Large deviations of Gaussian vectors

Let $X$ be a centered Gaussian random vector, on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with values in a real separable Banach space $E$ equipped with its Borel $\sigma$-algebra $\mathcal{B}$, and with norm $\| \cdot \|$.

It is a consequence of the sharp integrability of the norms of Gaussian random vectors (cf. [1]) that

$$\lim_{t \to \infty} t^2 \log \mathbb{P}(\|X\| \geq t) = -\frac{1}{2\sigma^2}$$

where

$$\sigma^2 = \sup_{\xi \in E^*, \|\xi\| \leq 1} E(\langle \xi, X \rangle^2).$$

This result is actually a particular case of a more general large deviation principle for the family of laws of $\varepsilon X$ as $\varepsilon \to 0$, providing further knowledge on tail behaviors.

The post briefly presents this large deviation theorem. General references on large deviations include [13, 8, 7, 11, 6] etc.

Table of contents

1. Rate function
2. The large deviation principle

References
1 Rate function

Given a centered Gaussian random vector $X$ on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $E$, its law $\mu$ on the Borel sets of $E$ gives rise to an abstract Wiener space structure $(E, \mathcal{H}, \mu)$, in which the Hilbert space $\mathcal{H} \subset E$, with scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, is the reproducing kernel Hilbert space associated to the covariance structure of $\mu$ (cf. [2]).

For the example of the Wiener measure $\mu$ on the Banach space $E = C([0, 1])$ of real continuous functions on $[0, 1]$, law of a standard Brownian motion or Wiener process $W = (W(t))_{t \in [0,1]}$, the reproducing kernel Hilbert space $\mathcal{H}$ is identified as the subspace of $E = C([0, 1])$ consisting of the absolutely continuous functions $h : [0, 1] \to \mathbb{R}$, with almost everywhere derivative $h'$ in $L^2([0, 1])$ (for the Lebesgue measure), and with

$$|h|_{\mathcal{H}} = \left( \int_0^1 (h'(t))^2 \, dt \right)^{1/2}.$$ 

The rate function $\mathcal{I} : E \to [0, +\infty]$ which will govern the large deviation properties of $\varepsilon X$ as $\varepsilon \to 0$ is defined as

$$\mathcal{I}(x) = \begin{cases} \frac{1}{2}|x|^2_{\mathcal{H}} & \text{if } x \in \mathcal{H}, \\ +\infty & \text{if } x \notin \mathcal{H}. \end{cases} \quad (3)$$

In the large deviation language, this rate function is a good rate function in the sense that its level sets $\{\mathcal{I} \leq a\}, a \geq 0$, are compact in $E$ (due to the compactness of the $\mathcal{H}$-balls in $E$).

2 The large deviation principle

Large deviations for Gaussian measures go back to M. Schilder [12] for the Wiener measure, and to M. Donsker and S. Varadhan [9] in general. The study of [9] actually addresses the large deviation principle for sums of independent Banach space valued random variables, the Gaussian case being a particular case.

In the context exposed in the first section, the following theorem presents the large deviation behavior of the law of $\varepsilon X$ as $\varepsilon \to 0$. For a subset $A$ of $E$, let

$$\mathcal{I}(A) = \inf_{x \in A} \mathcal{I}(x).$$

**Theorem 1** (The Gaussian large deviation principle). For any closed set $F$ in $E$,

$$\limsup_{\varepsilon \to 0} \varepsilon^2 \log \mathbb{P}(\varepsilon X \in F) \leq -\mathcal{I}(F). \quad (4)$$
For any open set \( O \) in \( E \),

\[
\liminf_{\varepsilon \to 0} \varepsilon^2 \log \mathbb{P}(\varepsilon X \in O) \geq -\I(O). \tag{5}
\]

Applied to complements of balls, this theorem easily produces the limit (1), together with the observation that \( \sigma = \sup_{|h|_{\mathcal{H}} \leq 1} ||h|| \).

The proof of the upper-bound (4) in Theorem 1 presented here relies on isoperimetric and concentration inequalities (cf. [3, 4]) which provide a very convenient tool to this task. The lower-bound (5) classically relies on the Cameron-Martin translation formula. The combined arguments actually produce a measurable version of the large deviation principle, without referring to any topology associated to the underlying abstract Wiener space (cf. [5, 10]).

**Proof.** A simple proof of the upper-bound (4) may therefore be provided by the Gaussian isoperimetric inequality (actually only the suitable concentration properties). Namely, let \( F \) be closed in \( E \), and take \( r \) such that \( 0 < r < \I(F) \). By the very definition of \( \I(F) \),

\[
F \cap \sqrt{2r} \mathcal{K} = \emptyset,
\]

where \( \mathcal{K} \) is the (closed) unit ball in \( \mathcal{H} \). Since \( F \) is closed and \( \mathcal{K} \) is compact in \( E \), there exists \( \eta > 0 \) such that it still holds true that

\[
F \cap [\sqrt{2r} \mathcal{K} + B_E(0, \eta)] = \emptyset
\]

where \( B_E(0, \eta) \) is the ball with center the origin and with radius \( \eta \) for the norm \( || \cdot || \) in \( E \). Clearly

\[
\lim_{\varepsilon \to 0} \mathbb{P}(\varepsilon X \in B_E(0, \eta)) = \lim_{\varepsilon \to 0} \mathbb{P}(X \in B_E(0, \frac{\eta}{\varepsilon})) = 1.
\]

Recall now the Gaussian isoperimetric inequality for the law of \( X \) (cf. [3]), expressing that, whenever \( \mathbb{P}(X \in A) \geq \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}x^2}dx \) for some \( a \in \mathbb{R} \),

\[
\mathbb{P}(X \in A + s \mathcal{K}) \geq \Phi(a + s)
\]

for every \( s \geq 0 \). For \( \varepsilon > 0 \) small enough, \( \mathbb{P}(X \in B_E(0, \frac{\eta}{\varepsilon})) \geq \frac{1}{2} = \Phi(0) \). Hence,

\[
\mathbb{P}(\varepsilon X \in F) \leq \mathbb{P}(\varepsilon X \notin \sqrt{2r} \mathcal{K} + B_E(0, \eta)) \leq 1 - \Phi\left(\frac{\sqrt{2r}}{\varepsilon}\right) \leq e^{-r/\varepsilon^2}.
\]

Therefore

\[
\limsup_{\varepsilon \to 0} \varepsilon^2 \log \mathbb{P}(\varepsilon X \in F) \leq -r,
\]

which is the result since \( r < \I(F) \) is arbitrary.
As mentioned above, the full strength of the Gaussian isoperimetric inequality is not really needed, and weaker concentration inequalities are enough to achieve the conclusion. For example, as emphasized in [4],

\[ \mathbb{P}(X \in A + sK) \geq 1 - e^{-\frac{1}{2}s^2 + \delta(\mu(A))s} \]

for every \( s \geq 0 \), where \( \delta(\mu(A)) \to 0 \) as \( \mu(A) \to 1 \), so that the proof may be developed similarly.

The proof of the lower-bound (5) is an application of the Cameron-Martin translation formula. Let \( h \in O \cap \mathcal{H} \). Since \( O \) is open, there exists \( \eta > 0 \) such that \( h + B_E(0, \eta) \subset O \), and thus

\[ \mathbb{P}(\varepsilon X \in O) \geq \mathbb{P}(\varepsilon X \in h + B_E(0, \eta)). \]

In the notation of [2], the Cameron-Martin translation formula yields that

\[ \mathbb{P}(\varepsilon X \in h + B_E(0, \eta)) = \mu\left(\frac{h}{\varepsilon} + B_E(0, \frac{\eta}{\varepsilon})\right) \]

\[ = \exp\left( - \frac{|h|_H^2}{2\varepsilon^2} \right) \int_{B_E(0, \frac{\eta}{\varepsilon})} \exp\left( - \frac{\tilde{h}}{\varepsilon} \right) d\mu, \]

where it is recalled that \( \tilde{h} \) is Gaussian under \( \mu \) with variance \( |h|_H^2 \) (\( \tilde{h} = \int_0^1 h'(t)dW(t) \) on the Wiener space). By Jensen’s inequality,

\[ \int_{B_E(0, \frac{\eta}{\varepsilon})} \exp\left( - \frac{\tilde{h}}{\varepsilon} \right) d\mu \geq \mu(B_E(0, \frac{\eta}{\varepsilon})) \exp\left( - \int_{B_E(0, \frac{\eta}{\varepsilon})} \frac{\tilde{h}}{\varepsilon} d\mu \right). \]

Now

\[ \int_{B_E(0, \frac{\eta}{\varepsilon})} \tilde{h} d\mu \leq \int_E |\tilde{h}| d\mu \leq \left( \int_E \tilde{h}^2 d\mu \right)^{1/2} = |h|_H. \]

For every \( \varepsilon > 0 \) small enough, \( \mu(B_E(0, \frac{\eta}{\varepsilon})) \geq \frac{1}{2} \) (for example). As a consequence of the various preceding lower-bounds,

\[ \mathbb{P}(\varepsilon X \in O) \geq \frac{1}{2} \exp\left( - \frac{|h|_H^2}{2\varepsilon^2} - \frac{2|h|_H}{\varepsilon} \right) \]

from which it follows that

\[ \liminf_{\varepsilon \to 0} \varepsilon^2 \log \mathbb{P}(\varepsilon X \in O) \geq -\frac{1}{2} |h|_H^2 = -\mathcal{I}(h). \]

This result for any \( h \in O \cap \mathcal{H} \) yields the announced lower-bound (5), and completing therefore the proof of Theorem 1.
References


