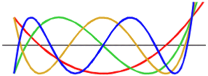
Small ball probabilities for Gaussian processes

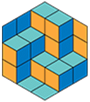
Yulia Petrova[1,2]

[1] St Petersburg State University,
Chebyshev Lab, St Petersburg, Russia

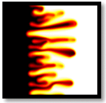
[2] IMPA, Instituto de Matematica
Pura e Aplicada, Rio de Janeiro, Brasil



https://yulia-petrova.github.io/



9 December 2021



This talk is a small overview, for more details see
M. Lifshits "Lectures on Gaussian processes", 2012, Springer

Small ball probabilities: definition

Let $(\mathcal{X}, \|\cdot\|)$ be a Banach space (f.e. $C[0,1]$ or $L^2[0,1]$).

Definition

An $\mathcal{X}$-valued random vector X is a measurable mapping

$$X : (\Omega, \mathbb{P}) \to \mathcal{X} \tag{1}$$

We will consider centered process, that is $\mathbb{E}X = 0$.

Definition

Small ball probability problem consists in finding the asymptotics

$$\mathbb{P}(\|X\| < \varepsilon) \qquad \textit{as} \quad \varepsilon \to 0 \tag{2}$$

Actually, it can be formulated as a problem in measure theory. Let P denote the distribution of X, that is a measure in $\mathcal{X}$, given by $P(A) = \mathbb{P}(X \in A)$, and let $U := \{x \in \mathcal{X} : \|x\| \leqslant 1\}$ be the unit ball in $\mathcal{X}$, then we want to study the measure of the small balls:

$$P(\varepsilon U), \quad \text{as} \quad \varepsilon \to 0.$$

Gaussian random vectors

Gaussian random vector extends the notion of a normally distributed random variable.

Definition

We call a random vector X, taking value in a linear topological space $\mathcal{X}$, Gaussian, if for every continuous linear functional $g \in \mathcal{X}^*$ the random variable $g(X)$ has a normal distribution.

The distribution of a Gaussian vector is uniquely determined by:

- means of $\{g(X) : g \in \mathcal{X}^*\}$;
- covariances of $\{g(X) : g \in \mathcal{X}^*\}$.

Main example

Wiener process $W(t)$ — a random element in $C[0,1]$ or in $L^2[0,1]$:

- $\mathbb{E}W(t) \equiv 0$;
- $\operatorname{cov}(W(s), W(t)) = \min(s,t)$.

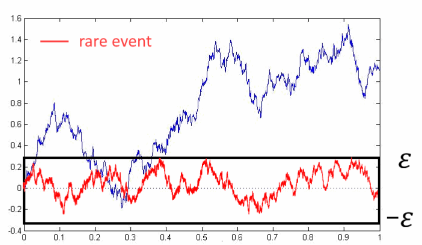
Example

Typical answer:

$$\mathbb{P}\left(\|X\| < \varepsilon\right) \sim D \cdot \varepsilon^{C} \cdot \exp\left(-B\varepsilon^{-A}\right), \qquad \varepsilon \to 0$$

A, B — *logarithmic* asymptotics; A, B, C, D — *exact* asymptotics

Example: $\mathcal{X} = C[0,1]$, $X = W(t)$ — Wiener process



$$\mathbb{P}\left(\sup_{0\leqslant t\leqslant 1} |W(t)| < \varepsilon\right) \sim \frac{4}{\pi}\, \exp\left(-\frac{\pi^2}{8}\,\varepsilon^{-2}\right)$$

Methods

"...there is no royal road to small ball probabilities..." M.A. Lifshits

Exist various methods, among others:

1. spectral method:
   - works for $\mathcal{X}$ being a Hilbert space
   - allows to get exact asymptotics
   - St Petersburg school:
     started by Ya. Nikitin, A. Nazarov, and followed by R. Pusev, A. Karol, N. Rastegaev, Yu. Petrova, etc
2. via metric entropy:
   - works for general classes of processes
   - allows to get only logarithmic asymptotics
   - M. Lifshits, F. Aurzada, I. Ibragimov, etc

Gaussian processes in Hilbert space

Karhunen-Loeve expansion (KL-expansion):

(K. Karhunen'1947, M. Loève'1948)
Let $\mathcal{X}$ be a separable Hilbert space with orthonormal basis (e_j). Then any Gaussian process X can be represented as

$$X(t) \stackrel{d}{=} \sum_{k=1}^{\infty} e_k \, \xi_k,$$

for ξ_k, $k \in \mathbb{N}$, independent and $\mathcal{N}(0, \sigma_k^2)$-distributed.

Main idea

All information about the process is in the variances σ_k^2

Hilbert structure $\Longrightarrow$ spectral problem

Karhunen-Loeve expansion (KL-expansion):

(K. Karhunen'1947, M. Loève'1948) Let $\mathcal{X} = L^2[0,1]$. Then

$$X(t) \stackrel{d}{=} \sum_{k=1}^{\infty} \sqrt{\mu_k}\, u_k(t)\, \xi_k$$

- ξ_k, $k \in \mathbb{N}$, — iid standard normal rv
- $u_k(t)$, μ_k — orthonormal eigenfunctions and positive eigenvalues of covariance operator $\mathbb{G}_X$:

$$\mu_k u_k = \mathbb{G}_X u_k \qquad \Longleftrightarrow \qquad \mu_k\, u_k(t) = \int_0^1 G_X(s,t)\, u_k(s)\, ds.$$

Small ball probability problem ($\varepsilon \to 0$):

$$\mathbb{P}(\|X\|_2 < \varepsilon) = \mathbb{P}\left(\sum_{k=1}^{\infty} \mu_k \xi_k^2 < \varepsilon^2\right).$$

Main idea

All information about the process is in spectrum of the covariance operator.

What is already known?

1974	—	G. Sytaya: implicit solution in terms of Laplace transform of the sum $\sum \mu_k \xi_k^2$
from 1974	—	V.M. Zolotarev, J. Hoffmann-Jorgensen , L. Shepp, R. Dudley, I. A. Ibragimov, M. A. Lifshits,. . . : simplification of the formula under different assumptions
1998	—	T. Dunker, M. A. Lifshits, W. Linde (DLL): rather simple formulas for

$$\mathbb{P}\left(\sum \mu_k \xi_k^2 < \varepsilon^2\right) \qquad \text{when}$$

- μ_k — decays, logarithmically convex
- $\mu_k = k^{-d}, \quad d > 0,$ — polynomial decay
- $\mu_k = A^{-k}, \quad A > 0,$ — exponential decay

Useful fact: Wenbo Li principle

Let $\widehat{\mu}_k \approx \mu_k$ — some approximation.
Question: How the following probabilities are connected

$$\mathbb{P}\left(\sum \mu_k \xi_k^2 < \varepsilon^2\right) \text{ and } \mathbb{P}\left(\sum \widehat{\mu}_k \xi_k^2 < \varepsilon^2\right)?$$

Theorem (Wenbo Li principle 1992, Gao et al. 2003)

Let μ_k, $\widehat{\mu}_k$ — two summable sequences. If

$$\prod_{k=1}^{\infty} \frac{\widehat{\mu}_k}{\mu_k} < \infty, \tag{3}$$

then as $\varepsilon \to 0$

$$\mathbb{P}\left(\sum_{k=1}^{\infty} \mu_k \xi_k^2 < \varepsilon^2\right) \sim \mathbb{P}\left(\sum_{k=1}^{\infty} \widehat{\mu}_k \xi_k^2 < \varepsilon^2\right) \cdot \left(\prod \frac{\widehat{\mu}_k}{\mu_k}\right)^{1/2}$$

General scheme

We are looking for small ball probabilities:

1. Consider a spectral problem for the covariance operator $\mathbb{G}_X$

$$\mu_k u_k = \mathbb{G}_X u_k \qquad \Longleftrightarrow \qquad \mu_k\, u_k(t) = \int_0^1 G_X(s,t)\, u_k(s)\, ds.$$

2. Find rather «good» approximation $\widehat{\mu}_k$ of eigenvalues such that

$$\prod_{k=1}^{\infty} \frac{\widehat{\mu}_k}{\mu_k} < \infty,$$

3. Use DLL theorem for $\widehat{\mu}_k$ and Wenbo Li principle

Example of a general theorem (Nazarov, Nikitin' 2004)

If eigenvalues μ_k have the asymptotics

$$\mu_k = (\vartheta(k + \delta + O(k^{-1})))^{-d},$$

then for the small deviation probabilities

$$\mathbb{P}(\|X\|_2 < \varepsilon) \sim D\varepsilon^{C} \exp(B\varepsilon^{A}), \qquad \varepsilon \to 0,$$

where $A = A(d)$, $B = B(d, \vartheta)$, $C = C(d, \vartheta, \delta)$, $D = D(\{\mu_k\})$:

$$A = -\frac{2}{d-1}, \quad B = -\frac{d-1}{2}\left(\frac{\pi/d}{\vartheta \sin(\pi/d)}\right)^{\frac{d}{d-1}}, \quad C = \frac{2-d-2\delta d}{2(d-1)}$$

Example: Durbin process for Gumbel distribution

Theorem (Yu. Petrova '2017)

For Durbin process $X(t)$ for Gumbel distribution,

$$G(s,t) = \min(s,t) - \psi(t)\psi(s), \qquad \psi(t) = C\, t \ln(t) \cdot \ln(-\ln(t))$$

eigenvalue asymptotics is as follows

$$\mu_k^{-1/2} = \pi k + \frac{\pi}{2} + (-1)^k \cdot 2 \operatorname{arctg}\left(\frac{1}{\ln(\ln(k)) + 1}\right) - \frac{1}{\ln(k)\ln(\ln(k))} + \\ + O\left(\frac{1}{\ln(k)(\ln(\ln(k)))^2}\right).$$

Small ball probability asymptotics

$$\mathbb{P}\left\{\|X\|_2 < \varepsilon\right\} \sim C \cdot \ln^{-1}(\ln(\varepsilon^{-1})) \cdot \varepsilon^{-1} \cdot \exp\left(-\frac{1}{8\varepsilon^2}\right)$$

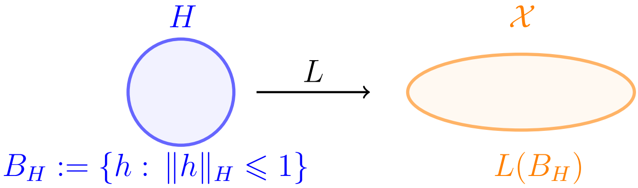
Summing up the fist part

- Hilbert space $\Longrightarrow$ spectral problem
- the whole sequence of eigenvalues μ_k is important (in contrast to large deviations where only the first eigenvalue is sufficient to know)
- very precise asymptotics can be obtained
  ... but it is quite sensitive to any perturbation of the process

Questions? Comments?

Linear operator

Consider an operator $L : H \to \mathcal{X}$ acting between normed spaces.

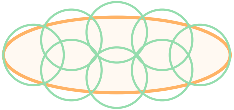


How to measure the "size" of the operator?

- The norm $\|L\|$ (half-diameter of $L(B_H)$) alone is not enough!
- We can use metric entropy

Covering numbers and entropy

One way to measure the compactness of operator $L: H \to \mathcal{X}$ is using metric entropy.



Covering numbers:

$$N_L(\varepsilon) = \inf \left\{ n : \exists \{x_j\}_{j \leqslant n}, \{Lh : \|h\|_H \leqslant 1\} \subset \cup_{j=1}^n B_\varepsilon(x_j) \right\}$$

Metric entropy: $\ln N_L(\varepsilon)$

Dyadic entropy numbers:

$$e_n(L) = \inf \{\varepsilon > 0 : N_L(\varepsilon) \leqslant 2^n\}$$

The main problem in operator language

Find the behavior of covering numbers $N_L(\varepsilon)$, as $\varepsilon \to 0$.

An example: integration operator

1. Let $H = L^2[0,1]$ and $\mathcal{X} = C[0,1]$, and let $L : L^2[0,1] \to C[0,1]$ be an integration operator:

$$L(f)(t) := \int_0^t f(s)\, ds, \qquad f \in L^2[0,1].$$

Then $e_n(L) \approx n^{-1}$.

2. Let $\alpha > 1/2$. Consider Riemann-Liouville fractional integration operator $L : L^2[0,1] \to C[0,1]$, defined by

$$L^\alpha(f)(t) := \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s)\, ds, \qquad f \in L^2[0,1].$$

Then $e_n(L) \approx n^{-\alpha}$.
Note that for $\alpha = 1$ this is the simple integration operator. Also there is a semigroup property: $L^\alpha \circ L^\beta = L^{\alpha+\beta}$.

Merging two stories: operators and processes

Any centered Gaussian vector in a separable Banach space $\mathcal{X}$ admits expansion

$$X = \sum_j \xi_j L(e_j), \qquad \text{almost surely,}$$

where ξ_j are iid standard normal rv, and $L : H \to \mathcal{X}$ an appropriate linear operator acting to a $\mathcal{X}$ from a Hilbert space H with basis (e_j).

Definition

Then the vector X and operator L are associated.

Note: the distribution of X doesn't depend on the basis (e_j),

Example of a random vector and an associated operator

Let $\mathcal{X} = C[0,1]$, $X = W$ — a Wiener process, $H = L^2[0,1]$. It turns out that an operator $L : L^2[0,1] \to C[0,1]$ that is associated to Wiener process is just an integration operator.

$$L(f)(t) = \int_0^t f(s)\,ds, \qquad f \in L^2[0,1].$$

Let us consider the cosine basis in $L^2[0,1]$, given by $e_0(s) := 1$ and

$$e_j(s) := \sqrt{2}\cos(\pi j s), \quad j \geqslant 1.$$

Integration yields $Le_0(t) = t$ and

$$Le_j(t) = \sqrt{2}\,\frac{\sin(\pi j t)}{\pi j}, \quad j \geqslant 1.$$

So we arrive at the expansion

$$W(t) = \xi_0 t + \sqrt{2}\sum_{j=1}^{\infty} \xi_j\,\frac{\sin(\pi j t)}{\pi j}.$$

Metric entropy and Gaussian small deviations

Let's concentrate on logarithmic small ball probabilities and define small deviation function by:

$$\varphi(\varepsilon) := -\ln \mathbb{P}(\|X\| < \varepsilon)$$

Relation between $\ln N_L(\varepsilon)$ and $\varphi(\varepsilon)$:

1. polynomial growth:
   Let $\beta \in (0, 2)$. Then

   $$\ln N_L(\varepsilon) \approx \varepsilon^{-\beta} \quad \Longleftrightarrow \quad \varphi(\varepsilon) \approx \varepsilon^{-\frac{2\beta}{2-\beta}}, \qquad \text{as } \varepsilon \to 0.$$

   Example: L — integration operator, W — Wiener process, $\beta = 1$.
   L — fractional integration operator, X — Riemann-Liouville process.
2. logarithmic growth:
   Let $\beta > 0$, $\gamma \in \mathbb{R}$. Then

$$\ln N_L(\varepsilon) \approx |\ln \varepsilon|^{\beta} \ln |\ln \varepsilon|^{\gamma} \quad \Longleftrightarrow \quad \varphi(\varepsilon) \approx |\ln \varepsilon|^{\beta} \ln |\ln \varepsilon|^{\gamma}, \quad \varepsilon \to 0.$$

General principles

The following properties are related:

- the small deviation probabilities $\mathbb{P}(\|X\| \leqslant \varepsilon)$ are not too small when $\varepsilon \to 0$;
- small deviation function $\varphi(\varepsilon) := -\ln \mathbb{P}(\|X\| \leqslant \varepsilon)$ is growing slowly when $\varepsilon \to 0$;
- sample paths of a process are rather smooth;
- X has good finite-rank approximations:

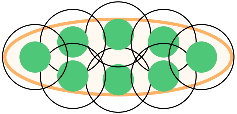
$$X \approx \sum_{j=1}^{n} \xi_j L(e_j), \qquad n \to \infty.$$

How the connection occurs?

We start with an operator $L : H \to \mathcal{X}$. Fix some R, ε. Take the image of the R-ball

$$L(B_R) = \{Lh : \|h\|_H < R\}$$

and construct a pairwise distant points: $h_1, h_2, \ldots$ such that $\|h_i\| < R$ and $\|Lh_i - Lh_j\| > \varepsilon$ for $i \neq j$.



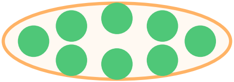
$L(B_R)$

Clearly, we can collect at least $N_{L(B_R)}(\varepsilon)$ points and

$$N_{L(B_R)}(\varepsilon) = N_{L(B_1)}(\varepsilon/R) = N_L(\varepsilon/R).$$

How the connection occurs? Continued

We have a picture from a former slide

$L(B_R)$ 

The green balls are $Lh_j + \frac{\varepsilon}{2}U$ where U is the unit ball in $\mathcal{X}$.
Christer Borell shift inequality: for every symmetric set $B \subset \mathcal{X}$ and every associated centered Gaussian vector X and operator L, and every $h \in H$

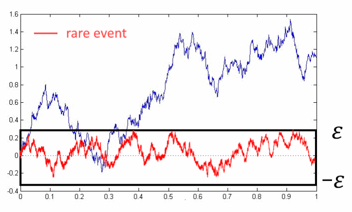
$$\mathbb{P}(X \in B + Lh) \geqslant \mathbb{P}(X \in B) \exp(-\|h\|_H^2/2).$$

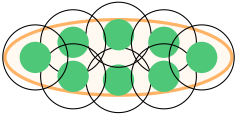
It follows that

$$1 \geqslant \mathbb{P}\big(X \in \cup_j \{Lh_j + \frac{\varepsilon}{2}U\}\big) = \sum_j \mathbb{P}(X \in \{Lh_j + \frac{\varepsilon}{2}U\})$$
$$\geqslant N_L(\varepsilon/R)\mathbb{P}(X \in \frac{\varepsilon}{2}U)e^{-R^2/2} = N_L(\varepsilon/R)\mathbb{P}(\|X\| < \frac{\varepsilon}{2})e^{-R^2/2}$$

This reads as $\mathbb{P}(\|X\| < \frac{\varepsilon}{2}) \leqslant e^{R^2/2} N_L(\varepsilon/R)^{-1}$. Optimize the RHS in R!

Thank you for your attention!





$L(B_R)$

Questions? Comments?

For any questions: https://yulia-petrova.github.io/