Small Deviations of Smooth Stationary Gaussian Processes

F. Aurzada, I.A. Ibragimov, M.A. Lifshits, and J.H. van Zanten

March 29, 2008

Abstract

We investigate the small deviation probabilities of a class of very smooth stationary Gaussian processes playing an important role in Bayesian statistical inference. Our calculations are based on the appropriate modification of the entropy method due to Kuelbs, Li, and Linde as well as on classical results about the entropy of classes of analytic functions. They also involve Tsirelson's upper bound for small deviations and shed some light on the limits of sharpness for that estimate.

1 Introduction

Let $X(t)$ be a centered stationary Gaussian process identified by its spectral measure $F(du)$. We restrict X on the interval $[0,1]$ and evaluate its small deviations with respect to the uniform norm $||\cdot||_\infty$ in terms of the small deviation function

$$\varphi(X,r) = -\log \mathbb{P}(||X||_\infty \le r), \qquad r \to 0.$$

See [8], [9] for many motivations for the study of small deviations and [10] for a complete bibliography on this subject.

In this note, we will be interested in the case of rather smooth processes. Namely, consider the family of processes X_ν corresponding to absolutely continuous spectral measures

$$F_\nu(du) = \exp\{-|u|^\nu\}du, \qquad 0 < \nu < \infty,$$

and a parallel family of periodic processes $\tilde{X}_\nu$ corresponding to discrete spectral measures

$$\tilde{F}_\nu(du) = \sum_{k=-\infty}^{\infty} \exp\{-|k|^\nu\}\delta_{2\pi k}, \qquad 0 < \nu < \infty.$$

The most interesting cases are $\nu = 1$ (exponential spectrum) and $\nu = 2$ (normal spectrum).

For exposition completeness, let us close the first family with

$$F_\infty(du) = \mathbf{1}_{[-1,1]}du.$$

Although the smoothness properties of X_ν and $\tilde{X}_\nu$ are the same, it turns out, quite surprisingly, that their small deviations behave differently.

An important motivation for this research comes from the recent work of A.W. van der Vaart and J.H. van Zanten [14], where such small deviations were considered in the context of Bayesian statistics. It was shown that they actually determine posterior convergence rates in nonparametric estimation problems. In particular the process X_2, which is known in the Bayesian and machine learning literature as the "squared exponential process", is a popular building block in the construction of prior distributions on functional parameters, cf. e.g. [13].

Before we state the results, let us fix some notation. We write $f(\cdot) \preceq g(\cdot)$ or $g(\cdot) \succeq f(\cdot)$ if $\limsup \frac{f}{g} < \infty$, while the equivalence $f \approx g$ means that we have both $f \preceq g$ and $g \preceq f$. Moreover, $f(\cdot) \lesssim g(\cdot)$ or $g(\cdot) \gtrsim f(\cdot)$ mean that $\limsup \frac{f}{g} \leq 1$. Finally, the strong equivalence $f \sim g$ means that $\lim \frac{f}{g} = 1$.

It was shown in [14], by using the RKHS-entropy method, that

$$\varphi(X_\nu, r) \preceq |\log r|^2, \qquad \nu \geq 1. \tag{1.1}$$

We will slightly improve this and obtain sharp bounds. Our main results are as follows.

Theorem 1.1 *We have*

$$\varphi(X_\nu, r) \approx \frac{|\log r|^2}{\log|\log r|}, \qquad 1 < \nu \leq \infty, \tag{1.2}$$

and

$$\varphi(X_\nu, r) \approx |\log r|^{1+\frac{1}{\nu}}, \qquad 0 < \nu \leq 1, \tag{1.3}$$

as $r \to 0$.

For the periodic processes the asymptotics is somewhat different.

Theorem 1.2 *We have*

$$\varphi(\tilde{X}_\nu, r) \approx |\log r|^{1+\frac{1}{\nu}}, \qquad \nu > 0, \tag{1.4}$$

as $r \to 0$.

Remark 1.3 The exponential discrete spectrum ($\nu = 1$) is well understood for L_2-norms where the estimate

$$\varphi(\tilde{X}_\nu, r) \sim C|\log r|^2, \qquad \nu = 1,$$

(and even more precise behavior) is obtained in the context of small deviations of the series (with exponentially decreasing coefficients), see [4], [3], or [2]. As usual (but not always), the small deviation rate is the same for the uniform and for the L_2-norm.

Remark 1.4 The radical difference of the two bounds (1.2) and (1.4) is that the first one does not depend on ν while the second one does. From this point of view, Theorem 1.1 provides a more surprising result than Theorem 1.2.

The authors were informed by A.I. Nazarov that the same phenomenon is well known for many years in the theory of integral operators. Generally speaking, smoother the kernel of a symmetric integral operator is, faster the eigenvalues decrease. However, there is a kind of barrier: the eigenvalues λ_k can not decrease faster than $\log \lambda_k \approx -n \log n$. Since behavior of the eigenvalues is tightly related to small deviations (once we consider the covariance operator of a Gaussian process) in L_2-norm, the bound for eigenvalues transforms in a bound for small deviations.

Remark 1.5 Notice that we do not have any general tools for tracing connections between small deviations for discrete and continuous spectra. The general feeling is that discrete spectrum provides larger small deviation probabilities.

In view of the applications in Bayesian nonparametrics we also provide upper bounds for the small deviations of rescaled versions of the processes X_ν. For a constant $c \leq 1$, define the rescaled process X_ν^c by setting $X_\nu^c(t) = X_\nu(t/c)$.

Theorem 1.6 *For all $c \leq 1$ we have*

$$\varphi(X_\nu^c, r) \preceq \frac{1}{c} \frac{|\log r|^2}{\log|\log r|}, \qquad \nu > 1, \tag{1.5}$$

$$\varphi(X_\nu^c, r) \preceq \frac{1}{c} |\log r|^{1+\frac{1}{\nu}}, \qquad 0 < \nu \leq 1, \tag{1.6}$$

as $r \to 0$.

2 RKHS tools

In this section, we recall a powerful approach to the study of Gaussian small deviations based on the entropy of the corresponding kernel (RKHS), suggested by J. Kuelbs and W. Li in [6]. In the literature, this approach is mainly applied to polynomial entropy, resp. small deviation function, while the results we need should handle slowly varying functions. Therefore, for the reader's convenience, we give here the complete proofs.

We work in a fairly general setting. Let X be a centered Gaussian vector in a separable Banach space $(E, \|\cdot\|)$. Then X generates a *kernel*, or RKHS, $\mathcal{H}$ which is a linear subspace of E equipped with the structure of a Hilbert space. For a detailed description of the RKHS we refer to [9]. We denote by $\mathcal{H}_1$ the unit ball of $\mathcal{H}$. Let the covering number $N(r)$ be defined as the minimal number of balls in the norm $\|\cdot\|$ of radius r that is needed to cover $\mathcal{H}_1$. Furthermore, let $H(r) := \log N(r)$ be the corresponding metric entropy of $\mathcal{H}_1$.

We still study the the behavior of small deviation function

$$\varphi(r) := \varphi(X, r) := -\log \mathbb{P}(||X|| \leq r), \qquad r \to 0.$$

Let us recall the central inequalities proved in [6].

Lemma 2.1 *Let $r > 0$ and $\lambda > 0$. Then*

$$H\left(\frac{2r}{\lambda}\right) \leq \varphi(r) + \lambda^2/2,$$

$$H\left(\frac{r}{\lambda}\right) \geq \varphi(2r) + \log \Phi(\lambda + \alpha_r),$$

where Φ is the distribution function of the standard normal law, and α_r is defined by $-\log \Phi(\alpha_r) = \varphi(r)$.

We obtain the following corollary from the first inequality in the case of a slowly varying entropy or small deviation function.

Corollary 2.2 *Let β be any real number and $\gamma, C > 0$. Then*

- $\varphi(r) \lesssim C|\log r|^\gamma(\log|\log r|)^\beta$ *implies* $H(r) \lesssim C|\log r|^\gamma(\log|\log r|)^\beta$.
- $H(r) \gtrsim C|\log r|^\gamma(\log|\log r|)^\beta$ *implies* $\varphi(r) \gtrsim C|\log r|^\gamma(\log|\log r|)^\beta$.

The relations also hold if $\lesssim$ and $\gtrsim$ are replaced by $\preceq$ and $\succeq$, respectively.

Proof: Simply set $\lambda = 2$ in the first inequality in Lemma 2.1. □

The arguments are slightly more involved when using the second inequality because of its implicit nature. First recall that

$$\log \Phi(x) \sim -x^2/2,$$

as $x \to -\infty$. This helps to simplify the second inequality in Lemma 2.1.

Lemma 2.3 *Let $\lambda = \lambda(r) > 0$ be a function such that $\lambda(r) \leq \sqrt{2\varphi(r)}$. Then, as $r \to 0$,*

$$H\left(\frac{r}{\lambda}\right) \gtrsim \varphi(2r) - \frac{1}{2}\left(\lambda - \sqrt{2\varphi(r)}\right)^2. \tag{2.1}$$

The usual choice in the regularly varying case is $\lambda = -\alpha_r \sim \sqrt{2\varphi(r)}$, which also works in the case of slow variation. The result reads as follows.

Corollary 2.4 *Let β be any real and $\gamma, C > 0$. Then*

- $H(r) \lesssim C|\log r|^\gamma(\log|\log r|)^\beta$ *implies* $\varphi(r) \lesssim C|\log r|^\gamma(\log|\log r|)^\beta$.
- *Assume that there is a constant $K > 0$ such that $\varphi(r/2) \leq K\varphi(r)$ for all $r \in (0,1)$. Then*

  $\varphi(r) \gtrsim C|\log r|^\gamma(\log|\log r|)^\beta$ *implies*

  $H(r) \gtrsim C\,(1 + \log K/(2\log 2))^{-\gamma}\,|\log r|^\gamma(\log|\log r|)^\beta$.

The relations also hold if $\lesssim$ and $\gtrsim$ are replaced by $\preceq$ and $\succeq$, respectively.

Proof: Let $\lambda := \sqrt{2\varphi(r)}$. For the first implication note that the assumption for H, relation (2.1), and the fact that $r/\lambda \to 0$ imply that

$$C|\log r - \log\sqrt{\varphi(r)}|^\gamma(\log|\log r/\sqrt{\varphi(r)}|)^\beta \gtrsim \varphi(2r). \tag{2.2}$$

The assumption for H furthermore implies that $H(r) \preceq r^{-\tau'}$ for any $\tau' > 0$. By Proposition 2.4 in [7], this yields

$$\varphi(r) \preceq r^{-\tau}, \qquad \text{for any } \tau > 0. \tag{2.3}$$

Thus

$$\limsup_{r\to 0} \frac{\log\sqrt{\varphi(r)}}{\varphi(2r)^{1/\gamma}} \leq \frac{\tau}{2} \limsup_{r\to 0} \frac{|\log r|}{\varphi(2r)^{1/\gamma}}.$$

Also (2.3) implies that $(\log|\log r/\sqrt{\varphi(r)}|)^\beta$ can be replaced by $(\log|\log r|)^\beta$ in (2.2). Therefore

$$\frac{1}{C} \leq \liminf_{r\to 0} \frac{|\log r - \log\sqrt{\varphi(r)}|^\gamma}{\varphi(2r)}(\log|\log r|)^\beta$$
$$= \liminf_{r\to 0} \left| \frac{|\log r|}{\varphi(2r)^{1/\gamma}} + \frac{\log\sqrt{\varphi(r)}}{\varphi(2r)^{1/\gamma}} \right|^\gamma (\log|\log r|)^\beta \leq \left(1+\frac{\tau}{2}\right)^\gamma \liminf_{r\to 0} \frac{|\log r|^\gamma}{\varphi(2r)}(\log|\log r|)^\beta.$$

Letting $\tau \to 0$ yields the assertion.

Let us come to the second implication. We may assume that $K \geq 1$. First note that the regularity assumption on φ implies that $\varphi(r) \leq K' r^{-h}$ with $h = \log K/\log 2$, $K' := \varphi(1)K$ and all $0 < r < 1$. Now if $\varphi(r) \gtrsim C|\log r|^\gamma(\log|\log r|)^\beta$, we obtain by (2.1) that

$$H\left(\frac{r}{\lambda}\right) \gtrsim C|\log r|^\gamma(\log|\log r|)^\beta.$$

We set $r' := r/\lambda$. We obtain, by the assumption on φ that $r' \geq r^{1+h/2}/\sqrt{2K'}$. Therefore,

$$H\left(r^{1+h/2}/\sqrt{2K'}\right) \geq H(r') = H\left(\frac{r}{\lambda}\right) \gtrsim C|\log r|^\gamma(\log|\log r|)^\beta.$$

In other words,

$$H(r) \gtrsim C|\log r^{1/(1+h/2)}|^\gamma(\log|\log r|)^\beta = \frac{C}{(1+\log K/(2\log 2))^\gamma}\,|\log r|^\gamma(\log|\log r|)^\beta.$$

□

Remark 2.5 Note that, as in the regularly varying case, one needs to know something about the maximal increase of φ in order to translate a lower bound for φ into a lower bound for H. If it is already known that φ behaves logarithmically, then the assumption holds for any $K > 1$ and one also obtains strong asymptotic equivalence.

As a particular case of Corollaries 2.2 and 2.4 we obtain the following.

Corollary 2.6 *Let β be any real and $\gamma > 0$. Then*

$$\varphi(r) \approx |\log r|^\gamma(\log|\log r|)^\beta \qquad \Leftrightarrow \qquad H(r) \approx |\log r|^\gamma(\log|\log r|)^\beta.$$

3 Entropy of stationary RKHS

Let now X be a complex valued centered stationary Gaussian process with spectral measure F and continuous sample paths. We consider X as a random element of $E = C[0,1]$. It is well known (see e.g. [9]) that the RKHS $\mathcal{H}$ admits the following representation: $h \in \mathcal{H}$ iff

$$h(t) = \int_{-\infty}^{\infty} \ell(u) e^{-itu} F(du), \qquad \ell \in L_2(\mathbb{R}, F), \tag{3.1}$$

for $0 \leq t \leq 1$ and

$$||h||_{\mathcal{H}} = \inf ||\ell||_{2,F}$$

where infimum is taken over all ℓ satisfying (3.1). In particular, $h \in \mathcal{H}_1$ (here, as above, $\mathcal{H}_1$ is the unit ball in $\mathcal{H}$) iff (3.1) holds with ℓ such that $||\ell||_{2,F} \leq 1$.

Now we specify this general scheme to the processes we are interested in and evaluate the entropy.

3.1 Continuous spectrum

Let now $F(du) = f_\nu(u)du$, where $f_\nu(u) = e^{-|u|^\nu}$, $\nu > 0$. We prove the following.

Proposition 3.1 *For any $\nu > 1$ it is true that*

$$H(\mathcal{H}_1, \varepsilon) \approx \frac{|\log \varepsilon|^2}{\log |\log \varepsilon|} \ .$$

Proof:

Upper bound. Let $h \in \mathcal{H}_1$. Then the representation (3.1) holds with some ℓ such that

$$||\ell||^2_{L_2(\mathbb{R},F)} = \int_{-\infty}^{\infty} |\ell(u)|^2 f_\nu(u) du \leq 1.$$

Clearly, h turns out to be an entire analytic function well defined on $\mathbb{C}$ by the same expression (3.1) and by Hölder's inequality

$$|h(z)| \leq \int_{-\infty}^{\infty} e^{|\operatorname{Im}(z)|\,|u|} |\ell(u)| f_\nu(u) du \leq \left(\int_{-\infty}^{\infty} e^{2|\operatorname{Im}(z)|\,|u|} f_\nu(u) du \right)^{1/2} := M_\nu(2|\operatorname{Im}(z)|).$$

Since

$$\log M_\nu(r) = \frac{1}{2} \ \log \int_{-\infty}^{\infty} e^{r\,|u| - |u|^\nu} du \sim \frac{(\nu-1)\, r^{\nu/(\nu-1)}}{2\, \nu^{\nu/(\nu-1)}}, \qquad \text{as } r \to \infty,$$

it follows that

$$|h(z)| \leq M_\nu(2|\operatorname{Im}(z)|) \leq C_1 \ \exp\{C_2 |\operatorname{Im}(z)|^{\nu/(\nu-1)}\}, \qquad \forall h \in H_1, z \in \mathbb{C}, \tag{3.2}$$

with appropriate constants $C_1 = C_1(\nu)$, $C_2 = C_2(\nu)$. It is known from Theorem XX of [5] that the entropy of the class of all entire analytic functions $\mathcal{A}(C_1, C_2, \nu)$ satisfying the even weaker condition

$$|h(z)| \leq C_1 \ \exp\{C_2 |z|^{\nu/(\nu-1)}\}, \qquad \forall z \in \mathbb{C}, \tag{3.3}$$

verifies

$$H(\mathcal{A}(C_1,C_2,\nu),\varepsilon) \approx \frac{|\log\varepsilon|^2}{\log|\log\varepsilon|}\ .$$

Since $\mathcal{H}_1 \subset \mathcal{A}(C_1,C_2,\nu)$, we obtain

$$H(\mathcal{H}_1,\varepsilon) \preceq \frac{|\log\varepsilon|^2}{\log|\log\varepsilon|}\ .$$

Lower bound. Here we will only need an inequality

$$f(u) \geq c_f, \qquad |u| \leq 1, \tag{3.4}$$

for a constant $c_f > 0$, which is fulfilled for all densities f_ν, $\nu > 0$.

We start with a construction of an auxiliary function and study its properties. Take any $\gamma \in (0,1)$ and let a sequence $(a_k)_{k\geq 1}$ be defined by $a_k = ck^{-1-\gamma}$ and normalized so that $\sum_{k=1}^{\infty} a_k = 1$. Let

$$G(z) = \prod_{k=1}^{\infty} \frac{\sin(a_k z)}{a_k z}, \qquad z \in \mathbb{C}.$$

Since

$$\frac{|\sin(z)|}{|z|} \leq \sum_{j=1}^{\infty} \frac{|z|^{j-1}}{j!} \leq e^{|z|},$$

we have

$$|G(z)| \leq \exp\left(\sum_{k=1}^{\infty} a_k |z|\right) = e^{|z|}. \tag{3.5}$$

The function G is rapidly decreasing on the real line. Namely, for any large $t \in \mathbb{R}$ choose a positive integer $\kappa = \kappa(t)$ such that $a_\kappa |t| \sim 2$, i.e. $\kappa \sim (c|t|/2)^{\frac{1}{1+\gamma}}$. Then

$$|G(t)| \leq \prod_{k=1}^{\kappa} |a_k t|^{-1} \leq 2^{-\kappa} \leq \exp\left(-C_G |t|^{\frac{1}{1+\gamma}}\right) \tag{3.6}$$

with appropriate $C_G > 0$. Finally, notice that

$$\theta_G := \inf_{0\leq t\leq 1} |G(t)| > 0, \tag{3.7}$$

since the smallest zero of G is attained at $\frac{\pi}{c} > \pi > 1$.

Now we start the entropy estimation. Consider a class Ψ_K of analytic functions ψ on complex plane satisfying

$$|\psi(z)| \leq K \exp\{|z|^{1/2}\}, \qquad z \in \mathbb{C}. \tag{3.8}$$

Again by Theorem XX in [5] it is true that

$$H(\Psi_K,\varepsilon) \approx \frac{|\log\varepsilon|^2}{\log|\log\varepsilon|}\ . \tag{3.9}$$

Next, consider a class of functions

$$B_K = \{b : b(z) = \psi(z)G(z),\ \psi \in \Psi_K, z \in \mathbb{C}\}.$$

With a minor abuse of notation, we do not distinguish the functions from B_K and their restrictions on $[0,1]$. Clearly,

$$H(B_K, \varepsilon) \geq H(\Psi_K, \theta_G^{-1}\varepsilon) \succeq \frac{|\log\varepsilon|^2}{\log|\log\varepsilon|} \ . \tag{3.10}$$

We will show now that for an appropriate choice of the parameter K it is true that $B_K \subset \mathcal{H}_1$. Let $b \in B_K$. Then by (3.8) and (3.5) we have

$$|b(z)| \leq K \exp\left\{|z| + |z|^{1/2}\right\}.$$

Moreover, by (3.8) and (3.6)

$$||b||^2_{L_2(\mathbb{R})} \leq K^2 \int_{-\infty}^{\infty} \exp\left\{2|t|^{1/2} - 2C_G|t|^{\frac{1}{1+\gamma}}\right\} dt := K^2 C^2_{G,2} < \infty.$$

By using these two properties, it follows from the classical Paley–Wiener theorem ([12] or [1], Chap. IV) that the Fourier transform

$$\hat{b}(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iut} b(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iut} \psi(t) G(t) dt$$

vanishes outside the interval $[-1,1]$.

On the other hand, we can write

$$\begin{aligned} b(t) &= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-iut} \hat{b}(u) du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-iut}\ \frac{\hat{b}(u)}{f(u)}\ f(u) du \\ &=: \int_{-\infty}^{\infty} e^{-iut} \ell(u)\ f(u) du. \end{aligned}$$

It remains to show that $||\ell||_{2,F} \leq 1$. By using (3.4) we have, indeed,

$$\begin{aligned} ||\ell||^2_{2,F} &= \frac{1}{2\pi} \int_{-1}^{1} \frac{|\hat{b}(u)|^2}{f(u)}\ du \\ &\leq \frac{1}{2\pi c_f} ||\hat{b}||^2_{L_2(\mathbb{R})} = \frac{1}{2\pi c_f} ||b||^2_{L_2(\mathbb{R})} \leq \frac{K^2 C^2_{G,2}}{2\pi c_f} \leq 1, \end{aligned}$$

whenever K is chosen sufficiently small (depending on c_f). Thus $B_K \subset \mathcal{H}_1$ and we obtain from (3.10)

$$H(\mathcal{H}_1, \varepsilon) \geq H(B_K, \varepsilon) \succeq \frac{|\log\varepsilon|^2}{\log|\log\varepsilon|}\ ,$$

as required. □

For small values of ν we only need the following upper bound.

Proposition 3.2 *For any $\nu \le 1$ it is true that*

$$H(\mathcal{H}_1, \varepsilon) \preceq |\log \varepsilon|^{1+1/\nu}.$$

Proof: The idea is simple: for $\nu = 1$ the result is already known from (1.1) and we reduce the general case to that one by truncation of the spectral measure. Namely, for any $\varepsilon > 0$ let $v = (3|\log \varepsilon|)^{1/\nu}$. Then by (3.1) the elements of $\mathcal{H}_1$ have the form

$$h(t) = \left(\int_{|u|\le v} + \int_{|u|>v} \right) \ell(u) e^{-itu} F(du) := h_v(t) + h^v(t), \qquad ||\ell||_{2,F} \le 1.$$

By the choice of v we have

$$|h^v(t)| \le ||\ell||_{2,F} \left(\int_{|u|>v} \exp(-|u|^\nu) du \right)^{1/2} \le C \exp(-|v|^\nu/2) v^{(1-\nu)/2} \le \varepsilon$$

for small ε. Therefore, we only need to study the entropy of the set $\mathcal{H}_1^v := \{h_v, h \in \mathcal{H}_1\}$. This will be done by means of the following result from [5] in the quantitative version of [14], Lemma 2.3.

Lemma 3.3 *Let F be a spectral measure and let a positive $\delta < 1$ be such that*

$$I := \int e^{\delta |u|} F(du) \le 1.$$

Then

$$H(\mathcal{H}_1, \varepsilon) \le C \ \frac{|\log \varepsilon|^2}{\delta} \ ,$$

where C is a numeric constant.

First, notice that if we drop the assumption $I \le 1$, then by scaling reasons we still have

$$H(\mathcal{H}_1, \varepsilon) \le C \ \frac{|\log(\varepsilon/\sqrt{I})|^2}{\delta} \ , \tag{3.11}$$

Apply this bound to our truncated measure $e^{-|u|^\nu} \mathbf{1}_{|u|\le v} du$ and $\delta = \theta |\log \varepsilon|^{1-1/\nu}$ with appropriately small parameter $\theta \le 3^{-1/\nu}$. Notice that $\delta |u| \le |u|^\nu$ whenever $|u| \le v$. Hence

$$I = \int_{|u|\le v} e^{\delta |u| - |u|^\nu} du \le 2v \approx |\log \varepsilon|^{1/\nu}.$$

We obtain from (3.11)

$$H(\mathcal{H}_1^v, \varepsilon) \preceq \frac{|\log \varepsilon|^2}{|\log \varepsilon|^{1-1/\nu}} = |\log \varepsilon|^{1+1/\nu},$$

as required. $\square$

3.2 Discrete spectrum

Let now $F(du) = \sum_{k=-\infty}^{\infty} \exp\{-|k|^\nu\}\delta_{2\pi k}$, $\nu > 0$. We prove the following.

Proposition 3.4 *For any $\nu > 0$ it is true that*

$$H(\mathcal{H}_1, \varepsilon) \preceq |\log \varepsilon|^{1+1/\nu}.$$

Proof: The reasoning goes along the same lines as that of the upper bound in the previous proposition. Let $h \in \mathcal{H}_1$. Then the representation (3.1) means that

$$h(t) = \sum_{k=-\infty}^{\infty} \ell_k e^{-ikt-|k|^\nu} \tag{3.12}$$

with some $\ell = (\ell_k)$ such that

$$||\ell||^2_{L_2(\mathbb{R},F)} = \sum_k |\ell_k|^2 \exp\{-k^\nu\} \leq 1.$$

Clearly, h turns out to be a periodic entire analytic function well defined on $\mathbb{C}$ by the same expression (3.12) and by the Hölder's inequality

$$|h(z)| \leq \sum_{k=-\infty}^{\infty} e^{|\operatorname{Im}(z)|\,|k|-|k|^\nu} |\ell_k| \leq \left(\sum_{k=-\infty}^{\infty} e^{2|\operatorname{Im}(z)|\,|k|-|k|^\nu} \right)^{1/2} := \tilde{M}_\nu(2|\operatorname{Im}(z)|).$$

It follows again that

$$|h(z)| \leq C_1 \ \exp\{C_2 |\operatorname{Im}(z)|^{\nu/(\nu-1)}\}, \qquad \forall h \in H_1, z \in \mathbb{C},$$

with appropriate constants $C_1 = C_1(\nu)$, $C_2 = C_2(\nu)$. It is known by Theorem XXI in [5] that the entropy of the class of all periodic entire analytic functions $\tilde{\mathcal{A}}(C_1, C_2, \nu)$ satisfying this condition verifies

$$H(\tilde{\mathcal{A}}(C_1, C_2, \nu), \varepsilon) \approx |\log \varepsilon|^{1+1/\nu}.$$

Hence

$$H(\mathcal{H}_1, \varepsilon) \preceq |\log \varepsilon|^{1+1/\nu}.$$

□

4 Proofs of main results

Proof of Theorem 1.2: The lower bound for small deviations follows immediately from Proposition 3.4 and Corollary 2.4.

For getting the upper bound we implement a simple but ingenious idea of B.S. Tsirelson initially designed for continuous spectra in [11]. Let l be an integer. Let us consider an auxiliary centered stationary Gaussian process $Y = Y_l(t)$ with the spectral measure

$$F_Y(du) = \exp\{-l^\nu\} \sum_{|k| \leq l} \delta_{2\pi k},$$

which minorates $\tilde{F}_\nu$. By the standard Anderson argument

$$\mathbb{P}(||\tilde{X}_\nu||_\infty \le r) \le \mathbb{P}(||Y||_\infty \le r) \qquad \forall r > 0.$$

The covariance of Y is

$$\mathbb{E}Y(t)Y(0) = \exp\{-l^\nu\} \sum_{|k|\le l} e^{itk} = \frac{4\exp\{-l^\nu\}}{|e^{it}-1|^2} \sin\left(\frac{(2l+1)t}{2}\right) \sin\left(\frac{t}{2}\right), \qquad t \ne 2\pi k, \tag{4.1}$$

while for the variance we have

$$\sigma^2 := \mathbb{E}|Y(t)|^2 = \exp\{-l^\nu\}\ (2l+1). \tag{4.2}$$

Define a grid step $\Delta = 2\pi/(2l+1)$. Observe from (4.1) that $(Y(k\Delta))_{k\in\mathbb{Z}}$ is a centered Gaussian non-correlated, hence independent, sequence with variance (4.2). For any $r > 0$ we get the bound

$$\begin{aligned}
\mathbb{P}(||Y||_\infty \le r) &\le \mathbb{P}(\sup_{0\le k\le 1/\Delta} |Y(k\Delta)| \le r) \\
&\le \mathbb{P}(\sigma|N| \le r)^{1/\Delta} \le \left(\sqrt{2/\pi}\ \frac{r}{\sigma}\right)^{1/\Delta} \le \left(\frac{r}{\sigma}\right)^{(2l+1)/2\pi} \\
&= \left(\frac{r}{(2l+1)\exp\{-l^\nu\}}\right)^{(2l+1)/2\pi} \le (r\exp\{l^\nu\})^{(2l+1)/2\pi}.
\end{aligned}$$

Next, an elementary optimization suggests to set

$$l \sim \left(\frac{|\log r|}{\nu+1}\right)^{1/\nu},$$

whereas

$$\begin{aligned}
\varphi(\tilde{X}_\nu, r) &\ge \varphi(Y,r) \ge -\frac{2l+1}{2\pi}\ \log(r\exp\{l^\nu\}) \sim \frac{l}{\pi}\ (|\log r| - l^\nu) \\
&\sim \frac{\nu l}{\pi(\nu+1)}\ |\log r| = \frac{\nu}{\pi(\nu+1)}\ \frac{|\log r|^{1+1/\nu}}{(\nu+1)^{1/\nu}} \\
&= \frac{\nu}{\pi(\nu+1)^{1+1/\nu}}\ |\log r|^{1+1/\nu},
\end{aligned}$$

and we arrive at the desired estimate. □

Proof of Theorem 1.1:

For $\nu > 1$ the result follows immediately from Proposition 3.1 and Corollary 2.6.

For $\nu \le 1$ the necessary upper bound follows immediately from Proposition 3.2 and Corollary 2.4.

The necessary lower bound

$$\varphi(X_\nu, r) \succeq |\log r|^{1+1/\nu} \tag{4.3}$$

holds for any $\nu > 0$ and can be obtained by Tsirelson's method, as described above. In this case, for any positive l we consider an auxiliary centered stationary Gaussian process $Y = Y_l(t)$ with the spectral density

$$f_Y(u) = \exp\{-l^\nu\}\ \mathbf{1}_{|u|\le l},$$

which minorates f_ν. Define a grid step $\Delta = \frac{2\pi}{l}$. It is easy to see again that $(Y(k\Delta))_{k\in\mathbb{Z}}$ is a centered Gaussian non-correlated, hence independent, sequence with variance

$$\sigma^2 := \mathbb{E}|Y(t)|^2 = 2l\exp\{-l^\nu\}.$$

and the final calculation leading to (4.3) goes through exactly as above. $\square$

Remark 4.1 We see from Theorem 1.1 that the estimate (4.3) is not sharp for $\nu > 1$. This is rather surprising since in the previously known examples (e.g. for polynomially decreasing spectral densities in [11]) Tsirelson's method always returned the right rates.

Proof of Theorem 1.6:
We prove (1.6), the proof of (1.5) is identical. Clearly,

$$\mathbb{P}(\|X_\nu^c\|_\infty \le r) = \mathbb{P}\Big(\sup_{t\in[0,1/c]} |X_\nu(t)| \le r\Big).$$

Let $\mathcal{H}_1^c$ be the unit ball of the RKHS of the process X_ν viewed as random element in $C[0,1/c]$, i.e. the class of functions on $[0,1/c]$ of the form

$$h(t) = \int_{-\infty}^{\infty} \ell(u)e^{-itu}\, dF_\nu(u), \qquad \|\ell\|_{2,F_\nu} \le 1.$$

Let n be the smallest integer larger or equal to $1/c$. Observe that if $h \in \mathcal{H}_1^c$, then for $k = 0,\dots,n-1$, the function $t \mapsto h(k+t)$ on $[0,1]$ belongs to the unit ball $\mathcal{H}_1$ of the RKHS of the process X_ν on $[0,1]$. Hence, if $h_1,\dots,h_N$ is an ε-net for $\mathcal{H}_1$, then the functions of the form

$$t \mapsto \sum_{k=0}^{n-1} h_{j_k}(t-k)\mathbf{1}_{[k,k+1)}(t)$$

form an ε-net for $\mathcal{H}_1^c$. There are at most N^n such functions. We keep only those for which there exists an element of $\mathcal{H}_1^c$ at uniform distance at most ε. The mentioned elements form a 2ε-net for $\mathcal{H}_1^c$. It follows that

$$H(\mathcal{H}_1^c, 2\varepsilon) \le nH(\mathcal{H}_1,\varepsilon) \le \frac{2}{c}\ H(\mathcal{H}_1,\varepsilon).$$

Now apply Proposition 3.4 and Corollary 2.4 to arrive at (1.6). $\square$

5 An open problem

It would be very interesting to extend our results to more general classes of smooth processes. Since Tsirelson's bound is sharp for spectral measures F_ν, $0 < \nu \le 1$, and in the case of polynomial spectral density $f(u) \approx |u|^{-1-\beta}$ it is also known to give a sharp bound $\varphi(X,r) \approx r^{-2/\beta}$, it is natural to conjecture that this bound is sharp in all intermediate cases, too. Our methods provide some reasonable bounds for general case but they should be at least enhanced in order to solve it properly. For example, on the test family of intermediate processes Y_α with spectral densities

$$f_\alpha(u) = \exp\{-(\log_+ |u|)^\alpha\}, \qquad \alpha > 1,$$

we get

$$|\log r|^{\frac{\alpha-1}{\alpha}} \ \exp\left\{(2|\log r|)^{1/\alpha}\right\} \preceq \varphi(Y_\alpha, r) \preceq |\log r| \ \exp\left\{(2|\log r|)^{1/\alpha} + \frac{5}{\alpha}|\log r|^{2/\alpha-1}\right\},$$

which is not as sharp as we would like.

References

[1] N.I. Achieser. Lectures on Approximation Theory. Nauka, Moscow, 1965 (Russian); Theory of approximation. Reprint of the 1956 English translation. Dover Publications, Inc., New York, 1992.

[2] F. Aurzada. A short note on small deviations of sequences of i.i.d. random variables with exponentially decreasing weights. To appear in: Statist. Probab. Lett. (2008),

[3] A.A. Borovkov and P.S. Ruzankin. On small deviations of independent positive random variables with weights close to exponential. To appear in: Siberian Adv. Math.

[4] T. Dunker, M.A. Lifshits, and W. Linde. Small deviations of sums of independent variables, In: High Dimensional Probability, Progress in Probability, vol. 43, Birkhäuser, Basel, 1998, 59–74.

[5] A.N. Kolmogorov and V.M. Tikhomirov. ε-entropy and ε-capasity of sets in functional space. *Amer. Math. Soc. Transl.* 2(17) (1961) 277–364. Translated from: *Uspehi Mat. Nauk* 14 (1959), No 2, 3–86 (Russian).

[6] J. Kuelbs and W. V. Li. Metric entropy and the small ball problem for Gaussian measures. *J. Funct. Anal.* 116 (1993), 133–157.

[7] W. V. Li and W. Linde. Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* 27 (1999), 1556–1578.

[8] W. V. Li and Q. Shao. Gaussian processes: inequalities, small ball probabilities and applications, In: Stochastic Processes: Theory and Methods, Handbook of Statistics (C.R. Rao and D. Shanbhag, eds.), vol. 19, North-Holland/Elsevier, Amsterdam, 2001, 533–597.

[9] M.A. Lifshits. Gaussian Random Functions, Kluwer, Dordrecht, 1995.

[10] M.A. Lifshits. Bibliography on small deviation probabilities. Compilation available from the site `www.proba.jussieu.fr/pageperso/smalldev/biblio.html`.

[11] M.A. Lifshits and B.S. Tsirelson. Small deviations of Gaussian fields. *Theor. Probab. Appl.*, 31 (1986), 557–558.

[12] R. Paley and N. Wiener. Fourier Transforms in the Complex Domain. Reprint of the 1934 original. AMS Colloquium Publications, 19. AMS, Providence, 1987.

[13] C.E. Rasmussen and C.K.I. Williams. Gaussian Processes for Machine Learning. MIT Press, 2006

[14] A.W van der Vaart and J.H. van Zanten. Bayesian inference with rescaled Gaussian process priors. *Electronic J. Statist.* 1 (2007), 433–448.