Fifty years ago,
a theorem by Xavier Fernique

(preliminary draft 10-07-2024)

Contents

We shall deal here with a remarkable result of Xavier Fernique about Gaussian processes, from the year 1974 [Fer$_2$]. In another of his articles, Fernique has written [Fer$_1$]:

Cet exposé ne contient à proprement parler aucun résultat nouveau. . . .

a sentence that would quite well apply to the present text of mine. But I wouldn't dare going on with Fernique and say:

. . . Il vise à réfléchir sur des résultats connus, à simplifier et unifier leurs démonstrations, à en dégager une ligne directrice qui semble prometteuse à l'auteur.

Let us begin with a short presentation of the result. A Gaussian process $(X_t)_{t\in T}$ consists of a collection of Gaussian random variables X_t, indexed by a non-empty set T and belonging to a *Gaussian space*, that is to say, a linear subspace of $L^2(\Omega, \mathrm{P})$ all of whose elements are Gaussian random variables, where (Ω, P) is some probability space. We shall restrict ourselves to *centered* processes, namely, the case when

$$\mathrm{E}\, X_t = 0, \quad \text{for all } \ t \in T.$$

The index set T will be equipped with the L^2-metric given by the distance in $L^2(\Omega, P)$ of the corresponding random variables,

$$d(s,t)^2 = \mathrm{E}(X_s - X_t)^2, \qquad s, t \in T.$$

It is well known that some *entropy conditions* for that metric on T allow one to control the supremum $\sup_{t\in T} X_t$ of the process [a]: given $\varepsilon > 0$, let $\mathcal{N}(T, \varepsilon)$ denote the minimal number of open balls of radius ε needed to cover T, and let Δ be the diameter of T. The so-called *Dudley's integral* is defined by

$$I_D(T) = \int_0^\Delta \sqrt{\ln \mathcal{N}(T,\varepsilon)}\, \mathrm{d}\varepsilon.$$

Notice that $\ln \mathcal{N}(T, \varepsilon) = 0$ when $\varepsilon > \Delta$, because $\mathcal{N}(T, \varepsilon) = 1$ is that case (one ball of that radius ε is enough). If the Dudley integral satisfies

$$I_D(T) < \infty, \quad \text{it follows that} \quad \mathrm{E}\Bigl(\sup_{t\in T} |X_t|\Bigr) < \infty.$$

The result of Fernique is that, under some group invariance of the process, one can prove the reverse implication. Perhaps surprisingly for whom is far from my own domain of interests, Fernique's theorem was one crucial piece for a theorem of Pisier [Pisi] on lacunary trigonometrical series, namely, the characterization of Sidon sets $\Lambda \subset \mathbb{Z}$ by the rate of growth as p tends to ∞ of the L^p-norm of the functions with spectrum in Λ. This rate of growth of order $\sqrt{p}$ for Sidon sets was established by Rudin [Rud].

The domination of the expectation (the integral) of the supremum of a Gaussian process by the Dudley integral was often attributed to Dudley [Dud$_1$], who himself, shortly after Sudakov died in 2016, pointed out [Dud$_2$] that Sudakov was actually the one to credit for that result (a result that, in essence, is far from being the hardest point in what will be recalled here from Sudakov's works and others).

1. Preliminaries

We start slowly, with elementary observations on the discretization of the Dudley integral, then with the definition of Gaussian random variables, and we proceed calmly toward the comparisons of Gaussian processes due to Slepian and Sudakov.

1.1. *Discrete versions of the Dudley integral*

We may discretize the Dudley integral, by choosing a radius $r_0 > \Delta$ and by letting for example $r_i = 3^{-i} r_0$ for every integer $i \geqslant 0$. Next, we introduce the series

$$\Sigma_1(T) = \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(T, r_{i+1})}. \tag{1}$$

For every $i \geqslant 0$, we see that

$$\int_{r_{i+1}}^{r_i} \sqrt{\ln \mathcal{N}(T, \varepsilon)}\, \mathrm{d}\varepsilon \leqslant r_i \sqrt{\ln \mathcal{N}(T, r_{i+1})},$$

and thus

$$I_D(T) = \sum_{i=0}^{\infty} \int_{r_{i+1}}^{r_i} \sqrt{\ln \mathcal{N}(T, \varepsilon)}\, \mathrm{d}\varepsilon \leqslant \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(T, r_{i+1})} = \Sigma_1(T).$$

The latter series $\Sigma_1(T)$ is of the sort that will appear several times later. Let us now write simply $\mathcal{N}(r_{i+1})$ instad of $\mathcal{N}(T, r_{i+1})$. With the present choice $r_{i+1} = r_i/3$, we also have conversely that

$$I_D(T) \geqslant \frac{2}{9} \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} = \frac{2}{9}\, \Sigma_1(T),$$

and a similar reverse inequality holds true for any choice where $r_0 > \Delta$ and $r_i = a^i r_0$, with $0 < a < 1$, or simply a choice where $r_0 > \Delta$ and $r_i - r_{i+1} \geqslant c\, r_{i-1}$, with $0 < c < 1$.

We may obtain yet another equivalent form for the Dudley integral as a series, by a sort of "change of variable": instead of fixing the radius r and looking for the number of balls of that radius necessary to cover T, we fix the number N of balls and look for a radius such that we can cover T by N balls with that radius. This second series (2) will not be mentioned later in the text, the reader can jump to the next section. In this change of variable $i \leftrightarrow k$, we think of k and i to be linked by

$$b^k \sim \ln \mathcal{N}(r_i),$$

for some real number $b > 1$.

To be more specific, choose b such that $b^{-1} < \ln 2$, suppose that $\Delta < r_0 < 3\Delta/2$ and let again $r_{i+1} = r_i/3$ for $i \geqslant 0$. With this choice, we have $\mathcal{N}(r_0) = 1$ and we see that $2 r_1 = 2 r_0/3 < \Delta$ so that $\mathcal{N}(r_1) \geqslant 2$. For simplicity, assume that $\mathcal{N}(\varepsilon)$ is unbounded as $\varepsilon \to 0$. For every integer $k \geqslant 0$, let $i(k)$ be the smallest $i \geqslant 0$ for which we have $b^{k-1} < \ln \mathcal{N}(r_{i+1})$; when $k = 0$ for example, we obtain that $i(0) = 0$ because we have $\ln \mathcal{N}(r_0) = 0 < b^{-1}$ and $\ln \mathcal{N}(r_1) \geqslant \ln 2 > b^{-1}$. Consider

$$\Sigma_2(T) = \sum_{k=0}^{\infty} r_{i(k)}\, b^{k/2}. \tag{2}$$

We will see that, up to a multiplicative constant depending only upon b, $\Sigma_2(T)$ is equivalent to the Dudley integral. Let $I \subset \mathbb{N}$ be the set of values $i(k)$, $k \geqslant 0$. For every $i \in I$, let $k(i)$ be the largest k such that $i(k) = i$. Summing geometric progressions, we get

$$\Sigma_2(T) = \sum_{i \in I} \Bigl(\sum_{i(k)=i} r_i\, b^{k/2} \Bigr) \leqslant \sum_{i \in I} r_i \frac{\bigl(\sqrt{b}\bigr)^{k(i)+1} - 1}{\sqrt{b} - 1} < \frac{b^{1/2}}{b^{1/2} - 1} \sum_{i \in I} r_i\, b^{k(i)/2}.$$

When $i(k) = i$ we have $b^{k-1} < \ln \mathcal{N}(r_{i+1})$, thus $b^{k(i)} < b \ln \mathcal{N}(r_{i+1})$ and

$$\Sigma_2(T) < \frac{b}{b^{1/2}-1} \sum_{i \in I} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} \leqslant \frac{b}{b^{1/2}-1} \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} = \frac{b}{b^{1/2}-1} \Sigma_1(T).$$

In the other direction, consider for each $k \geqslant 0$ the (perhaps empty) interval of integers

$$I_k = \bigl\{ i \in \mathbb{N} : b^{k-1} < \ln \mathcal{N}(r_{i+1}) \leqslant b^k \bigr\}.$$

These intervals cover $\mathbb{N}$, because $b^{-1} < \ln \mathcal{N}(r_1)$, so that $i = 0$ belongs to some I_k, and then every $i \geqslant 0$ does as well. Let $K \subset \mathbb{N}$ denote the set of k such that I_k is not empty. When $k \in K$, we see that $\min I_k = i(k)$, and we observe that

$$\sum_{i \in I_k} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} \leqslant \Bigl(\sum_{i \in I_k} r_i \Bigr) b^{k/2} < \frac{3}{2} r_{i(k)} \, b^{k/2}.$$

Then

$$\Sigma_1(T) = \sum_{i=0}^{\infty} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} = \sum_{k \in K} \sum_{i \in I_k} r_i \sqrt{\ln \mathcal{N}(r_{i+1})} < \frac{3}{2} \sum_{k \in K} r_{i(k)} \, b^{k/2} \leqslant \frac{3}{2} \Sigma_2(T).$$

By the definition of $i(k)$, we know that $\ln \mathcal{N}(r_{i(k)}) \leqslant b^{k-1} < \ln \mathcal{N}(r_{i(k)+1})$. Hence, for every integer $k \geqslant 0$, there exists a finite set $S_k \subset T$ such that $\ln |S_k| \leqslant b^{k-1}$ and such that the balls of radius $\rho_k = r_{i(k)}$ centered at the points of S_k cover T; for $k = 0$, we have $|S_0| \leqslant \exp(b^{-1}) < 2$ thus $|S_0| = 1$, and $\rho_0 = r_0$. If $\Sigma_2(T)$ is finite, we may summarize the situation as follows:

$$(3) \qquad |S_0| = 1; \;\; \ln |S_k| < b^k \text{ and } T = \bigcup_{s \in S_k} B(s, \rho_k) \text{ for all } k \geqslant 0; \quad \sum_{k=0}^{\infty} \rho_k \, b^{k/2} < \infty.$$

1.2. *Gaussian random variables*

This section is completely elementary and contains some basic definitions, together with the standard Lemmas 1 and 2, given here with explicit constants resulting from tedious calculations. A random variable X defined on a probability space (Ω, P) is said to be a $\mathrm{N}(0,1)$ *Gaussian random variable* when for every $x \in \mathbb{R}$, we have

$$\mathrm{P}(X > x) = \int_x^{\infty} \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}}.$$

We then get for the expectation $\mathrm{E}\, X$ and the variance $\mathrm{Var}\, X$ of X the values

$$\mathrm{E}\, X = \int_{\mathbb{R}} u \, \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} = 0,$$

and

$$\mathrm{Var}\, X := \mathrm{E}(X - \mathrm{E}\, X)^2 = \mathrm{E}\, X^2 = \int_{\mathbb{R}} u^2 \, \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} = 1.$$

So, the "0" and the "1" in $\mathrm{N}(0,1)$ refer to the expectation and variance of X.

Let $x > 0$ and observe that

$$(4) \qquad \mathrm{P}(X > x) = \int_x^{\infty} \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} < \int_x^{\infty} \frac{u}{x} \, \mathrm{e}^{-u^2/2} \, \frac{\mathrm{d}u}{\sqrt{2\pi}} = \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}}.$$

This estimate is essentially correct, for example because

$$\int_x^{\infty} \mathrm{e}^{-u^2/2} \, \mathrm{d}u \geqslant \int_x^{x+1/x} \mathrm{e}^{-u^2/2} \, \mathrm{d}u \geqslant \frac{1}{x} \, \mathrm{e}^{-(x+1/x)^2/2} = \frac{\mathrm{e}^{-(x^2/2)-1-1/(2x^2)}}{x},$$

so that we can infer that

$$x \geqslant 1 \quad \Rightarrow \quad \mathrm{P}(X > x) \geqslant \mathrm{e}^{-3/2}\, \frac{\mathrm{e}^{-x^2/2}}{x\sqrt{2\pi}} \cdot \tag{5}$$

When x goes to $+\infty$ we can do better, writing for $x > 0$ the equalities

$$\int_x^{\infty} \mathrm{e}^{-u^2/2}\, \mathrm{d}u = \int_x^{\infty} \frac{1}{u}\,(u\,\mathrm{e}^{-u^2/2})\,\mathrm{d}u = \frac{\mathrm{e}^{-x^2/2}}{x} - \int_x^{\infty} \frac{1}{u^2}\,\mathrm{e}^{-u^2/2}\,\mathrm{d}u$$

and repeating the trick,

$$\int_x^{\infty} \frac{1}{u^2}\,\mathrm{e}^{-u^2/2}\, \mathrm{d}u = \int_x^{\infty} \frac{1}{u^3}\,(u\,\mathrm{e}^{-u^2/2})\,\mathrm{d}u = \frac{\mathrm{e}^{-x^2/2}}{x^3} - \int_x^{\infty} \frac{3}{u^4}\,\mathrm{e}^{-u^2/2}\,\mathrm{d}u.$$

We obtain that

$$x > 0 \quad \Rightarrow \quad \int_x^{\infty} \mathrm{e}^{-u^2/2}\,\mathrm{d}u > \mathrm{e}^{-x^2/2}\Bigl(\frac{1}{x} - \frac{1}{x^3}\Bigr) = \frac{\mathrm{e}^{-x^2/2}}{x}\Bigl(1 - \frac{1}{x^2}\Bigr), \tag{6}$$

a certainly uninteresting assertion when $0 < x \leqslant 1$. It is easy to guess how to go on and produce an asymptotic expansion of $\mathrm{P}(X > x)$ in terms of the variable $x > 1$.

When g is a $\mathrm{N}(0,1)$ Gaussian random variable and $u > 0$, one has

$$P(|g| > u) \leqslant \mathrm{e}^{-u^2/2}\,. \tag{7}$$

We have to prove $\mathrm{P}(g > x) \leqslant \frac{1}{2}\,\mathrm{e}^{-x^2/2}$ when $x \geqslant 0$. We know it already for $x \geqslant \sqrt{2/\pi}$ by (4), and the remaining values of x are obtained by checking the sign of the derivative of the function $f : x \mapsto \mathrm{e}^{-x^2/2} - 2\,\mathrm{P}(g > x)$ on the segment $[0, \sqrt{2/\pi}]$, namely the sign of

$$f'(x) = \bigl(\sqrt{2/\pi} - x\bigr)\,\mathrm{e}^{-x^2/2}\,.$$

Inequality (7) holds *a fortiori* for any centered Gaussian random variable Y having variance $\leqslant 1$: we can write $Y = \theta g'$ with $\theta = (\mathrm{E}\,Y^2)^{1/2} \in [0,1]$ and with g' being a $\mathrm{N}(0,1)$ variable, therefore

$$\mathrm{E}\,Y^2 \leqslant 1 \quad \Rightarrow \quad \mathrm{P}(|Y| > u) \leqslant \mathrm{P}(|g| > u) \leqslant \mathrm{e}^{-u^2/2}\,.$$

Lemma 1. *If $N \geqslant 2$ and if $g_1, g_2, \ldots, g_N$ are $\mathrm{N}(0,1)$ Gaussian random variables, independent or not, we have*

$$\mathrm{E}\Bigl(\max_{1\leqslant i\leqslant N} |g_i|\Bigr) \leqslant \sqrt{2\ln N} + \frac{1}{\sqrt{2\ln N}} \cdot$$

It follows that for every $N \geqslant 1$, we have

$$\mathrm{E}\Bigl(\max_{1\leqslant i\leqslant N} |g_i|\Bigr) \leqslant 2\sqrt{\ln(N+1)}. \tag{8}$$

If $\sigma \geqslant 0$ and if $X_1, X_2, \ldots, X_N$ are centered Gaussian random variables, then

$$\max_{1\leqslant i\leqslant N} \mathrm{E}\,X_i^2 \leqslant \sigma^2 \quad \Rightarrow \quad \mathrm{E}\Bigl(\max_{1\leqslant i\leqslant N} |X_i|\Bigr) \leqslant 2\sigma\sqrt{\ln(N+1)}. \tag{9}$$

Proof. Let $N \geqslant 2$, $G_N^* = \max_{1\leqslant i\leqslant N} |g_i|$ and $x > 0$; by (7) we know that

$$P(|g_1| > x) = P(|g_2| > x) = \cdots = P(|g_N| > x) \leqslant \mathrm{e}^{-x^2/2},$$

and by the union bound inequality, we get

$$P(G_N^* > x) \leqslant N\,\mathrm{e}^{-x^2/2}\,.$$

Observe that if $x_0 = \sqrt{2 \ln N}$, then $N\,\mathrm{e}^{-x_0^2/2} = 1$, and $x_0 \geqslant \sqrt{2 \ln 2} > 0$. It follows that

$$\mathrm{E}\, G_N^* = \int_0^\infty \mathrm{P}(G_N^* > x)\,\mathrm{d}x \leqslant x_0 + N \int_{x_0}^\infty \mathrm{e}^{-x^2/2}\,\mathrm{d}x$$
$$\leqslant x_0 + N \int_{x_0}^\infty \frac{x}{x_0}\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x = x_0 + N\,\frac{\mathrm{e}^{-x_0^2/2}}{x_0} = x_0 + \frac{1}{x_0}\cdot$$

For the second inequality (8), we have first $\mathrm{E}\, G_1^* = \mathrm{E}\,|g_1| = \sqrt{2/\pi} < 1 < 2\sqrt{\ln 2}$. Next, we use

$$\sqrt{2 \ln y} + \frac{1}{\sqrt{2 \ln y}} < 2\sqrt{\ln(y+1)}$$

when $y \geqslant 2$, or equivalently the fact that

$$y \geqslant 2 \quad \Rightarrow \quad f(y) = 4 \ln(y+1) - 2 \ln y - 2 - \frac{1}{2 \ln y} > 0,$$

which is true because

$$f'(y) = \frac{4}{y+1} - \frac{2}{y} + \frac{1}{2y(\ln y)^2} > \frac{2y-2}{y(y+1)}$$

is > 0 when $y \geqslant 2$, and because $f(2) > \ln(81/4) - 3 > 0$.

If $\sigma \geqslant 0$ and if $X_1, X_2, \ldots, X_N$ are centered Gaussian such that $\mathrm{Var}\, X_i \leqslant \sigma^2$, the sequence has the same distribution as a sequence $\sigma_1 g_1, \sigma_2 g_2, \ldots, \sigma_N g_N$ with $0 \leqslant \sigma_i \leqslant \sigma$ and g_i a $\mathrm{N}(0,1)$ variable, therefore

$$\mathrm{E} \max_{1 \leqslant i \leqslant N} |X_i| = \mathrm{E} \max_{1 \leqslant i \leqslant N} |\sigma_i g_i| \leqslant \sigma\, \mathrm{E} \max_{1 \leqslant i \leqslant N} |g_i|$$

and the result (9) follows. □

Inequality (8) applies as well to *subgaussian variables* properly normalized, namely sequences $X_1, X_2, \ldots, X_N$ such that for each i and every $x > 0$, we have

$$\mathrm{P}(|X_i| > x) \leqslant \mathrm{e}^{-x^2/2},$$

as this is all that was used in the above proof.

Lemma 2. *If $N \geqslant 1$ and if $g_1, g_2, \ldots, g_N$ are independent $\mathrm{N}(0,1)$ Gaussian random variables, one has*

$$\mathrm{E}\Bigl(\max_{1 \leqslant i \leqslant N} g_i\Bigr) \geqslant \frac{1}{2}\sqrt{\ln N}. \tag{10}$$

If $\sigma \geqslant 0$ and if $X_1, X_2, \ldots, X_N$ are independent centered Gaussian random variables, then

$$\min_{1 \leqslant i \leqslant N} \mathrm{E}\, X_i^2 \geqslant \sigma^2 \;\Rightarrow\; \mathrm{E}\Bigl(\max_{1 \leqslant i \leqslant N} X_i\Bigr) \geqslant \frac{1}{2}\sigma\sqrt{\ln N}. \tag{11}$$

Furthermore, when the (g_i) are as above and N tends to ∞, one has

$$\frac{\mathrm{E}\Bigl(\max_{1 \leqslant i \leqslant N} g_i\Bigr)}{\sqrt{2 \ln N}} \longrightarrow 1.$$

Proof. Let

$$g_N^* = \max_{1 \leqslant i \leqslant N} g_i.$$

We have $\mathrm{E}\, g_1^* = \mathrm{E}\, g_1 = 0 \geqslant (1/2)\sqrt{\ln 1}$, it can be shown (b) easily that

$$\mathrm{E}\, g_2^* = 1/\sqrt{\pi}, \quad \text{and one checks that} \quad 1/\sqrt{\pi} > (1/2)\sqrt{\ln 2}.$$

Slightly less easy are the facts that

$$\mathrm{E}\, g_3^* = 3/(2\sqrt{\pi}), \qquad \text{and} \quad 3/(2\sqrt{\pi}) > (1/2)\sqrt{\ln 3},$$
$$\mathrm{E}\, g_4^* = 6 \arctan(\sqrt{2})/\pi^{3/2} > 1, \qquad \text{and} \quad 1 > (1/2)\sqrt{\ln 4}.$$

Our last effort has been to establish that

$$\mathrm{E}\, g_5^* = \frac{15}{\pi^{3/2}} \Bigl(\frac{\pi}{3} - \arcsin\Bigl(\frac{1}{\sqrt{3}} \Bigr) \Bigr) > 1.162 > 0.635 > \frac{1}{2}\sqrt{\ln 5}.$$

Taking this for granted, let $x_0 = 2\, \mathrm{E}\, g_5^* > 2.324$. Clearly $\mathrm{E}\, g_N^*$ increases with N, so we need only consider values of N such that

$$\frac{1}{2}\sqrt{\ln N} > \mathrm{E}\, g_5^* = \frac{x_0}{2} > 1.162,$$

or $\ln N > x_0^2 > 5.4$ and $N > 221$. Hence, we shall restrict our study to integers N such that $\sqrt{\ln N} \geqslant x_0$. The function $y \mapsto y - \ln(2y)$ is increasing when $y > 1$, thus

$$\ln N - \ln(2 \ln N) \geqslant x_0^2 - \ln(2x_0^2).$$

Let $u = x_0^2 - \ln(2x_0^2)$. One can check that $u > 3$. Let

$$s = \sqrt{2 \ln N - \ln(2 \ln N) - u}.$$

We have

$$x_0 \leqslant \sqrt{\ln N} \leqslant s \leqslant \sqrt{2 \ln N}. \tag{12}$$

Next we use (6), then we notice that $x_0^2 > 5$ and we obtain

$$\begin{aligned} \mathrm{P}(g_1 > s) &\geqslant \frac{\mathrm{e}^{-s^2/2}}{s\sqrt{2\pi}} \Bigl(1 - \frac{1}{s^2}\Bigr) \geqslant \frac{\mathrm{e}^{-s^2/2}}{\sqrt{2 \ln N}\sqrt{2\pi}} \Bigl(1 - \frac{1}{x_0^2}\Bigr) \\ &= \frac{1}{N} \frac{\mathrm{e}^{u/2}}{\sqrt{2\pi}} \Bigl(1 - \frac{1}{x_0^2}\Bigr) > \frac{1}{N} \frac{\mathrm{e}^{u/2}}{\sqrt{2\pi}} \frac{4}{5} > \frac{1.40}{N}. \end{aligned}$$

It follows that

$$\mathrm{P}(g_N^* < s) = \mathrm{P}(g_1 < s)^N \leqslant (1 - 1.4/N)^N \leqslant \mathrm{e}^{-1.4} < 1/4.$$

We need to take care of the rare but possible negative values of g_N^*. Let

$$\nu(\omega) = \min\bigl(g_N^*(\omega), 0\bigr) \leqslant 0, \quad \omega \in \Omega.$$

We see that

$$\nu \geqslant \min(g_N, 0)\, \mathbf{1}_{g_{N-1}^* < 0},$$

thus by independence

$$\mathrm{E}\, \nu \geqslant \Bigl(\mathrm{E} \min(g_N, 0)\Bigr) \Bigl(\mathrm{E}\, \mathbf{1}_{g_{N-1}^* < 0}\Bigr) = -\frac{1}{\sqrt{2\pi}}\, 2^{-N+1} > -2^{-N}.$$

We know that $N > 221$ and $x_0 > 1$, therefore

$$2^{-N} < 2^{-221} x_0 \leqslant 2^{-221} \sqrt{\ln N}.$$

Finally, using (12),

$$\begin{aligned} \mathrm{E}\, g_N^* &\geqslant s\, \mathrm{P}(g_N^* > s) + \mathrm{E}\, \nu \geqslant (1 - \mathrm{e}^{-1.4})\sqrt{\ln N} - 2^{-N} \\ &> \Bigl(\frac{3}{4} - 2^{-221}\Bigr) \sqrt{\ln N} > \frac{1}{2} \sqrt{\ln N}. \end{aligned}$$

The claim about non $\mathrm{N}(0,1)$ variables is proved as before. Let us explain rapidly the last sentence. Lemma 1 implies that the limsup of the quotient $\mathrm{E}\, g_N^*/\sqrt{\ln N}$ is $\leqslant \sqrt{2}$. For the other direction, fix u rather large and $\varepsilon > 0$ small. When N goes to ∞, we certainly have

$$\varepsilon \ln N - \ln(2 \ln N) - u \geqslant 0.$$

Then

$$\sqrt{2 \ln N} > s = \sqrt{2 \ln N - \ln(2 \ln N) - u} \geqslant \sqrt{(2-\varepsilon) \ln N}.$$

Because $s > 1$ we may use (5) and write

$$\mathrm{P}(g_1 > s) \geqslant \mathrm{e}^{-3/2} \frac{\mathrm{e}^{-s^2/2}}{s\sqrt{2\pi}} \geqslant \frac{\mathrm{e}^{-(s^2+3)/2}}{\sqrt{2 \ln N}\sqrt{2\pi}} = \frac{1}{N} \frac{\mathrm{e}^{(u-3)/2}}{\sqrt{2\pi}} =: \frac{\alpha}{N},$$

but now α can be made as large as we wish, and

$$\mathrm{E}\, g_N^* \geqslant s\, \mathrm{P}(g_N^* > s) - 2^{-N} \geqslant (1 - \mathrm{e}^{-\alpha} - 2^{-N}) \sqrt{(2-\varepsilon) \ln N}. \qquad \square$$

Numerical experiments suggest that the quotient $\mathrm{E}\, g_N^*/\sqrt{\ln N}$ is actually increasing with $N \geqslant 2$. If this were true, the correct constant $c > 1/2$ in the inequality (10) for all $N \geqslant 2$ would simply be the value at $N = 2$, namely $c = 1/\sqrt{\pi \ln 2} > 0.67 > 2/3$.

1.3. *The comparison result*

The next comparison result plays a major rôle in what follows.

Proposition 1. *Let $(X_t)_{t \in T}$ and $(Y_t)_{t \in T}$ be two centered Gaussian processes indexed by the same set T. If we have*

$$\mathrm{E}(Y_s - Y_t)^2 \leqslant \mathrm{E}(X_s - X_t)^2$$

for all $s, t \in T$, we conclude that

$$\mathrm{E}\Bigl(\sup_{t \in T} Y_t\Bigr) \leqslant \mathrm{E}\Bigl(\sup_{t \in T} X_t\Bigr).$$

The centering is necessary here, as the simple example $Y_t = 1 + X_t$ shows.

The proof of this result is not as elementary as what we have seen so far, it will not be given here. A first comparison result goes back to Slepian [Slep] in 1962, and it applies to comparing the *distributions* of the suprema: if in addition to the hypothesis of Proposition 1 one adds that $\mathrm{E}\, Y_t^2 = \mathrm{E}\, X_t^2$ for every $t \in T$, then for every x real one has

$$\mathrm{P}\Bigl(\sup_{t \in T} Y_t > x\Bigr) \leqslant \mathrm{P}\Bigl(\sup_{t \in T} X_t > x\Bigr).$$

Under this additional assumption, we see that Slepian's lemma implies the conclusion of Proposition 1.

The comparison result in Proposition 1 was announced in a Note without proof by Sudakov; a complete proof can be found in Sudakov's book [Sud]. The result was also approached by Chevet [Chev], and given by Fernique [Fer_1] ([c]).

A more general version of Proposition 1 is due to Gordon [Gor], and deals with a mixture of min and max; a proof can be found in Chap. 8 of Li–Queffélec [LiQ]. Gordon's result is extremely useful for estimating the invertibility of Gaussian random maps between finite dimensional normed spaces, as estimating the norm of the inverse of a Gaussian random map $T_\omega : E \to F$ involves estimating the inf of norms $\|T_\omega(x)\|_F$ of the images $T_\omega(x)$ of norm one vectors $x \in E$, where each norm $\|T_\omega(x)\|_F$ is a max of Gaussian random variables of the form $< y^*, T_\omega(x) >$, and the y^* are all the norm one linear functionals on the target space F.

1.3.1. *The Sudakov entropy bound*

Suppose that $\operatorname{E}\sup_{t\in T} X_t$ is finite. It follows from Proposition 1 that for every $\varepsilon > 0$, we can cover T with a finite number of balls of radius ε: indeed, if $t_1, t_2, \ldots, t_n$ in T have mutual distances larger than ε, we shall compare the processes $(X_{t_j})_{j=1}^n$ and $(Y_{t_j})_{j=1}^n$ where the Y_{t_j} are independent centered Gaussian random variables of variance $\varepsilon^2/2$. We have when $i \neq j$

$$\operatorname{E}(Y_{t_j} - Y_{t_i})^2 = \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2} \leqslant d(t_j, t_i)^2 = \operatorname{E}(X_{t_j} - X_{t_i})^2,$$

hence by Proposition 1 and Inequality (11),

$$\operatorname{E}\Bigl(\sup_{t\in T} X_t\Bigr) \geqslant \operatorname{E}\Bigl(\sup_{1\leqslant i\leqslant n} X_{t_i}\Bigr) \geqslant \operatorname{E}\Bigl(\sup_{1\leqslant i\leqslant n} Y_{t_i}\Bigr) \geqslant \frac{\varepsilon}{2\sqrt{2}}\sqrt{\ln n}.$$

This provides a bound on n, namely,

$$\mathcal{N}(T, \varepsilon) \leqslant \exp\biggl(\frac{8}{\varepsilon^2}\Bigl(\operatorname{E}\sup_{t\in T} X_t\Bigr)^2\biggr).$$

It follows that when the expectation of $\sup_{t\in T} X_t$ is finite, the closure in $L^2(\Omega, \mathrm{P})$ of the set of variables $(X_t)_{t\in T}$ is a compact subset of L^2.

Given $\delta > 0$, we say that a subset $S \subset T$ is *δ-separated* when any two of its points are δ-far apart,

$$s_1, s_2 \in S,\ s_1 \neq s_2 \quad \Rightarrow \quad d(s_1, s_2) \geqslant \delta.$$

Given a subset $A \subset T$, a *δ-net for A* is a maximal δ-separated subset S of A: maximality implies that no point of A can be added to S and keep it δ-separated: for every $a \in A$, there is some $s \in S$ such that $d(a, s) < \delta$, in other words,

$$A \subset \bigcup_{s\in S} B(s, \delta).$$

We will need to consider the supremum of the process when the index t ranges, not only in the whole of T, but also in balls. For this we introduce when $s \in T$, $r > 0$, the quantity

$$\varphi_X(s, r) = \operatorname{E}\Bigl(\sup_{t\in B(s,r)} X_t\Bigr). \tag{13}$$

Under the assumption of Proposition 1, we have

$$\varphi_Y(s, r) \leqslant \varphi_X(s, r)$$

for all $s \in T$, $r > 0$: we just observe that the assumption of Proposition 1 holds for the two processes restricted to the ball $B(s, r)$. When only one process (X_t) appears in the discussion, we shall simply write $\varphi(s, r) = \varphi_X(s, r)$.

It is obvious that $\varphi(t, r)$ is non-decreasing in r, perhaps not continuous in r: if t_0 is isolated in T, with $B(t_0, r_0) = \{t_0\}$ and $X_{t_0} \neq 0$, and if there are non zero random variables X_s in the process with $d(t_0, s) = r_0$, then $\varphi(t_0, r) = \operatorname{E} X_{t_0}$ is 0 when $r < r_0$ but jumps at r_0 to a non-zero value larger than or equal to $\operatorname{E}\max(X_{t_0}, X_s) > 0$.

1.3.2. *A convex digression*

We may use *autoindexation* for the process, by considering that the indexing set T is precisely the subset of $L^2(\Omega, P)$ consisting of the variables in the process. This would not take care of situations where perhaps $s \neq t$ but $X_s = X_t$; they are anyway irrelevant when dealing with the supremum of a process. This approach of considering subsets of $L^2(\Omega, \mathrm{P})$ is the one taken by Sudakov in his book [Sud].

Using autoindexing, we understand that passing from $T \subset L^2(\Omega, P)$ to the convex hull $\mathrm{conv}(T)$ of T in $L^2(\Omega, \mathrm{P})$ will not change the supremum: for every $\omega \in \Omega$, we have

$$\sup_{t \in T} X_t(\omega) = \sup_{s \in \mathrm{conv}(T)} X_s(\omega).$$

If we use autoindexation, we may assume that T is a convex set in $L^2(\Omega, \mathrm{P})$; however, when $T \subset L^2$ and $t \in T$ we will prefer writing X_t than just $t = X_t$. If $s \in T$, $\theta \in (0,1)$ and $t \in B(s,r)$, then $(1-\theta)s + \theta t \in B(s, \theta r)$ hence

$$\sup_{u \in B(s,\theta r)} X_u \geqslant (1-\theta) X_s + \theta \sup_{t \in B(s,r)} X_t,$$

and since $\mathrm{E}\, X_s = 0$, we see that

$$\varphi(s, \theta r) \geqslant \theta \varphi(s, r).$$

If $s, t \in T$ are such that $d(s,t) < \delta$, then $B(t,r) \subset B(s, r+\delta)$, therefore

$$\varphi(t,r) \leqslant \varphi(s, r+\delta) \leqslant \frac{r+\delta}{r}\, \varphi(s,r).$$

It follows that $\varphi(t,r)$ is continuous in $t \in T$ in the convex case.

1.4. *First applications*

We begin with a simple lemma.

Lemma 3. *Let (X_i), $i \in I$, and $(Y_{i,j})$, $i \in I$ and $j \in J_i$, be two independent families of integrable real random variables. One has that*

$$\mathrm{E}\Big(\sup_{i,j} (X_i + Y_{i,j}) \Big) \geqslant \mathrm{E}\Big(\sup_{i \in I} X_i \Big) + \inf_{i \in I} \mathrm{E}\Big(\sup_{j \in J_i} Y_{i,j} \Big),$$

where we must agree that $(+\infty) + (-\infty) = -\infty$ if it appears in the sum above.

Proof. Let

$$Y_* = \inf_{i \in I} \mathrm{E}\Big(\sup_{j \in J_i} Y_{i,j} \Big) \in [-\infty, \infty].$$

By independence, we may think of the X_i as functions of a variable u while the $Y_{i,j}$ are functions of a different variable v. We have

$$\mathrm{E}_v \sup_{i,j} \big(X_i + Y_{i,j} \big) = \mathrm{E}_v \sup_{i \in I} \sup_{j \in J_i} \big(X_i + Y_{i,j} \big) \geqslant \sup_{i \in I} \mathrm{E}_v \sup_{j \in J_i} \big(X_i + Y_{i,j} \big),$$

then

$$\mathrm{E}_v \sup_{j \in J_i} \big(X_i(u) + Y_{i,j}(v) \big) = X_i(u) + \mathrm{E} \sup_{j \in J_i} Y_{i,j} \geqslant X_i(u) + Y_*$$

and

$$\sup_{i \in I} \mathrm{E}_v \sup_{j \in J_i} \big(X_i + Y_{i,j} \big) \geqslant \sup_{i \in I} X_i(u) + Y_*.$$

Integrating in u concludes the proof. □

The next Lemma does most of the serious job in what follows([d]).

Lemma 4. *Suppose that* $(X_t)_{t\in A}$ *is a centered Gaussian process, that* ρ, δ, λ *are positive real numbers such that* $\rho^2 + \lambda^2 \leqslant \delta^2/2$ *and that* $(A_i)_{i=1}^N$ *are subsets of* A *satisfying*

– *there exists* $a_i \in A$ *such that* $A_i \subset B(a_i, \rho)$, *and*

– *for all* $t_i \in A_i$, $t_j \in A_j$ *and* $i \neq j$ *we have* $d(t_i, t_j) \geqslant \delta$ *(we may say that the sets* A_i *and* A_j *are* δ-separated*).*

It follows that

$$\mathrm{E}\Bigl(\sup_{t\in A} X_t\Bigr) \geqslant (\lambda/2)\sqrt{\ln N} + \min_{1\leqslant i\leqslant N} \mathrm{E}\Bigl(\sup_{t\in A_i} X_t\Bigr).$$

Proof. Let $(X_t^{(i)})_{i=1}^N$ be N independent copies of the process (X_t), let $g_1, g_2, \ldots, g_N$ be independent $\mathrm{N}(0,1)$ Gaussian variables, that are also independent from the $(X_t^{(i)})_{i=1}^N$, and set

$$A_* = \bigcup_{i=1}^N A_i \subset A.$$

Let us define another Gaussian process $(Y_t)_{t\in A_*}$ as

$$Y_t = X_t^{(i)} - X_{a_i}^{(i)} + \lambda g_i \quad \text{when} \quad t \in A_i.$$

We want to check that

$$\mathrm{E}(Y_s - Y_t)^2 \leqslant \mathrm{E}(X_s - X_t)^2$$

for all $s, t \in A_*$, in order to apply the Sudakov–Slepian lemma; there are two cases to consider: if there is an index i such that $s, t \in A_i$, we have

$$Y_s - Y_t = X_s^{(i)} - X_t^{(i)},$$

hence

$$\mathrm{E}(Y_s - Y_t)^2 = \mathrm{E}(X_s - X_t)^2$$

in this first case. If now $s \in A_i$, $t \in A_j$ and $i \neq j$, we see that

$$Y_s - Y_t = (X_s^{(i)} - X_{a_i}^{(i)}) + \lambda g_i - (X_t^{(j)} - X_{a_j}^{(j)}) - \lambda g_j\,;$$

the random variables $X_s^{(i)} - X_{a_i}^{(i)}$, g_i, $X_t^{(j)} - X_{a_j}^{(j)}$ and g_j are centered and independent, hence orthogonal, and $d(s, a_i) \leqslant \rho$, $d(t, a_j) \leqslant \rho$, therefore

$$\mathrm{E}(Y_s - Y_t)^2 \leqslant 2\,\rho^2 + 2\,\lambda^2 \leqslant \delta^2 \leqslant \mathrm{E}(X_s - X_t)^2,$$

because we know that $d(s,t) \geqslant \delta$. Using the Sudakov–Slepian Lemma we obtain

$$\mathrm{E}\Bigl(\sup_{t\in A_*} Y_t\Bigr) \leqslant \mathrm{E}\Bigl(\sup_{t\in A_*} X_t\Bigr)$$

and as $A_* \subset A$ we have obviously that

$$\mathrm{E}\Bigl(\sup_{t\in A_*} X_t\Bigr) \leqslant \mathrm{E}\Bigl(\sup_{t\in A} X_t\Bigr).$$

It remains to apply Lemma 3 and the lower bound (10) to

$$\begin{aligned}
\mathrm{E}\Bigl(\sup_{t\in A} X_t\Bigr) &\geqslant \mathrm{E}\sup_{t\in A_*} Y_t = \mathrm{E}\sup_i \sup_{t\in A_i}\bigl((X_t^{(i)} - X_{a_i}^{(i)}) + \lambda g_i\bigr)\\
&\geqslant \mathrm{E}\max_{1\leqslant i\leqslant N}(\lambda g_i) + \min_i \mathrm{E}\sup_{t\in A_i}(X_t^{(i)} - X_{a_i}^{(i)})\\
&\geqslant (1/2)\,\lambda\sqrt{\ln N} + \min_i\bigl(\mathrm{E}\sup_{t\in A_i} X_t - \mathrm{E}\,X_{a_i}\bigr)\\
&= (\lambda/2)\sqrt{\ln N} + \min_i \mathrm{E}\sup_{t\in A_i} X_t. \qquad \square
\end{aligned}$$

Corollary 1. *Suppose that $(X_t)_{t\in T}$ is a centered Gaussian process, that $S \subset T$ is a finite 2δ-separated set contained in a ball $B(t_0, r)$, where $t_0 \in T$, $\delta, r > 0$. It follows that*

$$\mathrm{E}\Bigl(\sup_{t\in B(t_0, r+\delta/2)} X_t\Bigr) \geqslant (\delta/4)\sqrt{\ln |S|} + \min_{s\in S} \mathrm{E}\Bigl(\sup_{u\in B(s,\delta/2)} X_u\Bigr),$$

or, using Notation (13) and letting $N = |S|$ denote the cardinality of S,

$$\varphi(t_0, r+\delta/2) \geqslant (\delta/4)\sqrt{\ln N} + \min_{s\in S} \varphi(s, \delta/2).$$

Proof. We apply Lemma 4 with $\lambda = \rho = \delta/2$, $A = B(t_0, r+\delta/2)$ and $A_i = B(s_i, \rho)$, where $S = \{s_1, s_2, \ldots, s_N\}$. Then $\rho^2 + \lambda^2 = \delta^2/2$, and by the triangle inequality the balls $B(s_i, \rho) = B(s_i, \delta/2)$ are δ-separated and contained in A. □

1.4.1. *Trees and branches*

We shall deal with rooted trees $\mathcal{X}$; we consider that $\mathcal{X}$ is a set of *nodes*, and that to each node $x \in \mathcal{X}$ is associated a finite subset $C(x)$ of nodes in $\mathcal{X}$ that are the *children* of x, with $x \notin C(x)$ of course. We assume that this *child relation* defines a rooted tree, the root of $\mathcal{X}$ will be called x_0. For each child $y \in C(x)$ we say that x is the *parent* of y. The *descendants* of $x \in \mathcal{X}$ are the elements of one the sets $C^{(k)}(x)$, $k \geqslant 1$, inductively defined by letting $C^{(1)}(x) = C(x)$ and

$$C^{(k+1)}(x) = \bigcup_{y\in C(x)} C^{(k)}(y), \quad k \geqslant 1.$$

For $x \in \mathcal{X}$, we shall pay attention to the number of *siblings* of x: except if x is the root, this is the number of children of the parent of x; this number is set equal to 1 for the root. We shall indicate the successive levels in the tree by successive nonnegative integers: level 0 for the root, level 1 for the children of the root, 2 for the grandchildren of the root, and so on. The level of a node x in the tree will be denoted by $\ell(x) \in \mathbb{N}$,

$$y \in C(x) \ \Rightarrow\ \ell(y) = \ell(x) + 1.$$

The *maximal branches* of the tree are sequences $\mathbf{x} = (x_i)_{i\in\mathbb{N},\, i<\ell}$ of nodes in the tree, such that x_0 is the root of the tree, and x_{i+1} is a child of x_i whenever $i \in \mathbb{N}$ and $i+1 < \ell$, with ℓ finite or $\ell = +\infty$. When ℓ is finite, the node $x_{\ell-1}$ has no child, it is a *leaf* of the tree. We shall prefer avoiding leaves: the maximal branches of our trees will all be infinite, this will help us to keep a somewhat unified treatment. An infinite maximal branch $\mathbf{x}$ will look like

$$\mathbf{x} = (x_0,\ x_1,\ x_2,\ \ldots,\ x_j,\ \ldots).$$

A *path* in the tree will be a portion of a branch, namely, a finite sequence of nodes

$$x_i,\ x_{i+1},\ \ldots,\ x_k,$$

where x_{j+1} is a child of x_j for every j such that $i \leqslant j < k$.

2. The Fernique theorem

Let G be a compact abelian additive group and let $(X_g)_{g\in G}$ be a Gaussian process indexed by G. We suppose that this process is *invariant* under the group action, meaning that for every finite sequence $\{h, g_1, \ldots, g_k\}$ of elements of G, the two k-tuples

$$(X_{g_1}, X_{g_2}, \ldots, X_{g_k}) \quad \text{and} \quad (X_{g_1+h}, X_{g_2+h}, \ldots, X_{g_k+h})$$

have the same joint distribution. We of course assume that the family $(X_g)_{g\in G}$ is not reduced to a single random variable. We may modify the point of view, and consider that the group acts on the set $\{X_g : g \in G\}$ of random variables, letting

$$h.X_g = X_{g+h}.$$

The action is transitive: given $g_0, g_1 \in G$, we have with $h = g_1 - g_0$ that

$$h.X_{g_0} = X_{g_1}.$$

We now arrive to our definitive setting: we assume that $(X_t)_{t\in T}$ is a centered Gaussian process, that G is a multiplicative group acting transitively on T, and that for every integer $k \geqslant 1$, every sequence $\{t_1, \ldots, t_k\}$ of elements of T and every $g \in G$, the two k-tuples

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \quad \text{and} \quad (X_{g.t_1}, X_{g.t_2}, \ldots, X_{g.t_k}) \tag{14}$$

have the same joint distribution. It follows that for $t_1, t_2 \in T$ and $g \in G$,

$$d(t_1, t_2) = d(g.t_1, g.t_2).$$

If $B(t, r)$ is a ball in T and $g \in G$, we see that

$$g.B(t, r) \subset B(g.t, r) \quad \text{therefore} \quad g.B(t, r) = B(g.t, r)$$

by letting the inverse g^{-1} of g act on $B(g.t, r)$. If $s_1, s_2, \ldots, s_n$ are arbitrary elements in $B(t, r)$, then $g.s_1, g.s_2, \ldots, g.s_n$ are in $B(g.t, r)$, and the distribution invariance implies that

$$\mathrm{E}\Bigl(\sup_{1\leqslant i\leqslant n} X_{s_i}\Bigr) = \mathrm{E}\Bigl(\sup_{1\leqslant i\leqslant n} X_{g.s_i}\Bigr).$$

Invariance and transitivity of the group action imply that for any t_0, t_1 in T,

$$\varphi(t_0, r) = \mathrm{E}\Bigl(\sup_{s\in B(t_0,r)} X_s\Bigr) = \mathrm{E}\Bigl(\sup_{s\in B(t_1,r)} X_s\Bigr) = \varphi(t_1, r). \tag{15}$$

In this invariant setting (e), we simply let

$$\varphi(r) = \varphi(t_0, r)$$

for $r > 0$ and for an arbitrary fixed $t_0 \in T$.

We shall prove the Fernique theorem under the form that follows.

Theorem 1. *Let $(X_t)_{t\in T}$ be a centered Gaussian process that satisfies the invariance condition (14). If the expectation of the supremum is finite, then the Dudley integral is finite, and more precisely*

$$I_D(T) \leqslant 432\, \mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr).$$

We can claim with absolute certainty that the constant 432 that appears above is not optimal.

2.1. *Looking for an entropy-like condition*

We assume that we have an invariant centered Gaussian process $(X_t)_{t\in T}$ such that

$$E^* := \mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr) < \infty.$$

Due to invariance, we can set $\varphi(r) = \varphi(t, r)$ for every $t \in T$ and every radius $r > 0$, and we know that $\varphi(r) \leqslant E^* < +\infty$. We shall construct a rooted tree $\mathcal{X}$ whose nodes have the form $x = (t, r, N)$, where $t \in T$, $r > 0$, and the integer N is an upper bound for the number of siblings of x. The maximal branches of this tree will all be infinite, at the cost of doing things in a way that is slightly unnatural (f), but that helps us to give a moderately unified treatment, making no difference between the case of a *perfect* set T or that of a finite set T.

Let $\Delta > 0$ denote the diameter of T, let t_0 be an arbitrary point in T and let r_0 satisfy $\Delta < r_0 < 4\Delta/3$, so that $B(t_0, r_0) = T$. The root of the tree is $x_0 = (t_0, r_0, N_0)$ with $N_0 = 1$. The inductive construction of the tree goes as follows:

suppose that a node $x = (t, r, N)$ has already been introduced in the tree $\mathcal{X}$. Let

$$\widetilde{r} = r/3 \ \text{ and let } \ \widetilde{S} \ \text{ be a } \widetilde{r}\text{-net for } \ B(t, r).$$

The size of $\widetilde{S}$ will play a rôle in what follows. The case $|\widetilde{S}| > N^3$ corresponds to when $B(t, r)$ is "not too small", and we call this the "rich case", the case $|\widetilde{S}| \leqslant N^3$ being of course the "poor" one. In the rich case, we let

$$\widetilde{N} = |\widetilde{S}|, \ \text{ and in the poor case } \ \widetilde{N} = N^3, \ \text{ so that } \ \widetilde{N} = \max(N^3, |\widetilde{S}|).$$

For each $s \in \widetilde{S}$, we define a new node $y = (s, \widetilde{r}, \widetilde{N})$ of the tree, that is declared to be a child of x. Let us say things a little differently, mentioning the successive generations in the tree: if $i \geqslant 0$ is an integer and $x_i = (t_i, r_i, N_i)$ a node at level i in the tree, then let

$$r_{i+1} = r_i/3 \ \text{ and let } \ S_{i+1} \ \text{ be a } r_{i+1}\text{-net for } \ B(t_i, r_i).$$

Define $N_{i+1} = \max(N_i^3, |S_{i+1}|)$. The children of x_i have the form $x_{i+1} = (s, r_{i+1}, N_{i+1})$, where s varies in S_{i+1}. There is at least one child, and at most N_{i+1} children.

A few comments are in order.

— We see that r_i is just $3^{-i} r_0$, but I find that r_i is better looking than $3^{-i} r_0$: I will keep r_i throughout.

— Being an r_{i+1}-net for $B(t_i, r_i)$, the set S_{i+1} is r_{i+1}-separated and

$$B(t_i, r_i) \subset \bigcup_{s \in S_{i+1}} B(s, r_{i+1}).$$

— When $i = 0$ and when $x = x_0 = (t_0, r_0, 1)$ is the root, we have $\Delta < r_0 < 4\Delta/3$ and $B(t_0, r_0) = T$; we certainly know that $\mathcal{N}(T, r_0/3) > 1$ since $2(r_0/3) < \Delta$, implying that $B(t, r_0/3) \neq T$ for any $t \in T$. We have thus $N_1 \geqslant 2 > N_0^3 = N_0 = 1$.

— In the poor case, we have $N_{i+1} = N_i^3$ and N_{i+1} is merely an *upper bound* for the number of children of x_i. We may encounter a degenerate case $x = (t, r, N)$ where the ball $B(t, r)$ is a finite set with less than N^3 points. This may be said "very poor" (g).

— If x_i is a node at level i, all the children (s, r_{i+1}, N_{i+1}) of x_i have the same value of N_{i+1}, that is thus function of x_i only, $N_{i+1} = N_{i+1}(x_i)$.

— We shall have $r_{i+1} = r_i/3$ for $i \geqslant 0$. In order to avoid repeating: "when $i > 0$, we have that $r_{i-1} = 3r_i$", we shall decide that $r_{-1} = 3r_0$ and have $r_{i-1} = 3r_i$ for all $i \geqslant 0$.

— Finally, observing that $N_1 \geqslant 2$ and $N_{i+2} \geqslant N_{i+1}^3 > 2N_{i+1}$ we get

$$N_i \geqslant 2^i, \quad \sum_{j=1}^{\infty} N_j^{-1} < 1, \quad \sum_{j=0}^{\infty} N_j^{-1} < 2 \tag{16}$$

Let us fix a node x_i in the ith generation. Consider passing from x_i to its children $x_{i+1} = (t_{i+1}, r_{i+1}, N_{i+1})$, where $t_{i+1} \in S_{i+1}$; we have $r_{i+1} = r_i/3 = 3^{-i-1} r_0$ and N_{i+1} depends only upon x_i that is fixed:

— suppose first that we are in the "rich" case: then $N_{i+1} = |S_{i+1}|$. Let $2\delta = r_{i+1}$, so that S_{i+1} is 2δ-separated, and $\delta/2 = r_{i+1}/4 = r_i/12$. Applying Corollary 1 and (15) we get

$$\varphi(r_i + r_i/12) \geqslant (1/2)(r_i/12)\sqrt{\ln N_{i+1}} + \varphi(r_i/12),$$

and carelessly writing $r_i + r_i/12 < 2r_i$ we arrive at

$$\varphi(2r_i) \geqslant (r_i/24)\sqrt{\ln N_{i+1}} + \varphi(r_i/12); \tag{17}$$

— otherwise, in the "poor case" for x_i, we know that $N_{i+1} = N_i^3$. We essentially do nothing in this case, it is not the right time, we just observe that

$$r_i\sqrt{\ln N_{i+1}} = \sqrt{3}\, r_i\sqrt{\ln N_i} = \frac{\sqrt{3}}{3} r_{i-1}\sqrt{\ln N_i} \leqslant \frac{2}{3} r_{i-1}\sqrt{\ln N_i}, \tag{18}$$

because $r_i = r_{i-1}/3$ and $\sqrt{3} < 2$, hence

$$(r_i/24)\sqrt{\ln N_{i+1}} \leqslant \frac{2}{3}(r_{i-1}/24)\sqrt{\ln N_i},$$

and obviously we have that

$$\varphi(r_i/12) \leqslant \varphi(2r_i).$$

Adding these two informations and letting $\kappa = 1/24$ we obtain

$$\varphi(2r_i) + \frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i} \geqslant \kappa r_i\sqrt{\ln N_{i+1}} + \varphi(r_i/12), \tag{19}$$

and we observe that this inequality is clearly valid also in the "rich" case, simply because we have (17) and $\kappa r_{i-1}\sqrt{\ln N_i} \geqslant 0$ (if $i = 0$, then $\kappa r_{-1}\sqrt{\ln N_0} = 0$ as $N_0 = 1$).

Let $\mathbf{x} = (x_j)_{j\geqslant 0}$ denote a maximal branch in the tree $\mathcal{X}$: the node x_0 in the branch $\mathbf{x}$ is the root of $\mathcal{X}$, and x_{i+1} is a child of x_i for every $i \geqslant 0$. We may consider the elements in the branch as functions of $\mathbf{x}$ and write $x_i = x_i(\mathbf{x}) = (t_i(\mathbf{x}), r_i, N_i(\mathbf{x}))$ for the node x_i at level i in the branch $\mathbf{x}$. Now, for any branch $\mathbf{x}$ and every $i \geqslant 0$, Equation (19) gives

$$\varphi(2r_i) + \frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i(\mathbf{x})} \geqslant \kappa r_i\sqrt{\ln N_{i+1}(\mathbf{x})} + \varphi(r_i/12). \tag{20}$$

We will remember that N_i is a *function* of branches but we shall omit the $\mathbf{x}$ variable. Summing (20) from $i = 0$ to k, using $r_{-1}\sqrt{\ln N_0} = 0$ because $N_0 = 1$ and reorganizing the log terms we get

$$\sum_{i=0}^{k} \varphi(2r_i) \geqslant \frac{\kappa}{3}\sum_{i=0}^{k} r_i\sqrt{\ln N_{i+1}} + \sum_{i=0}^{k}\varphi(r_i/12),$$

for any infinite branch. Observe that

$$\varphi(r_i/12) \geqslant \varphi(2r_{i+3}) \tag{21}$$

because $2r_{i+3} = 2r_i/27 < r_i/12$. Hence

$$\sum_{i=0}^{k}\varphi(r_i/12) \geqslant \sum_{j=3}^{k}\varphi(2r_j) \quad \text{thus} \quad \sum_{i=0}^{2}\varphi(2r_i) \geqslant \frac{\kappa}{3}\sum_{i=0}^{k} r_i\sqrt{\ln N_{i+1}}.$$

Finally, for every infinite branch $\mathbf{x}$ we have that

$$\sum_{i=0}^{\infty} r_i\sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{9}{\kappa}\,\varphi(2r_0) = 216\ \mathrm{E}\Big(\sup_{t\in T} X_t\Big).$$

The series above is very similar to the series $\Sigma_1(T)$ in (1). We let

$$\sigma_1(\mathbf{x}) = \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})},$$

a function of the branches $\mathbf{x}$.

Let us review for future use the properties of our tree, some being expressed in terms of properties of functions of branches, or functions of nodes. For every node $x = (t, r, N)$ in $\mathcal{X}$ let us say that $B(t, r)$ is the *ball associated to* x, and denote it by $\beta(x) = B(t, r)$.

$\mathbf{a_0}$ — The root $x_0 = (t_0, r_0, N_0)$ of the tree is such that $t_0 \in T$, $\Delta < r_0 < 4\Delta/3$, hence we have $\beta(x_0) = B(t_0, r_0) = T$. Furthermore, $N_0 = 1$.

$\mathbf{a_1}$ — For every node $x = (t, r, N) \in \mathcal{X}$, the ball $\beta(x) = B(t, r)$ associated to x is covered by the balls associated to the children of x,

$$\beta(x) \subset \bigcup_{y \in C(x)} \beta(y).$$

$\mathbf{a_2}$ — We have $r_i = r_{i-1}/3$ for $i \geqslant 0$ (h).

$\mathbf{a_3}$ — For every $i \geqslant 0$ and every branch $\mathbf{x}$, $N_{i+1}(\mathbf{x})$ depends only upon $x_i(\mathbf{x})$ and we have $N_{i+1}(\mathbf{x}) \geqslant N_i(\mathbf{x})^3$, in other words, for every node x we have

$$y, y' \in C(x) \;\Rightarrow\; N(y) = N(y'); \qquad y \in C(x) \Rightarrow N(y) \geqslant N(x)^3.$$

$\mathbf{a_4}$ — For every $i > 0$, $N_i(\mathbf{x})$ is an upper bound for the number of siblings of $x_i(\mathbf{x})$,

$$|C(x_{i-1}(\mathbf{x}))| \leqslant N_i(\mathbf{x}), \qquad \text{or equivalently:} \quad y \in C(x) \Rightarrow |C(x)| \leqslant N(y).$$

We did not yet make a full use of the group invariance condition (14). When introducing the children of $(t, r, N) \in \mathcal{X}$, we covered $B(t, r)$ by balls $B(s, r/3)$, without observing that we may start by covering $B(t_0, r)$ with N balls $B(s, r/3)$, for t_0 fixed in T, and then use the transitive group action in order to cover each ball of radius r in T with the same number N of balls of that radius $r/3$. Doing this from the root x_0 on, we may replace the functions $N_i(\mathbf{x})$ by *constant* values, and write the conclusion of the tree construction as

$$\sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} < +\infty. \tag{22}$$

2.2. *The entropy-like condition is sufficient*

We assume that the tree $\mathcal{X}$ satisfies the five properties $\mathbf{a_0}$ to $\mathbf{a_4}$, where r_i, N_i do not depend on the branches, and we want to explain that conversely, the condition (22) allows one to bound the expectation of the supremum of the process $(X_t)_{t \in T}$.

We start from the root $(t_0, r_0, 1)$, that was chosen so that $B(t_0, r_0) = T$ (see the condition $\mathbf{a_0}$), and we move toward an arbitrary point $t \in T$ in successive steps in the tree, that will form an infinite branch $\mathbf{x}(t)$. First, the point t is contained in one of the balls $B(s, r_1)$ corresponding to the children of x_0, that cover $B(t_0, r_0)$ (condition $\mathbf{a_1}$); we let $x_1(t) = (t_1(t), r_1, N_1)$ be a child of $x_0(t) = x_0$ such that $t \in B(t_1(t), r_1)$. We go on choosing successive nodes $x_{i+1}(t) = (t_{i+1}(t), r_{i+1}, N_{i+1})$ such that $x_{i+1}(t)$ is a child of $x_i(t)$ and such that $t \in B(t_{i+1}(t), r_{i+1})$, using $\mathbf{a_1}$ again. So the step from $t_i(t)$ to $t_{i+1}(t)$ has size $\leqslant r_i + r_{i+1} < 2r_i$. The sequence $(t_i(t))$ tends to t in T because r_i tends to 0 (i).

There are N_1 possible paths from the root $(t_0, r_0, 1)$ to a node in the first generation, and in general, there are at most N_i possibilities for moving from a given node x_{i-1} in the $(i-1)$th generation to all the children of that node in the ith generation. By condition $\mathbf{a_4}$ applied in the constant case, we have when $i > 0$ that

$$N_{i-1} \leqslant N_i^{1/3}.$$

When t varies in T and $i > 0$, the global number M_i of paths from x_0 to $x_i(t)$ is(j) bounded by

$$M_i \leqslant N_i N_{i-1} \dots N_1 \leqslant N_i N_i^{1/3} N_i^{1/9} \dots N_i^{3^{-i+1}} < N_i^{3/2} < N_i^2. \tag{23}$$

By (9) and because $\mathrm{Var}(X_{t_{i+1}(t)} - X_{t_i(t)}) = d(t_{i+1}(t), t_i(t))^2 \leqslant 4r_i^2$, we have for each integer $i \geqslant 0$ that

$$\mathrm{E} \sup_{t \in T} |X_{t_{i+1}(t)} - X_{t_i(t)}| \leqslant 2r_i.2\sqrt{\ln(M_{i+1}+1)} \leqslant 4r_i\sqrt{\ln N_{i+1}^2} < 6r_i\sqrt{\ln N_{i+1}}.$$

It follows that

$$\mathrm{E} \sup_{t \in T} |X_t - X_{t_0}| \leqslant \sum_{i=0}^{\infty} \mathrm{E} \sup_{t \in T} |X_{t_{i+1}(t)} - X_{t_i(t)}| \leqslant 6 \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}}$$

thus, knowing that $\mathrm{E}\, X_{t_0} = 0$, we conclude that

$$\mathrm{E} \sup_{t \in T} X_t \leqslant 6 \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}}, \qquad \mathrm{E} \sup_{t \in T} |X_t| \leqslant \mathrm{E}\,|X_{t_0}| + 6 \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}}.$$

Of course what we just did is nothing but a variant of proving the entropy integral criterion that involves Dudley's integral. We could actually have easily related directly the condition

$$\sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} < +\infty$$

to Dudley's integral. Indeed, we have seen at (23) that there are $M_i \leqslant N_i^2$ paths from the root x_0 to the nodes of the ith generation in the tree; in other words, the set T_i of points $s \in T$ appearing in the nodes (s, r_i, N_i) of that ith generation contains at most N_i^2 points, and we know by iterating $\mathbf{a_1}$ that the balls $B(s, r_i)$ centered at those points $s \in T_i$ cover T. This means that

$$\mathcal{N}(r_i) \leqslant M_i \leqslant N_i^2,$$

where $\mathcal{N}(\varepsilon) = \mathcal{N}(T, \varepsilon)$ denotes the minimal number of open balls of radius ε needed to cover T. When $r_{i+1} \leqslant \varepsilon \leqslant r_i$ we thus have

$$\mathcal{N}(\varepsilon) \leqslant \mathcal{N}(r_{i+1}) \leqslant N_{i+1}^2,$$

hence

$$\int_{r_{i+1}}^{r_i} \sqrt{\ln \mathcal{N}(\varepsilon)}\, \mathrm{d}\varepsilon \leqslant r_i \sqrt{\ln(N_{i+1}^2)} \leqslant 2r_i \sqrt{\ln N_{i+1}}$$

and because $r_0 > \Delta$ and $\lim_i r_i = 0$, we conclude that

$$I_D(T) = \int_0^{\Delta} \sqrt{\ln \mathcal{N}(\varepsilon)}\, \mathrm{d}\varepsilon \leqslant 2 \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}} \leqslant 432\ \mathrm{E}\Big(\sup_{t \in T} X_t\Big).$$

This transition series–integral is pretty much the same as what we have seen before, when we discretized the Dudley integral.

2.3. *Beyond invariance*

Consider again a centered Gaussian process $(X_t)_{t\in T}$ such that

$$E^* := \mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr) < \infty.$$

If we give up invariance(k) for the process $(X_t)_{t\in T}$, we lose Equality (15), and now, for each point $s \in T$ and $r > 0$, we can do nothing but consider

$$\varphi(s,r) = \mathrm{E}\Bigl(\sup_{t\in B(s,r)} X_t\Bigr) \leqslant E^*.$$

Applying Corollary 1 without invariance, it only remains from (17), when $i = 0$ for example, that

$$\begin{aligned}\varphi(t_0, 2r_0) = \mathrm{E}\Bigl(\sup_{t\in B(t_0,2r_0)} X_t\Bigr) &\geqslant (r_0/24)\sqrt{\ln N_1} + \min_{s\in S_1} \mathrm{E}\Bigl(\sup_{t\in B(s,r_0/12)} X_t\Bigr)\\ &= (r_0/24)\sqrt{\ln N_1} + \min_{s\in S_1} \varphi(s, r_0/12).\end{aligned}$$

This cannot be used, unless we can make sure that the different values of $\varphi(s, r_0/12)$, for $s \in S_1$, are essentially the same. For this we need a quite natural extra step in the construction of the tree. After finding generation i and before defining the $(i+1)$th, we have to make a division into "zones" where the function φ will be controlled in a suitable way. The collection of children of a node x_i in the ith generation will possibly be split into different sub-collections corresponding to these "zones". Also, we shall need an improving control over φ as i increases, by locating values of φ in smaller and smaller intervals; in order to avoid multiplying unnecessarily the number of parameters of the construction, we shall use the rapidly increasing numbers N_i to this end, locating values of φ at level i in small intervals of size N_i^{-1}.

We shall thus divide the successive sets obtained in the invariant case —there, they were the balls $B(t_i, r_i)$— into "homogeneous zones". We shall put one more information in the nodes of the tree: the tree $\mathcal{X}$ will now consist of quadruples $x = (t, r, N, V)$ where we have $t \in T$, $r > 0$, and N an integer as before; additionally, we introduce a non-empty subset V of T, a "homogeneous zone" where a certain variation of φ will be controlled; the integer N will no longer be a bound for the number of siblings of x, but rather a bound on the entropy of V. We shall note $t(x)$, $r(x)$, $N(x)$ and $V(x)$, so that we could (uselessly) write any node x of the tree as $x = (t(x), r(x), N(x), V(x))$.

The construction of the tree goes as follows: we suppose of course that T has at least two points, so that its diameter Δ is > 0. The root of our tree is $x_0 = (t_0, r_0, 1, T)$, with r_0 such that $\Delta < r_0 < 4\Delta/3$ and t_0 given in T, hence $T = B(t_0, r_0)$; we let $N_0 = 1$. So far this is exactly as before, but for the addition of $V_0 = T$. We also let $r_{-1} = 3r_0$ as before. Let us describe the modified construction step.

Suppose that a node $x = (t, r, N, V)$ has already been introduced in the tree $\mathcal{X}$. The first part of the construction step is essentially identical to what was done in the invariant case, except that the set V replaces now what was the ball $B(t, r)$ before: let

$$\widetilde{r} = r/3 \text{ and let } \widetilde{S} \text{ be a } \widetilde{r}\text{-net for } V,$$

that is to say, the set $\widetilde{S} \subset V$ is a $\widetilde{r}$-separated set such that

$$V \subset \bigcup_{s\in\widetilde{S}} B(s, \widetilde{r}). \tag{24}$$

We let $\widetilde{N} = \max(N^3, |\widetilde{S}|)$. Again, the *rich case* is when $\widetilde{N} = |\widetilde{S}| > N^3$.

Here comes the difference. For every integer α such that $0 \leqslant \alpha < \widetilde{N}$ and for every point $s \in \widetilde{S}$, we consider the (possibly empty) subset of $B(s, \widetilde{r})$ defined by

$$\widetilde{U}_\alpha(s, \widetilde{r}, \widetilde{N}) = \{t \in B(s, \widetilde{r}) : \varphi(t, \widetilde{r}/12) \in Z(\alpha, \widetilde{N})\} \subset B(s, \widetilde{r})$$

where

$$Z(\alpha, \widetilde{N}) = [\alpha E^*/\widetilde{N}, (\alpha+1)E^*/\widetilde{N}].$$

The set $\widetilde{U}_\alpha(s, \widetilde{r}, \widetilde{N})$ is one of the "homogeneous zones" that were mentioned earlier. Notice that

$$B(s, \widetilde{r}) = \bigcup_{0 \leqslant \alpha < \widetilde{N}} \widetilde{U}_\alpha(s, \widetilde{r}, \widetilde{N})$$

since the segments $Z(\alpha, \widetilde{N})$ cover $[0, E^*]$, and as the balls $B(s, \widetilde{r})$ cover V by (24), we have

$$V \subset \bigcup \{\widetilde{U}_\alpha(s, \widetilde{r}, \widetilde{N}) : s \in \widetilde{S}, \, 0 \leqslant \alpha < \widetilde{N}\}. \tag{25}$$

Each non-empty subset $V \cap \widetilde{U}_\alpha(s, \widetilde{r}, \widetilde{N})$ of V produces a child $\widetilde{x}$ of x. The number of the children of x is less than or equal to $|\widetilde{S}| \, \widetilde{N} \leqslant \widetilde{N}^2$. We give those children of x the form $\widetilde{x} = (t, \widetilde{r}, \widetilde{N}, \widetilde{V})$, where t is a point chosen in the non-empty set $\widetilde{V}$ and

$$\widetilde{r} = r/3, \quad \widetilde{V} = V \cap \widetilde{U}_\alpha(s, \widetilde{r}, \widetilde{N}) \neq \emptyset, \quad s \in \widetilde{S}, \quad 0 \leqslant \alpha < \widetilde{N}.$$

By (25), the sets $\widetilde{V}$ corresponding to all children of x cover V, and we have $\widetilde{V} \subset V$ by construction.

Let us repeat things with a mention to the level $i \geqslant 0$ in the new tree of the parent node $x = x_i = (t_i, r_i, N_i, V_i)$. We let $r_{i+1} = \widetilde{r} = r_i/3$, we have a set $S_{i+1} = \widetilde{S}$ that is a r_{i+1}-net for V_i, and $N_{i+1} = \widetilde{N} = \max(|S_{i+1}|, N_i^3)$. Then V_i is divided into homogeneous zones V_{i+1}, a point t_{i+1} is chosen in the non-empty set V_{i+1}. The children of x_i have the form $x_{i+1} = (t_{i+1}, r_{i+1}, N_{i+1}, V_{i+1})$, with

$$V_{i+1} = V_i \cap U_\alpha(s, r_{i+1}, N_{i+1}) \neq \emptyset, \quad s \in S_{i+1}, \quad 0 \leqslant \alpha < N_{i+1}.$$

By (25), the sets V_{i+1} corresponding to all the children of x_i cover V_i. We have

$$V_{i+1} \subset B(t_{i+1}, 2r_{i+1})$$

because $V_{i+1} \subset B(s, r_{i+1})$ for some $s \in S_{i+1}$ and $t_{i+1} \in V_{i+1}$. By construction, when v runs in the set V_{i+1}, the values of $\varphi(v, r_{i+1}/12)$ stay in a segment of length E^*/N_{i+1}. Let

$$z(x_i) = \inf \{\varphi(v, r_i/12) : v \in V_i\}. \tag{26}$$

We have just said that when $i > 0$, we have

$$v \in V_i \;\Rightarrow\; z(x_i) \leqslant \varphi(v, r_i/12) \leqslant z(x_i) + E^*/N_i,$$

and it is also correct when $i = 0$ because $N_0 = 1$ and $\varphi \leqslant E^*$. When $i = 0$, we have that $V_0 = T$ and $\mathcal{N}(V_0, r_0/3) > 1$, thus $N_1 \geqslant 2 > N_0^3 = 1$.

In the invariant case, we know that the function $t \mapsto \varphi(t, r_{i+1}/12)$ is constant and the above procedure in that situation produces exactly *one* homogeneous zone inside the ball $B(s, r_{i+1})$, for each $s \in S_{i+1}$, namely, the ball $B(s, r_{i+1})$ itself: there is only one set $V_{i+1} = V_i \cap B(s, r_{i+1})$ for each $s \in S_{i+1}$.

2.3.1. *Estimates*

Suppose that $i \geqslant 0$ and that $x_i = (t_i, r_i, N_i, V_i)$ is a node in the ith generation of the new tree $\mathcal{X}$. There are as before two possibilities:

— in the "rich case", we have $N_{i+1} = |S_{i+1}| > N_i^3$ and we know that the set S_{i+1} is r_{i+1}-separated, contained in $B(t_i, 2r_i)$ because $S_{i+1} \subset V_i \subset B(t_i, 2r_i)$; hence, by Corollary 1 applied with $2\delta = r_{i+1}$ (thus $\delta/2 = r_{i+1}/4 = r_i/12$) we have

$$\begin{aligned}\varphi(t_i, 2r_i{+}r_i/12) &= \mathrm{E}\Bigl(\sup_{t\in B(t_i,2r_i+r_i/12)} X_t\Bigr)\\ &\geqslant (r_i/24)\sqrt{\ln N_{i+1}} + \min_{s\in S_{i+1}} \mathrm{E}\Bigl(\sup_{t\in B(s,r_i/12)} X_t\Bigr)\\ &\geqslant \kappa r_i\sqrt{\ln N_{i+1}} + \min_{v\in V_i}\varphi(v, r_i/12) \geqslant \kappa r_i\sqrt{\ln N_{i+1}} + z(x_i),\end{aligned}$$

by the definition (26) of $z(x_i)$;

— otherwise, we are in the poor case, thus $N_{i+1} = N_i^3$. On one hand, we use as before $\frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i} \geqslant \kappa r_i\sqrt{\ln N_{i+1}}$ —see Equation (18)—; on the other hand, let $v \in V_i$; we have

$$z(x_i) \leqslant \varphi(v, r_i/12) \leqslant \varphi(t_i, 2r_i{+}r_i/12)$$

because $B(v, r_i/12) \subset B(t_i, 2r_i{+}r_i/12)$ since $v \in V_i \subset B(t_i, 2r_i)$.

In both cases, realizing that $2r_i{+}r_i/12 = 25r_i/12$, we conclude that

$$\varphi(t_i, 25r_i/12) + \frac{2}{3}\kappa r_{i-1}\sqrt{\ln N_i} \geqslant \kappa r_i\sqrt{\ln N_{i+1}} + z(x_i). \tag{27}$$

This replaces Inequality (19) from the invariant case.

2.3.2. *Summing up*

Let us consider a branch $\mathbf{x} = (x_i)_{i\geqslant 0}$ of the tree, with $x_i = (t_i, r_i, N_i, V_i)$ for each $i \geqslant 0$, and write the functions $t_i(\mathbf{x})$, $N_i(\mathbf{x})$ without their $\mathbf{x}$ variable. Adding (27) from 0 to k and reorganizing the log terms we get

$$\sum_{i=0}^{k}\varphi(t_i, 25r_i/12) \geqslant \frac{\kappa}{3}\sum_{i=0}^{k} r_i\sqrt{\ln N_{i+1}} + \sum_{i=0}^{k} z(x_i). \tag{28}$$

Inequality (21) has to be revised. It reads now as follows: let $x_i = (t_i, r_i, N_i, V_i)$ be a node of $\mathbf{x}$ at level $i \in \mathbb{N}$; the set $V_i \subset B(t_i, 2r_i)$ is an "homogeneous" subset, meaning precisely that $z(x_i) \leqslant \varphi(v, r_i/12) \leqslant z(x_i) + E^*/N_i$ for all $v \in V_i$. One has

$$\varphi(t_{i+3}, 25r_{i+3}/12) \leqslant z(x_i) + E^*/N_i\,; \tag{29}$$

indeed, we know that $t_{i+3} \in V_{i+3} \subset V_i$ ([1]), we see that $25r_{i+3}/12 < 27r_{i+3}/12 = r_i/12$, therefore

$$\varphi(t_{i+3}, 25r_{i+3}/12) \leqslant \varphi(t_{i+3}, r_i/12) \leqslant z(x_i) + E^*/N_i.$$

It follows from (29) that

$$\sum_{i=0}^{k} z(x_i) + \sum_{i=0}^{k} E^*/N_i \geqslant \sum_{j=3}^{k}\varphi(t_j, 25r_j/12).$$

From (16) and from (28) we see that

$$3\,\varphi(t_0, 25r_0/12) + 2E^* \geqslant \sum_{i=0}^{2}\varphi(t_i, 25r_i/12) + \sum_{i=0}^{k} E^*/N_i \geqslant \frac{\kappa}{3}\sum_{i=0}^{k} r_i\sqrt{\ln N_{i+1}}$$

and because $\varphi \leqslant E^*$ we conclude that

$$\sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant \frac{15}{\kappa}\, E^* = 360\, \mathrm{E}\Bigl(\sup_{t\in T} X_t\Bigr)$$

for every branch $\mathbf{x}$ of the tree.

Let $\mathbf{x} = (x_j)_{j\in\mathbb{N}}$ be an infinite branch in the new tree. There are several functions of the variable $\mathbf{x}$: for every j, $x_j(\mathbf{x})$ is the node of $\mathbf{x}$ at level $j \in \mathbb{N}$; the function values $t_j(\mathbf{x})$, $N_j(\mathbf{x})$ and $V_j(\mathbf{x})$ are defined by the equality $x_j(\mathbf{x}) = (t_j(\mathbf{x}), r_j, N_j(\mathbf{x}), V_j(\mathbf{x}))$. Let us review the properties of the tree. For every node $x = (t, r, N, V)$ in $\mathcal{X}$ we shall say that the set V is the *region associated to the node x*.

$\mathbf{a_0^*}$ — (compare to $\mathbf{a_0}$) The root $x_0 = (t_0, r_0, N_0, V_0)$ of the tree is such that $t_0 \in T$, we have $\Delta < r_0 < 4\Delta/3$ hence $B(t_0, r_0) = T$. Furthermore, $N_0 = 1$ and $V_0 = T$.

$\mathbf{a_1^*}$ — (compare to $\mathbf{a_1}$) For every node $x = (t, r, N, V) \in \mathcal{X}$, the region V associated to x is the union of the regions associated to the children of x: we have

$$V(x) = \bigcup_{y\in C(x)} V(y),$$

that we may write more explicitely as follows: for every $j \in \mathbb{N}$ and every node x_j in the jth generation, we have

$$V_j(x_j) = \bigcup_{y\in C(x_j)} V_{j+1}(y).$$

In particular, we have $V_{j+1}(\mathbf{x}) \subset V_j(\mathbf{x})$ for every branch $\mathbf{x}$.

$\mathbf{a_2^*}$ — (see $\mathbf{a_2}$) For every integer $i \geqslant 0$ and $x_i = (t_i, r_i, N_i, V_i) \in \mathcal{X}$, we have

$$r_i = r_{i-1}/3 \quad \text{and} \quad t_i \in V_i \subset B(t_i, 2r_i),$$

or in other words, for every node $x \in \mathcal{X}$,

$$t(x) \in V(x) \subset B(t(x), 2r(x)).$$

$\mathbf{a_3^*}$ — (see $\mathbf{a_3}$) For every branch $\mathbf{x}$ and every $i \geqslant 0$, the value $N_{i+1}(\mathbf{x})$ depends only upon $x_i(\mathbf{x})$ and we have $N_{i+1}(\mathbf{x}) \geqslant N_i(\mathbf{x})^3$, in other words,

$$y, y' \in C(x) \Rightarrow N(y) = N(y'), \quad y \in C(x) \Rightarrow N(y) \geqslant N(x)^3.$$

Furthermore, we have that $N_1 \geqslant 2$. It follows that $N_i \geqslant 2^i$ (see (16)).

$\mathbf{a_4^*}$ — (see $\mathbf{a_4}$) For every integer $i \geqslant 0$, the square $N_{i+1}(\mathbf{x})^2$ is an upper bound for the number of children of $x_i(\mathbf{x})$,

$$|C(x_i(\mathbf{x}))| \leqslant N_{i+1}(\mathbf{x})^2, \quad \text{or equivalently:} \;\; y \in C(x) \Rightarrow |C(x)| \leqslant N(y)^2.$$

2.4. *Suppose we have that nice tree*

Assume that for some constant K and for every branch $\mathbf{x}$ in a tree $\mathcal{X}$, we have

$$\sigma_1(\mathbf{x}) = \sum_{i=0}^{\infty} r_i \sqrt{\ln N_{i+1}(\mathbf{x})} \leqslant K,$$

for a tree satisfying the properties $\mathbf{a_0^*}$ to $\mathbf{a_4^*}$. We want to use this information in order to bound the expectation of the supremum of the process. For every node $x \in \mathcal{X}$, we write $x = (t(x), r(x), N(x), V(x))$, and also $x_j = (t_j, r_j, N_j, V_j)$ for nodes x_j at level j, for every $j \in \mathbb{N}$.

When $\tau \in T$, we can find a branch $\mathbf{x}$ such that $t_i(\mathbf{x})$ tends to τ, and more precisely, such that $\tau \in V_i \subset B(t_i, 2r_i)$ for every integer $i \geqslant 0$: we let $x_0(\mathbf{x}) = x_0$ be the root of the tree; then $\tau \in V_0 = T$. Assuming that $x_i = x_i(\mathbf{x})$ has been found such that $\tau \in V_i(x_i)$, we can find a child x_{i+1} of x_i such that $\tau \in V_{i+1}(x_{i+1})$, because these sets V_{i+1} cover $V_i(x_i)$, by condition $\mathbf{a_1^*}$.

Let us fix a level $i \in \mathbb{N}$ in the tree. Consider a node $x_i = (t_i, r_i, N_i, V_i)$ from the ith generation. Let us first estimate the probability of a (relatively) large jump when passing from x_i to a specific fixed child $x_{i+1} = (t_{i+1}, r_{i+1}, N_{i+1}, V_{i+1})$ of x_i, that is to say, the probability of having a big absolute value for the difference $X_{t_{i+1}} - X_{t_i}$. Consider $u > 0$ and let the size of the jump be

$$s_i(u, x_i) = \bigl(u + \sqrt{8 \ln N_{i+1}}\,\bigr) 2r_i .$$

We know that $t_{i+1} \in V_{i+1} \subset V_i \subset B(t_i, 2r_i)$ by $\mathbf{a_2^*}$. We have therefore $d(t_{i+1}, t_i) \leqslant 2r_i$, so the variance of $(X_{t_{i+1}} - X_{t_i})/(2r_i)$ is $\leqslant 1$, yielding

$$\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > \bigl(u + \sqrt{8 \ln N_{i+1}}\,\bigr) 2r_i\bigr) \leqslant \mathrm{P}\bigl(|g| > u + \sqrt{8 \ln N_{i+1}}\,\bigr)$$

where g is a $\mathrm{N}(0,1)$ Gaussian random variable, and by (7) we have

$$\begin{aligned} \mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > s_i(u, x_i)\bigr) &\leqslant \mathrm{P}\bigl(|g| > u + \sqrt{8 \ln N_{i+1}}\,\bigr) \leqslant \exp\bigl(-\bigl(u + \sqrt{8 \ln N_{i+1}}\,\bigr)^2/2\bigr) \\ &\leqslant \exp\bigl(-u^2/2 - 4 \ln N_{i+1}\bigr) = \mathrm{e}^{-u^2/2}\, N_{i+1}^{-4}. \end{aligned}$$

We know that $N_k \geqslant 2^k$ by condition $\mathbf{a_3^*}$, we rewrite the last line as

$$\mathrm{P}\bigl(|X_{t_{i+1}} - X_{t_i}| > s_i(u, x_i)\bigr) \leqslant \mathrm{e}^{-u^2/2}\, N_{i+1}^{-1}\, N_{i+1}^{-3} \leqslant \mathrm{e}^{-u^2/2}\, 2^{-i-1}\, N_{i+1}^{-3} = c_i(u) N_{i+1}^{-3}$$

where we have set $c_i(u) = 2^{-i-1}\, \mathrm{e}^{-u^2/2}$. Note that N_{i+1} is actually a function of x_i only (see $\mathbf{a_3^*}$). The total probability $p_i^{(i+1)}(x_i)$ of having a jump of that size $s_i(u, x_i)$ when passing from x_i to *some* of its children in the $(i{+}1)$th generation (there are at most N_{i+1}^2 of them by $\mathbf{a_4^*}$) is thus bounded by

$$p_i^{(i+1)}(x_i) \leqslant N_{i+1}^2 . c_i(u) N_{i+1}^{-3} \leqslant c_i(u) N_{i+1}^{-1}(x_i).$$

Using $N_{k+1} \geqslant N_k^3$ for every integer $k \geqslant 0$ (condition $\mathbf{a}_3^*$) we conclude that

$$p_i^{(i+1)}(x_i) \leqslant c_i(u)\, N_i^{-3}(x_{i-1}).$$

Next, let $p_j^{(i+1)}(x_j)$ denote the probability that starting from $x_j \in \mathcal{X}$, with $0 \leqslant j \leqslant i$, along any possible path in the tree ending at a node in the $(i{+}1)$th generation, we have a jump of size $\geqslant s_i(u, x_i)$ between the levels i and $(i{+}1)$. In the same way, we see that

$$p_{i-1}^{(i+1)}(x_{i-1}) \leqslant N_i^2 . c_i(u)\, N_i^{-3} \leqslant c_i(u)\, N_i^{-1} \leqslant c_i(u)\, N_{i-1}^{-3}(x_{i-2}).$$

We can write $N_{i-1} = N_{i-1}(x_{i-2})$ as long as $i - 2 \geqslant 0$. If $i = 1$, then $N_0 = 1$ is a constant and we need not mention a variable. By induction, we obtain that

$$p_0^{(i+1)}(x_0) \leqslant c_i(u) N_0^{-3} = c_i(u) = 2^{-i-1}\, \mathrm{e}^{-u^2/2} .$$

This is a bound for the probability of a jump of size $s_i(u, x_i)$ between the ith generation and the next, for any path from the root to a node in the $(i{+}1)$th generation. Starting from the root, the first jump to consider is $s_0(u, x_0)$ between x_0 and its children x_1, so the probability $p(u)$ of finding for some $i \geqslant 0$ a jump of size $s_i(u, x_i)$ between the levels i and $(i{+}1)$ in the tree is bounded above by

$$p(u) \leqslant \sum_{i=0}^{\infty} p_0^{(i+1)}(x_0) \leqslant \mathrm{e}^{-u^2/2} \sum_{i=0}^{\infty} 2^{-i-1} < \mathrm{e}^{-u^2/2} .$$

Except for a set $S(u) \subset \Omega$ having probability $\leqslant p(u)$, we obtain that *all* steps along *any* branch $\mathbf{x}$ are less than $s_i(u, \mathbf{x}) = s_i(u, x_i(\mathbf{x}))$, so that outside $S(u)$, we have

$$\begin{aligned} \sum_{i=0}^{\infty} |X_{t_{i+1}(\mathbf{x})} - X_{t_i(\mathbf{x})}| &\leqslant \sum_{i=0}^{\infty} s_i(u, \mathbf{x}) \\ &\leqslant 2 \Bigl(\sum_{i=0}^{\infty} r_i \Bigr) u + 2 \sum_{i=0}^{\infty} r_i \sqrt{8 \ln N_{i+1}(\mathbf{x})} \\ &< 3 r_0 \, u + 6K \leqslant 4 \Delta u + 6K. \end{aligned}$$

It follows that for every $u > 0$, we have

$$\mathrm{P}\bigl(\sup_{t \in T} |X_t - X_{t_0}| > 4\Delta u + 6K \bigr) \leqslant \mathrm{P}(S(u)) \leqslant \mathrm{e}^{-u^2/2} .$$

This certainly implies that

$$\mathrm{E} \Bigl(\sup_{t \in T} |X_t - X_{t_0}| \Bigr) < \infty,$$

and more precisely, letting $X^* = \sup_{t \in T} |X_t - X_{t_0}|$, we obtain that

$$\begin{aligned} \mathrm{E}\, X^* &= \int_0^{\infty} \mathrm{P}(X^* > v) \, \mathrm{d}v \leqslant 6K + \int_{6K}^{\infty} \mathrm{P}(X^* > v) \, \mathrm{d}v \\ &= 6K + \int_0^{\infty} \mathrm{P}(X^* > 6K + v) \, \mathrm{d}v = 6K + 4\Delta \int_0^{\infty} \mathrm{P}(X^* > 6K + 4\Delta u) \, \mathrm{d}u \\ &\leqslant 6K + 4\Delta \int_0^{\infty} \mathrm{e}^{-u^2/2} \, \mathrm{d}u = 6K + 4 \sqrt{\frac{\pi}{2}} \, \Delta. \end{aligned}$$

Also,

$$\Delta \sqrt{\ln 2} < r_0 \sqrt{\ln N_1} < K$$

so that

$$\mathrm{E}\, X^* \leqslant 6K + 4 \sqrt{\frac{\pi}{2}} \, \Delta \leqslant \Bigl(6 + 4 \sqrt{\frac{\pi}{2 \ln 2}} \Bigr) K < 13\, K.$$

Notes

(a) There are obvious and well known difficulties when one is trying to consider

$$\sup_{t \in T} X_t(\omega),$$

where T is *uncountable* but each X_t is merely a *class* of random variables (an uncountable union of negligible sets is no longer negligible...). This is not our main concern here; we shall be happy enough to be able to deal with countable index sets T.

(b) If *Mathematica* or another software of the kind does not give you immediately the results claimed here, you can try checking my ugly calculations at

https://webusers.imj-prg.fr/~bernard.maurey/articles/MaxiGauss.pdf

(c) Simone Chevet in [Chev] does not study the *expectation* of the supremum, but its *distribution*, and she writes a proof for the Slepian lemma; however, Fernique in [Fer$_1$] claims to see between the lines of Chevet's article a proof for Proposition 1; he himself writes on page 63 a one line justification for that Proposition, namely:

$$4 \frac{d}{d\alpha} \bigl[\{ \sup_{t \in T} Z_\alpha(t) \} \bigr] = \sum_{\substack{s,t \in T \times T \\ s \neq t}} \frac{d}{d\alpha} [\Delta_{Z_\alpha}(s,t)] \int \frac{dx}{dx_s \, dx_t} \int g_\alpha(x) \, du,$$

only precising that the last integral is done on the domain $x_s = x_t = \sup x_i = u$. If I try be a little understandable, I have to add that

$$Z_\alpha = \sqrt{\alpha}\, X + \sqrt{1-\alpha}\, Y, \quad 0 \leqslant \alpha \leqslant 1,$$

where X and Y are independent copies of the two processes from Proposition 1, and say that Δ_Z denotes the L^2-metric associated to a process Z. Also, T is supposed to be finite here, and g_α is the function on $\mathbb{R}^T$ equal to the density (supposed to exist) of the distribution of Z_α.

It seems that Fernique meant to have an expectation in the left-hand side of the main equality above.

(d) This rather simple lemma does not appear in the original proofs by Fernique. I learned about it by attending some lectures about the so-called generic chaining method of Talagrand, given at Marne-la-Vallée around 2002 by (by then) young researchers there. This method leads to the results of section 2.3 for the non invariant case.

(e) We do not actually need an explicit group action; all we need is that for any two balls $B(t_1, r)$ and $B(t_2, r)$ in T, with T considered as a subset of L^2, there is an onto affine isometry of L^2 between them, sending t_1 to t_2. From this, the properties of the Gaussian vectors will imply that the distributions of $(X_t - X_{t_1})_{t \in B(t_1, r)}$ and $(X_t - X_{t_2})_{t \in B(t_2, r)}$ are the same.

(f) If T has no isolated point, all branches produced by our process will be "naturally" infinite. We could arrange to work with a set T with no isolated point, for example by replacing T by its convex hull in $L^2(\Omega, \mathrm{P})$, but the added complexity would ruin the small simplification of not having to deal with isolated points. What we do is allowing for an infinite branch to have the form $\mathbf{x} = ((t_i, r_i, N_i))_{i \geqslant 0}$ with r_i decreasing to 0 but possibly with $t_i = t$ for $i \geqslant i_0$, where t is isolated in T.

(g) Note that in this situation, each $s \in B(t_1, r_1)$ is isolated in T. However, this fact does not play a rôle in the further discussion.

(h) Remember that we decide to set $r_{-1} = 3$. In the construction of the tree we insisted that $r_{i+1} = r_i/3$, but this will not be used in the reverse direction that comes next: only $\lim_i r_i = 0$ will be used.

(i) When T is finite, we have of course $t_i(t) = t$ when $i \geqslant i_0$. Our "unnatural" treatment was meant to avoid considering the finite case separately.

(j) It is a remarkable and very important fact that, due to the absolutely huge growth of the constants N_i, the logarithm of the number M_i of points in the ith generation is comparable to the logarithm of the number N_i of children of a *single node* in the $(i-1)$th generation.

(k) The results for the non-invariant case have been obtained by Talagrand.

(l) In the invariant case, we did not insist that the ball $B(t_{i+1}, r_{i+1})$ associated to a child x_{i+1} of $x_i = (t_i, r_i, N_i)$ be *contained* in the ball $B(t_i, r_i)$ associated to its parent x_i. But here, we need that for $k > 1$, the points in the regions V_{i+k} of the next generations will still satisfy the homogeneity condition that was set before for the points of V_{i+1}. We ensure it by imposing that $V_{i+1} \subset V_i$ for every $i \geqslant 0$.

References

[Chev] S. Chevet, *Mesures de Radon sur* $\mathbb{R}^n$ *et mesures cylindriques*, Ann. Fac. Sci. Univ. Clermont **43** (6) (1970), 91–158.

[Dud_1] R. M. Dudley, *Sample functions of the Gaussian process*, Ann. Probab. **1** (1973), 66–103.

[Dud_2] R. M. Dudley, *V. N. Sudakov's Work on Expected Suprema of Gaussian Processes*, High Dimensional Probability VII, Progress in Probability 71 (2016), 37–43.

[Fer_1] X. Fernique, *Minorations des fonctions aléatoires gaussiennes*, Ann. Inst. Fourier, Grenoble **24**, 2 (1974), 61–66.

[Fer_2] X. Fernique, *Régularité des trajectoires des fonctions aléatoires gaussiennes*, École d'Été de Probabilités de Saint-Flour IV – 1974. Lecture Notes in Math. 480, Springer, Berlin, 1975, p. 1–96.

[Gor] Y. Gordon, *Some inequalities for Gaussian processes and applications*, Israel J. of Math. **50** (1985), 265–289.

[LiQ] D. Li & H. Queffélec, *Introduction à l'étude des espaces de Banach*, Société Mathématique de France, Paris 2004. Augmented English translation: *Introduction to Banach Spaces: Analysis and Probability*, 2 vol., Cambridge Univ. Press, Cambridge 2018.

[Pisi] G. Pisier, *Ensembles de Sidon et processus gaussiens*, C. R. Acad. Sci. Paris **286** (1978), 671–674.

[Rud] W. Rudin, *Trigonometric series with gaps*, J. Math. Mech. **9** (1960), 203–227.

[Slep] D. Slepian, *The one-sided barrier problem for Gaussian noise*, Bell. Syst. Techn. J. **41** (1962), 463–501.

[Sud] V. N. Sudakov, *Geometric problems in the Theory of Infinite-Dimensional Probability Distributions*, (in Russian). Trudy Mat. Inst. Steklov 141, 191 pp. (1976); English cover to cover translation, Proc. Steklov Inst. Math. 1979 (2), American Mathematical Society.