Dirichlet Sobolev Spaces on $[0, 1]$
and the Brownian Bridge

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA
Version: February 11, 2014, 15:56

1 Motivation: Donsker’s Theorem

If $\{X_i\}$ are iid random variables with common CDF

$$F(x) := P[X_i \leq x],$$

then the empirical CDF

$$F_n(x) := \frac{\#\{i \leq n : X_i \leq x\}}{n}$$

converges pointwise almost-surely to $F(x)$ by the strong law of large numbers. For finite $n$ the probability distribution of

$$n F_n(x) := \#\{i \leq n : X_i \leq x\}$$

is $\mathcal{B}(F(x), n)$, so also

$$\sqrt{n} \left[ F_n(x) - F(x) \right] \Rightarrow \mathcal{N}(0, F(x)[1 - F(x)])$$

by the central limit theorem. This is helpful but not quite enough to help us study the probability distribution of functionals like the Kolmogorov-Smirnov statistic

$$D_n := \sup_{-\infty < x < \infty} |F_n(x) - F(x)|, \quad (1)$$

a commonly used measure of the distance between the actual distribution of $\{X_i\}$ and the distribution with CDF $F(x)$. Independently Kolmogorov (1933) and Smirnov (1939) showed that the distribution of $D_n$ is the same for all possible continuous CDFs $F$. To prove this, replace $X_i \sim F(x)$ with $U_i := F(X_i) \sim \mathcal{U}(0, 1)$ and note $F_n(x) = G_n(F(x))$, where $G_n$ is the empirical CDF for uniforms:

$$G_n(u) := \frac{\#\{i \leq n : U_i \leq u\}}{n}, \quad \{U_i\} \overset{iid}{\sim} \mathcal{U}(0, 1). \quad (2a)$$

Later Donsker (1952), building on work of Doob (1949), showed that the probability distribution of the function-valued random variable

$$\Delta_n(t) := \sqrt{n} \left[ G_n(t) - t \right], \quad 0 \leq t \leq 1 \quad (2b)$$
converged to that of the “Brownian Bridge” $B(t)$, a mean-zero Gaussian stochastic process on the unit interval that we will construct below, and hence that for any continuous CDF $F(\cdot)$, $\Delta_n(t)$ converges in distribution to $B(t)$ and $\sqrt{n}D_n$ to $\sup_{0 < t < 1} |B(t)|$.

For $0 < s < t < 1$ the random vector $Y = (Y_1, Y_2, Y_3)$ with $Y_1 = nG_n(s)$, $Y_2 = n[G_n(t) - G_n(s)]$, and $Y_3 = n[1 - G_n(t)]$ has the multinomial distribution $Y \sim MN(n; p)$ with probabilities $p_1 = s$, $p_2 = (t - s)$, and $p_3 = (1 - t)$, so $V(Y_1) = ns(1 - s)$, $\text{Cov}(Y_1, Y_2) = -ns(t - s)$, and $\Delta(t)$ has covariance

$$\mathbb{E}\Delta_1 \Delta_t = \frac{1}{n} \text{Cov}(Y_1, Y_1 + Y_2) = s(1 - s) - s(t - s) = s(1 - t), \quad 0 < s < t < 1$$

or, for any $0 < s, t < 1$, $\mathbb{E}\Delta_s \Delta_t = (s \wedge t)(1 - s \vee t) = s \wedge t - st$. Now we consider the properties of a mean-zero Gaussian process with that covariance.

2 The Brownian Bridge

The Brownian Bridge is a stochastic process $B_t$ defined on the unit interval $[0, 1]$ and characterized in any of the following ways:

- A continuous-path Gaussian process with mean zero and covariance $\gamma(s, t) = \mathbb{E}B_s B_t$ given by

$$\gamma(s, t) := s \wedge t - st \quad (3a)$$

- A Diffusion with initial value $B_0 = 0$, drift $\alpha_t(x) = \frac{x}{1 - t}$, and diffusion rate $\beta^2_t(x) = 1$;

- A process related to Brownian Motion $W(t)$ on $\mathbb{R}_+$ by the relations

$$B_t = (1 - t)W\left(\frac{t}{1 - t}\right) \quad W_s = (1 + s)B\left(\frac{s}{1 + s}\right) \quad (3b)$$

- A process related to Brownian Motion $W(t)$ on $[0, 1]$ by the relation

$$B_t = W(t) - tW(1)$$

- A process related to Brownian Motion $W(t)$ on $[0, 1]$ by the relation

$$B_t \sim W(t) \mid W(1) = 0. \quad (3c)$$

We will construct it quite explicitly in Section (3.1) below, so for now we can view these as heuristic and see where Eqn (3a) leads us.

Denote by $\mathcal{H}_0$ the real Hilbert space of all square-integrable real-valued Borel functions on the unit interval, with inner product

$$\langle f, g \rangle_0 := \int f(t)g(t) \, dt \quad (4)$$

(where this and all similar unlabeled integrals below are taken over the unit interval $[0, 1]$) and norm

$$|f|_0 := \langle f, f \rangle_0^{1/2}.$$
All continuous functions on the compact set \([0,1]\) are bounded and so are in \(H_0\), and the space \(C_c^\infty\) of smooth compactly-supported functions \(\phi: (0,1) \to \mathbb{R}\) is dense in \(H_0\) (so is the larger space \(C([0,1])\) of continuous functions, of course).

Let \(t \sim B(t)\) be a “Brownian Bridge”, a normally-distributed random element of \(C([0,1])\) with mean \(\mathbb{E}B(t) = 0\) and covariance \(\mathbb{E}B(s)B(t) = \gamma(s,t)\) of Eqn (3a). We are now going to define a “generalized random process” or “random field” based on \(B\), a linear transformation that takes elements \(\phi \in C_c^\infty\) into the Gaussian random variables

\[
B[\phi] := \int \phi(t)B(t)dt = \langle \phi, B \rangle_0
\]

by integrating against the Brownian Bridge \(B(t)\). These each have mean zero, and covariance which we will denote by

\[
\langle \phi, \psi \rangle_{-1} := \mathbb{E}B[\phi]B[\psi] = \int \int \phi(s)\gamma(s,t)\psi(t)dsdt.
\]

Let \(H_{-1}\) denote the Hilbert space completion of \(C_c^\infty\) in the inner product \(\langle \cdot, \cdot \rangle_{-1}\), with norm

\[
|\phi|_{-1} := \langle \phi, \phi \rangle_{-1}^{1/2}.
\]

Evidently we may write the \(H_{-1}\) inner product in the form

\[
\langle \phi, \psi \rangle_{-1} = \int \phi(s) \left\{ \int \gamma(s,t)\psi(t)dt \right\} ds
= \langle \phi, \mathcal{G}\psi \rangle_0,
\]

where “\(\mathcal{G}\)” denotes the linear operator given by

\[
\mathcal{G}\psi(s) := \int \gamma(s,t)\psi(t)dt
= \int_0^s t(1-s)\psi(t)dt + \int_s^1 s(1-t)\psi(t)dt.
\]

It will prove useful to explore properties of the operator \(\psi \sim \mathcal{G}\psi\). First, note from Eqn (8) that for any \(\psi \in H_0\) the function \(g(s) := \mathcal{G}\psi(s)\) satisfies the Dirichlet boundary conditions \(g(0) = 0\) and \(g(1) = 0\). It is also continuous and differentiable (even if \(\psi\) wasn’t), with

\[
g'(s) = s(1-s)\psi(s) - \int_0^s t\psi(t)dt - s(1-s)\psi(s) + \int_s^1 (1-t)\psi(t)dt
= -\int_0^s t\psi(t)dt + \int_s^1 (1-t)\psi(t)dt
\]

for every \(s\) and, for any point \(s\) where \(\psi\) is continuous,

\[
g''(s) = -s\psi(s) - (1-s)\psi(s) = -\psi(s).
\]

An “eigenfunction” for the operator \(\mathcal{G}\) is a function \(\psi\) for which there is some constant \(\lambda \in \mathbb{C}\) such that \(\mathcal{G}\psi(s) = \lambda \psi(s)\) for all \(s\), similar to eigenvectors for matrices. Since \(\langle \psi, \psi \rangle_{-1} = \langle \psi, \mathcal{G}\psi \rangle_0 = \)
\[ \lambda \langle \psi, \psi \rangle_0 > 0, \text{ necessarily } \lambda > 0. \] Such an eigenfunction must satisfy \( \psi(0) = 0 \) and \( \psi(1) = 0 \), since \( g = \mathfrak{g} \psi \) does, and \( \lambda \psi'' = -\psi \). The solutions to \( \lambda \psi'' = -\psi \) for \( \lambda > 0 \) are all of the form \( \psi(t) \propto \sin(\alpha t + b) \) with \( \alpha^2 = 1/\lambda \) and \( b \in \mathbb{R} \) arbitrary. For \( \psi(0) = 0 \) we need \( b = 0 \), while for \( \psi(1) = 0 \) we need \( a \in \mathbb{N} \), and hence \( \lambda = (n^2 \pi^2)^{-1} \) for some \( n \in \mathbb{N} \). The solutions with unit \( \mathcal{H}_0 \) norm are:

\[ \psi_n(t) := \sqrt{2} \sin(n\pi t). \] (10)

In fact, these form a complete orthonormal system in \( \mathcal{H}_0 \): \( \langle \psi_n, \psi_m \rangle_0 \) is zero if \( n \neq m \) and one if \( n = m \) (so \( |\psi_n|_0 = 1 \)), and every element \( \phi \in \mathcal{H}_0 \) has a unique convergent expansion

\[ \phi = \sum_{n \in \mathbb{N}} a_n \psi_n \]

with coefficients given by

\[ a_n = \langle \phi, \psi_n \rangle_0 = \int \phi(t) \psi_n(t) \, dt, \]

called the Fourier sine series. To see this, start with the trig identities \( \cos(x \pm y) = \cos(x) \cos(y) \mp \sin(x) \sin(y) \) (which follow easily from Euler’s formula \( \exp(i\theta) = \cos(\theta) + i \sin(\theta) \)), we find that \( 2 \sin(x) \sin(y) = \{\cos(x - y) - \cos(x + y)\} \) and so for \( n, m \in \mathbb{N} \),

\[ \langle \psi_n, \psi_m \rangle_0 = \int 2 \sin(n\pi t) \sin(m\pi t) \, dt \]

\[ = \int \{\cos[(n - m)\pi t] - \cos[(n + m)\pi t]\} \, dt \]

\[ = 1 \text{ if } n = m, \text{ and } 0 \text{ if } n \neq m. \]

Now comes the fun part. If we expand each of two elements \( \phi, \psi \) of \( \mathcal{H}_0 \) in their Fourier sine series

\[ \phi = \sum_{n \in \mathbb{N}} a_n \psi_n \quad \psi = \sum_{m \in \mathbb{N}} b_m \psi_m, \]

then multiply these and integrate, we see that both their inner product and their norms

\[ \langle \phi, \psi \rangle_0 := \int \phi(t) \psi(t) \, dt \]

\[ = \sum_{n \in \mathbb{N}} \sum_{m \in \mathbb{N}} a_n b_m \int \psi_n(t) \psi_m(t) \, dt \]

\[ = \sum_{n \in \mathbb{N}} a_n b_n \]

\[ |\phi|_0 = \langle \phi, \phi \rangle_0^{1/2} = \left\{ \sum_{n \in \mathbb{N}} |a_n|^2 \right\}^{1/2} \]
can be evaluated in terms of the Fourier coefficients. Moreover, since $\mathcal{G}\psi_n = (n^2 \pi^2)^{-1} \psi_n,$
\[
\mathcal{G}\psi = \sum_{m \in \mathbb{N}} b_m \mathcal{G}\psi_m
\]
\[
= \sum_{m \in \mathbb{N}} b_m (m^2 \pi^2)^{-1} \psi_m
\]

and so the $\mathcal{H}_{-1}$ inner product (and covariance) can also be computed from the coefficients, as
\[
\langle \phi, \psi \rangle_{-1} = \mathbb{E}[B[\phi] B[\psi]] \\
= \langle \phi, \mathcal{G}\psi \rangle_0 \\
= \sum_{n \in \mathbb{N}} (\pi^2 n^2)^{-1} a_n b_n.
\]

3 Dirichlet Sobolev Spaces

We have seen two inner products on spaces of functions on the unit interval, each expressible in terms of the Fourier sine coefficients:
\[
\langle \phi, \psi \rangle_0 = \int \phi(t) \psi(t) \, dt = \sum_{n \in \mathbb{N}} a_n b_n
\]
\[
\langle \phi, \psi \rangle_{-1} = \int \int \phi(s) \gamma(s, t) \psi(t) \, ds \, dt = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^{-1} a_n b_n
\]

This suggests a generalization:

**Definition 1** (Sobolev Space). For $s \in \mathbb{R}$, let $\mathcal{H}_s$ be the Hilbert-space completion of $C_c^\infty$ in the inner product
\[
\langle \phi, \psi \rangle_s = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^s a_n b_n
\]

and norm
\[
|\phi|_s = \langle \phi, \phi \rangle_s^{1/2} = \left\{ \sum_{n \in \mathbb{N}} (\pi^2 n^2)^s |a_n|^2 \right\}^{1/2},
\]

where $a_n = \langle \phi, \psi_n \rangle_0 = \int \phi(t) \psi_n(t) \, dt$ and $b_m = \langle \psi, \psi_m \rangle_0 = \int \psi(t) \psi_m(t) \, dt$.

These are the “Dirichlet Sobolev spaces” of order $s$ on the unit interval $[0, 1]$. Roughly speaking, as we’ll see below, they are the spaces of of $s$-times differentiable functions on the interval that vanish at the end-points. We also define
\[
\mathcal{H}_\infty = \bigcap_{s \in \mathbb{R}} \mathcal{H}_s \\
\mathcal{H}_{-\infty} = \bigcup_{s \in \mathbb{R}} \mathcal{H}_s;
\]
each of these is a complete separable metric space (a Fréchet space, in fact), but neither is a Hilbert or Banach space. The elements of $\mathcal{H}_\infty$, called “test functions”, are $C^\infty$ functions that vanish to all orders at the boundary of $[0,1]$, while $\mathcal{H}_{-\infty}$ is a collection of interesting objects called (alas) “distributions” by the mathematical community. We’ll look at these more in Section (3.2). For all $-\infty < s < t < \infty$ the spaces are nested:

$$C_c^\infty \subset \mathcal{H}_\infty \subset \mathcal{H}_t \subset \mathcal{H}_s \subset \mathcal{H}_{-\infty}$$

One can construct similar spaces with other boundary conditions, such as Neumann $\phi'(0) = \phi'(1) = 0$ (where $\psi_n(x) \propto \cos(n\pi x)$ for $n \in \mathbb{Z}_+ = \{0,1,2,\ldots\}$) or Periodic $\phi(0) = \phi(1)$, or (the most widely studied) Free boundary conditions on all of $\mathbb{R}$.

### 3.1 Sobolev Representation of the Brownian Bridge

If a Brownian Bridge $B(t)$ has continuous sample paths with mean zero and covariance function $\gamma(s,t)$, then its Fourier sine coefficients

$$\zeta_n = B[\psi_n] = \langle B, \psi_n \rangle_0 = \int B(t)\psi_n(t)\,dt$$

will be Gaussian random variables with mean zero and covariance (see Eqn (6))

$$\mathbb{E} \zeta_m \zeta_n = \langle \psi_m, \psi_n \rangle_{-1} = \langle \psi_m, \psi_n \rangle_0 = \langle \psi_m, \frac{1}{\pi n^2} \psi_n \rangle_0 = (\pi^2 n^2)^{-1}$$

if $m = n$, and 0 otherwise.

Thus $\{Z_n := n\pi \zeta_n\} \text{iid } \mathbb{N}(0,1)$ and we can recover $B(t)$ as the $L_2$-convergent sum

$$B(t) = \sum_{n \in \mathbb{N}} \frac{\sqrt{2}}{\pi n} Z_n \sin(n\pi t). \quad (14a)$$

We take this as our constructive definition of the Brownian Bridge (we’ll verify path continuity below). The sample paths lie in $\mathcal{H}_s$ almost surely for each $s < 1/2$, because

$$\mathbb{E}|B|^2_s = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^{2} \frac{2}{\pi^2 n^2} \mathbb{E} Z_n^2 \quad (14b)$$

$$= 2 \sum_{n \in \mathbb{N}} (\pi^2 n^2)^{2} \langle \psi \rangle_{-1}$$

$$< \infty \text{ for } s < 1/2.$$ 

By Eqns (14a, 14b) we may think of the Brownian Bridge $t \mapsto B(t)$ as a random element of $\mathcal{H}_s$ for any $s < 1/2$. Its (probability) distribution is the Borel measure $\mathbb{P}[B \in A]$ for Borel sets $A \subset \mathcal{H}_s$. By Eqn (11) we can also view $B$ as the “unit Gaussian process” on the Hilbert space $\mathcal{H}_{-1}$, i.e., a linear isomorphism $B : \mathcal{H}_{-1} \to L_2(\Omega, \mathcal{F}, \mathbb{P})$ from $\mathcal{H}_{-1}$ into a space of square-integrable random variables with the property that inner products are preserved:

$$\mathbb{E} \{B[\phi] \cdot B[\psi]\} = \langle \phi, \psi \rangle_{-1}$$
Using Eqn (3b) we can construct the Wiener process \( W(t) \) on all of \( \mathbb{R}_+ \) from \( B \) on \([0, 1]\); we can also construct \( W(t) \) on \([0, 1]\) in a manner similar to that above with mixed boundary conditions \( g(0) = 0 = g'(1) \), leading to

\[
W(t) = \sum_{n \in \mathbb{N}} \frac{\sqrt{2}}{\pi(n - 1/2)} Z_n \sin \left( (n - 1/2)\pi t \right).
\]

### 3.1.1 Testing \( H_0 : X_i \overset{iid}{\sim} F(x) \)

If \( \{X_j\} \overset{iid}{\sim} F(x) \) for a continuous CDF \( F(x) \) we’ve seen that the empirical CDF \( G_n(u) \) for \( \{U_j := F(X_j)\} \) satisfies

\[
\Delta_n(t) := \sqrt{n} [G_n(t) - t] \approx B(t),
\]

approximately distributed as a Brownian Bridge for large \( n \). On the other hand, if the \( \{X_j\} \) are iid with some other continuous CDF \( \tilde{F} \) that is close to \( F \), write that CDF in the form

\[
\tilde{F}(x) = F(x) + \epsilon(F(x))/\sqrt{n}
\]

for some small function \( \epsilon : [0, 1] \to \mathbb{R} \) with \( \epsilon(0) = \epsilon(1) = 0 \). Under this alternate, the random function \( \Delta_n(t) \) is approximately distributed as a Gaussian process with mean \( \epsilon(t) \) and with covariance

\[
\tilde{\gamma}(s, t) = [s + \epsilon(s)/\sqrt{n}] \wedge [t + \epsilon(t)/\sqrt{n}] - [s + \epsilon(s)/\sqrt{n}] [t + \epsilon(t)/\sqrt{n}], = \gamma(s, t) + O(n^{-1/2})
\]

which will be close to the Brownian Bridge’s \( \gamma(s, t) \) if \( \epsilon \) is small or \( n \) is large. For \( \epsilon \in \mathcal{H}_1 \), the Fourier sine coefficients would then be distributed approximately as

\[
\zeta_j := \langle \Delta_n, \psi_j \rangle_0 \overset{ind}{\sim} \mathcal{N}(a_j, \frac{1}{j^2 \pi^2})
\]

where \( \{a_j := \langle \epsilon, \psi_j \rangle_0\} \) are the coefficients of \( \epsilon(t) \). For large sample size \( n \), the likelihood ratio against \( H_0 : X_i \overset{iid}{\sim} F(x) \) on the basis of the first \( J \) coefficients is then:

\[
\Lambda_n = \prod_{j \leq J} \left( \frac{\pi^2 j^2}{2 \pi} \right)^{1/2} \exp \left\{ -\pi j^2 (\zeta_j - a_j)^2 / 2 \right\}
\]

\[
= \exp \left\{ \sum_{j \leq J} \left( \frac{\pi^2 j^2}{2 \pi} \right)^{1/2} \exp \left\{ -n(\pi j)^2 (\zeta_j)^2 / 2 \right\} \right\}
\]

\[
\rightarrow \exp \left\{ \langle \Delta_n, \epsilon \rangle_1 - \frac{1}{2} \| \epsilon \|^2_1 \right\} \quad \text{as} \ J \to \infty.
\]

Thus the Neyman-Pearson likelihood ratio test of \( H_0 : \{X_i\} \sim F \) against \( H_1 : \{X_i\} \sim \tilde{F} \) would reject \( H_0 \) for large values of

\[
\langle \Delta_n, \epsilon \rangle_1 = n^{-1/2} \sum_{j \leq n} \epsilon'(F(X_j)).
\]
3.2 More about Sobolev Spaces

Because $\psi_n(t) := \sqrt{2} \sin(n\pi t)$ satisfies $\psi_n''(t) = -(n^2\pi^2)\psi_n(t)$, the negative second derivative $-\Delta \psi$ of any function $\psi \in C_c^\infty$ with Fourier coefficients $b_n = \langle \psi, \psi_n \rangle_0$ is

$$-\Delta \psi = \sum_{n \in \mathbb{N}} b_n [-\Delta \psi_n]$$
$$= \sum_{n \in \mathbb{N}} b_n (n^2\pi^2)\psi_n,$$

so by Eqn (12) the $\mathcal{H}_1$ inner product of $\phi \in C_c^\infty$ with Fourier sine coefficients $\{a_n\}$ with $\psi$ is

$$\langle \phi, \psi \rangle_1 = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^1 a_n b_n$$
$$= \langle \phi, -\Delta \psi \rangle_0$$
$$= -\int \phi(s) \psi''(s) ds; \text{ integrating by parts},$$
$$= -\phi(s)\psi'(s)|^1_0 + \int \phi'(s) \psi'(s) ds$$
$$= \langle \phi', \psi' \rangle_0. \quad (15)$$

Thus $\mathcal{H}_1$ consists precisely of those continuous functions that vanish at the boundary of $[0, 1]$ whose first derivative is square-integrable. Similarly $\mathcal{H}_2$ consists of functions whose second derivative is in $\mathcal{H}_0 = L_2$ and, for all integers $s > 0$, $\mathcal{H}_s$ consists of $s$-times differentiable functions. For any $-\infty < s < \infty$ can write

$$\langle \phi, \psi \rangle_s = \langle \phi, (-\Delta)^s \psi \rangle_0 \quad (16)$$
for the fractional derivative operator $(-\Delta)^s$ given by

$$(-\Delta)^s \phi(s) = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^s \langle \phi, \psi_n \rangle_0 \psi_n(s).$$

In particular, for $s = -1$, we see that

$$\langle \phi, \psi \rangle_{-1} = \langle \phi, (-\Delta)^{-1} \psi \rangle_0$$

By Eqn (7) we already had

$$\langle \phi, \psi \rangle_{-1} = \langle \phi, \mathcal{G} \psi \rangle_0,$$

so we now recognize

$$\mathcal{G} \psi(s) := \int \gamma(s, t)\psi(t) dt$$
$$= (-\Delta)^{-1} \psi(s) \quad (17)$$

as the solution $\mathcal{G} \psi = g$ to the equation $(-\Delta)g = \psi$, and $\mathcal{G} = (-\Delta)^{-1}$ is the inverse Laplace operator. More specifically, it is the Dirichlet inverse Laplacian that gives the solution $g = \mathcal{G} \phi$
of \(-\Delta g = \phi\) that satisfies \(g(0) = g(1) = 0\). Adding any linear function \(g(x) + a + bx\) would give another solution of \(-\Delta g = \phi\) with different BC.

The operators \((-\Delta)\) and \(\Psi\) are continuous isomorphisms of \(\mathcal{H}_s\) onto \(\mathcal{H}_{s-2}\) and \(\mathcal{H}_{s+2}\), respectively, for every \(s \in \mathbb{R}\). The function \(\gamma(s, t)\) is called the Dirichlet “Green’s function” for \((-\Delta)\), a solution with Dirichlet boundary conditions to the formal equation \(-\Delta g(s, t) = \delta(s - t)\) for the Dirac delta function (more on that below).

### 3.2.1 Duality
For each \(s > 0\) there is a natural duality between \(\mathcal{H}_s\) and \(\mathcal{H}_{-s}\): elements of \(\mathcal{H}_{-s}\) may be viewed as continuous linear functionals on \(\mathcal{H}_s\), so \(\mathcal{H}_{-s} = \mathcal{H}_s^*\) and \(\mathcal{H}_s = (\mathcal{H}_{-s})^*\) as Banach spaces. The pairing of a linear functional \(\xi\) with coefficients \(b_n = \xi[\psi_n]\) with an element \(\phi \in \mathcal{H}_s\) with coefficients \(\{a_n\}\) is given by

\[
\langle \phi, \xi \rangle = \sum_{n \in \mathbb{N}} a_n b_n = \sum_{n \in \mathbb{N}} (n^2 \pi^2)^s a_n \frac{1}{n^2 \pi^2} b_n = \langle (-\Delta)^s \phi, (-\Delta)^{-s} \xi \rangle_0
\]

that satisfies

\[
|\langle \phi, \xi \rangle| \leq |\phi|_s |\xi|_{-s}.
\]

Of course \(\mathcal{H}_s\), a Hilbert space, is also self-dual in the sense that each continuous linear functional \(\xi\) is given by \(\xi[\phi] = \langle \phi, \psi \rangle_s\) for some \(\psi \in \mathcal{H}_s\), with coefficients \(\{c_n\}\). The connection is given by

\[
\langle \phi, \psi \rangle_s = \sum_{n \in \mathbb{N}} (n^2 \pi^2)^s a_n c_n = \sum_{n \in \mathbb{N}} a_n b_n = \langle \phi, \xi \rangle,
\]

so \(b_n = (n^2 \pi^2)^s c_n\) and \(\xi = (-\Delta)^s \psi\).

### 3.2.2 Examples
Here we look at some examples of functions and distributions, and see in which spaces \(\mathcal{H}_s\) they lie.

#### Constants
The constant function \(\phi(s) \equiv 1\), for example, has coefficients

\[
a_n = \int \phi(t) \sqrt{2} \sin(n \pi t) \, dt = \frac{-\sqrt{2}}{n \pi} \cos(n \pi t) \bigg|_0^1 = 2\sqrt{2}/n \pi \text{ for odd } n, \text{ and } 0 \text{ for even } n
\]
and so, by Eqn (13), squared $\mathcal{H}_s$ norm

$$
|\phi|^2 = \sum_{\text{odd } n} (n^2 \pi^2)^s \frac{8}{\pi^2 n^2} = 8 \sum_{\text{odd } n} (n^2 \pi^2)^{(s-1)} < \infty \text{ if and only if } s < 1/2.
$$

The function $\phi(t) \equiv 1$ is as smooth as possible, but doesn’t vanish at $t = 0$ and $t = 1$ and so it won’t be in all of the $\{\mathcal{H}_s\}$. Every element $\psi \in \mathcal{H}_s$ with $s > 1/2$ must be continuous and satisfy those boundary conditions, because then

$$
|\psi|^2 = \sum_{n \in \mathbb{N}} (n^2 \pi^2)^s |a_n|^2 < \infty
$$

and so, by the Cauchy-Schwartz inequality (see Prop. (1) on p. 25),

$$
\sum_{n \in \mathbb{N}} |a_n| = \sum_{n \in \mathbb{N}} |a_n|(n\pi)^s(n\pi)^{-s} \leq \left\{ \sum_{n \in \mathbb{N}} |a_n|^2 (n^2 \pi^2)^s \right\}^{1/2} \left\{ \sum_{n \in \mathbb{N}} (n^2 \pi^2)^{-s} \right\}^{1/2} \leq |\phi|_s \pi^{-s} (1 - 1/2s)^{-1/2} < \infty
$$

for $s > 1/2$. It follows that

$$
\phi(t) = \sum_{n \in \mathbb{N}} a_n \sqrt{2} \sin(n\pi t)
$$

converges uniformly to a continuous function $\phi$ which, like each $\sin(n\pi t)$, must vanish at $t = 0$ and $t = 1$. 

10
Splines

A spline is a piecewise-polynomial function of some specified degree $d$, whose first $(d - 1)$ derivatives are continuous at the \textquotedblleft knots\textquotedblright{} (the boundaries of the intervals on which the function has polynomial form). They pop up a lot in Sobolev theory and function approximation. The simplest non-trivial one would be a linear spline like the tent function

$$h(t) = t \wedge (1 - t),$$

a continuous function that rises linearly from zero to a maximum of $h\left(\frac{1}{2}\right) = 1/4$, then falls linearly to $h(1) = 0$. Its Fourier coefficients $a_n = \int h(t) \psi_n(t) \, dt$ vanish for even $n$ (by symmetry) and for odd $n$ are $a_n = (-1)^{(n-1)/2} \sqrt{8/n \pi}$, so its squared Sobolev norm

$$\|h\|_s^2 = 8 \sum_{\text{odd } n} (\pi^2 n^2)^{s-2}$$

is finite for $s < 3/2$— e.g., $|h|_1 = |h'|_0 = 1$. Every linear spline vanishing at the boundary is in $\mathcal{H}_s$ for $s < 3/2$; every quadratic (or cubic) spline vanishing to order two (or three) at the boundary is in $\mathcal{H}_s$ for $s < 5/2$ (or $s < 7/2$), respectively, and so forth.

The derivative $h'$ of a linear spline $h$ is piecewise-constant and orthogonal to the constant function $1$, to ensure that the boundary condition $h(1) = 0$ is met for the function $h(t) = \int_0^t h'(x) \, dx$. Similarly the $d$th derivative of a rank-$d$ spline is piecewise-constant and orthogonal to polynomials of degree $\leq d$; in fact that is a succinct characterization of all splines on $[0,1]$ with Dirichlet boundary conditions.

Dirac Delta Functions

For $t_* \in (0,1)$ the evaluation functional

$$\delta_{t_*} [\phi] := \phi(t_*), \quad (18)$$

has coefficients

$$a_n = \phi_n(t_*) = \langle \delta_{t_*}, \psi_n \rangle_0 = \sqrt{2} \sin(n \pi t_*)$$

that satisfy $|a_n| \leq \sqrt{2}$ for all $n$, and hence

$$\|\delta_{t_*}\|_s^2 = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^s |a_n|^2$$

$$\leq 2 \sum_{n \in \mathbb{N}} (\pi^2 n^2)^s$$

$$< \infty \text{ for } s < -1/2.$$  

This functional, called the \textquotedblleft Dirac delta distribution,\textquotedblright{} is a well-defined object in $\mathcal{H}_s$ for $s < -1/2$ but is not a function like the elements of $\mathcal{H}_s$ for $s \geq 0$. By the same reasoning, every finite measure $\mu$ on $[0,1]$ is in $\mathcal{H}_s$ for $s < -1/2$— for example,

$$\|\mu\|_{-1}^2 = \langle \mu, \mu \rangle_{-1} = \int \int \mu(ds) \gamma(s,t) \mu(dt) \leq \mu([0,1])^2/4 < \infty.$$
Derivatives of the Dirac Delta Function

For \( t_s \in (0, 1) \) the derivative evaluation functional
\[
\delta_{t_s}^{I} [\phi] := \phi'(t_s)
\]
has coefficients
\[
a_n = \phi_n'(t_s) = \langle \delta_{t_s}^{I}, \psi_n \rangle_0 = n\pi \sqrt{2} \cos(n\pi t_s)
\]
that satisfy \(|a_n| \leq n\pi \sqrt{2}\) for all \( n \), and hence
\[
\|\delta_{t_s}^{I}\|^2 = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^s |a_n|^2
\]
\[
\leq 2 \sum_{n \in \mathbb{N}} (\pi^2 n^2)^{(s+1)}
\]
\[
< \infty \quad \text{for} \quad s < -3/2.
\]

This distribution is in \( \mathcal{H}_s \) for \( s < -3/2 \) but is not even a measure (i.e., is not continuous as a linear functional on \( \mathcal{C}([0,1]) \)). In a similar way one can construct derivatives \( \delta_{t_s}^{[k]} \) in \( \mathcal{H}_s \) for \( s < -k - 1/2 \) of all orders \( k \in \mathbb{N} \) or, similarly, derivatives of any function \( \phi \) or finite measure \( \mu \).

**Sobolev’s Lemma**

For integers \( k \geq 0 \) denote by \( \mathcal{C}_0^k \) the linear space of continuous functions \( f(t) \) that vanish to order \( k \) at the boundary \( t \in [0,1] \), i.e., that satisfy
\[
\lim_{t \to 0^+} t^{-k} |f(t)| = 0 \quad \text{and} \quad \lim_{t \to 1^-} (1 - t)^{-k} |f(t)| = 0
\]

For \( k = 0 \) the space \( \mathcal{C}_0^0 \) consists of all the continuous functions that take the value \( f(0) = f(1) = 0 \), but to be in \( \mathcal{C}_0^k \) for \( k > 0 \) both \( f(t) \) and \( f(1 - t) \) must be \( o(t^k) \) at \( t = 0 \). What is the relation of these functions, with \( k \) continuous derivatives, and \( \mathcal{H}_k \), whose \( k \)th derivative is in \( L_2 \)? Sobolev’s Lemma gives the answer:

**Theorem 1** (Sobolev’s Lemma). For each integer \( k \geq 0 \) and number \( s > k + 1/2 \), the space \( \mathcal{C}_0^k \) of \( k \)-times differentiable functions \( \phi \) that vanish to order \( k \) at the boundary \( \partial [0,1] \) and the Sobolev spaces satisfy
\[
\mathcal{H}_s \subset \mathcal{C}_0^k \subset \mathcal{H}_k
\]
with each inclusion continuous.

So, every element of \( \mathcal{H}_s \) for \( s > 1/2 \) is a continuous function; every element of \( \mathcal{H}_s \) for \( s > 3/2 \) is continuously differentiable, and so on.

**Proof.** First, to illustrate the idea, here’s a simple proof of a weaker result, that \( \mathcal{H}_1 \subset \mathcal{C}_0^0 \subset \mathcal{H}_0 \). For \( f \in \mathcal{C}_0^\infty \) and \( t \in [0,1] \), \( f(t) = \int_0^t f'(x) \, dx = \langle f', 1_{[0,x] \leq t} \rangle_0 \), so the Cauchy-Schwarz inequality \( E[XY] \leq \|X\|_2 \|Y\|_2 \) (see p. 25) gives:
\[
|f(t)| = \left| \int_0^t f'(x) \, dx \right| \leq \left\{ \int_0^1 f'(x)^2 \, dx \right\}^{1/2} \left\{ \int_0^1 \left(1_{[0,x] \leq t}\right)^2 \, dx \right\}^{1/2} = |f|_1 \|t\|^{1/2}
\]
for every $0 < t < 1$, and so $\sup_{t} |f(t)| \leq |f|_1$. Thus the embedding of $C^\infty_c \subset \mathcal{H}_1$ (with the $\mathcal{H}_1$ norm) into $C^0_0$ (in the sup norm) is continuous, and can be continuously extended to all of $\mathcal{H}_1$, and $\mathcal{H}_1 \subset C^0_0$ as claimed.

Conversely, each element $f \in C^0_0$ is continuous on the compact set $[0,1]$ and so is bounded there by $\sup |f(t)|$ and Borel measurable, so $f \in H_0 = L_2$ with norm $|f|_0 \leq \sup |f(t)|$ so also $C^0_0 \subset H_0$.

\[\]

**Proof.** Now, using a little more Sobolev technology, we can show the stronger result that $\mathcal{H}_s \subset C^0_0$ for any $s > \frac{1}{2}$. First calculate the Fourier coefficients for the function $1_{[x \leq t]}$:

\[
c_n := \langle 1_{[x \leq t]}, \psi_n \rangle_0 = \int_0^t \sqrt{2} \sin(n\pi x) \, dx = \sqrt{2} [1 - \cos(n\pi t)] / n\pi \quad \text{so} \quad |c_n|^2 \leq 8/n^2\pi^2.
\]

Now, express $f \in C^\infty_c \subset \mathcal{H}_s$ for $s > \frac{1}{2}$ and its derivative $f' \in \mathcal{H}_{s-1}$ in Fourier form

\[
f(t) = \sum_{n \in \mathbb{N}} a_n \sqrt{2} \sin(n\pi t)
\]
\[
f'(t) = \sum_{n \in \mathbb{N}} b_n \sqrt{2} \sin(n\pi t)
\]

with coefficients $\{a_n = \langle f, \psi_n \rangle_0\}$ and $\{b_n = \langle f', \psi_n \rangle_0\}$ satisfying

\[
\infty > |f|^2_s = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^s |a_n|^2
\]
\[
= |f'|^2_{(s-1)} = \sum_{n \in \mathbb{N}} (\pi^2 n^2)^{(s-1)} |b_n|^2.
\]

For any $0 < t < 1$, then, again by Cauchy-Schwarz and the duality of Section (3.2.1),

\[
|f(t)|^2 = |\langle f', 1_{[x \leq t]} \rangle_0|^2
\]
\[
= \left| \sum_{n \in \mathbb{N}} b_n c_n \right|^2
\]
\[
= \left| \sum_{n \in \mathbb{N}} [b_n (n\pi)^{s-1}] : [c_n (n\pi)^{1-s}] \right|^2
\]
\[
\leq \left\{ \sum_{n \in \mathbb{N}} |b_n|^2 (n^2 \pi^2)^{s-1} \right\} \left\{ \sum_{n \in \mathbb{N}} |c_n|^2 (n^2 \pi^2)^{1-s} \right\}
\]
\[
\leq |f'|^2_{(s-1)} \times \sum_{n \in \mathbb{N}} (8/n^2\pi^2)(n^2 \pi^2)^{1-s}
\]
\[
= 8 |f|^2_s \sum_{n \in \mathbb{N}} (n^2 \pi^2)^{-s}
\]

which is finite for $s > \frac{1}{2}$, so for some finite $c$ we have $\sup |f(t)| \leq c |f|^s_s$. \[\]

The extension from $C^0_0$ to $C^k_0$ is left to the reader.
3.3 Extensions

In some sense the Dirichlet Sobolev spaces \( \mathcal{H}_s \) are built around the positive definite operator \((-\Delta)\), the Dirichlet Laplacian on \( L_2([0,1]) \), and its inverse operator \( \mathcal{E} \). Similar spaces and theory can be built around other positive definite operators on other spaces \( L_2(T) \). For example, the negative Laplacian on \( \mathbb{R}^n \) with free boundary conditions is only semi-definite, but adding a positive constant \( m^2 \) will make it positive definite and one can build Sobolev spaces

\[
\mathcal{H}_s = \left\{ h : \int (m^2 + |x|^2)^s |\hat{h}(x)|^2 \, dx < \infty \right\}
\]

where \( \hat{h} \) is the Fourier transform of \( h \); for positive integers \( s = k \) these will be functions all of whose \( k \)-fold derivatives are in \( L_2(\mathbb{R}^d) \) or, for all \( s \), those for which \( (-\Delta + m^2)^s \hat{h} \in L_2 \). For \( d = 1 \) the unit Gaussian process on \( \mathcal{H}_{-1} \) is the Ornstein-Uhlenbeck (“OU”) velocity process, while in \( d > 1 \) it’s a special case of the Matérn family.

Another interesting choice is the circle \( S^1 \) (where instead of the Brownian Bridge we would construct closed Brownian Motion) or the sphere \( S^{d-1} \) in \( \mathbb{R}^d \) (there might be applications for this in astronomy or geology or meteorology with \( d = 3 \)) where again the Laplacian is only semi-definite (since constant functions are harmonic, \( i.e., \) satisfy \( \Delta h \equiv 0 \)), so again it’s necessary to take \( m > 0 \) for \( (-\Delta + m^2) \). In mathematics it’s conventional to take \( m = 1 \); in quantum field theory applications, \( m \) denotes the mass of a particle (hence the notation).

There is also a well-developed theory of \( L_p \) Sobolev space for \( p \neq 2 \), with some applications in the theory of differential equations and some use in wavelets, but the \( L_2 \) theory is simpler and more useful.

4 Reproducing Kernel Hilbert Spaces

For fixed \( 0 \leq s \leq 1 \) the function

\[
\gamma_s(t) := \gamma(s,t) = (s \land t) - st
\]

is continuous on \([0,1]\), and so is in \( \mathcal{H}_1 \) with coefficients

\[
a_n = \int \gamma_s(t) \psi_n(t) \, dt = \frac{1}{n^2 \pi^2} \sqrt{2} \sin(n\pi s)
\]

Expanding, we find the representation

\[
\gamma(s,t) = \sum_{n \in \mathbb{N}} \frac{2}{n^2 \pi^2} \sin(n\pi s) \sin(n\pi t).
\]

For any \( \phi \in \mathcal{H}_1 \) with coefficients \( \{ b_n \} \),

\[
\langle \gamma_s, \phi \rangle_1 = \sum_{n \in \mathbb{N}} (n^2 \pi^2)^{\frac{1}{2}} \frac{1}{n^2 \pi^2} \sqrt{2} \sin(n\pi s) b_n
\]

\[
= \sum_{n \in \mathbb{N}} b_n \psi_n(s)
\]

\[
= \phi(s),
\]

14
i.e., the $\mathcal{H}_1$ inner product of $\gamma_s$ with any $\phi \in \mathcal{H}_1$ is the evaluation functional $\phi(s)$—inner products with $\gamma_s$ “reproduce” elements of $\mathcal{H}_1$. For this reason $\gamma(\cdot, \cdot)$ is called a reproducing kernel on $\mathcal{H}_1$, and $\mathcal{H}_1$ itself is called a “reproducing kernel Hilbert space” or “RKHS”. This could be seen more easily (but more abstractly) from Eqns (16, 17) by writing

$$
\phi(s) = \Theta(-\Delta)\phi(s) = \langle \gamma_s, (-\Delta)\phi \rangle_0 = \langle \gamma_s, \phi \rangle_1.
$$

**Definition 2** (RKHS). A Reproducing Kernel Hilbert Space is a Hilbert space $\mathcal{H}$ of functions on some set $\mathcal{T}$ along with a positive-definite function $k : \mathcal{T} \times \mathcal{T} \to \mathbb{C}$ with the property that for each $s \in \mathcal{T}$ the function $k_s(t) := k(s, t)$ lies in $\mathcal{H}$ and satisfies

$$
\langle k_s, h \rangle = h(s)
$$

for every $h \in \mathcal{H}$ and $s \in \mathcal{T}$.

Not every Hilbert space of functions can be a RKHS—evaluation at a point is a continuous linear operator on any RKHS, but it isn’t on, e.g., $\mathcal{H}_0 = L_2[0, 1]$. In any RKHS $\mathcal{H}$ the linear span of $\{k_s\}$ is dense in $\mathcal{H}$. For any CONS $\{e_n(s)\}$ on a RKHS the kernel can be written in the form

$$
k(s, t) = \sum_{n \in \mathbb{N}} e_n(s)e_n(t)
$$

with the sum converging at every point and also in the $\mathcal{H} \oplus \mathcal{H}$ Hilbert space norm. In our Dirichlet unit interval example one particular CONS in $\mathcal{H}_1$ is given by $e_n(t) = (n\pi)^{-1}\psi_n(t)$, leading to the representation of Eqn (19) for the Brownian Bridge covariance.

Every positive-definite function $k(s, t)$ on any set $\mathcal{T}$ determines a RKHS $\mathcal{H}(k)$ of functions on $\mathcal{T}$ and also determines a Gaussian stochastic process $Z_s$ indexed by $s \in \mathcal{T}$ with mean zero and covariance $k(s, t)$. RKHSs are helpful in problems where interpolation is needed: one way to construct a function $g : \mathcal{T} \to \mathbb{C}$ attaining specified values $g(s_i) = y_i$ for finite set $\{(s_i, y_i)\} \subset \mathcal{T} \times \mathbb{C}$ is to set

$$
g(s) = \sum \beta_k k(s, s_k)
$$

where $\{\beta_k\}$ is the vector solution of the matrix equation

$$
y_k = \sum_j k(s_i, s_j)\beta_j,
$$

and this will be the element $g \in \mathcal{H}(k)$ of minimum norm that satisfies $g(s_i) = y_i$. It is also identical to the Bayesian posterior expectation

$$
g(s) = \mathbb{E}[Z(s) \mid Z(s_i) = y_i, i \in I]
$$

for the Gaussian Process (“GP”) model $Z \sim \text{GP}(0, k)$ with mean zero and covariance $k(\cdot, \cdot)$ (non-zero means are easy to accommodate too).

There are deep connections between splines, GPs, and RKHSs. See Wahba (1990, 2003) for more about this, or just Google RKHS to find lots of interesting references.
5 Mercer Kernels and Karhunen-Loève Expansions

Mercer’s Theorem (and a few variations) give conditions to guarantee that a covariance function will have a representation like Eqn (20); Karhunen and Loève use that to find a representation for Gaussian processes similar to that of Eqn (14a).

**Theorem 2** (Mercer). Let $\mathcal{T}$ be a compact Hausdorff space, let $\mu: B(\mathcal{T}) \to \mathbb{R}_+$ be a Borel measure on $\mathcal{T}$, and let $k: \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ be a symmetric positive-definite continuous function on $\mathcal{T}$. Then there exists an orthonormal set $\{e_n\} \subset L_2(\mathcal{T}, d\mu)$ of continuous functions on $\mathcal{T}$ and a summable sequence $\{\lambda_n\} \subset \mathbb{R}_+$ of positive numbers for which

$$k(s, t) = \sum_{n \in \mathbb{N}} \lambda_n e_n(s)e_n(t) \quad (22)$$

for every $s, t \in \mathcal{T}$. The sum converges uniformly and absolutely.

The functions $e_n(\cdot)$ are in fact the eigenfunctions for the integral operator $K: \phi \mapsto K[\phi] \in L_2$ given by

$$K[\phi](s) = \int_{\mathcal{T}} k(s, t) \phi(t) \mu(dt)$$

for $s \in \mathcal{T}$, $\phi \in L_2$, with positive eigenvalues $\{\lambda_n\}$ whose sum is

$$\sum \lambda_n = \int_{\mathcal{T}} k(t, t) \mu(dt) < \infty$$

In the Brownian Bridge example above we had the compact unit interval $\mathcal{T} = [0, 1]$ with Lebesgue measure $\mu(dx) = dx$ and continuous kernel $k(s, t) = \gamma(s, t) = s \wedge t - st$, but here we see many of the same ideas would work in more general settings.

Further generalizations are possible—for example, $k$ need not be continuous as long as

$$\int_{\mathcal{T}} k(s, s) \mu(ds) < \infty \quad \text{and} \quad \iint_{\mathcal{T} \times \mathcal{T}} k^2(s, t) \mu(ds) \mu(dt) < \infty,$$

but then the eigenfunctions $e_n(s)$ may not be continuous, and the convergence of Eqn (22) may not be uniform.

**Theorem 3** (Karhunen-Loève). Let $X_t$ be a square-integrable mean-zero stochastic process indexed by a compact Hausdorff space $\mathcal{T}$, let $\mu: B(\mathcal{T}) \to \mathbb{R}_+$ be a Borel measure on $\mathcal{T}$, and suppose that the covariance function $k(s, t) = EX_tX_s$ is jointly continuous. Then there exists an orthonormal set $\{e_n\} \subset L_2(\mathcal{T}, d\mu)$ of continuous functions on $\mathcal{T}$, a summable sequence $\{\lambda_n\} \subset \mathbb{R}_+$ of positive numbers, and a collection $\{Z_n\}$ of uncorrelated random variables with mean zero and variance one such that

$$X_t = \sum_{n \in \mathbb{N}} \sqrt{\lambda_n} Z_n e_n(s).$$

The $\{Z_n\}$ are given by

$$Z_n = \int_{\mathcal{T}} X_t e_n(t) \mu(dt),$$

i.e., the inner product $Z_n = \langle X, e_n \rangle$ in $L_2(\mathcal{T}, d\mu)$. The sum converges uniformly and absolutely. In case the process $X_t$ is Gaussian, the $\{Z_n\}$ will be iid $\mathcal{N}(0, 1)$. 

16
6 Tightness & Compactness

6.1 The Kolmogorov Continuity Theorem

**Definition 3** (Hölder continuity). A function \( f : S \to \mathbb{R}^d \) is Hölder continuous with index \( \alpha > 0 \) on the set \( S \subset \mathbb{R} \) if

\[
(\exists C < \infty)(\forall s, t \in S) \| f(s) - f(t) \| \leq C |s - t|^\alpha.
\]

A function with bounded continuous first derivative will be Hölder continuous with index \( \alpha = 1 \), and one whose first \( k \) derivatives are continuous and bounded will be Hölder continuous of index \( \alpha = k \); this generalizes that idea of smoothness to fractional degrees. Any Hölder continuous function is, in particular, continuous.

**Definition 4** (Version). A stochastic process \( \{\bar{X}_t\}_{t \in \mathcal{T}} \) is called a version or a modification of a process \( \{X_t\}_{t \in \mathcal{T}} \) if, for every \( t \in \mathcal{T} \), \( P[\bar{X}_t = X_t] = 1 \).

For countable index sets \( \mathcal{T} \) this is the same as \( P[(\forall t \in \mathcal{T}) \bar{X}_t = X_t] \) but not for uncountable index sets (give an example!).

**Theorem 4** (Kolmogorov Continuity Theorem). Let \( \alpha, \epsilon, c > 0 \). If a \( d \)-dimensional stochastic process \( X_t : [0, 1] \to L_0(\Omega, \mathcal{F}, P) \) satisfies the bound

\[
E\|X_s - X_t\|^\alpha \leq c|t - s|^{1+\epsilon}
\]

uniformly for all \( s, t \in [0, 1] \), then there exists a version \( \bar{X}_t \) of \( X_t \) whose sample paths are Hölder continuous of index \( \gamma \) for every \( 0 < \gamma < \epsilon / \alpha \).

**Proof.** Take \( d = 1 \) (it’s easy to extend to \( d > 1 \)). Denote by

\[
Q_n := \{j/2^n : 0 \leq j \leq 2^n\}
\]

the dyadic rationals of order \( n \) and by \( Q := \bigcup Q_n \) their union, the dense set of all dyadic rationals in the unit interval. For \( 0 < \gamma < \epsilon / \alpha \) Markov’s inequality (Lemma 2 on p. 26) gives

\[
P\left[ \max_{1 \leq j \leq 2^n} |X_{j/2^n} - X_{(j-1)/2^n}| > 2^{-\gamma n} \right] = P\left[ \bigcup_{1 \leq j \leq 2^n} |X_{j/2^n} - X_{(j-1)/2^n}| > 2^{-\gamma n} \right] \leq \sum_{j=1}^{2^n} P\left\{ |X_{j/2^n} - X_{(j-1)/2^n}| > 2^{-\gamma n} \right\} \leq \sum_{j=1}^{2^n} E\left| X_{j/2^n} - X_{(j-1)/2^n} \right|^\alpha / 2^{-\alpha \gamma n} \leq 2^nc2^{-n(1+\epsilon)} / 2^{-\alpha \gamma n} = c2^{-n(\epsilon - \alpha \gamma)}.
\]
Since this is summable in $n$, the Borel-Cantelli Lemma (lemma 1 on p. 26) asserts that outside some null set $\mathcal{N}$ there exists a random $N < \infty$ such that for all $n \geq N$

$$\max_{1 \leq j \leq 2^n} |X_{j/2^n} - X_{j-1/2^n}| \leq 2^{-\gamma n}$$

By looking at the maximum over the finitely-many $n < N$, we can find a finite random number $C < \infty$ such that

$$\max_{1 \leq j \leq 2^n} |X_{j/2^n} - X_{(j-1)/2^n}| \leq C2^{-\gamma n}$$

for every $n \in \mathbb{N}$. Now let’s show that, outside the null set $\mathcal{N}$, the paths are $\gamma$-Hölder continuous on $\mathbb{Q}$.

For any $s, t \in \mathbb{Q}$ let $n = \lfloor (-\log_2 |s - t|) \rfloor$ (where $\lfloor x \rfloor$ denotes the least integer $\geq x$), so

$$2^{-n} \leq |s - t| < 2^{1-n}.$$ 

The sequence $s_k := \max\{q \in \mathbb{Q}_k : q \leq s\}$ for $k \geq n$ increases monotonically to $s$ and satisfies $0 \leq (s_k - s_{k-1}) \leq 1/2^k$; indeed each such term is exactly zero or $1/2^k$, and only finitely-many are nonzero because $s \in \mathbb{Q}$. Similarly set $t_k := \max\{q \in \mathbb{Q}_k : q \leq t\}$ for $k \geq n$. Then

$$X_t - X_s = \sum_{i > n} (X_{s_i} - X_{s_{i-1}}) + \sum_{i > n} (X_{t_i} - X_{t_{i-1}}) + (X_{s_n} - X_{t_n})$$

where only finitely-many terms in each sum are non-zero. Also, $|s_n - t_n| \leq |s - t| + 2^{-n} \leq 3 \cdot 2^{-n}$. Summing,

$$|X_t - X_s| \leq \sum_{i > n} |X_{s_i} - X_{s_{i-1}}| + \sum_{i > n} |X_{t_i} - X_{t_{i-1}}| + |X_{s_n} - X_{t_n}|$$

$$\leq 2 \sum_{j > n} C2^{-\gamma j} + 3C2^{-\gamma n}$$

$$= C \left( \frac{2}{2^{\gamma - 1}} + 3^{\gamma} \right) 2^{-\gamma n} \leq C \left( \frac{2}{2^{\gamma - 1}} + 3^{\gamma} \right) |t - s|^\gamma.$$ 

Now define the modification $\tilde{X}_t$ to be zero on $\mathcal{N}$, and elsewhere

$$\tilde{X}_t := \lim_{n \to \infty} \{X_{q_n} : q_n = \max(\mathbb{Q}_n \cap [0, t])\},$$

well-defined and continuous since $X_t$ is uniformly continuous on $\mathbb{Q}$.

For $\alpha = 2$ the Brownian Bridge satisfies

$$\mathbb{E}|B_s - B_t|^2 = |s - t|(1 - |s - t|) \leq |s - t|$$

which is bounded by $|s - t|^1$ but not by $c|s - t|^{1+\epsilon}$ for any $\epsilon > 0$. For $\alpha = 4$ we do better, though, with

$$\mathbb{E}|B_s - B_t|_4 \leq 3|s - t|^2$$

proving that some version of $B_s$ has Hölder-continuous sample paths of any index $\gamma < 1/4$. A similar argument shows for even integers $\alpha \geq 2$ that

$$\mathbb{E}|B_s - B_t|^\alpha = \sqrt{\pi}2^{\alpha/2}\Gamma\left(\frac{\alpha + 1}{2}\right)|t - s|^{\alpha/2}[1 - |t - s|]^{\alpha/2}$$

$$\leq c|t - s|^{\alpha/2};$$
so the paths are Hölder continuous of any index

$$\gamma < \frac{\alpha/2 - 1}{\alpha} = \frac{1}{2} - \frac{1}{\alpha}$$

or, taking the supremum as $\alpha \to \infty$, of any index $\gamma < \frac{1}{2}$. It turns out that this bound is sharp—paths are not continuous with index $\gamma \geq \frac{1}{2}$ and, in particular, are not differentiable.

### 6.2 Compactness in Sobolev Spaces

The normalized difference $\Delta_n$ between the empirical and actual CDFs for $n$ iid uniforms of Eqn (2) may be written

$$\Delta_n = n^{-\frac{1}{2}} \sum_{i \leq n} \{1_{\{V_i \leq t\}} - t\}$$

$$= n^{-\frac{1}{2}} \sum_{i \leq n} \phi_i(t)$$

where

$$\phi_i(t) = \{1_{\{V_i \leq t\}} - t\} = \begin{cases} 
-t & t < U_i \\
1 - t & t > U_i 
\end{cases}$$

has Fourier sine coefficients

$$\langle \phi_i, \psi_n \rangle_0 = \int \phi_i(t) \sqrt{2} \sin(n\pi t) dt$$

$$= \sqrt{2} \cos(n\pi U_i) / n\pi$$

and hence squared Sobolev norm

$$\| \phi_i \|_s^2 = \sum_{n \in \mathbb{N}} (\frac{\pi^2 n^2}{2})^s \frac{2}{\pi^2 n^2} \cos^2(n\pi U_i)$$

with expectation

$$\mathbb{E} \| \phi_i \|_s^2 = \sum_{n \in \mathbb{N}} (\frac{\pi^2 n^2}{2})^s - 1. $$

This is finite for $s < \frac{1}{2}$, so also $\Delta_n \in \mathcal{H}_s$ for $s < \frac{1}{2}$. By independence, for $s < \frac{1}{2}$

$$\mathbb{E} \| \Delta_n \|_s^2 = \frac{1}{n} \sum_{i \leq n} \mathbb{E} \| \phi_i \|_s^2 + \frac{1}{n} \sum_{i \neq j} \mathbb{E} \langle \phi_i, \phi_j \rangle_s$$

$$= \sum_{n \in \mathbb{N}} (\frac{\pi^2 n^2}{2})^s - 1.$$
uniformly in $n$. By Markov’s inequality (Lemma 2, on p. 26), for any $s < \frac{1}{2}$ and $\epsilon > 0$ we can find $r < \infty$ such that for all $n$

$$P[|\Delta_n|_s \leq r] \geq 1 - \epsilon,$$

i.e., such that the ball $B_s(r)$ of radius $r$ in $\mathcal{H}_s$ has probability exceeding $(1 - \epsilon)$ for the distribution measure $\mu_n$ of each $\Delta_n$. Although $B_s(r)$ isn’t compact in $\mathcal{H}_s$, it is compact in $\mathcal{H}_{s-1}$—hence $\{\mu_n\}$ is tight as a family of measures on any $\mathcal{H}_r$ for $r < -\frac{1}{2}$ (for example, $\mathcal{H}_{-1}$). This sequence of probability measures converges to the distribution $\mu$ for the Brownian Bridge, and the distribution of any continuous functional of $\Delta_n$ converges as well.

I’d like to build that into a Sobolev-based argument that the distribution of

$$\sqrt{n}D_n = \sup_{0 \leq t \leq 1} |\Delta_n(t)|$$

converges to that of $\sup_{0 \leq t \leq 1} |B_t|$, but it doesn’t seem to quite work because “sup” is only continuous as a functional on $\mathcal{H}_s$ for $s > \frac{1}{2}$ (try to prove that) but we’ve only shown $\Delta_n \in \mathcal{H}_s$ for $s < \frac{1}{2}$. Close, but not quite good enough. Kolmogorov’s criterion will still get tightness on $C([0,1])$ and sup is continuous on that, proving that $\sqrt{n}D_n \Rightarrow \sup_{0 \leq t \leq 1} |B_t|$, but I’d still like to get a Sobolev argument to work.

7 Reflecting on Brownian Bridge Extremes

We motivated a study of Dirichlet Sobolev spaces in Section (1) by looking at Donsker’s Theorem and the Kolmogorov-Smirnov statistic. In this brief section we derive a closed-form representation of the CDF of the statistic $D_n$ of Eqn (1), based on representation Eqn (3c) and a reflection argument. The approach relies on a feature of Brownian Motion (the “strong Markov property”) that we will assert but not prove; ask me for details or references if you would like to look at this more carefully.

First let $W(t)$ be a Wiener process or Brownian Motion, and fix $T > 0$ and $u > 0$. Set $\tau_u := \inf\{t : W(t) \geq u\}$, the first time $W(t)$ exceeds $u$. Then

$$P[\sup_{0 \leq t \leq T} W(t) \geq u] = P[\tau_u \leq T]$$

Conditionally on the event $[\tau_u \leq T]$, we have $W(\tau_u) = u$ by path continuity and so, by the strong Markov property, the Brownian path on the (random) time interval $[\tau_u,T]$ is equally likely to increase to $W(T) > u$ or decrease to $W(T) < u$. Thus

$$P[\sup_{0 \leq t \leq T} W(t) \geq u] = 2P[\tau_u \leq T \cap W(T) > u]$$

$$= 2P[W(T) > u]$$

$$= 2\Phi(-u/\sqrt{T}),$$

since necessarily $\tau_u < T$ if $W(T) > u$ and since $W(T) \sim \mathcal{N}(0, T)$. This argument is called the “reflection principle,” since it relies on the identical distributions of $[W(t) - u]$ and its “reflection” $[u - W(t)]$ on the time interval $[\tau_u, T]$. A conditional version of the same argument will help us with the Brownian Bridge.
7.1 Distribution of the First Hitting Time

The collection of random variables \( \{ \tau_u \} \) themselves make up a stochastic process, indexed by \( u \geq 0 \), with interesting properties. From the argument above, \( \tau_u \) has CDF

\[
F_u(t) = P[\tau_u \leq t] = 2\Phi(-u t^{-1/2})1_{\{t > 0\}}
\]

and hence pdf

\[
f_u(t) = 2\phi(-u t^{-1/2})[u t^{-3/2}]1_{\{t > 0\}}
= \frac{u}{\sqrt{2\pi}} t^{-3/2} e^{-u^2/2t}1_{\{t > 0\}},
\]

rising from zero at \( t = 0 \) to a peak of \( u^{-2} [27/2\pi e^3]^{1/2} \) at \( t = u^2 / 3 \), then falling at rate \( f_u(t) \propto t^{-3/2} \) as \( t \to \infty \). Here’s a plot for \( u = 1 \):

![Plot](image.png)

Because this decreases more slowly than \( t^{-2} \), \( \tau_u \) has infinite mean \( E\tau_u = \infty \). This distribution is often called the inverse Gaussian, or sometimes the Lévy distribution, denoted \( iG(u) \) or \( L\nu(u) \); it is a special case of the \( \alpha \)-Stable family, the fully-skewed \( \alpha \)-Stable with \( \alpha = \frac{1}{2} \). That can be verified by computing its characteristic function:

\[
\chi_u(\omega) = E[e^{i\omega \tau_u}]
= \frac{u}{\sqrt{2\pi}} \int_0^\infty t^{-3/2} e^{i\omega t-u^2/2t} dt
= \frac{u}{\sqrt{2\pi}} \int_0^\infty z^{-3/2} e^{-u^2/2z} \omega^2 u^2 dz
= \chi_1(\omega u^2), \text{ where}
\]

\[
\chi_1(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^\infty z^{-3/2} e^{i\omega z-1/2z} dz
\]

does not depend on \( u \). But also the sum of independent random variables \( \tau_u \sim iG(u) \) and \( \tau_v \sim iG(v) \) has distribution \( \tau_u + \tau_v \sim iG(u+v) \), because \( W_t \) can only reach level \( u+v \) by first reaching level \( u \) and then rising an additional height \( v \), so

\[
\chi_{u+v}(\omega) = \chi_1(\omega(u+v)^2) = \chi_u(\omega)\chi_v(\omega) = \chi_1(\omega u^2)\chi_1(\omega v^2);
\]

each of these is of the form

\[
\chi_1(\omega) = e^{-\gamma|\omega|^{1/2}}
\]
for some \( \gamma \geq 0 \). Evaluating the integral above shows that \( \gamma = 1 \), so the inverse Gaussian has ch.f.

\[
\chi_u(\omega) = e^{-u|\omega|^{1/2}},
\]

the ch.f. of the fully-skewed \( \alpha \)-Stable distribution: \( iG(u) = St(\alpha = 1/2, \beta = 1, \gamma = u, \delta = 0) \).

### 7.2 Conditional Reflection

Now let \( x \in \mathbb{R} \) and fix a small \( \epsilon > 0 \), and consider the conditional probability

\[
P[\sup_{0 \leq t \leq T} W(t) \geq u \mid \{|W(T) - x| < \epsilon\}] = P[\tau_u \leq T \mid \{|W(T) - x| < \epsilon\}]
\]

This probability is one if \( x > u + \epsilon \); for \( x < u \), and \( \epsilon < (u - x) \),

\[
\begin{align*}
&= \frac{P[\tau_u \leq T \cap |W(T) - x| < \epsilon]}{P[|W(T) - x| < \epsilon]} \\
&= \frac{P[\tau_u \leq T \cap |W(T) - 2u + x| < \epsilon]}{P[|W(T) - x| < \epsilon]}
\end{align*}
\]

because \( |W(t) - u| \) and \( |u - W(t)| \) have the same distribution on \([\tau_u, T]\)

\[
= \frac{P[|W(T) - 2u + x| < \epsilon]}{P[|W(T) - x| < \epsilon]}
\]

because \( \tau_u < T \) if \( W(T) > 2u - x - \epsilon > u \) since \( \epsilon < (u - x) \)

\[
\begin{align*}
&= \frac{\Phi \left( \frac{2u-x+\epsilon}{\sqrt{T}} \right) - \Phi \left( \frac{2u-x-\epsilon}{\sqrt{T}} \right)}{\Phi \left( \frac{x+\epsilon}{\sqrt{T}} \right) - \Phi \left( \frac{x-\epsilon}{\sqrt{T}} \right)} \\
&\to \frac{\phi \left( \frac{2u-x}{\sqrt{T}} \right)}{\phi \left( \frac{x}{\sqrt{T}} \right)} \quad \text{as} \quad \epsilon \to 0
\end{align*}
\]

\[
= \left\{ \frac{1}{\sqrt{2\pi T}} e^{-(2u-x)^2/2T} \right\} \left\{ \frac{1}{\sqrt{2\pi T}} e^{-x^2/2T} \right\}^{-1}
\]

\[
= \exp \left\{ \frac{-(2u-x)^2 + x^2}{2T} \right\}
\]

\[
= \exp \left\{ \frac{-2u(u-x)}{T} \right\}
\]

Taking \( T = 1 \) and \( x = 0 \), we find for the Brownian Bridge that

\[
P[\sup_{0 \leq t \leq 1} B_t > u] = e^{-2u^2}
\]

and so

\[
P[\sup_{0 \leq t \leq 1} |B_t| > u] \approx 2e^{-2u^2},
\]
a strict upper bound with very small error for \( u \gg 0 \). To do better, we can use the reflection principle multiple times to try to find (with some informality about the notation)

\[
P[\sup_{0 \leq t \leq 1} |B_t| > u] = P[\tau_u \leq \tau_{-u} \land 1 \mid W(1) \in 0 + dx] + P[\tau_{-u} \leq \tau_u \land 1 \mid W(1) \in 0 + dx]
\]

\[
= 2P[\tau_u \leq \tau_{-u} \land 1 \mid W(1) \in 0 + dx]
\]

\[
= 2P[\tau_u \leq \tau_{-u} \land 1 \cap W(1) \in 0 + dx]/P[W(1) \in 0 + dx]
\]

where

\[
P[W(1) \in 0 + dx] = \phi(0) dx
\]

and, by reflection (draw picture),

\[
P[\tau_u \leq \tau_{-u} \land 1 \cap W(1) \in 0 + dx] = P[W(1) \in 2u + dx] - P[W(1) \in 4u + dx] + \ldots
\]

\[
= \phi(2u) dx - \phi(4u) dx + \ldots,
\]

and so

\[
P[\sup_{0 \leq t \leq 1} |B_t| > u] = 2 \frac{\phi(2u) - \phi(4u) + \ldots}{\phi(0)}
\]

\[
= 2 \left[ e^{-2u^2} - e^{-8u^2} + \ldots \right]
\]

\[
= 2 \sum_{n=1}^{\infty} (-1)^{n-1} e^{-2n^2u^2}.
\]

In a bit more detail,

\[ A_n := \{ \exists 0 < t_1 < \cdots < t_n < 1 : B(t_j) = (-1)^j u \} , \]

the probability that \( B_n \) crosses the endpoints of the interval \([-u, u]\) at least \( n \) times starting with \( +u \). By symmetry,

\[
P[\sup_{0 \leq t \leq 1} |B_t| \geq u] = 2P[A_1 \cap \tau_u < \tau_{-u}]
\]

and by symmetry, for all \( n \in \mathbb{N} \),

\[
P[A_{n+1} \cap \tau_u < \tau_{-u}] = P[A_n] - P[A_n \cap \tau_{-u} < \tau_u]
\]

\[
= P[A_n] - P[A_{n+1} \cap \tau_u < \tau_{-u}].
\]

By induction, for each \( n \in \mathbb{N} \),

\[
P[A_1 \cap \tau_u < \tau_{-u}] = P[A_1] - P[A_2] + P[A_3] - \cdots + (-1)^{n-1} P[A_n \cap \tau_u < \tau_{u-1}]
\]

\[
= \sum_{n \geq 1} (-1)^{n-1} P[A_n]
\]

in the limit as \( n \to \infty \) since \( P[A_n] \to 0 \). By reflecting each time \( W(t) \) hits \( \pm u \),

\[
P[A_n] = \lim_{\epsilon \to 0} \frac{P[|W_1 - 2nu| < \epsilon]}{P[|W_1| < \epsilon]} = e^{-2n^2u^2}.
\]

\[ ^1 \text{Note: The sum converges by Leibniz's test because successive terms decrease in absolute value and alternate in sign. It also converges absolutely, because } P[A_n] = e^{-2n^2u^2} \text{ decreases faster than geometrically.} \]
A Appendix

Here we collect definitions and a few properties of some of the kinds of spaces needed above, and some standard results from real and functional analysis. For more details consult any standard text on functional analysis, such as (Reed and Simon, 1980) or (Rudin, 1973). Classic sources for more details about distributions and test functions are (Gel’fand and Vilenkin, 1964) and (Schwartz, 1966).

Definition 5 (Hilbert Space). A Hilbert Space is a complete separable normed linear vector space \( \mathcal{H} \) over either the real numbers \( \mathbb{R} \) or the complex numbers \( \mathbb{C} \) whose norm is given by \( \|x\| = \langle x, x \rangle^{1/2} \) for an inner product \( \langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C} \) that satisfies for \( x, y \in \mathcal{H} \) and \( a, b \in \mathbb{C} \):

\[
\langle y, x \rangle = \overline{\langle x, y \rangle}
\]

\( \langle x, x \rangle \geq 0 \), with \( \langle x, x \rangle = 0 \iff x = 0 \)

\( \langle ax + bw, y \rangle = a \langle x, y \rangle + b \langle x, y \rangle \)

The inner product may be recovered from the norm from the “polarization identity”

\[
4 \langle x, y \rangle = \|x + y\|^2 - \|x - y\|^2
\]

for real Hilbert spaces or, for complex ones,

\[
= \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2.
\]

This will be a valid inner product (and hence the normed linear space will actually be a Hilbert space) if and only if the norm obeys the (Pythagorean!) parallelogram law,

\[
\|x + y\|^2 + \|x - y\|^2 = 2 (\|x\|^2 + \|y\|^2)
\]

(prove this!) Familiar examples include \( \mathbb{R}^n \) and \( \mathbb{C}^n \), any \( L_2 \) space, and sequence spaces

\[
\ell_2 = \left\{ a : \sum |a_n|^2 < \infty \right\}
\]

where the sum may go over any countable set like \( \mathbb{N} \) or \( \mathbb{Z} \).

Every Hilbert space \( \mathcal{H} \) admits a complete orthonormal system (CONS), a countable collection \( \{x_n\} \) with the properties

- \( \langle x_m, x_n \rangle = 0 \) if \( m \neq n \), and 1 if \( n = m \);
- \( \langle \forall n \rangle \langle x, x_n \rangle = 0 \) if and only if \( x = 0 \);
- Each \( x \in \mathcal{H} \) admits a unique convergent expansion \( x = \sum a_n x_n \) with coefficients \( a_n = \langle x, x_n \rangle \);
- \( \langle x, y \rangle = \sum a_n \overline{b}_n \) if \( \{a_n\} \) and \( \{b_n\} \) are the coefficients corresponding to \( x \) and \( y \), respectively.

The basis \( \{x_n\} \) isn’t uniquely determined, just as there are many different possible bases for \( \mathbb{R}^n \) or \( \mathbb{C}^n \). In fact any linear independent collection \( \{y_n\} \) can be turned into an orthonormal system by the Gram-Schmidt process, and extended if necessary to form a CONS.
**Definition 6** (Banach Space). A Banach Space is a complete separable normed linear vector space $\mathcal{B}$ over either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$ whose norm satisfies the conditions

$$
\|x\| \geq 0, \quad \text{with} \quad \|x\| = 0 \iff x = 0
$$

$$
\|ax\| = |a| \|x\|
$$

$$
\|x + y\| \leq \|x\| + \|y\|
$$

Every Hilbert space is also a Banach space; additional examples include the continuous functions on any compact set $\mathcal{C}(K)$ with $\|f\| = \sup_{x \in K} |f(x)|$ and any $L_p$ space with $1 \leq p \leq \infty$. If $\mathcal{B}$ is a Banach space then a "continuous linear functional" on $\mathcal{B}$ is a linear mapping

$$
\ell : \mathcal{B} \to \mathbb{C}
$$

that is "continuous" in the sense that there exists some constant $c < \infty$ for which

$$
|\ell(x)| \leq c \|x\|
$$

for every $x \in \mathcal{B}$. The smallest such constant $c$ is the norm of $\ell$,

$$
\|\ell\| = \inf \{c < \infty : (\forall x \in \mathcal{B}) \ |\ell(x)| \leq c \|x\| \}
$$

$$
= \sup_{0 \neq x \in \mathcal{B}} |\ell(x)|/\|x\|.
$$

The space of all such continuous linear functionals is itself a Banach space with this norm, called the "dual space" to $\mathcal{B}$ and denoted by $\mathcal{B}^*$. Every Hilbert space is "self-dual" in that $\mathcal{H}^* = \mathcal{H}$, with each continuous linear functional given by the inner-product with some element of $\mathcal{H}$: $\ell(x) = \langle x, y \rangle$ for some $y \in \mathcal{H}$ (a result called the Hilbert-space "Reisz Representation Theorem" or the "Fréchet-Riesz Theorem"). The dual space of the continuous functions $\mathcal{B} = \mathcal{C}(K)$ on a compact set $K$ is the space of finite Borel measures on $K$, with linear functionals $\ell(f) = \int_K f(x) \mu(x) \, dx$ (a result also called the "Reisz representation theorem," or the "Riesz-Markov-Kakutani representation theorem"). The dual space of any $L_p$ space is $L_q$, with conjugate exponent $q = p/(p - 1)$ (so $\frac{1}{p} + \frac{1}{q} = 1$); the spaces $L_1$ and $L_\infty$ are mutually dual as well.

**Proposition 1** (Cauchy Schwarz). Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle x, y \rangle$ and norm $\|x\| = \langle x, x \rangle^{1/2}$. Then for all $x, y \in \mathcal{H}$,

$$
|\langle x, y \rangle| \leq \|x\| \|y\|.
$$

**Proof.** Write $x = \|x\| x^*$ and $y = \|y\| y^*$ for $x^* := x/\|x\|$ and $y^* := y/\|y\|$; each of norm one. From the positivity of

$$
\|x^* + y^*\|^2 = \|x^*\|^2 + 2 \langle x^*, y^* \rangle + \|y^*\|^2 = 2 \left[1 + \langle x^*, y^* \rangle\right]
$$

and

$$
\|x^* - y^*\|^2 = \|x^*\|^2 - 2 \langle x^*, y^* \rangle + \|y^*\|^2 = 2 \left[1 - \langle x^*, y^* \rangle\right]
$$

it follows that $|\langle x^*, y^* \rangle| \leq 1$ and hence that

$$
|\langle x, y \rangle| = \|x\| \|y\| \|\langle x^*, y^* \rangle\| \leq \|x\| \|y\|.
$$
Definition 7 (Some properties).

- A metric space $X$ is complete if every Cauchy sequence converges—i.e., if $\{x_n\}$ has the property that
  $$(\forall \epsilon > 0)(\exists N < \infty)(\forall n, m \geq N) \quad d(x_n, x_m) < \epsilon$$
then there exists $x \in X$ such that $x_n \to x$. The spaces $\mathbb{R}^n$ and $\mathcal{C}^n$ are complete; the rationals $\mathbb{Q}$ are not. Every function in $X = L_2([0, 1], dx)$ is also in $L_1([0, 1], dx)$, so $X$ with distance $d_1(x, y) = \|x - y\|_1$ is a metric space, but it is not complete in that metric (although it is complete in the metric $d_2(x, y) = \|x - y\|_2$).

- A metric space $X$ is separable if it contains a countable dense set $\{x_n\}$, i.e., one for which
  $$(\forall x \in X)(\forall \epsilon > 0)(\exists n \in \mathbb{N}) \exists d(x, x_n) < \epsilon$$
  The spaces $L_p$ are separable for $p < \infty$ but not for $p = \infty$.

Lemma 1 (Borel-Cantelli). Let $\{A_n\}$ be events whose probabilities are summable,

$$\sum_{n \in \mathbb{N}} P(A_n) < \infty.$$ 
Then with probability one only finitely-many $\{A_n\}$ occur; i.e.,

$$P\left[ \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \right] = 0$$

Proof. Denote by $A := \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m$ the event that infinitely-many $A_n$ occur. By subadditivity, for any finite $n \in \mathbb{N},$

$$P[A] \leq P\left[ \bigcup_{m \geq n} A_m \right] \leq \sum_{m \geq n} P[A_m] \to 0$$
as $n \to \infty$, since $\sum_{n \in \mathbb{N}} P[A_n] < \infty$. $\Box$

Lemma 2 (Markov’s Inequality). Let $Y$ be a random variable and $a > 0$ a positive constant. Then

$$P[Y > a] \leq \frac{EY_+}{a}$$
where $Y_+ := 0 \vee Y$ is the maximum of zero and $Y$.

Proof. Let $Z := a \mathbf{1}_{\{Y > a\}}$. Then $Z \leq Y_+$, since either $Y \geq a$ in which case $Z = a \leq Y_+$ or $Y < a$ in which case $Z = 0 \leq Y_+$. Hence

$$aP[Y > a] = EZ \leq EY_+.$$ 
Dividing by $a$ completes the proof. Note this implies that

$$P[X \geq a] \leq \frac{E\phi(X)}{\phi(a)}$$
for any random variable $X$, real number $a$, and nonnegative function $\phi$ that is monotone non-decreasing on the interval $[a, \infty)$. For example, $P[X > a] \leq (\sigma^2 + \mu^2)/a^2$ for any $a > 0$. $\Box$
References


