Lecture 6

Concentration for the maximum

In this lecture we will establish bounds on the maximum of Gaussian random variables which will not be based on comparisons but rather on the behavior of the covariance kernel. The first result to be proved here is the Borell-Tsirelson-Ibragimov-Sudakov inequality on concentration of the maximum. Any use of this inequality will inevitably require controlling the expected maximum, which we do by way of Fernique's majorization technique. Once these are stated and proved, we will infer some standard but useful consequences concerning boundedness and continuity of centered Gaussian processes.

6.1. Inheritance of Gaussian tails

Much of present day probability hinges on the phenomenon of *concentration of measure*. For Gaussian random variables this is actually a very classical subject. The relevant inequality that comes out of this is the content of:

Theorem 6.1 [Borell-TIS inequality] *Let X be a centered Gaussian on $\mathbb{R}^n$ and set*

$$\sigma_X^2 := \max_{i=1,\dots,n} E(X_i^2). \tag{6.1}$$

Then for each $t > 0$,

$$P\Big(\big|\max_{i=1,\dots,n} X_i - E(\max_{i=1,\dots,n} X_i)\big| > t\Big) \le 2\mathrm{e}^{-\frac{t^2}{2\sigma_X^2}}. \tag{6.2}$$

This result can be verbalized as: *The tail of the maximum of Gaussian random variables is no worse than the worst tail seen among these random variables*. Of course, the maximum is no longer centered (cf Exercise 5.11) and so any use of this bound requires information on the expectation as well.

The original proof of this result was given by Borell using a Gaussian isoperimetric inequality. We will instead proceed using the ideas of hypercontractivity whose main output is encapsulated in:

Lemma 6.2 *Let $X_1,\dots,X_n$ be i.i.d. copies of $\mathcal{N}(0,1)$ and let $f,g\in C^1(\mathbb{R}^n)$ be such that $\nabla f,\nabla g\in L^2(\mathrm{e}^{-|x|^2/2}\mathrm{d}x)$. Then*

$$\operatorname{Cov}\big(f(X),g(X)\big)=\int_0^1\mathrm{d}t\,E\Big(\nabla f(X)\cdot\nabla g\big(\sqrt{1-t^2}X+tY\big)\Big),\tag{6.3}$$

where $Y\overset{\text{law}}{=}X$ with $Y\perp\!\!\!\perp X$ on the right-hand side.

Proof. Since this is an equality between fairly manageable expressions for two functions, the identity can be proved by checking that it holds for a sufficiently large class of functions (e.g., $x\mapsto\mathrm{e}^{k\cdot x}$) and then use extension arguments. We will instead proceed by Gaussian integration by parts.

For X and Y as above, let $Z_t:=tX+\sqrt{1-t^2}\,Y$. By approximation arguments may assume $g\in C^2$ with subgaussian tail. Then

$$\begin{aligned}\operatorname{Cov}\big(f(X),g(X)\big)&=E\Big(f(X)\big[g(Z_1)-g(Z_0)\big]\Big)\\&=\int_0^1\mathrm{d}t\,E\Big(f(X)\frac{\mathrm{d}}{\mathrm{d}t}g(Z_t)\Big)\\&=\int_0^1\mathrm{d}t\sum_{i=1}^n E\bigg(\Big[X_i-\frac{t}{\sqrt{1-t^2}}Y_i\Big]f(X)\frac{\partial g}{\partial x_i}(Z_t)\bigg)\end{aligned}\tag{6.4}$$

The integration by parts will result in two contributions depending on whether the derivative hits f or (the partial derivative of) g. In light of the i.i.d. nature of the random variables, the contribution of the latter option cancels out and we then readily get the result. $\square$

As a side note, we notice that this implies:

Corollary 6.3 [Gaussian Poincaré inequality] *For $X_1,\dots,X_n$ i.i.d. copies of $\mathcal{N}(0,1)$ and any $f\in C^1(\mathbb{R}^n)$ with $\nabla f\in L^2(\mathrm{e}^{-|x|^2/2}\mathrm{d}x)$,*

$$\operatorname{Var}\big(f(X)\big)\le E\big(|\nabla f(X)|^2\big).\tag{6.5}$$

An important feature of this bound is that it is completely dimension-less — i.e., with no n dependence of the (implicit) constant on the right-hand side.

Moving along with the proof of the Borell-TIS inequality, next we will prove:

Lemma 6.4 [Concentration for Lipschitz functions] *Let $X_1,\dots,X_n$ be i.i.d. copies of $\mathcal{N}(0,1)$ and let $f\colon\mathbb{R}^n\to\mathbb{R}$ be Lipschitz in the sense that, for some $M\in(0,\infty)$,*

$$\big|f(x)-f(y)\big|\le M|x-y|,\quad x,y\in\mathbb{R}^n,\tag{6.6}$$

where $|\cdot|$ on the right-hand side is the Euclidean norm. Then for each $t>0$,

$$P\big(f(X)-Ef(X)>t\big)\le\mathrm{e}^{-\frac{t^2}{2M^2}}\tag{6.7}$$

Proof. By approximation we may assume that $f \in C^1$ with ∇f having Euclidean norm at most M. Shifting f appropriately, we may also assume $Ef(X) = 0$. (This does not effect the bound (6.7).) Chebyshev's inequality shows

$$P\big(f(X) - Ef(X) > t\big) \le \mathrm{e}^{-\lambda t} E\big(\mathrm{e}^{\lambda f(X)}\big) \tag{6.8}$$

for any $\lambda \ge 0$ and so we need to bound the expectation on the right.

Here we note that Lemma 6.2 with $g(x) := \mathrm{e}^{\lambda f(x)}$ implies

$$E\big(f(X)\mathrm{e}^{\lambda f(X)}\big) = \int_0^1 \mathrm{d}t\, \lambda E\Big(\nabla f(X) \cdot \nabla f(Z_t) \mathrm{e}^{\lambda f(Z_t)}\Big) \overset{\lambda \ge 0}{\le} \lambda M^2 E\big(\mathrm{e}^{\lambda f(X)}\big) \tag{6.9}$$

As the left-hand side is the derivative of the expectation on the right-hand side, the function

$$h(\lambda) := E\big(\mathrm{e}^{\lambda f(X)}\big), \tag{6.10}$$

obeys the differential inequality

$$h'(\lambda) \le \lambda M^2 h(\lambda), \qquad \lambda \ge 0. \tag{6.11}$$

As $h(0) = 1$, this is readily solved to give

$$E\big(\mathrm{e}^{\lambda f(X)}\big) \le \mathrm{e}^{\frac{1}{2}\lambda^2 M^2}. \tag{6.12}$$

Inserting this into (6.8) and optimizing over $\lambda \ge 0$ then yields the claim. □

We are now ready to prove the Borell-TIS inequality, we will also need:

Exercise 6.5 *Denote $f(x) := \max_{i=1,\dots,n} x_i$. Prove that for any $n \times n$-matrix A,*

$$\big|f(Ax) - f(Ay)\big| \le \sqrt{\max_{i=1,\dots,n} (A^T A)_{ii}}\; |x - y|, \quad x, y \in \mathbb{R}^n, \tag{6.13}$$

with $|x - y|$ denoting the Euclidean norm of $x - y$ on the right-hand side.

Proof of Theorem 6.1. Let X be the centered Gaussian on $\mathbb{R}^n$ from the statement and let C denote its covariance matrix. In light of positive semi-definiteness of C, there is an $n \times n$-matrix A such that $C = A^{\mathrm{T}} A$. If $Z = (Z_1, \dots, Z_n)$ are i.i.d. copies of $\mathcal{N}(0,1)$, then

$$X \overset{\text{law}}{=} AZ. \tag{6.14}$$

Denoting $f(x) := \max_{i=1,\dots,n} x_i$, Exercise 6.5 implies that $x \mapsto f(Ax)$ is Lipschitz with Lipschitz constant σ_X^2. Using this in combination with (6.14), the claim follows from (6.7) and a union bound. □

6.2. Fernique majorization

Our next task will be to introduce a method for estimating the expected maximum of Gaussian random variables. We will actually do this for the supremum over a countable family of such variables as that requires no additional effort. A principal notion here is that of the canonical (pseudo)metric ρ_X associated with the Gaussian process $\{X_t \colon t \in T\}$ on any set T. Our principal result here is:

Theorem 6.6 [Fernique majorization] *There is $K \in (0,\infty)$ such that the following holds for any Gaussian process $\{X_t : t \in T\}$ over a countable set T for which (T, ρ_X) is totally bounded: For any probability measure μ on T, we have*

$$E\big(\sup_{t\in T} X_t\big) \le K \sup_{t\in T} \int_0^\infty \mathrm{d}r \sqrt{\log \frac{1}{\mu(B(t,r))}} \tag{6.15}$$

where $B(t,r) := \{s \in T : \rho_X(t,s) < r\}$.

The integral exists in Riemann sense as the integrand is non-increasing and left-continuous while the domain of integration is bounded because $\mu(B(t,r)) = 1$ whenever r exceeds the ρ_X-diameter of T.

The above theorem takes its origin in:

Theorem 6.7 [Dudley's inequality] *For the same setting as in the previous theorem, there is a universal constant $K \in (0,\infty)$ such that*

$$E\big(\sup_{t\in T} X_t\big) \le K \int_0^\infty \mathrm{d}r \sqrt{\log N_X(r)}\,, \tag{6.16}$$

where $N_X(r)$ is the minimal number of ρ_X-balls of radius r that are needed to cover T.

The advantage of Fernique's bound over Dudley's inequality is that it allows optimizing over the probability measure μ. By a celebrated result due to Talagrand, the optimal choice of μ in fact leads to an *sharp* bound on the expected maximum; i.e., one where the integral bounds the expectation from below modulo a universal multiplicative constant. One way to think of the optimizers (although this has not been made rigorous) of the above integral is as the distribution of t where the maximum is achieved.

As we will expound on later, the setting of the above theorems is so general that they fairly seamlessly connect boundedness of Gaussian processes to sample-path continuity. Here is an exercise in this vain:

Exercise 6.8 *Apply Dudley's inequality to the process $X_{t,s} := X_t - X_s$ to prove*

$$E\Big(\sup_{\substack{t,s\in T\\ \rho_X(t,s)\le R}} |X_t - X_s|\Big) \le K' \int_0^R \mathrm{d}r \sqrt{\log N_X(r)}\,, \tag{6.17}$$

where K' is again a universal constants. Conclude that if $r \mapsto \sqrt{\log N_X(r)}$ is integrable at zero, then $t \mapsto X_t$ has a version with (uniformly) ρ_X-continuous sample paths a.s.

To see this exercise in action, it is instructive to solve:

Exercise 6.9 *Use Dudley's inequality to prove the existence of a ρ_X-continuous version for the following Gaussian processes:*

(1) the standard Brownian motion, i.e., a centered Gaussian process $\{B_t : t \in [0,1]\}$ with $E(B_t B_s) = t \wedge s$,

(2) *the Brownian sheet, i.e., a centered Gaussian process* $\{W_t \colon t \in [0,1]^d\}$

$$E(W_t W_s) = \prod_{i=1}^{d} (t_i \wedge s_i) \tag{6.18}$$

(3) *any centered Gaussian process* $\{X_t \colon t \in [0,1]\}$ *such that*

$$E\big([X_t - X_s]^2\big) \le c[\log(1/|t-s|)]^{-1-\delta} \tag{6.19}$$

for some $\delta > 0$ *and* $c > 0$ *and* $|t-s|$ *sufficiently small.*

A proof of continuity of these processes is just as well provided by the Komogorov-Čenstov condition. As can be also checked, both techniques give a way to prove uniform Hölder continuity of these processes as well.

6.3. Proof of Fernique's estimate

Here we will give the proof of Fernique's bound but before we set out to do so, let us outline its main idea. The basic strategy is simple: we identify an auxiliary centered Gaussian process $\{Y_t \colon t \in T\}$ whose intrinsic distance function ρ_Y dominates ρ_X. The Sudakov-Fernique inequality then bounds the expected supremum of X by that of Y.

For the reduction to be useful, the Y-process must be constructed with a lot of independence built in from the start. This is achieved by a method called *chaining*. First we organize points in T in a kind of tree structure by defining, for each $n \in \mathbb{N}$, a map $\pi_n \colon T \to T$ whose image is a finite set such that the ρ_X-distance between t and $\pi_n(t)$ is well controlled and tending to zero *exponentially* fast with $n \to \infty$, uniformly in t. A Borel-Cantelli estimate then allows us to write

$$X_t - X_s = \sum_{n=1}^{\infty} \Big(\big[X_{\pi_n(t)} - X_{\pi_{n-1}(t)}\big] - \big[X_{\pi_n(s)} - X_{\pi_{n-1}(s)}\big] \Big), \tag{6.20}$$

with the sum converging a.s. To define Y we may for instance replace the increments $X_{\pi_n(t)} - X_{\pi_{n-1}(t)}$ by independent random variables with a similar variance. A key point to note is that the n-th summand above will be non-zero only if $\pi_k(t) \neq \pi_k(s)$ for at least one of $k = n, n-1$. Matters will be arranged in such a way that this forces n to be (roughly) $\log 1/\rho_X(t,s)$.

We will now begin with the actual proof:

Proof of Theorem 6.6. Assume the setting of the theorem and fix a probability measure μ on T. The proof (which follows the corresponding proof in a book by R. Adler) comes in five steps.

STEP 1: Reduction to unit diameter. As the case $D := \mathrm{diam}(T)$ vanishes is that of a single random variable for which the statement holds trivially, we may assume that $D > 0$. One can then check that the process $\widetilde{X}_t := D^{-1/2} X_t$ has a unit diameter. In light of $\rho_{\widetilde{X}}(s,t) = D^{-1/2}\rho_X(s,t)$ the $\rho_{\widetilde{X}}$-ball of radius r centered at t coincides with $B(t, D^{-1/2}r)$. One can then check that, passing from X to $\widetilde{X}$ in (6.15), both sides scale by factor $\sqrt{D}$.

STEP 2: Construction of the tree structure. Next we will define the aforementioned maps π_n subject to properties that will be needed later. This is the content of:

Lemma 6.10 *For each $n \in \mathbb{N}$ there is $\pi_n \colon T \to T$ such that*

(1) $\pi_n(T)$ is finite,

(2) for each $t \in T$, $\rho(t, \pi_n(t)) < 2^{-n}$,

(3) for each $t \in T$,
$$\mu\big(B(\pi_n(t), 2^{-n-2})\big) \geq \mu\big(B(t, 2^{-n-3})\big). \tag{6.21}$$

(4) the sets in $\{B(t, 2^{-n-2}) \colon t \in \pi_n(T)\}$ are (pairwise) disjoint.

Proof. Fix $n \in \mathbb{N}$ and, using the assumption of total boundedness, let $t_1, \dots, t_{r_n}$ be point such that
$$\bigcup_{i=1}^{r_n} B(t_i, 2^{-n-3}) = T. \tag{6.22}$$
Assume that the points were ordered in such a way that
$$i \mapsto \mu\big(B(t_i, 2^{-n-2})\big) \quad \text{is non-increasing.} \tag{6.23}$$
We will now identify a disjoint subcollection $\{C_k\} \subset \{B(t_i, 2^{-n-3}) \colon i = 1, \dots, r_n\}$ by progressively dropping balls that have non-empty intersection with one of the previous ones. To give a formal definition, set
$$C_1 := B(t_1, 2^{-n-2}) \tag{6.24}$$
and, assuming that $C_1, \dots, C_i$ have already been defined, let
$$C_{i+1} := \begin{cases} B(t_{i+1}, 2^{-n-2}), & \text{if } B(t_{i+1}, 2^{-n-2}) \cap \bigcup_{j=1}^{i} C_j = \emptyset, \\ \emptyset, & \text{else.} \end{cases} \tag{6.25}$$
Now we define π_n as the composition of two maps described, somewhat informally, as follows: Using the ordering of $t_1, \dots, t_{r_n}$ as induced by (6.23), first assign t to the smallest point t_i such that $t \in B(t_i, 2^{-n-3})$. Then assign this t_i to the largest t_j from $t_1, \dots, t_i$ such that $B(t_i, 2^{-n-2}) \cap C_j \neq \emptyset$. Formally, let
$$\begin{aligned} i = i(t) &:= \min\big\{i = 1, \dots, r_n \colon t \in B(t_i, 2^{-n-3})\big\} \\ j = j(t) &:= \max\big\{j = 1, \dots, i(t) \colon B(t_{i(t)}, 2^{-n-2}) \cap C_j \neq \emptyset\big\}, \end{aligned} \tag{6.26}$$
where we notice that, by the construction of $\{C_k\}$, the set in the second line is always non-empty. We then define
$$\pi_n(t) := t_j \quad \text{for} \quad j = j(t). \tag{6.27}$$

This means $\pi_n(T) \subseteq \{t_1, \dots, t_{r_n}\}$ and so $\pi_n(T)$ is indeed finite, proving (1). For (2), using i and j for the given t as above, the construction implies

$$\begin{aligned} \rho_X(t, \pi_n(t)) &= \rho_X(t, t_j) \\ &\le \rho_X(t, t_i) + \rho_X(t_i, t_j) \\ &\le 2^{-n-3} + 2\,2^{-n-2} < 2^{-n}. \end{aligned} \tag{6.28}$$

For (3) we note that

$$B(t, 2^{-n-3}) \subseteq B(t_i, 2^{-n-2}) \tag{6.29}$$

and, by (6.23),

$$\mu\big(B(t_i, 2^{-n-2})\big) \le \mu\big(B(t_j, 2^{-n-2})\big). \tag{6.30}$$

Finally, $t \in \pi_n(T)$ only if $C_j \neq \emptyset$ at which point $C_j = B(t_j, 2^{-n-2})$. The construction ensures that the C_j's are disjoint from each other thus proving (4). □

STEP 3: Auxiliary process. We are now ready to defined the aforementioned process $\{Y_t \colon t \in T\}$. For this, consider a collection $\{Z_n(t) \colon n \in \mathbb{N},\, t \in \pi_n(T)\}$ of i.i.d. standard normals and set

$$Y_t := \sum_{n \ge 1} 2^{-n} Z_n\big(\pi_n(t)\big). \tag{6.31}$$

The sum converges a.s. for each t due to the fact that the maximum of the first n terms in a sequence of i.i.d. standard normals grows at most like a constant times $\sqrt{\log n}$. We now state:

Lemma 6.11 *For any $t, s \in T$,*

$$E\big([X_t - X_s]^2\big) \le 6E\big([Y_t - Y_s]^2\big) \tag{6.32}$$

In particular,

$$E\big(\sup_{t \in T} X_t\big) \le \sqrt{6}\, E\big(\sup_{t \in T} Y_t\big)\,. \tag{6.33}$$

Proof. We may assume $\rho_X(t,s) > 0$ as otherwise there is nothing to prove. Since we know that $\operatorname{diam}(T) = 1$, there is $N \in \mathbb{N}$ such that $2^{-N} < \rho_X(t,s) \le 2^{-N+1}$. Lemma 6.10(2) and the triangle inequality then show

$$\pi_n(t) \neq \pi_n(s), \qquad n \ge N+1. \tag{6.34}$$

This is quite relevant because the independence built into Y_t yields

$$E\big([Y_t - Y_s]^2\big) = \sum_{n \ge 1} 2^{-2n} E\Big(\big[Z_n(\pi_n(t)) - Z_n(\pi_n(s))\big]^2\Big) \tag{6.35}$$

and the expectation on the right vanishes unless $\pi_n(t) \neq \pi_n(s)$. As the expectation is either zero or 2, this shows

$$E\big([Y_t - Y_s]^2\big) \ge 2 \sum_{n \ge N+1} 2^{-2n} = 2\frac{4^{-(N+1)}}{3/4} = \frac{1}{6} 4^{-N+1} \ge \frac{1}{6} E\big([X_t - X_s]^2\big)\,, \tag{6.36}$$

where the last inequality follows from the definition of N. This is (6.32); the second conclusion then follows from the Sudakov-Fernique inequality. $\square$

STEP 4: Majorizing $E(\sup_{t\in T} Y_t)$. For the following argument it will be convenient to have a random variable τ, taking values in T, that identifies the maximizer of $t \mapsto Y_t$. Such a random variable can certainly be defined when T is finite. For T infinite, one has to work with approximate maximizers only. To this end we pose:

Exercise 6.12 *Suppose that there is $M \in (0,\infty)$ such that $E(Y_\tau) \le M$ holds for any T-valued random variable measurable with respect to $\{Y_t \colon t \in T\}$. Prove that then also $E(\sup_{t\in T} Y_t) \le M$.*

It thus suffices to estimate $E(Y_\tau)$ for any T-valued random variable τ. For this we first partition the expectation according to the values of $\pi_n(\tau)$ as

$$E(Y_\tau) = \sum_{n\ge 1} 2^{-n} \sum_{t\in\pi_n(T)} E\big(Z_n(t)1_{\{\pi_n(\tau)=t\}}\big)\,. \tag{6.37}$$

We now estimate the expectation on the right as follows: Set $g(a) := \sqrt{2\log(1/a)}$ and note that, for $Z = \mathcal{N}(0,1)$ and any $a > 0$,

$$E\big(Z1_{\{Z>g(a)\}}\big) = \frac{1}{\sqrt{2\pi}}\int_{g(a)}^\infty x\,\mathrm{e}^{\frac12 x^2}\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac12 g(a)^2} = \frac{a}{\sqrt{2\pi}}. \tag{6.38}$$

Therefore, using the notation from the proof of Lemma 6.10,

$$\begin{aligned} E\big(Z_n(t)1_{\{\pi_n(\tau)=t\}}\big) &\le E\big(Z_n(t)1_{\{Z_n(t)>g(a)\}}\big) + g(a)P\big(\pi_n(\tau)=t\big) \\ &= \frac{a}{\sqrt{2\pi}} + g(a)P\big(\pi_n(\tau)=t\big)\,. \end{aligned} \tag{6.39}$$

Now set $a := \mu(B(t,2^{-n-2}))$ in this term and perform the sum over t and n. In the first term we use the disjointness claim from Lemma 6.10(4) to get

$$\sum_{t\in\pi_n(T)} \mu(B(t,2^{-n-2})) \le 1 \tag{6.40}$$

while in the second term we note that

$$g\big(\mu(B(\pi_n(t),2^{-n-2})\big) \le g\big(\mu(B(\tau,2^{-n-3})\big) \tag{6.41}$$

by Lemma 6.10(3) and the fact that g is non-increasing. Hence

$$\begin{aligned} &\sum_{n\ge1} 2^{-n} \sum_{t\in\pi_n(T)} g\big(\mu(B(t,2^{-n-2})\big)P\big(\pi_n(\tau)=t\big) \\ &\qquad = E\bigg[\sum_{n\ge1} 2^{-n} g\big(\mu(B(\pi_n(\tau),2^{-n-2})\big)\bigg] \\ &\qquad \le E\bigg[\sum_{n\ge1} 2^{-n} g\big(\mu(B(\tau,2^{-n-3})\big)\bigg] \le \sup_{t\in T}\sum_{n\ge1} 2^{-n} g\big(\mu(B(t,2^{-n-3})\big)\,. \end{aligned} \tag{6.42}$$

Using the monotonicity of g,

$$2^{-n} g\bigl(\mu(B(t,2^{-n-3})\bigr) \le 16 \int_{2^{-n-4}}^{2^{-n-3}} g\bigl(\mu(B(t,r)\bigr)\,\mathrm{d}r, \tag{6.43}$$

and so the last sum in (6.42) can now be dominated by 16-times the integral in the statement of the theorem. Putting the contribution of both terms on the right of (6.39) together, we thus conclude

$$E(Y_\tau) \le \frac{1}{\sqrt{2\pi}} + 16 \sup_{t\in T} \int_0^1 g\bigl(\mu(B(t,r)\bigr)\,\mathrm{d}r. \tag{6.44}$$

STEP 5: A final touch. In order to finish the proof, we need to show that the term $1/\sqrt{2\pi}$ is dominated by, and can thus be absorbed into, the integral. Here we use the fact that, since $\operatorname{diam}(T) = 1$, there is $t \in T$ such that $\mu(B(t,1/2)) \le 1/2$. The supremum on the right of (6.44) is then at least $\frac{1}{2}\sqrt{\log 2}$. The claim follows with, e.g., $K := 17\sqrt{6}$. $\square$

6.4. Consequences for continuity

As already alluded to after the statement of Dudley's inequality, the generality of the setting in which Fernique's inequality was proved permits a rather easy extension to a criterion for continuity. The relevant statement is as follows:

Theorem 6.13 *There is a universal constant $K' \in (0,\infty)$ such that the following holds for every Gaussian process $\{X_t \colon t \in T\}$ on a countable set T such that (T,ρ_X) is totally bounded: For any probability measure μ on T and any $R > 0$,*

$$E\Bigl(\sup_{\substack{t,s\in T\\ \rho_X(t,s)\le R}} |X_t - X_s|\Bigr) \le K' \sup_{t\in T} \int_0^R \mathrm{d}r \sqrt{\log \frac{1}{\mu(B(t,r))}} \tag{6.45}$$

Proof. We will reduce this to Theorem 6.6 but that requires some preparations. Let

$$U \subset \bigl\{(t,s) \in T\times T \colon \rho_X(t,s) \le R\bigr\} \tag{6.46}$$

be a finite and symmetric set. Denote $Y_{s,t} := X_t - X_s$ and notice that

$$\rho_Y\bigl((s,t),(s',t')\bigr) := \sqrt{E\bigl([Y_{s,t} - Y_{s',t'}]^2\bigr)} \tag{6.47}$$

obeys

$$\rho_Y\bigl((s,t),(s',t')\bigr) \le \begin{cases} \rho(s,s') + \rho(t,t'), \\ \rho(s,t) + \rho(s',t'). \end{cases} \tag{6.48}$$

Writing B_Y for the balls (in $T\times T$) in the ρ_Y-metric and B_X for the balls (in T) in ρ_X-metric, the first line then implies

$$B_Y\bigl((s,t),r\bigr) \supseteq B_X(s,r/2) \times B_X(t,r/2) \tag{6.49}$$

while the second line shows

$$\operatorname{diam}_{\rho_Y}(U) \le 2R. \tag{6.50}$$

Now define $f\colon T\times T \to U$ by

$$f(y) := \begin{cases} y, & \text{if } y \in U, \\ \operatorname{argmin}_U \rho_Y(y,\cdot), & \text{else,} \end{cases} \tag{6.51}$$

where the minimizer in the second line is chosen minimal in some prior ordering of U. This f is clearly measurable (this is where it helps to have U finite) and so, given a probability measure μ on $T\times T$

$$\nu(A) := \mu\otimes\mu\big(f^{-1}(A)\big)\,. \tag{6.52}$$

defines a probability measure on U. Theorem 6.6 then yields

$$E\Big(\sup_{(s,t)\in U} Y_{s,t}\Big) \le K \sup_{(s,t)\in U}\int_0^{2R}\sqrt{\log\frac{1}{\nu(B_Y((t,s),r))}}\,\mathrm{d}r\,. \tag{6.53}$$

Our task is now to bring the integral on the right to the form in the statement. First observe that if $x\in U$ and $y\in B(x,r)$, then

$$\rho_Y\big(x,f(y)\big) \le \rho_Y(x,y) + \rho_Y\big(y,f(y)\big) \overset{x\in U}{\le} 2\rho_Y(x,y)\,. \tag{6.54}$$

Hence we get

$$B_Y(x,r) \subseteq f^{-1}\big(B_Y(x,2r)\big), \qquad x\in U, \tag{6.55}$$

and so, in light of (6.49),

$$\begin{aligned} \nu\big(B((s,t),2r)\big) &= \mu\otimes\mu\big(f^{-1}\big(B((s,t),2r)\big)\big) \\ &\ge \mu\otimes\mu\big(B((s,t),r)\big) \\ &\ge \mu\big(B(s,r/2)\big)\mu\big(B(t,r/2)\big)\,. \end{aligned} \tag{6.56}$$

Using $\sqrt{a+b}\le\sqrt{a}+\sqrt{b}$ and some elementary calculus, we conclude

$$E\Big(\sup_{(s,t)\in U} X_{t,s}\Big) \le 4K\sup_{t\in T}\int_0^{R}\sqrt{\log\frac{1}{\mu(B_X(t,r))}}\,\mathrm{d}r\,. \tag{6.57}$$

Increasing U to $U_R := \{(s,t)\in T\times T\colon \rho_X(s,t)\le R\}$ and invoking the Monotone Convergence Theorem, the bound holds for $U := U_R$ as well. Noting that

$$E\Big(\sup_{\substack{t,s\in T\\ \rho_X(t,s)\le R}} |X_t - X_s|\Big) = E\Big(\sup_{(s,t)\in U_R} Y_{s,t}\Big)\,, \tag{6.58}$$

the claim follows with $K' := 4K$, where K is as in Theorem 6.6. $\square$

The above criterion gives continuity with respect to the intrinsic metric. However, more often than not, T has its own private metric structure and continuity is desired in the topology thereof. Here the following exercise helps:

Exercise 6.14 *Suppose (T,ρ) is a metric space, $\{X_t\colon t\in T\}$ a Gaussian process and ρ_X the intrinsic metric on T induced thereby. Assume*

(1) (T,ρ) is totally bounded, and

(2) $s,t\mapsto \rho_X(s,t)$ is uniformly ρ-continuous on $T\times T$.

Prove that, if there is a probability measure μ on T such that

$$\lim_{R\downarrow 0}\sup_{t\in T}\int_0^R\sqrt{\log\frac{1}{\mu(B(t,r))}}\,\mathrm{d}r=0, \tag{6.59}$$

then X admits (uniformly) ρ-continuous sample paths on T, a.s.

We note that condition (2) is necessary for sample path continuity, but definitely not sufficient (and that not even for Gaussian processes). To see this, solve:

Exercise 6.15 *Given a measure space $(\mathscr{X},\mathcal{F},\nu)$ with ν finite, consider the (centered) Gaussian white-noise process $\{W(A)\colon A\in\mathcal{F}\}$ defined by*

$$E\big(W(A)W(B)\big)=\nu(A\cap B). \tag{6.60}$$

This corresponds to the intrinsic metric $\rho_W(A,B)=\sqrt{\nu(A\triangle B)}$. Give a (simple) example of $(\mathscr{X},\mathcal{F},\nu)$ for which $A\mapsto W(A)$ does not admit ρ_W-continuous sample paths.

As our last item of concern in this lecture, we return to the problem of uniform continuity of the binding field for the DGFF and its (resulting) coupling to the its continuum counterpart. The relevant bounds are stated in:

Lemma 6.16 *Let $\widetilde D, D\in\mathfrak{D}$ obey $\widetilde D\subset D$ with $\mathrm{Leb}(D\smallsetminus\widetilde D)=0$. For $\delta>0$, denote $\widetilde D^\delta:=\{x\in D\colon \operatorname{dist}(x,D^{\mathrm{c}})>\delta\}$. Then for each $\epsilon,\delta>0$,*

$$\lim_{r\downarrow 0}P\Bigg(\sup_{\substack{x,y\in\widetilde D^\delta\\ |x-y|<r}}\big|\Phi^{D,\widetilde D}(x)-\Phi^{D,\widetilde D}(y)\big|>\epsilon\Bigg)=0. \tag{6.61}$$

Similarly, given an admissible sequence $\{D_N\colon N\ge 1\}$ of approximating domains and denoting $\widetilde D_N^\delta:=\{x\in\widetilde D_N\colon \operatorname{dist}(x,\widetilde D_N^{\mathrm{c}})>\delta N\}$, for each $\epsilon,\delta>0$,

$$\lim_{r\downarrow 0}\limsup_{N\to\infty}P\Bigg(\sup_{\substack{x,y\in\widetilde D_N^\delta\\ |x-y|<rN}}\big|\varphi_x^{D_N,\widetilde D_N}-\varphi_y^{D_N,\widetilde D_N}\big|>\epsilon\Bigg)=0. \tag{6.62}$$

Proof of (6.61). Consider the set $\widetilde D_1^\delta:=\{x\in\mathbb{C}\colon \operatorname{dist}(x,\widetilde D^\delta)<\delta/2\}$. The intrinsic metric associated with $\{\Phi^{D,\widetilde D}(x)\colon x\in\widetilde D_1^\delta\}$ is given by

$$\rho_\Phi(x,y)=\sqrt{C^{D,\widetilde D}(x,x)+C^{D,\widetilde D}(y,y)-2C^{D,\widetilde D}(x,y)} \tag{6.63}$$

Since $x \mapsto C^{D,\widetilde{D}}(x,y)$ is harmonic on $\widetilde{D}$, it is continuously differentiable and thus uniformly Lipschitz on $\widetilde{D}_1^\delta$. It follows that, for some constant $L = L(\delta) < \infty$,

$$\rho_\Phi(x,y) \le L\sqrt{|x-y|}, \qquad x,y \in \widetilde{D}_1^\delta. \tag{6.64}$$

Let $B(x,r) := \{y \in \mathbb{C} \colon |x-y| < r\}$ and denote $B_\Phi(x,r) := \{y \in \widetilde{D}_1^\delta \colon \rho_\Phi(x,y) < r\}$ and let μ be the normalized Lebesgue measure on $\widetilde{D}_1^\delta$. Then

$$B_\Phi(x, L\sqrt{r}) \supseteq B(x,r), \qquad x \in \widetilde{D}_1^\delta \tag{6.65}$$

while (by the choice of $\widetilde{D}_1^\delta$),

$$\mu\big(B(x,r)\big) \ge cr^2, \quad x \in \widetilde{D}_1^\delta, \tag{6.66}$$

for some $c = c(\delta) > 0$. Hence, $\mu(B_\Phi(x,r)) \ge cL^{-2}r^4$. As $r \mapsto \log(1/r^4)$ is integrable at zero, (6.61) follows from Theorem 6.13, Exercise 6.14 and Markov inequality. $\square$

Exercise 6.17 *Using an analogous argument with the normalized counting measure replacing the Lebesgue measure, prove* (6.62).