1. Upper and lower bounds on the supremum of a Gaussian process

Let $X = (X_t)_{t\in T}$ be a centered Gaussian process. Endow T with the metric $\tau(s,t) = \sqrt{\mathbf{E}[(X_t - X_s)^2]}$. This just means viewing X as a "curve in the Hilbert space" $L^2(\Omega, \mathcal{F}, \mathbf{P})$ and its image $\{X_t\}$ as a metric space inheriting the metric from $L^2(\Omega, \mathcal{F}, \mathbf{P})$. Assume that T is bounded in τ (in what follows we shall in fact have to assume that T is totally bounded).

Let $X^* = \sup_{t\in T} X_t$ and $\|X\| = \sup_t |X_t|$. The goal is to get upper bounds on X^*, more precisely, on quantities like $\mathbf{E}[X^*]$ and $\mathbf{P}\{X^* \geq t\}$. The various bounds one may consider are closely related.

(1) As a standing convention, assume that X^* is measurable so that quantities such as $\mathbf{P}(X^* > u)$ and $\mathbf{E}[X^*]$ make sense. Alternately, simply assume that T is countable, in which case X^* is indeed measurable.

(2) Firstly, bounds on tail of X^* and bounds on the expected value of X^* imply each other as follows.
- We have $\mathbf{E}[(X^*)_+] = \int_0^\infty \mathbf{P}\{X^* > u\}du$. The lower side never poses a problem since $X^* \geq X_{t_0}$ (for any $t_0 \in T$ and hence $\mathbf{E}[(X^*)_-] \leq \mathbf{E}[(X_{t_0})_-] \leq \sigma_T/\sqrt{2\pi}$. Thus, bounds on the tail probability give a bound for the mean.
- For the other direction, by Markov's inequality, $\mathbf{P}\{X^* \geq u\} \leq \mathbf{E}[X^*_+]/u$.

These two observations are general, but for Gaussian processes one can say much more. Going back to the proof of Borell's inequality, it can be seen that if X^* is finite with positive probability, then it is finite with probability 1, has sub-gaussian tails like $2\exp\{-x^2/2\sigma_T^2\}$, and has finite expectation. Of course finiteness of expectation implies that the random variable is itself finite a.s.

(3) Secondly, bounds for X^* and for $\|X\|$ are equivalent as clear from the following exercise (taken from Talagrand's book).

Exercise 1. Fix any t_0. Then $\mathbf{E}[\|X\|] \leq 2\mathbf{E}[X^*] + \mathbf{E}[|X_{t_0}] \leq 3\mathbf{E}[\|X\|]$.

Now we get to work and try to bound $\mathbf{E}[X^*]$. Although we have talked about the supremum of a Gaussian process in earlier lectures, all that we have said is essentially this:

(1) X^* is well concentrated about its mean or median.

(2) If certain correlation inequalities hold, then $\mathbf{E}[X^*]$ can be bounded by $\mathbf{E}[Y^*]$.

We have not calculated how large $\mathbf{E}[X^*]$ is except in one case, that of i.i.d. Gaussians. From that calculation, we extract the following general bound.

Lemma 2. *If X is a centered Gaussian process on T, then* $\mathbf{E}[X^*_+] \leq 10\sigma_T\sqrt{\log|T|}$.

Proof. For any $t \in T$ and any $u > 0$ we have $\mathbf{P}\{X_t \geq u\} \leq 2e^{-u^2/2\sigma_T^2}$. Hence by the union bound, $\mathbf{P}\{X^*_+ \geq u\} \leq 2|T|e^{-u^2/2\sigma_T^2}$. We use this bound when it is better than the trivial bound 1 for a probability, i.e., for $u \geq \sigma_T\sqrt{2\log(2|T|)}$. By integrating we get $\mathbf{E}[X^*_+] \leq 5\sigma_T\sqrt{\log|T|}$. ■

Clearly this alone will not suffice. If we try to approximate T by large finite sets for which we use the above bound, then the bound gets worse and worse with the cardinality of the finite set. But analysing why the above bound fails is instructive. The union bound is good when X_t are independent (recall that for i.i.d. $N(0,1)$ we proved that a lower bound of the order $\sqrt{\log|T|}$ also). But when X_t are strongly correlated (as an extreme case think of $X_t = \xi$ for all t), then the union bound greatly overestimates the actual probability. The following lemma confirms our intuition that this is the only problem. It also sets a benchmark for what to expect in the upper bound.

Theorem 3 (Sudakov minoration)**.** *Let X be a centered Gaussian process on T. If $\tau(t,s) \geq \varepsilon$ for all $t \neq s$, then* $\mathbf{E}[X^*] \geq 0.1\varepsilon\sqrt{\log|T|}$.

Proof. Let ξ_t be i.i.d $N(0,1)$ and let $Y_t = \frac{1}{\sqrt{2}}\varepsilon\xi_t$. Then $\tau_Y(t,s) = \varepsilon^2 \leq \tau_X(t,s)$ for all $t \neq s$. By Sudakov-Fernique inequality, $\mathbf{E}[X^*] \geq \mathbf{E}[Y^*] \geq 0.1\varepsilon\sqrt{\log|T|}$. ■

As an immediate corollary we get a lower bound for $\mathbf{E}[X^*]$ in general.

Corollary 4 (Fernique)**.** *Let X be a centered Gaussian process on T. Let N_ε denote the smallest size of an ε-net in (T,τ_X). Then, (if X^* is measurable) $\mathbf{E}[X^*] \geq 0.1\liminf_{\varepsilon\to 0}\varepsilon\sqrt{\log N_\varepsilon}$. In particular, if the latter quantity is infinite, then any version of the Gaussian process must be unbounded w.p.1.*

The theorem shows us that the union bound only loses when comparing Gaussians that are almost equal (must have $\tau(t,s)$ close to zero). In addition, the corollary also sets a benchmark for how small an upper bound for $\mathbf{E}[X^*]$ can be.

2. THEOREMS OF DUDLEY, FERNIQUE AND TALAGRAND ON THE SUPREMUM

Here is the basic lemma by the method of *generic chaining*[11].

Setting: Let $X = (X_t)_{t\in T}$ be a stochastic process with zero mean random variables indexed by a metric space (T,τ) such that $\mathbf{P}\{|X_t - X_s| \ge u\tau(t,s)\} \le 2\exp\{-\frac{1}{2}u^2\}$ for all $u > 0$ and for all $t,s \in T$. A particular case is of a centered Gaussian process with $\tau(s,t) := \sqrt{\mathbf{E}[|X_t - X_s|^2]}$.

Lemma 5 (The generic chaining bound)**.** *Let T be finite or countable. Fix $t_0 \in T$ and numbers $u_k \ge 1$. Choose any subsets $T_k \subseteq T$ with $T_0 = \{t_0\}$ and such that each $t \in T$ is contained in T_k for all large k. Then for any $x > 0$ we have*

$$\mathbf{P}\{X^* - X_{t_0} \ge (x+1)A\} \le Qe^{-x^2/2}$$

where $A = \sup_{t\in T} \sum_{k=1}^{\infty} \tau(t,T_k)$ and $Q = 2\sum_{k=1}^{\infty} |T_k|\cdot|T_{k-1}|e^{-u_k^2/2}$.

Proof. First take $x = 0$. We have $X_t - X_{t_0} = \sum_{k=1}^{\infty} X_{\pi_t^k} - X_{\pi_{k-1}^t}$ where π_k^t is (one of) the closest point to t in T_k. If $X_t - X_{t_0} > \sum_{k=1}^{\infty} \tau(t,T_k)$, then the kth summand in $X_t - X_{t_0}$ must exceed the $\tau(t,\pi_k^t)$ for at least one k. For any k, the probability that $|X_t - X_s| \ge u\sqrt{\tau(t,s)}$ for some $t \in T_k$ and $s \in T_{k-1}$, is bounded by the kth summand in Q (by the union bound). Put everything together and use triangle inequality liberally.

To get the inequality for general x, just replace u_k by $u_k(1+x)$. ■

Motivated by this bound, we define two fundamental quantities associated to a metric space.

Definition 6. For a metric space (T,τ), define

(1) *Talagrand's γ_2-functional:* $\gamma_2(T) := \inf_{\{T_k\}} \sup_t \sum_{k=0}^{\infty} 2^{k/2}\tau(t,T_k)$.

(2) *Dudley's integral:* $\mathcal{D}(T) := \inf_{\{T_k\}} \sum_{k=0}^{\infty} 2^{k/2} \sup_t \tau(t,T_k)$.

Both infima are over all choices of the sets $\{T_k\}$ subject to the condition $|T_k| = 2^{2^k}$.

This definition amounts to choosing $u_k = 2^{k/2}$ in the lemma. In class I discussed why that is the correct choice, but I don't want to write it in detail here. Clearly $\gamma_2(T) \le \mathcal{D}(T)$. The latter quantity is achieved by choosing T_k to be a set with cardinality 2^{2^k} so that it is an ε-net with the smallest possible ε. The reason for calling $\mathcal{D}(T)$ an integral is that it is bounded from above and below by constant multiples of $\int_0^{\infty} \sqrt{\log N(\varepsilon)}d\varepsilon$, where $N(\varepsilon)$ is the smallest size of an ε-net in T (show this!).

Theorem 7 (Dudley's integral)**.** *In the above setting, let $N(\varepsilon)$ be the smallest size of an ε-net for (T,τ). Let $\mathcal{D}_\tau := \int_0^{\infty} \sqrt{\log N(\varepsilon)}d\varepsilon$.*

$$\mathbf{E}\left[\sup_{t\in T} X_t - X_{t_0}\right] \lesssim \mathcal{D}_\tau \text{ and}$$

$$\mathbf{P}\left\{\sup_{t\in T} X_t - X_{t_0} \ge (x+1)\mathcal{D}_\tau\right\} \lesssim e^{-x^2/2}.$$

There is an easy lower bound due to Fernique.

Theorem 8 (Fernique)**.** *Let $\mathcal{F}(T) := \sup_{\varepsilon} \varepsilon\sqrt{\log N(\varepsilon)}$. Then $\mathcal{F}(T) \lesssim \mathbf{E}[\sup_t X_t - X_{t_0}]$.*

---

[11] This topic is beautifully explained in Talgrand's book *Generic chaining*. We do not repeat the arguments in detail.

Dudley's upper bound and Fernique's lower bound are almost tight, but not quite. For example, if $N(\varepsilon) = \exp\{-\varepsilon^{-c}\}$, then for $c < 1$ both are finite, while for $c > 1$ both are infinite. Thus, it is essentially when $N(\varepsilon) = \exp\{-\varepsilon^{-1}\}$ (with lower order corrections) that an ambiguity arises. However, there is such an ambiguity. In addition, even when both $\mathcal{F}(T)$ and $\mathcal{D}(T)$ are finite, they may be of different orders of magnitude, and it is not clear whether $\mathbf{E}[\sup X_t - X_{t_0}]$ is like one or the other or somewhere in between.

The quantity which exactly characterizes the expected supremum is the γ_2 functional. The upper bound was proved by Fernique. The lower bound was conjectured by Fernique and proved by Talagrand.

Theorem 9 (Fernique-Talagrand). $\mathbf{E}[\sup_t X_t - X_{t_0}] \asymp \gamma_2(T)$.

We have already proved the upper bound $\mathbf{E}[X^* - X_{t_0}] \lesssim \gamma_2(T)$ (the first lemma!). We do not prove the lower bound as I don't understand it yet.

3. Some examples

<u>**Independent Gaussians:**</u> Let $X_k \sim N(0, \sigma_k^2)$ be independent. Assume that σ_k^2 decreases to 0. Then $\tau(m,n) = \sqrt{\sigma_n^2 + \sigma_m^2}$. Note that for $m < n$ we have $\sigma_m \le \tau(m,n) \le \sigma_m\sqrt{2}$. For simplicity let us pretend that $\tau(m,n) = \sigma_{m\wedge n}$ (we leave it as an exercise to make appropriate modifications).

If $0 < \varepsilon < \sigma_1$, then there is a unique n such that $\sigma_n \le \varepsilon < \sigma_{n-1}$. Then $\{1, 2, \ldots, n\}$ is an ε-net whence $N(\varepsilon) \le n$. Since $\tau(i,j) > \sigma_{n-1} > \varepsilon$ for $i, j \le n-1$, it is clear that $N(\varepsilon) \ge n-1$. Thus the Dudley integral is (as always ignoring constant factors) $\mathcal{D} = \sum_{k=2}^{\infty} (\sigma_{k-1} - \sigma_k)\sqrt{\log k}$.

On the other hand, we may write $X_n = \sigma_n \xi_n$ where ξ_n are i.i.d. $N(0,1)$ variables. Recall that $\limsup\limits_{n\to\infty} \frac{\xi_n}{\sqrt{2\log n}} = 1$ a.s. (if not clear, provide a proof!). Thus, $\sup_n X_n < \infty$ a.s. (recall that this also implies that the supremum has finite expectation and Gaussian tail decay) whenever $\limsup \sigma_n \sqrt{\log n}$ is finite.

Thus, by choosing, for example, $\sigma_n = \frac{1}{\sqrt{\log n \log\log n}}$, we see that the Dudley integral may diverge but the supremum is finite.

Exercise 10. Compute γ_2 or at lease verify that it is finite for this choice of σ_ns.

<u>**Trees:**</u> Let $\mathcal{T}$ be a rooted leafless tree. The boundary of the tree is defined as the set of all infinite, simple paths emanating from the root[12]. That is,

$$\partial\mathcal{T} = \{\mathbf{v} : \mathbf{v} = (v_0, v_1, \ldots), v_0 \text{ is the root and } v_{i+1} \text{ is a child of } v_i\}.$$

Fix $\lambda > 1$. Let ξ_v be i.i.d. $N(0,1)$ random variables indexed by the vertices of the tree. Then, define $X(\mathbf{v}) := \sum_{k=0}^{\infty} \xi_{v_k} \lambda^{-k/2}$.

Exercise 11. Show that X is a centered Gaussian process on $\partial\mathcal{T}$ with $\tau(\mathbf{v},\mathbf{w}) := C_\lambda \lambda^{-|\mathbf{v}\wedge\mathbf{w}|/2}$ where $\mathbf{v}\wedge\mathbf{w}$ is the last common vertex in $\mathbf{v}$ and $\mathbf{w}$ and $|\mathbf{v}\wedge\mathbf{w}|$ is its graph distance from the root.

As the constant C_λ is unimportant, we define the metrics $\tau_\lambda(\mathbf{v},\mathbf{w}) := \lambda^{-|\mathbf{v}\wedge\mathbf{w}|/2}$ for $\mathbf{v}, \mathbf{w} \in \partial\mathcal{T}$.

Exercise 12. The boundary $\partial\mathcal{T}$ is a compact metric space under τ_λ.

What is the Dudley integral for τ_λ? If $0 < \varepsilon < 1$, then there is a unique $k \ge 1$ such that $\lambda^{-k/2} \le \varepsilon < \lambda^{-(k-1)/2}$. Then $N(\varepsilon) = |\mathcal{T}_{[k]}|$, the cardinality of the kth generation of $\mathcal{T}$. To see this, take a collection of paths, one

[12] The lower bounds here are all due to Fernique. Our presentation is essentially from Kahane's book *Some random series of functions*. He does not mention trees but what we call spherically symmetric trees are referred to as generalized Cantor sets there. But the essence is the same.

passing through each vertex of $\mathcal{T}_{[k]}$. All these paths are at distance $\lambda^{-k/2}$ from each other and form a $\lambda^{-k/2}$ net for $\partial\mathcal{T}$. Thus, the Dudley integral is

$$\mathcal{D}_\lambda := (\sqrt{\lambda}-1)\sum_{k=1}^{\infty}\lambda^{-k/2}\sqrt{\log|\mathcal{T}_{[k]}|}.$$

Spherically symmetric trees: Suppose that all vertices in $\mathcal{T}_{[k]}$ have M_k children, for some positive integers M_k, $k \geq 1$. Such a tree is called *spherically symmetric*. Its special feature is that it has many automorphisms. For example, we can permute all the subtrees emanating from any given vertex. The automorphism group preserves each $\mathcal{T}_{[k]}$ and in fact acts transitively on each generation. It extends naturally to a transitive group action on $\partial\mathcal{T}$. Thus, $\partial\mathcal{T}$ "looks the same everywhere".

One of Fernique's discoveries was that in such cases the Dudley integral provided the right lower bound! We shall see a few theorems of this nature.

Assumption: $M_k \geq k^a$ for some $a > 0$. A weaker growth condition will suffice as will be clear from the proof.

Indeed, from the previous computation, and using $\sqrt{\log|\mathcal{T}_{[k]}|} = \sqrt{\log M_0 + \log M_1 + \ldots + \log M_{k-1}}$ which is at most $\sqrt{\log M_0} + \ldots + \sqrt{\log M_{k-1}}$ we see that

$$\mathcal{D}_\lambda \leq C_\lambda \sum_{k=0}^{\infty}\lambda^{-k/2}\sqrt{\log M_k}.$$

On the other hand, we can get a lower bound for $\sup_{\mathbf{v}} X(\mathbf{v})$ by taking the "greedy path" $\mathbf{v}^*$ defined by letting v_0^* to be the root and v_{i+1}^* to be the child of v_i^* with the largest value of ξ (i.e., $\xi_w \leq \xi_{v_{i+1}^*}$ for all $w \leftarrow v_i^*$).

Let E be the event that $\xi_{v_k^*} \geq \sqrt{\log M_k}$ for all $k \geq 0$. If the event E occurs, then $X(\mathbf{v}^*) \geq \sum_{k=0}^{\infty}\lambda^{-k/2}\sqrt{\log M_k}$.

To estimate $\mathbf{P}(E)$ we shall use the following exercise.

Exercise 13. Let ξ_i be i.i.d $N(0,1)$ random variables. For any $\delta > 0$, there exists $c_\delta > 0$ such that $\mathbf{P}\{\max_{i\leq n}\xi_i \geq \sqrt{2(1-\delta)\log n}\} \geq 1 - e^{-c_\delta n^{-\delta/2}}$ for any $n \geq 1$.

Successively conditioning on the ξ-values on the first few steps of $\mathbf{v}^*$ and using independence of ξ_vs, from the above exercise we deduce that

$$\mathbf{P}(E) \geq \prod_{k=1}^{\infty}\left(1 - e^{-cn^{-1/2}}\right) > 0.$$

The strict positivity of the product comes from our assumption that $M_j \geq j^a$.

Thus, we have shown that $X(\mathbf{v}^*) \geq c\mathcal{D}_\lambda$ with probability at least c (for some small enough c). In particular, $\mathbf{E}[X(\mathbf{v}^*) - X(\mathbf{v}_0)] \geq c\mathcal{D}_\lambda$ (we subtract $X(\mathbf{v}_0)$ to make the supremum non-negative). Now we would like to say that $\mathbf{E}[\sup_{\mathbf{v}} X(\mathbf{v}) - X(\mathbf{v}_0)] \geq c\mathcal{D}_\lambda$, which is of course obvious, provided it makes sense (the supremum may not be measurable). Whenever the supremum is measurable, the Dudley integral gives the expectation of the supremum (in this class of examples).

Exercise 14. Exact homogeneity is not needed. Let M_k be as above and suppose each vertex in generation k has between $\sqrt{M_k}$ and M_k^2 children. The tree is no longer spherically symmetric, but show Dudley integral is a lower bound for the expected supremum.

Stationary processes: If G is a group and X is a centered Gaussian process indexed by G, then we say that X is *left-stationary* if $(X_{hg})_g \stackrel{d}{=} (X_g)_{g\in G}$ for any $h \in G$. For Gaussian processes, this just means that $\tau(hg, hg') = \tau(g, g')$ for all g, g', h.

Theorem 15 (Fernique)**.** *Let G be a locally compact group and let X be a centered stationary Gaussian process on G. Then, for any compact $K \subseteq G$ and $g_0 \in K$, we have $\mathbf{E}[\sup_{g\in K} X_g - X_{g_0}] \geq c\mathcal{D}_K$. If G itself is compact, we may of course take $K = G$.*

Sketch of the proof. Without loss of generality, assume that $\mathrm{dia}_\tau(G) = 1$. Let $S_0 = \{1\}$ and for $k \geq 1$ let S_k be a maximal 2^{-k}-separated set in $B_\tau(1, 2^{-k+1})$ (open ball centered at the identity). All finite products of the form $g_1 g_2 \ldots g_m$ with $g_i \in S_i$, are distinct (even the union over all m). Therefore, we get a spherically symmetric tree $\mathcal{T}$ embedded in G by taking all such finite products as vertices, the identity element as the root, and declaring $g_1 \ldots g_{m+1}$ to be a child of $g_1 \ldots g_m$ whenever $g_i \in S_i$. Some observations.

(1) The vertices in $\mathcal{T}_{[m]}$ form a 2^{-m}-net for G.

(2) If $g_i \in S_i$, then $g_1 \ldots g_m$ is a Cauchy sequence in G. Thus, for every path $\mathbf{v} \in \partial\mathcal{T}$, we can associate an element of G, namely $g_{\mathbf{v}} := \lim v_i$. Indeed, by the first point, every element of G is of the form $g_{\mathbf{v}}$ for some $\mathbf{v}$. However, it is possible to have $g_{\mathbf{v}} = g_{\mathbf{w}}$ for $\mathbf{v} \neq \mathbf{w}$.

(3) More generally, taking $\lambda = 4$ (we fix this choice till the end of the proof), we see that $3\tau_4(\mathbf{v}, \mathbf{w}) \geq \tau(g_{\mathbf{v}}, g_{\mathbf{w}})$ but no inequality holds the other way.

The following example illustrates these points in a familiar situation.

Example 16. Let in $\mathbb{R}\backslash\mathbb{Z} \cong [0,1]$ (cyclically identify 0 to , we may take $S_k = \{0, 2^{-k}\}$. Then $\mathcal{T}_{[m]} = \{k/2^m : k \leq 2^m\}$ are just dyadic rationals of the first m generations. Clearly these are all distinct. However, $\partial\mathcal{T}$ will be the whole of $[0,1]$ (in general vertices of $\mathcal{T}$ will be dense in G) and some points have multiple binary expansion, for example, $2^{-1} + 2^{-2} + \ldots$ is equal to 0 in this case.

A related problem is that for any λ, the metric τ_λ on $\partial\mathcal{T}$ is quite different from the metric τ. In the example here, 0.4999 and 0.50001 are close in τ but far in τ_λ. If τ_λ were to agree with τ, we would simply invoke the theorem on Gaussian processes on spherically symmetric trees.

Coming back to the proof, the third point listed is not good for us, since we want to get a lower bound for the Gaussian process on G by comparing it to a Gaussian process on $\partial\mathcal{T}$ (we already know that Dudley's integral is a lower bound on the boundary of spherically symmetric trees). However, comparison theorems require the opposite inequality! This can be arranged by pruning the tree a little (but not too much, as we want the Dudley integral on $\partial\mathcal{T}$ to be only a constant factor smaller than the Dudley integral on G!).

First, the Dudley integral. Since $\mathcal{T}_{[m]}$ is a 2^{-m} net for G, we see that $\mathcal{D}_G \lesssim \sum_{k=0}^{\infty} 4^{-k/2}\sqrt{\log|S_k|}$. Alternately, we could have said that since $\tau \leq 3\tau_4$, the Dudley integral on G is bounded by (a constant times) the Dudley integral on $\partial\mathcal{T}$ which we had computed earlier to be $\sum_{k=0}^{\infty} 4^{-k/2}\sqrt{\log|S_k|}$. We now want to prune the tree so that the Dudley integral stays about the same. For this, observe that one of the ten sums $\sum_k 4^{-(10k+r)/2}\sqrt{\log|S_{10k+r}|}$ as r ranges from 0 to 9, must be more than 1/10 of the total sum. For simplicity of notation, let us assume that $r = 0$ works. Then $\mathcal{D}_G \lesssim \sum_k 4^{-10k/2}\sqrt{\log|S_{10k}|}$.

The pruning. Introduce the following family planning scheme. In generation number $0, 10, 20, \ldots$, no change is made. For vertices in all other generations, keep only the first child and delete the rest (along with the entire subtree beneath them). The resulting spherically symmetric tree has $M_j = |S_j|$ if $j = 0 \pmod{10}$ and $M_j = 1$ otherwise. Thus we get a reduced tree $\tilde{\mathcal{T}}$ which is still embedded in G. The nice thing about it is that $\tau_4(\mathbf{v}, \mathbf{w}) \leq 100\tau(\mathbf{v}, \mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in \partial\tilde{\mathcal{T}}$.

Note that $\partial\tilde{\mathcal{T}}$ is a subset of G on which we have two Gaussian processes corresponding to the metrics τ (the original stationary process X) and τ_4 (the tree process Y). By Sudakov-Fernique inequality, $\mathbf{E}[\sup_{\partial\tilde{\mathcal{T}}} Y] \lesssim \mathbf{E}[\sup_{\partial\tilde{\mathcal{T}}} X]$. Of course the latter is smaller than $\mathbf{E}[\sup_G X]$. From the lower bound on spherically symmetric trees, $\mathcal{D}_{\partial\tilde{\mathcal{T}}} \lesssim \mathbf{E}[\sup_{\partial\tilde{\mathcal{T}}} Y]$ (strictly speaking, the assumption that we made there that $M_j \geq j^a$ is not satisfied. We leave it as an exercise to check that the conclusions are still okay). But we already pruned in such a way that $\mathcal{D}_G \lesssim \mathcal{D}_{\partial\tilde{\mathcal{T}}}$. This completes the proof that $\mathcal{D}_G \lesssim \mathbf{E}[\sup_G X]$. ■