle X_n, x\rangle \xrightarrow{\mathcal{D}} \langle X, x\rangle \quad \text{for all } x \in \mathbb{R}^d.$$

This is called the Cramér-Wold theorem. Equivalently, one can replace the above with the condition that the characteristic

Altenative proof

It is almost true! There is one property which is missing. It is a concept called *tightness*. A sequence of random

$$P(X_n \in A) \geq 1 - \varepsilon \qquad \text{for all } n.$$

This property is fairly manageable in Hilbert spaces as one can usually tie it to the behavior of the covariance operators. Much If the sequence is tight, then one can apply the Cramer-Wold theorem, otherwise, it need not be true.

Example

Let e_n be an orthonormal basis. Set $X_n = e_n$ (with probability 1). Then one has that, for any $x \in \mathcal{H}$

$$\langle X_n, x \rangle = \langle e_n, x \rangle \to 0.$$

However, X_n does not converge in distribution (or in probability) to zero. The sequence is not tight!

Laws of large numbers

We won't prove them here, but one also has the laws of large numbers. If X_n is an iid sequence of strongly integrable functions in $\mathcal{H}$ then one has

$$\text{Weak: } N^{-1}\sum_{n=1}^{N} X_n \xrightarrow{\mathcal{P}} \mu,$$

$$\text{Strong: } N^{-1}\sum_{n=1}^{N} X_n \xrightarrow{\text{as}} \mu.$$

Functional PCA

FPCA

The goal of FPCA is to approximate infinite dimensional objects with finite dimensional ones, whose dimension is as low as possible. The typical introduction to any FPCA is the following. Suppose that $X \in \mathcal{H}$ is a mean zero square-integrable random function. Then we aim to find a set of orthonormal functions $\{u_1, \ldots, u_d\}$ such that

$$S(u_1, \ldots, u_p) = \mathrm{E} \left\| X - \sum_{k=1}^{p} \langle X, u_k \rangle u_k \right\|^2,$$

is as small as possible. Notice that $\sum_{k=1}^{p} \langle X, u_k \rangle u_k$ is just a projection of X onto the subspace spanned by the $\{u_1, \ldots, u_p\}$.

FPCA cont

We can rephrase S as

$$\mathrm{E}\left\| X - \sum_{k=1}^{p} \langle X, u_k \rangle u_k \right\|^2 = \mathrm{E}\,\|X\|^2 - \sum_{k=1}^{p} \mathrm{E}\langle X, u_k \rangle^2.$$

(See Chapter 11.4 for details) So, minimizing S is equivalent to maximizing

$$\sum_{k=1}^{p} \mathrm{E}\langle X, u_k \rangle^2 = \sum_{k=1}^{p} \langle C(u_k), u_k \rangle,$$

subject to the constraint that the u_k are orthonormal. According to our previous statements about eigenfunctions, the maximum occurs when $u_i = v_i$, the ith eigenfunction of C.

Explained variance

Plugging the eigenfunctions in for u_k, we get a pretty clean way of quantifying the variance of X explained by the projections:

$$\mathrm{E}\,\|X\|^2 - \sum_{k=1}^{p} \mathrm{E}\langle X, v_k\rangle^2 = \sum_{k=1}^{\infty} \lambda_k - \sum_{k=1}^{p} \lambda_k = \sum_{k=p+1}^{\infty} \lambda_k.$$

The quantity above represents the *approximation error* between the infinite and finite-dimensional objects. The explained variance is then defined as

$$\frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k}.$$

Karhunen-Loéve expansion

Expanding X with respect to the v_i basis we have the KL expansion:

$$X = \mu + \sum_{k=1}^{\infty} \xi_k v_k \quad \text{where } \xi_k = \langle X - \mu, v_k \rangle,$$

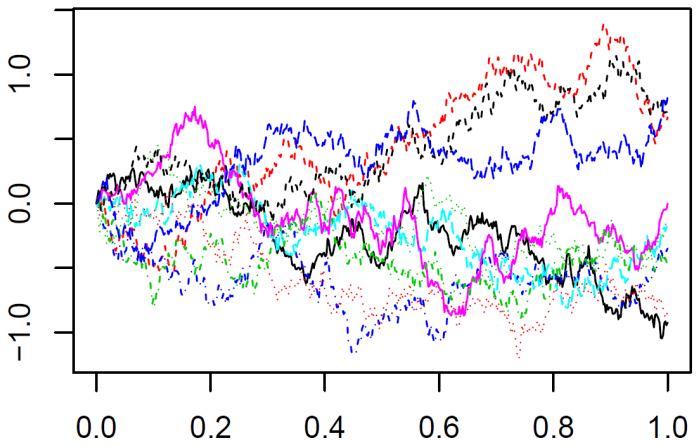
which is the foundation for FPCA. Recall that

$$\mathrm{E}[\xi_k \xi_j] = \lambda_k 1_{k=j}.$$

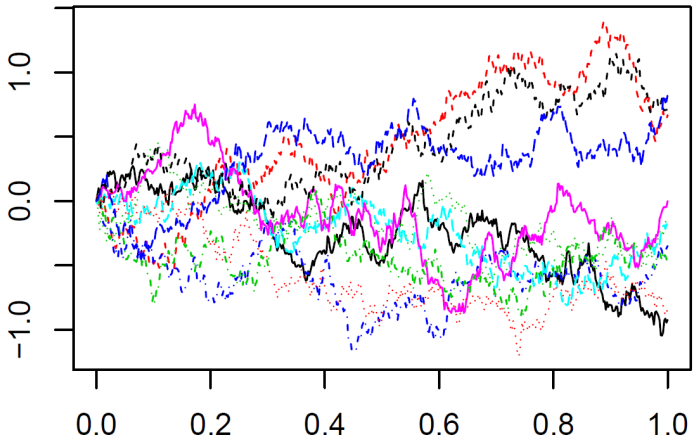
For Gaussian processes, what else can we say about the scores?

Example - Brownian motion

We say that $X \in L^2[0,1]$ is a Brownian motion if it is Gaussian with mean function 0 and covariance function $\mathrm{Cov}(X(t), X(s)) = \min\{t, s\}$.



Example - Brownian motion



Later, you can generate random functions from a given covariance function.

Properties of Brownian motion

Below are several properties of Brownian motion

- $X(0) = 0$,
- For $s \leq t$, $X(t) - X(s)$ is normal with mean 0 and variance $t - s$,
- For $t_1 \leq \cdots \leq t_k$, $\{X(t_k) - X(t_{k-1})\}$ are independent.

The above can be taken as the definition of Brownian motion.

Brownian motion - KL expansion

One can explicitly calculate the eigenfunctions and values for Brownian motion:

$$v_j(t) = \sqrt{2}\sin\left((j-1/2)\pi t\right) \qquad \lambda_j = \frac{1}{(j-1/2)^2\pi^2}.$$

This means that

$$X(t) = \sum_{j=1}^{\infty} \frac{1}{(j-1/2)\pi} Z_j \sqrt{2}\sin\left((j-1/2)\pi t\right),$$

where Z_j are iid standard normal. How can we show this?

Properties of Brownian motion

Show (by using integration by parts)

$$\int_0^1 v_k(s) \min(t,s)\ ds = \frac{1}{(j-1/2)^2\pi^2}\sqrt{2}\sin\left((j-1/2)\pi t\right)$$

Proof

By definition, we want to show that

$$\lambda_k v_j(t) = \int_0^1 v_k(s) \min(t,s)\ ds = \int_0^t v_j(s) s\ ds + t \int_t^1 v_k(s) t\ ds.$$

The second term is a bit easier

$$\begin{aligned} & t\sqrt{2} \int_t^1 \sin\left((j-1/2)\pi s\right)\ ds \\ &= -t\sqrt{2} \frac{\cos\left((j-1/2)\pi\right) - \cos\left((j-1/2)\pi t\right)}{(j-1/2)\pi} \\ &= \frac{t\sqrt{2}}{(j-1/2)\pi} \cos\left((j-1/2)\pi t\right). \end{aligned}$$

Proof

We can handle the second term using integration by parts:

$$\begin{aligned}
&\sqrt{2}\int_0^t s\sin\left((j-1/2)\pi s\right)\ ds \\
&= \sqrt{2}\left[\frac{-s\cos\left((j-1/2)\pi s\right)}{(j-1/2)\pi}\right]_0^t - \sqrt{2}\int_0^t \left[\frac{-\cos\left((j-1/2)\pi s\right)}{(j-1/2)\pi}\right]\ ds \\
&= \sqrt{2}\frac{-t\cos\left((j-1/2)\pi t\right)}{(j-1/2)\pi} + \sqrt{2}\left[\frac{\sin\left((j-1/2)\pi s\right)}{(j-1/2)^2\pi^2}\right]_0^t.
\end{aligned}$$

The first piece above cancels with our previous calculation and we get that

$$\int_0^1 v_k(s)\min(t,s)\ ds = \frac{1}{(j-1/2)^2\pi^2}\sqrt{2}\sin\left((j-1/2)\pi t\right),$$

as desired.

Non-differentiability of BM

Notice that

$$\mathrm{E}\,\|X\|^2 = \sum \frac{1}{(j-1/2)^2\pi^2} < \infty,$$

but if we try to take a derivative, then we get something like

$$\mathrm{E}\,\|X'\|^2 = C \sum \frac{(j-1/2)^2\pi^2}{(j-1/2)^2\pi^2} = \infty.$$

Example - Brownian Bridge

Recall that if $X(t)$ is a Brownian motion, then $B(t) = X(t) - tX(1)$ is a Brownian bridge. The covariance is given by

$$\mathrm{Cov}(B(t), B(s)) = s(1-t) \quad s \leq t.$$

The eigenvalues and functions of the covariance are

$$v_j(t) = \sqrt{2}\sin(j\pi t) \qquad \lambda_j = \frac{1}{j\pi}.$$

Example: BM v.s. BB