III. PRELUDE TO CONTINUITY

1. Boundedness and Continuity.

The principle aim of this section is to show that for Gaussian processes the question of sample path continuity is intimately related to the boundedness of the supremum. In one direction this is obvious and, indeed, non-probabilistic. If the parameter space T is compact, then the a.s. continuity of X implies the a.s. boundedness of $\sup_{t\in T}|X_t|$. Thus the problem is essentially to find conditions under which processes with bounded suprema are also continuous.

Recall that we treat only centered processes, and measure continuity in terms of the canonical metric $d(s,t) = (E(X_s - X_t)^2)^{1/2}$ on T.

The first result we need is the following easy lemma, which tells us that as far as a.s. boundedness is concerned it is irrelevant whether we work with $\sup X_t$ or $\sup|X_t|$.

3.1 LEMMA. *For X centered, Gaussian, on T, and $t_o \in T$*

$$E\sup_{t\in T} X_t \;\le\; E\sup_{t\in T}|X_t| \;\le\; E|X_{t_o}| + 2E\sup_{t\in T} X_t .$$

PROOF: Only the rightmost inequality needs proving. Note the trivial inequalities that for any $t, t_o \in T$

$$X_t - X_{t_o} \;<\; \sup_{t\in T}(X_t - X_{t_o}), \qquad X_{t_o} - X_t \;<\; -\inf_{t\in T}(X_t - X_{t_o}).$$

Furthermore, both $\sup_{t\in T}(X_t - X_{t_o}) \ge 0$ and $-\inf_{t\in T}(X_t - X_{t_o}) \ge 0$. Applying the relationships $\max(a,-a) = |a|$ and $\max(a,b) \le a+b$ if $a,b \ge 0$, it follows from the above two inequalities that

$$\begin{aligned} |X_t| &\le |X_{t_o}| + |X_t - X_{t_o}| \\ &\le |X_{t_o}| + \sup_{t\in T}(X_t - X_{t_o}) - \inf_{t\in T}(X_t - X_{t_o}). \end{aligned}$$

Taking a supremum over the left hand side leaves the right side unchanged. Now take expectations, and note that symmetry gives us that $E\sup_T X_t = -E\inf_T X_t$ to complete the proof. ∎

The next result is somewhat more interesting and important, and is a good example of the power of Borell's inequality. The original proof of this result, which appears as Exercise 1.1, is due to Fernique (1978), and involves a comparatively long and sophisticated calculation.

3.2 THEOREM. *For X centered, Gaussian,*

$$P\{\sup_{t\in T} X_t < \infty\} = 1 \iff E\sup_{t\in T} X_t < \infty \tag{3.1}$$
$$\iff Ee^{\alpha\|X\|^2} < \infty$$

for sufficiently small α.

REMARK: Theorem 3.2 can be strengthened considerably. For the best possible results, which tell you exactly how large α can be for something like (3.1) to hold, see Talagrand (1984). The proof will not use the fact that $\|\ \|$ is the supremum function, and so the result will hold in the same generality as Borell's inequality itself.

PROOF OF THEOREM 3.2: The existence of the exponential moments of $\|X\|$ implies the existence of $E\|X\|$, and this in turn implies the a.s. finiteness of $\|X\|$. Furthermore, since by Theorem 2.1 we already know that the a.s. finiteness of $\|X\|$ entails that of $E\|X\|$, all that remains is to prove is that the a.s. finiteness of $\|X\|$ also implies the existence of exponential moments.

But this is an easy consequence of Borell's inequality, since, with both $\|X\|$ and $E\|X\|$ now finite,

$$\begin{aligned} Ee^{\alpha\|X\|^2} &= \int_0^\infty P\left\{e^{\alpha\|X\|^2} > \lambda\right\} d\lambda \\ &= E\|X\| + \int_{E\|X\|}^\infty P\{\|X\| > \sqrt{\log\lambda^{1/\alpha}}\}\, d\lambda \\ &\le E\|X\| \\ &\quad + 2\int_{E\|X\|}^\infty \exp\left\{-\tfrac{1}{2}\left(\sqrt{\log\lambda^{1/\alpha}} - E\|X\|\right)^2/\sigma_T^2\right\} d\lambda \\ &\le E\|X\| \\ &\quad + 4\alpha\int_0^\infty u\exp\left\{-\tfrac{1}{2}\left(u - E\|X\|\right)^2/\sigma_T^2\right\}\exp\left\{\alpha u^2\right\} du, \end{aligned}$$

which is clearly finite for small enough α. ■

The next theorem is the first to relate continuity to boundedness, and, in essence, is the result that will enable us to concentrate only on the expected values of suprema in the future, and derive all other results from results on these.

3.3 THEOREM. *Let X be a.s. bounded on T and let τ be a metric on T such that the canonical metric d is τ-uniformly continuous. Then X is τ-uniformly continuous with probability one if, and only if, $\lim_{\eta\to 0}\phi_\tau(\eta) = 0$, where ϕ_τ is given by*

$$\phi_\tau(\eta) = E\sup_{\tau(s,t)<\eta}(X_s - X_t). \tag{3.2}$$

PROOF: We start with necessity. For each ω we have

$$\lim_{\eta \to 0} \sup_{\tau(s,t)<\eta} |X_s(\omega) - X_t(\omega)| = 0,$$

so that the fact that $\lim_{\eta \to 0} \phi_\tau(\eta) = 0$ follows from dominated convergence (c.f. Theorem 3.2).

Conversely, since d is τ-uniformly continuous, we can find a sequence η_n with $\phi_\tau(\eta_n) \le 2^{-n}$, such that $\tau(s,t) < \eta_n$ implies $d(s,t) < 2^{-n}$. Consider the event

$$A_n = \{ \sup_{\tau(s,t)<\eta_n} |X_s - X_t| > 2^{-n/2} \}. \tag{3.3}$$

Since X is, by assumption, a.s. bounded, we can apply Borell's inequality (Theorem 2.1) to obtain that, for $n \ge 3$,

$$\begin{aligned} P\{A_n\} &\le 2 \exp\left(-\tfrac{1}{2}(2^{-n/2} - 2^{-n})^2 / 2^{-2n} \right) \\ &\le K \exp\left(-2^{n-1} \right). \end{aligned}$$

Since $P\{A_n\}$ is an admirably summable series, Borel-Cantelli gives us that X is a.s. uniformly τ-continuous, as required. This completes the proof. ■

The proof of Theorem 3.3 actually yields more than what was claimed in the statement, which is why we departed from our regular policy of stating everything in terms of the canonical metric only. In fact, if we denote the τ-modulus of (uniform) continuity of X by

$$W_\tau(\eta) = \sup_{\tau(s,t)<\eta} |X_s - X_t|, \tag{3.4}$$

then the above calculations actually give us substantial information on the size of W_τ. Thus we have that not only is the comparatively simple question of sample path continuity inextricably tied up with the question of boundedness, but much finer information on moduli of continuity comes for free in this formulation.

3.4 COROLLARY. *Assume that the conditions of Theorem 3.3 hold and that* $\lim_{\eta \to 0} \phi_\tau(\eta) = 0$. *Then, for all* $\epsilon > 0$ *there exists an a.s. finite random variable* $\delta = \delta(\omega)$ *such that, for almost all* ω,

$$W_\tau(\eta) \le \phi_\tau(\eta) |\log \phi_\tau(\eta)|^\epsilon, \tag{3.5}$$

for all $\eta \le \delta(\omega)$. *That is,* $\phi_\tau(\cdot)|\log \phi_\tau(\cdot)|^\epsilon$ *is a uniform sample modulus for* X *in the metric* τ.

REMARK: The ϵ in (3.5) is, as are all such infinitesimals, exceedingly irritating, and one would like to dispose of it. That this can in fact be done will be shown in the following chapter. As you read the following proof, note that if only (3.6) could be replaced with $d_\tau(\eta) = o(\phi_\tau(\eta))$ for small η, then virtually the same proof would suffice to establish the stronger result. In order to obtain such an inequality, however, we have to have sharp tools for calculating $E\|X\|$. But this is precisely what Chapter 4 is about.

PROOF: Set

$$\begin{aligned} d_\tau(\eta) &:= \sup_{\tau(s,t)\le\eta} d(s,t) \\ &= \sup_{\tau(s,t)\le\eta} \left(E|X_t - X_s|^2\right)^{\frac{1}{2}}, \end{aligned}$$

and note the trivial inequality

$$\begin{aligned} d_\tau(\eta) &= \sup_{\tau(s,t)<\eta} \left(E|X_t - X_s|^2\right)^{\frac{1}{2}} \\ &= \sqrt{2\pi} \sup_{\tau(s,t)<\eta} E|X_t - X_s| \\ &\le 2\sqrt{2\pi}\, E \sup_{\tau(s,t)<\eta} (X_t - X_s) \\ &= 2\sqrt{2\pi}\, \phi_\tau(\eta), \end{aligned} \tag{3.6}$$

where the second line is a standard Gaussian result and the inequality follows from Lemma 3.1, applied to the two-parameter process $Y(s,t) = X(s) - X(t)$ with "t_o" $= (t,t)$ for some $t \in T$.

Since, by assumption, ϕ_τ is τ-continuous we can define, for each $n \ge 1$,

$$\eta_n = \inf\{\eta : \phi_\tau(\eta) = e^{-n}\}.$$

Define also

$$B_n = \Big\{ \sup_{\tau(s,t)<\eta_n} |X_s - X_t| > \phi_\tau(\eta_n)|\log\phi_\tau(\eta_n)|^{\epsilon/2} \Big\}.$$

By two (because of the absolute value sign) applications of Borell's inequality,

$$\begin{aligned} P\{B_n\} &\le 4\exp\left\{ -\tfrac{1}{2}\left(|\log\phi_\tau(\eta_n)|^{\epsilon/2} - 1\right)^2 \frac{\phi_\tau^2(\eta_n)}{d_\tau^2(\eta_n)} \right\} \\ &\le K_1 \exp\left\{ -K_2 n^\epsilon \right\} \end{aligned}$$

by (3.6) and the definition of η_n.

Since $\sum_n P\{B_n\} < \infty$, we have that for $n \ge N(\omega)$

$$W_\tau(\eta_n) \le \phi_\tau(\eta_n)|\log\phi_\tau(\eta_n)|^{\epsilon/2}.$$

Monotonicity of W_τ and d, along with separability, complete the proof. ■

Another easy corollary of the proof of Theorem 3.3 is the following.

3.5 COROLLARY. *Let X be as in Theorem 3.3 and for $t \in T$ set*

$$\phi_\tau^t(\eta) = E \sup_{s:\tau(s,t)<\eta} (X_s - X_t).$$

Then X is a.s. τ-continuous at t if, and only if, $\lim_{\eta\to 0} \phi_\tau^t(\eta) = 0$.

The proof is almost verbatim that of the theorem, so we shall not give it. Note that a local modulus of τ-continuity based on ϕ_τ^t instead of ϕ_τ follows exactly as in Corollary 3.4.

At first sight, Corollary 3.5 seems to be of far less interest than Theorem 3.3, since it deals only with continuity at a specific point, rather than throughout T. That this is *not* the case is a consequence of the following unexpected and important result.

2. Zero-One Laws and Continuity.

3.6 THEOREM. *A Gaussian process X on T has continuous sample paths with probability one if, and only if, it is continuous at each fixed $t \in T$ with probability one; i.e.*

$$(3.7) \qquad P\{\lim_{s\to t} X_s = X_t \text{ for all } t \in T\} = 1$$
$$\iff \quad P\{\lim_{s\to t} X_s = X_t\} = 1, \text{ for each } t \in T.$$

Necessity is obvious. It is the sufficiency of (3.7) that is the big surprise. There is something very special about Gaussian processes that makes Theorem 3.6 hold. A similar result for, for example, Poisson processes is palpably false.

We shall need a number of preliminary results before we can prove Theorem 3.6. Most of these are of significant interest in their own right and that, in fact, is the main reason for bringing them at this point. However, they, are not necessary for the understanding of the main results of these notes, and so can be skipped at the first reading.

We start with the *reproducing kernel Hilbert space* (RKHS) of a Gaussian process with covariance function R.

In essence, this is made up of functions that have about the same smoothness properties that $R(s,t)$ has, as a function in t for fixed s, or vice versa. Start with

$$S = \left\{f: T \to \Re : f(\cdot) = \sum_{i=1}^{n} a_i R(s_i, \cdot),\ a_i \text{ real},\ s_i \in T,\ n \geq 1\right\}.$$

Define an inner product on S by

$$(3.8)\qquad (f,g)_H = \Big(\sum_{i=1}^{n} a_i R(s_i,\cdot), \sum_{j=1}^{m} b_j R(s_j,\cdot)\Big)_H = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j R(s_i,s_j).$$

The fact that R is non-negative definite implies $(f,f)_H \geq 0$ for all $f \in S$. Furthermore, note that the inner product (3.8) has the following unusual property:

$$(3.9)\qquad f(t) = \sum_{i=1}^{n} a_i R(s_i,t) = \Big(\sum_{i=1}^{n} a_i R(s_i,\cdot), R(t,\cdot)\Big)_H = (f, R(t,\cdot))_H .$$

We refer to (3.9) as the "reproducing kernel" property. From this it follows that for $f \in S$, $t \in T$

$$(3.10)\qquad |f(t)|^2 = |(f,R(t,\cdot))_H|^2 \leq (f,f)_H \left(R(t,\cdot),R(t,\cdot)\right)_H ,$$

the inequality being merely the Schwartz inequality for semi-inner products, which holds as long as $(f,f)_H \geq 0$. Thus, if $(f,f) = 0$, then (3.10) implies that $f(t) = 0$ for all $t \in T$. Consequently, (3.8) defines a proper inner product on S, and so we thus obtain a norm $\|f\|_H = (f,f)_H^{1/2}$. For $\{f_n\}_{n\geq 1}$ a sequence in S we have

$$(3.11)\qquad \begin{aligned} |f_n(t) - f_m(t)|^2 &= |(f_n - f_m, R(t,\cdot))_H|^2 \\ &\leq \|f_n - f_m\|^2 \, \|R(t,\cdot)\|_H^2 \\ &\leq \|f_n - f_m\|_H^2 \, R(t,t), \end{aligned}$$

the last line following directly from (3.8). Thus it follows that if $\{f_n\}$ is Cauchy in $\|\cdot\|_H$ then it is pointwise Cauchy. The closure of S under this norm is a space of real-valued functions, denoted by $H(R)$, and called the RKHS of X or of R, since every $f \in H(R)$ satisfies (3.9) by the separability of $H(R)$. Since T is separable, and R continuous, it follows that $H(R)$ is also separable. In a moment we shall look at two examples of these spaces, and more can be found among the exercises. In general, however, the RKHS is a rather nebulous concept, good more for proving theorems than anything else. The principle exception is the case $T = \Re^1$, in which case the RKHS has been exploited as an important tool in the detection and estimation problems of communication theory. It is also in this case that the RKHS is comparatively easy to identify. ■

For the first example, however, take $T = \{1, \dots, N\}$, finite, and X centered Gaussian with covariance matrix $R = (r_{ij})$. Let $R^{-1} = (r^{ij})$ denote the inverse of R. Then the RKHS of X is made up of all N-dimensional vectors $f = \langle f_1, \dots, f_N \rangle$ with inner product

$$(f, g)_H = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i r^{ij} g_j.$$

To prove this, we need only check that the reproducing kernel property (3.9) holds. But, with $\delta(i, j)$ the Kronecker delta function, and R_k denoting the k-th row of R,

$$\begin{aligned}(f, R_k)_H &= \sum_{i=1}^{N} \sum_{j=1}^{N} f_i r^{ij} r_{kj} \\ &= \sum_{i=1}^{N} f_i \delta(i, k) \\ &= f_k,\end{aligned}$$

as required. ■

For a slightly more interesting example, take X to be standard Brownian motion on $T = [0, 1]$, so that $R(s, t) = \min(s, t)$. Note that the function $R(s, \cdot)$ is differentiable everywhere except at s, so that following the heuristics developed above we expect that $H(R)$ should be made up of a subset of functions that are differentiable almost everywhere.

To both make this statement more precise, and prove it, we start by looking at what the space S looks like. Thus let

$$f(t) = \sum_{i=1}^{n} a_i R(s_i, t), \qquad g(t) = \sum_{i=1}^{n} b_i R(t_i, t)$$

be two elements of S. According to (3.8), the S-inner product between them is given by

$$(f, g)_H = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j \min(s_i, t_j).$$

Note that the derivative of $R(s, t)$ with respect to t is given by $1_{[0,s]}(t)$, where 1_A is the indicator function of $A \subset T$, and so the derivative of f is given by $\sum_{i=1}^n a_i 1_{[0,s_i]}(t)$. Therefore, we can rewrite the above as follows:

$$\begin{aligned}(f,g)_H &= \sum_{i=1}^n \sum_{j=1}^n a_i b_j \int_0^1 1_{[0,s_i]}(t) 1_{[0,t_j]}(t)\, dt \\ &= \int_0^1 \sum_{i=1}^n a_i 1_{[0,s_i]}(t) \sum_{j=1}^n b_j 1_{[0,t_j]}(t)\, dt \\ &= \int_0^1 \dot f(t)\dot g(t)\, dt.\end{aligned}$$

We can now go about treating the general case. Set

$$H = \left\{f: f(t) = \int_0^t \dot f(s)\, ds, \quad \int_0^1 \left(\dot f(s)\right)^2 ds < \infty\right\}, \tag{3.12}$$

and define the following inner product on H:

$$(f,g)_H = \int_0^1 \dot f(s)\dot g(s)\, ds. \tag{3.13}$$

Since it is immediate that $R(s,\cdot) \in H$ for $t \in [0,1]$, and

$$\left(f, R(t,\cdot)\right)_H = \int_0^1 \dot f(s) 1_{[0,t]}(s)\, ds,$$

it now follows that $H = H(R)$. That is, the RKHS is, in this case, determined by the space (3.12) and the inner product (3.13). ∎

More examples can be found in the exercises. Now, however, we shall look at the RKHS from a slightly different viewpoint, and see how to use it in a very practical fashion.

Since $H(R)$ is a separable Hilbert space, it must have a countable orthonormal basis. This will be extremely important for us. Sometimes this basis is easy to find, particularly in the case $T = [0,1]^k$, in which case it leads us to the *Karhunen-Loève expansion* of X, a topic we shall cover in the following section. Sometimes, we have to be content with merely knowing it exists. In an abstract setting, however, even this is useful, as we shall soon see.

Define $\mathcal{H}^1(X)$, the so-called "linear part" of the $\mathcal{L}^2$ space of the process X, as the closure in $\mathcal{L}^2(P) = \mathcal{L}^2(\Omega, \mathcal{F}, P)$ of

$$\left\{\sum_{i=1}^n a_i X(t_i), \ a_i \text{ real}, \ t_i \in T, \ n \geq 1\right\}, \tag{3.14}$$

thinned out by identifying all elements indistinguishable in $\mathcal{L}^2(P)$. (i.e. elements U, V for which $E(U-V)^2 = 0$.) This contains all distinguishable random variables, with finite variance, obtainable as linear combinations of values of the process. There is a linear, one-one mapping between the space S of real-valued functions on T to this $\mathcal{L}^2$ space, defined by

$$\Theta(f) = \Theta\Big(\sum_{i=1}^n a_i R(t_i,\cdot)\Big) = \sum_{i=1}^n a_i X(t_i).$$

Note that Θ is clearly norm preserving, and so extends to all of $H(R)$ with range equal to all of $\mathcal{H}^1(X)$. The extension is called the *canonical isomorphism* between these spaces.

Since $H(R)$ is separable, we now also know that $\mathcal{H}^1(X)$ is. We can use this to build an orthonormal basis for $\mathcal{H}^1(X)$, for if $\{\phi_n\}_{n\geq 1}$ is an orthonormal basis for $H(R)$, then setting $\xi_n = \Theta(\phi_n)$ gives $\{\xi_n\}_{n\geq 1}$ as an orthonormal basis for $\mathcal{H}^1(X)$. In particular, we must have $E\xi_n = 0$ for all $n \geq 1$, and

$$X_t = \sum_{n=1}^{\infty} \xi_n E(X_t \xi_n), \tag{3.15}$$

where the series converges in $\mathcal{L}^2(P)$. Since Θ was an isometry, it follows from (3.15) that

$$\begin{aligned} EX_t\xi_n &= (R(t,\cdot),\phi_n)_H \\ &= \phi_n(t) \end{aligned} \tag{3.16}$$

the last equality coming from the reproducing kernel property of $H(R)$. Putting (3.16) together with (3.15) is almost enough to give the following central result.

3.7 THEOREM. *If $\{\phi_n\}_{n\geq 1}$ is an orthonormal basis for $H(R)$, then X has the $\mathcal{L}^2$-representation*

$$X_t = \sum_{n=1}^{\infty} \xi_n \phi_n(t), \tag{3.17}$$

where $\{\xi_n\}_{n\geq 1}$ is the orthonormal sequence of centered Gaussian variables given by $\xi_n = \Theta(\phi_n)$.

PROOF: We have proven everything other than the fact that the ξ_n are Gaussian. But this follows from standard properties of Gaussian random variables since a countable, $\mathcal{L}^2$ convergent sum of Gaussian random variables will itself always be Gaussian. ■

The equivalence in (3.17) is only in $\mathcal{L}^2$; i.e. whereas X_t is defined with probability one, the sum is, in general, convergent only in mean square. The following result is indicative of how much we get for free once we have an a.s. continuous process.

3.8 THEOREM. *If X has continuous sample paths with probability one, then the sum in (3.17) converges uniformly on T with probability one.*

There is also a converse result, that the a.s. uniform convergence of a sum like (3.17) implies the continuity of X. But since we shall soon have better ways of establishing sample function continuity, we have no need of the converse. We shall require one non-standard result from probability theory in order to prove Theorem 3.8. It is due to Itô and Nisio (1968).

3.9 LEMMA. *Let $\{Z_n\}_{n\geq 1}$ be a sequence of symmetric independent random variables, taking values in a separable, real Banach space B, equipped with the norm topology. Let $X_n = \sum_{i=1}^n Z_i$. Then X_n converges with probability one if, and only if, there exists a B-valued random variable X such that $\langle x^*, X_n\rangle \to \langle x^*, X\rangle$ in probability for every $x^* \in B^*$, the topological dual of B.*

We also require another preliminary result, of considerable intrinsic interest.

3.10 LEMMA. *If $\{\phi_n\}_{n\geq 1}$ is an orthonormal basis for $H(R)$, then*

$$\sum_{n=1}^{\infty} \phi_n^2(t)$$

converges uniformly in $t \in T$ to $R(t,t)$.

PROOF: By the orthonormal expansion and the reproducing kernel property,

$$\begin{aligned} R(t,\cdot) &= \sum_{n=1}^{\infty} \phi_n(\cdot)\big(R(t,\cdot),\phi_n\big)_H \\ &= \sum_{n=1}^{\infty} \phi_n(\cdot)\phi_n(t), \end{aligned} \tag{3.18}$$

convergence of the sum being in the $\|\cdot\|_H$ norm. Hence, $\sum_{n=1}^{\infty} \phi_n^2(t)$ converges to $R(t,t)$ for every $t \in T$. Furthermore, the convergence is monotone, and so it follows that it is also uniform ($\equiv$ Dini's theorem). ∎

We can now turn to the

PROOF OF THEOREM 3.8: We know that, for each $t \in T$, $\sum_{n=1}^{\infty} \xi_n \phi_n(t)$ is a sum of independent variables converging in $\mathcal{L}^2(P)$. Thus, by Lemma 3.9, applied to real-valued random variables, it converges with probability one to a limit we denote by X_t. Since the limit process is, by assumption, continuous, this defines the same process as that appearing in the $\mathcal{L}^2$ sense in Theorem 3.7.

Now, consider both X and each function $\xi_n\phi_n(\cdot)$ as elements of the Banach space $C(T)$, with sup-norm topology, and define

$$X_n(\cdot) = \sum_{i=1}^{n} \xi_i\phi_i(\cdot) = \sum_{i=1}^{n} \Theta(\phi_i)\phi_i(\cdot).$$

By Lemma 3.9, it suffices to show that for every $x^* \in C^*(T)$ the random variables $\langle x^*, X_n\rangle$ converge in probability to $\langle x^*, X\rangle$.

Recall that every x^* in the topological dual of $C(T)$ is a finite, signed, Borel measure on T. Thus,

$$\begin{aligned} E|\langle x^*, X_n\rangle - \langle x^*, X\rangle| &= E|\int_T (X_n(t) - X(t))x^*(dt)| \\ &\leq \int_T E|X_n(t) - X(t)| \cdot |x^*|(dt) \\ &\leq \int_T \left[E(X_n(t) - X(t))^2\right]^{\frac{1}{2}} |x^*|(dt) \\ &= \int_T \Big(\sum_{j=n+1}^{\infty} \phi_j^2(t)\Big)^{\frac{1}{2}} |x^*|(dt), \end{aligned}$$

where $|x^*|(A)$ is the total variation of x^* on $A \subset T$.

Since $\sum_{j=n+1}^{\infty} \phi_j^2(t) \to 0$ uniformly in $t \in T$ by Lemma 3.10, the last expression above tends to zero as $n \to \infty$. Since this implies the convergence in probability of $\langle x^*, S_n\rangle$ to $\langle x, X\rangle$, we are done. ■

Theorem 3.8 is nice to have, (especially since it comes for free with a.s. continuity) but it is really Theorem 3.7 that is the more important. The fact that X can be written as an infinite sum of independent random variables implies that any property of X that is a "tail event", in that its occurrence or non-occurrence does not depend on "the first n among the ϕ_j", can be expected to have probability zero or one only. This is in fact the case, as we shall see in a moment. Firstly, however, we need a definition.

The *oscillation function* w_X of X is defined as

$$w_X(t) = \lim_{\epsilon \downarrow 0} \sup_{u,v \in B(t,\epsilon)} |X_u - X_v|,$$

where $B(t, \epsilon)$ is the d-ball of radius epsilon centered on $t \in T$. Here is a key result:

3.11 THEOREM. *There exists a (non-random) $\bar{\Re}$-valued upper-semicontinuous function $h(t)$ on T such that*

$$P\{w_X(t) = h(t), \text{ for all } t \in T\} = 1. \tag{3.19}$$

(Recall that $f: T \to \bar{\Re}$ is called upper semi-continuous if $\overline{\lim}_{n\to\infty} f(t_n) = f(t)$ for every sequence $\{t_n\}$ with $\lim_{n\to\infty} t_n = t$.)

Before we give a proof of Theorem 3.11, consider some of the consequences of this result.

PROOF OF THEOREM 3.6: Since h is non-random, the condition of a.s. continuity of X at each point $t \in T$ implies that $h(t) \equiv 0$. Theorem 3.11 allows us to move the "each point" inside the probability statement. ∎

3.12 THEOREM.

$$P\{X \textit{ is continuous for all } t \in T\} = 0 \textit{ or } 1.$$

PROOF: Let h be the function of Theorem 3.11. If $h(t) > 0$ for some t, then, by Theorem 3.11, X is discontinuous at that point, and so $P\{X$ is continuous for all $t \in T\} = 0$. If $h(t) = 0$ for all t, then Theorem 3.11, following the lines of the above proof, also shows that X is continuous with probability one. ∎

The theory of Gaussian processes is rich in such zero-one laws, the most famous perhaps being Belyaev's (1961) dichotomy of stationary process as being either a.s. continuous or a.s. unbounded on every open subset of T, a result which follows from arguments similar to those above. In Exercise 2.3 you can work through a simple proof of a very general zero-one law for vector valued Gaussian variables, that has an elegant extension to processes.

Here is a very general zero-one law for Gaussian processes, that we shall not, however, prove. (c.f. Kallianpur (1969, 1970), Jain (1971) and Cambanis and Rajput, (1973) for details.)

As always, X_t, $t \in T$ is a centered Gaussian process on a probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{F}$ is assumed to be P-complete. Now let Λ be a space of functions which contains, with probability one, the sample paths of X. Let $\mathcal{B}(\Lambda)$ be the cylindrical σ-algebra of subsets of Λ, define $\Phi: (\Omega, \mathcal{F}, P) \to (\Lambda, \mathcal{B}(\Lambda))$ as

$$\Phi(\omega) = \begin{cases} X(\cdot,\omega) & \text{if } X(\cdot,\omega) \in \Lambda, \\ 0 & \text{if } X(\cdot,\omega) \notin \Lambda. \end{cases}$$

Then Φ clearly induces a probability measure μ on $(\Lambda, \mathcal{B}(\Lambda))$. Let $\bar{\mathcal{B}}(\Lambda)$ be the completion of $\mathcal{B}(\Lambda)$ with respect to μ.

We assume the following two conditions:

(A) Λ is a linear function space under addition of functions and multiplication by scalars.

(B) $H(R) \subset \Lambda$.

3.13 THEOREM. *If (A) and (B) are satisfied, and G is a $\bar{\mathcal{B}}(\Lambda)$-measurable subgroup of Λ, then $\mu(G) = 0$ or 1.*

An immediate corollary of this result is the following:

3.14 COROLLARY. *Let Λ be a set of functions which contains the paths of X, P a.s., and satisifies (A) and (B) above. If $F \in \mathcal{F}$ is such that $F = \Phi^{-1}(G)$, where G is a $\bar{\mathcal{B}}(\Lambda)$-measurable subgroup of Λ, then $P(F) = 0$ or 1.*

You can find a number of applications of this important result in Exercise 2.4. Among them is an alternative proof of Theorem 3.12.

Now, however, we return to a

PROOF OF THEOREM 3.11: Let $B \subseteq T$ be closed, and set

$$w_X(B) = \sup_{s,t \in B} |X_t - X_s|. \tag{3.20}$$

Separability ensures that the $w_X(t)$ of the theorem, as well as $w_X(B)$, are well defined random variables. Let

$$X^{(n)}(t) = \sum_{j=n+1}^{\infty} \xi_j \phi_j(t).$$

Since the ϕ_j are continuous, so must be $\sum_{j=1}^{n} \xi_j \phi_j$, for any $n \geq 1$. It thus follows that for each $w \in \Omega$, $n \geq 1$, and any $B \subset T$,

$$w_{X^{(n)}}(B, w) = w_X(B, w).$$

Since $w_{X^{(n)}}(S)$ is measurable with respect to the σ-algebra generated by $\xi_{n+1}, \xi_{n+2}, \ldots$, it follows by Kolmogorov's zero-one law that, for some fixed number $h = h(B)$,

$$P\{w_X(B, w) = h(B)\} = 1.$$

This defines a set indexed version of h. We still need to exhibit a point indexed version, and to show that it is upper semi-continuous. To this end, let $\mathcal{B}$ be a countable open basis for the topology on T generated by open balls with respect to d, and for $t \in T$ set

$$h(t) = \inf_{B \in \mathcal{B}\,:\, t \in B} h(B).$$

It is immediate from the definition that h is upper-semicontinuous, and not hard to see that

$$\begin{aligned} w_X(t) &= \inf_{B \in \mathcal{B}\,:\, t \in B} w_X(B) \\ &\stackrel{\text{as}}{=} \inf_{B \in \mathcal{B}\,:\, t \in B} h(B) \\ &= h(t), \end{aligned}$$

which is all we need to complete the proof. ∎

3. The Karhunen-Loève Expansion.

The Karhunen-Loève expansion of a continuous Gaussian process is really just a special case of the expansion (3.17) when the parameter space T is a compact subset of $\Re^k$. In this case it is possible to choose particularly convenient orthonormal bases for $H(R)$ and $\mathcal{H}^1(X)$.

Thus, let $T = [0,1]^k$, and X a centered Gaussian process on T with continuous covariance function $R(s,t)$. Let $\lambda_1, \lambda_2, \ldots$, and $\psi_1, \psi_2, \ldots$, be, respectively, the eigenvalues and normalised eigenfunctions of the operator $\mathcal{R}: \mathcal{L}^2(T) \to \mathcal{L}^2(T)$ defined by $\mathcal{R}\psi(t) = \int_T R(s,t)\psi(s)\,ds$. That is, the λ_n and ψ_n solve the integral equation

$$\int_T R(s,t)\psi(s)\,ds \;=\; \lambda\psi(t), \qquad \text{for all } t \in T, \tag{3.23}$$

and

$$\int_T \psi_n(t)\psi_m(t)\,dt \;=\; \begin{cases} 1 & \text{for } n = m, \\ 0 & \text{for } n \neq m. \end{cases}$$

Since it involves no loss of generality, we shall assume that $\lambda_1 \geq \lambda_2 \geq \ldots$.

The following result can be found, for example, in Riesz and Sz-Nagy (1955) when $k = 1$ or Zaanen (1956) for general k.

3.15 THEOREM (MERCER). *Let R, $\{\lambda_n\}_{n\geq 1}$ and $\{\psi_n\}_{n\geq 1}$ be as above. Then*

$$R(s,t) \;=\; \sum_{n=1}^{\infty} \lambda_n \psi_n(s)\psi_n(t)\,, \tag{3.24}$$

where the series converges absolutely and uniformly on $[0,1]^k \times [0,1]^k$.

The claims made above are best summarised as

3.16 THEOREM. *Let R, $\{\lambda_n\}_{n\geq 1}$ and $\{\psi_n\}_{n\geq 1}$ be as defined above. Then $\{\sqrt{\lambda_n}\psi_n\}$ is a complete orthormal system in $H(R)$.*

PROOF: Set $\phi_n = \sqrt{\lambda_n}\psi_n$ and define

$$H \;=\; \Big\{ f\colon f(t) = \sum_{n=1}^{\infty} a_n \phi_n(t),\; t \in [0,1]^k,\; \sum_{n=1}^{\infty} a_n^2 < \infty \Big\}.$$

Give H the inner product

$$(f,g)_H \;=\; \sum_{n=1}^{\infty} a_n b_n\,,$$

where $f = \sum a_n \phi_n$ and $g = \sum b_n \phi_n$.

To check that H has the reproducing kernel property, note that

$$\begin{aligned}
\big(f(\cdot), R(t,\cdot)\big)_H &= \Big(\sum_{n=1}^{\infty} a_n \phi_n(\cdot), \sum_{n=1}^{\infty} \sqrt{\lambda_n}\psi_n(t)\phi_n(\cdot)\Big) \\
&= \sum_{n=1}^{\infty} \sqrt{\lambda_n} a_n \psi_n(t) \\
&= f(t).
\end{aligned}$$

It remains to be checked that H is in fact a Hilbert space, and that $\{\sqrt{\lambda_n}\psi_n\}$ is both complete and orthonormal. But all this is standard, given Mercer's theorem, and so is left to you. ■

Remaining with the basic notation of Mercer's theorem, we thus have that the RKHS, $H(R)$, consists of all square integrable functions f on $[0,1]^k$ for which

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} \left| \int_T f(t)\psi_n(t)\, dt \right|^2 < \infty,$$

with inner product

$$(f,g)_H = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \int_T g(t)\psi_n(t)\, dt \int_T g(t)\psi_n(t)\, dt.$$

The Karhunen-Loève expansion of X is obtained by setting $\phi_n = \lambda_n^{\frac{1}{2}} \psi_n$ in the orthonormal expansion (3.17), so that

$$X_t = \sum_{n=1}^{\infty} \lambda_n^{\frac{1}{2}} \xi_n \psi_n(t), \tag{3.25}$$

where the ξ_n are orthonormal Gaussian.

We shall give one, classic, example – that of standard Brownian motion on $[0,1]$. Unfortunately, not too many examples are known, since the integral equation (3.23) is generally not easy to solve. Nevertheless, the fact that it is always possible to solve (3.23) numerically implies that Karhunen-Loève expansions are of substantial practical importance in a variety of applied settings, most notably communication theory.

For Brownian motion (3.23) becomes

$$\begin{aligned}
\lambda\psi(t) &= \int_0^1 \min(s,t)\psi(s)\, ds \\
&= \int_0^t s\psi(s)\, ds + t\int_t^1 \psi(s)\, ds.
\end{aligned}$$

Differentiating both sides with respect to t gives

$$\begin{aligned}\lambda\psi'(t) &= \int_t^1 \psi(s)\,ds,\\ \lambda\psi''(t) &= -\psi(t),\end{aligned}$$

together with the boundary condition $\psi(0) = 0$.

The solutions of this pair of differential equations are given by

$$\psi_n(t) = \sqrt{2}\sin\left(\tfrac{1}{2}(2n+1)\pi s\right), \qquad \lambda_n = \left(\frac{2}{(2n+1)\pi}\right)^2,$$

as is easily verified by substitution. Thus, the Karhunen-Loève expansion of Brownian motion on $[0,1]$ is given by

$$W_t = \frac{\sqrt{2}}{\pi}\sum_{n=0}^{\infty}\xi_n\left(\frac{2}{2n+1}\right)\sin\left(\tfrac{1}{2}(2n+1)\pi t\right),$$

where (ξ_n) is an orthonormal Gaussian sequence.

4. Exercises.

SECTION 3.1:

1.1 Fernique's (1978) original proof of Theorem 3.2 is of intrinsic interest. His argument for showing that the finiteness of $\|X\|$ implies the existence of exponential moments is roughly as follows. You should fill in the details for yourself.

Let X^1 and X^2 be two independent copies of X. Then, for every pair (a,b) of reals

$$\begin{aligned}&P\{\|X^1\| \le a\}\,P\{\|X^2\| > b\}\\ &\quad = P\{\|X^1 + X^2\| \le \sqrt{2}a,\ \|X^1 - X^2\| > \sqrt{2}b\},\end{aligned}$$

from which it follows that

$$P\{\|X^1\| \le a\}P\{\|X^2\| > b\} \le \left(P\{\|X\| > (b-a)/\sqrt{2}\}\right)^2.$$

Choose $a > 0$ such that $q := P\{\|X\| \le a\} \in (\tfrac{1}{2}, 1)$, set

$$b_0 = a, \qquad b_{n+1} = a + b_n\sqrt{2}, \qquad n \ge 0,$$

as well as $x_0 = (1-q)/q < 1$, and define x_n, $n \ge 1$ by

$$P\{\|X\| > b_n\} = qx_n.$$

It follows by induction that

$$P\{\|X\| > b_n\} \;\le\; q\Big(\frac{1-q}{q}\Big)^{2^n}.$$

Since the recursive definition of the b_n implies that

$$\begin{aligned} b_n &= a\big(1+\sqrt{2}+\cdots+(\sqrt{2})^n\big) \\ &= a(\sqrt{2}+1)(2^{(n+1)/2}-1), \end{aligned}$$

it follows that

$$Ee^{\alpha\|X\|^2} \le q\Big(e^{\alpha a^2} + \sum_{n=0}^{\infty}\Big(\frac{1-q}{q}\Big)^{2^n} \exp\big(\alpha(\sqrt{2}+1)^2(2^{(n+2)/2}-1)^2a^2\big)\Big),$$

which converges for small enough α.

1.2 An interesting side result, that is any easy consequence of the above calculations, is the following:

If X is centered Gaussian, then there exists a universal constant C such that

$$P\{\|X\| \le \lambda\} \ge \tfrac{1}{2} \qquad \Longrightarrow \qquad E\|X\| \le C\lambda.$$

Prove this.

SECTION 3.2:

2.1 Let X be an Ornstein-Uhlenbeck process on $T = [a,b]$, $-\infty < a, b < \infty$; i.e. the centered, stationary Gaussian process on T with covariance function $R(s,t) = e^{-\alpha|s-t|}$. Show that the RKHS of X is made up of all absolutely continuous functions with inner product

$$\begin{aligned} (f,g)_H &= f(a)g(a) + \frac{1}{2\alpha}\int_T \big(f'(t)+\alpha f(t)\big)\big(g'(t)+\alpha g(t)\big)\,dt \\ &= \tfrac{1}{2}\big(f(a)g(a)+f(b)g(b)\big) + \frac{1}{2\alpha}\int_T \big(f'(t)g'(t)+\alpha^2 f(t)g(t)\big)\,dt. \end{aligned}$$

What happens when $T = \Re$?

2.2 Find the RKHS, along with the appropriate inner product, for the Brownian sheet on $[0,1]^k$.

2.3 Here is a very basic zero-one law for vector valued Gaussian variables.

Let X be a $\Re^k$-valued, centered, Gaussian variable, and E a subspace of $\Re^k$. Let X_1 and X_2 be independent copies of X, and for $\theta \in [0, \pi/2]$ set

$$A(\theta) = \{X_1\cos\theta + X_2\sin\theta \in E,\; X_1\sin\theta - X_2\cos\theta \notin E\}.$$

(i) Show that $P\{A(\theta)\}$ is independent of θ.
(ii) Show that if $\theta_1 \neq \theta_2$, but $x_1, x_2 \in \Re^k$ are such that both $x_1 \cos\theta_1 + x_2 \sin\theta_1 \in E$ and $x_1 \cos\theta_2 + x_2 \sin\theta_2 \in E$, then both $x_1 \in E$ and $x_2 \in E$, and so $x_1 \sin\theta_1 - x_2 \cos\theta_1 \in E$. Conclude that the events $A(\theta)$ are thus disjoint.
(iii) Show that $P\{X \in E\} = 0$ or 1.
(iv) Let $\|X\| = \sup_{i=1,\dots,k} |X(i)|$. Show that $P\{\|X\| < \infty\} = 0$ or 1.
(v) Note how easy it is to extend the above to X taking values in any vector space Λ on which $\| \ \|$ is a pseudo-semi-norm. (i.e. a mapping from Λ to $(\bar{\Re}, \mathcal{B}(\bar{\Re}))$ for which $(\| \ \|)^{-1}(\Re)$ is a subspace of Λ on which $\| \ \|$ induces a semi-norm.)

2.4 Apply Corollary 3.14 to show that if X is a Gaussian process on a bounded $T \subset \Re$ then with probability one or zero the paths of X
(i) are bounded on T,
(ii) are continuous on T,
(iii) are free of oscillatory discontinuities on T,
(iv) satisfy a Hölder condition on T.

2.5 Complete the proof of the sufficiency part of Theorem 1.5.

SECTION 3.3:

3.1 Let X_t be the centered Gaussian process on $\Re$ with covariance function $R(s,t) = R(t-s) = \cos 2\pi(t-s)$. Find the Karhunen-Loève expansion of X and show that it has only two terms.