1. Basic theorems for Gaussian processes

We now prove concentration inequalities and comparison theorems for Gaussian processes. The difficult and interesting part of the work was already done in the finite dimensional case. We simply deduce the general case from the finite case.

Comparison theorems: We discuss the Sudakov-Fernique inequality. Let X and Y be Gaussian processes on T. Suppose they have the same mean function and $\tau_X(t,s) \le \tau_Y(t,s)$ for all $t,s \in T$ (we shall henceforth use the notation $\tau_X(t,s) := \sqrt{\mathbf{E}[(X_t - X_s)^2]}$). Then is it true that $\mathbf{E}[\sup_{t\in T} X_t] \le \mathbf{E}[\sup_{t\in T} Y_t]$? The answer is essentially yes, with a few caveats.

Firstly, the question of measurability of $X^* = X^*_T := \sup_{t\in T} X_t$. If T is countable, X^*_T is trivially measurable. But if T is uncountable this is not necessarily the case. Indeed, in the cylinder sigma-algebra this is not measurable at all, so the only hope is when X has some more structure. Therefore we make an assumption.

Separability assumption: There exists a countable subset T' of T such that $\sup_{t\in T} X_t = \sup_{t\in T'} X_t$. For example, this satisfied if T is a separable metric space and X is $C(T)$-valued. Under this assumption, X^*_T is a random variable (measurable).

Let F_n be finite sets that increase to T'. Then $\mathbf{E}[X^*_{F_n}] \le \mathbf{E}[Y^*_{F_n}]$ and $X^*_{F_n} \uparrow X^*_{T'}$ a.s. and similarly for Y. By the monotone convergence theorem we get $\mathbf{E}[X^*] \le \mathbf{E}[Y^*]$.

Observe that monotone convergence theorem holds when $X^*_{F_n}$ (respectively $Y^*_{F_n}$) are bounded from below by an integrable random variable. In this case we can take and $t_0 \in T$ and use X_{t_0} as a lower bound for $X^*_{F_n}$. Same story for Y. For this same reason, $\mathbf{E}[X^*_T]$ has a well-defined value in $(-\infty, +\infty]$. Unlike in finite dimensional settings, the supremum could very well be infinite.

Exercise 1. Check with care that Slepian's and Gordon's inequalities also remain valid for general Gaussian processes.

Concentration theorems: Let X be a centered Gaussian process on T. For finite T, we saw that $\mathbf{P}\{|X^* - M| \ge t\} \le 2\overline{\Phi}(t/\sigma_T)$ where M is a median of X^* and $\sigma_T^2 = \sup_{t\in T} \mathbf{E}[X_t^2]$. Does this remain valid for general T?

We again make the separability assumption. In addition assume that $X^* < \infty$ a.s. Then as before there exist increasing finite sets F_n such that $X^*_{F_n} \uparrow X^*$ a.s. Argue that the medians of $X^*_{F_n}$ converge to that of X^*. By adding more points to F_n if necessary, we may assume that $\sigma^2_{F_n} \to \sigma^2_T$.

From $\mathbf{P}\{|X^*_{F_n} - M| \ge t\} \le 2\overline{\Phi}(t/\sigma_{F_n})$ educe that the concentration inequality holds. Quite often Borell's isoperimetric inequality refers to this inequality -

$$\mathbf{P}\{|X^* - M| \ge t\} \le 2\overline{\Phi}(t/\sigma_T) \le 2e^{-t^2/2\sigma_T^2}. \tag{1}$$

Thus the supremum of a Gaussian process (if it is finite!) has tails that decay no slower than the Gaussian in the process with maximal variance. Historically the following consequence was a predecessor to Borell's inequality (and perhaps inspired the research that led to it?)

Exercise 2. Let X be a centered Gaussian process on T. Assume separability and that X^* is finite a.s. Show that $\lim_{x\to+\infty} \frac{1}{x^2} \log \mathbf{P}\{X^*_T \ge x\} = -\frac{1}{2\sigma_T^2}$.

How to check that X^* is finite a.s.? If T is a compact metric space and X is $C(T)$-valued, then it is clear that X^* is finite. But checking that X is continuous on T is no easier than checking the boundedness. Indeed both problems are closely related and we shall discuss them later. As remarked in an earlier lecture, this is a fundamental (and completely solved) problem in Gaussian processes.

Exercise 3. If X is as above and $X^* < \infty$ a.s. then $\mathbf{E}[X^*] < \infty$.

Isoperimetric inequality: Let $X = (X_1, X_2, \ldots)$ where X_i are i.i.d. $N(0,1)$. Then X is a Gaussian process on $\mathbb{N}$ and its distribution is denoted $\gamma_{\mathbb{N}}$. What form does isoperimetric inequality take?

From the finite dimensional case, $\Phi^{-1}(\gamma_n(A_\varepsilon)) \geq \Phi^{-1}(\gamma_n(A)) + \varepsilon$, where γ_n is the distribution of $(X_1, \ldots, X_n)$ (push forward $\gamma_{\mathbb{N}}$ by projection on the first n coordinates). Note that A_ε is the ε neighbourhood in Euclidean metric on $\mathbb{R}^n$.

Now if A is a cylinder set in $\mathbb{R}^{\mathbb{N}}$, then $A = B \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \ldots$ for some large n and for some $B \subseteq \mathbb{R}^n$ (Borel). Therefore, by the previous paragraph we easily get $\Phi^{-1}(\gamma_{\mathbb{N}}(A_\varepsilon)) \geq \Phi^{-1}(\gamma_{\mathbb{N}}(A)) + \varepsilon$, where $A_\varepsilon = \{\mathbf{x} \in \mathbb{R}^{\mathbb{N}} : \|\mathbf{x} - \mathbf{y}\|_2 \leq \varepsilon$ for some $\mathbf{y} \in A\}$ with $\|\cdot\|_2$ being the ℓ^2 norm.

If A is an arbitrary measurable set in $\mathbb{R}^{\mathbb{N}}$ then there exist finite dimensional cylinder sets whose measures decrease to A. Deduce that the isoperimetric inequality continues to remain valid.

Exercise 4. Let $X = (X_1, X_2, \ldots)$ be a centered Gaussian process on $\mathbb{N}$ with $\sigma^2 = \sup_n \mathbf{E}[X_n^2] < \infty$. Let μ be the distribution of X. Show that $\Phi^{-1}(\mu(A_{\varepsilon\sigma})) \geq \Phi^{-1}(\mu(A)) + \varepsilon$ for any measurable $A \subseteq \mathbb{R}^{\mathbb{N}}$.

A zero-one law: In proving concentration for the maximum, we assumed that X^* is finite almost surely. Full discussion of when that happens will come later. For now we show that $\mathbf{P}\{X^* < \infty\}$ is either 0 or 1, perhaps not unexpected. As always, we assume separability.

By moving to the countable subset T', we may assume that $T = \mathbb{N}$. Suppose $\mathbf{P}\{X^* = \infty\} < 1$. Then there exists $A < \infty$ and $u > -\infty$ such that $\mathbf{P}\{X^* < A\} \geq \Phi(u)$. By Exercise **??**, we get $\mathbf{P}\{X^* \leq A + \sigma\varepsilon\} \geq \Phi(u + \varepsilon)$ which converges to 1 as $\varepsilon \to 0$. Thus, $X^* < \infty$ a.s.

To summarize, if X^* is finite with positive probability, then it is finite a.s. and has finite expectation and sub-Gaussian tails!

Banach space valued Gaussian random variable: Let $(B, \|\cdot\|)$ be a Banach space with a separable dual B^*. A B-valued random variable X is said to be Gaussian if its distribution is Radon (i.e., $\mathbf{P}\{X \in A\} = \sup_K \mathbf{P}\{X \in K\}$ where the supremum is over all compact subsets $K \subseteq A$) and for every $L \in B^*$ we have that $L(X)$ is $N(0, \sigma_L^2)$ for some σ_L^2.

Consider a Gaussian process $X = (X_t)_{t \in T}$. Suppose we have some regularity on sample paths, for example, suppose that T is a separable metric space and X is $C(T)$ valued. Then X is a $C(T)$-valued Gaussian (of course $C(T)$ is a Banach space if T is compact and a locally convex space even if not).

Conversely, if X is a B-valued Gaussian random variable, and T is a countable dense subset of the unit ball of B^*, then by setting $X_L := L(X)$ for $L \in T$, we get a Gaussian process indexed by T.

Many books talk in the language of Banach space valued Gaussians. I have not understood what advantage there is in talking about B-valued random variables instead of simply Gaussian processes. I will avoid the language of Banach space valued random variables till I understand it myself!

1. AN APPLICATION TO RANDOM MATRICES

Here we present[10] few applications of the basic results on Gaussian processes, namely concentration of measure and comparison theorems.

Extreme singular values of a rectangular Gaussian matrix: Let $A_{m,n} = (a_{i,j})_{i \le m, j \le n}$ be a matrix whose entries are i.i.d. $N(0,1)$. We assume $m \le n$ and denote the singular values of A by $s_1 \le s_2 \le \ldots \le s_m$ (by definition s_i^2 are the eigenvalues of AA^t). The following result gives bounds for the smallest and largest singular values.

Theorem 1 (Gordon). *With A as above, $\mathbf{E}[s_1] \ge \sqrt{n} - \sqrt{m}$ and $\mathbf{E}[s_n] \le \sqrt{n} + \sqrt{m}$.*

Proof. For $(u,v) \in T := S^{m-1} \times S^{n-1}$ define $X(u,v) = u^t A v = \sum_{i=1}^m \sum_{j=1}^n a_{i,j} u_i v_j$. It has zero mean and $\mathbf{E}[|X_{u,v} - X_{u',v'}|^2] = 2 - 2\langle u, u' \rangle \langle v, v' \rangle$ (check!).

Consider a different Gaussian process on the same index set defined by $Y(u,v) = \sum_{i=1}^m u_i \xi_i + \sum_{j=1}^n v_i \eta_i$ where ξ_i, η_i are i.i.d. $N(0,1)$. Then $\mathbf{E}[|Y_{u,v} - Y_{u',v'}|^2] = |u-u'|^2 + |v-v'|^2 = 4 - 2\langle u,u' \rangle - 2\langle v, v' \rangle$. Both X and Y are continuous on T and hence the comparison theorems are applicable.

Thus,

$$\mathbf{E}[|Y_{u,v} - Y_{u',v'}|^2] - \mathbf{E}[|X_{u,v} - X_{u',v'}|^2] = 2(1 - \langle u, u' \rangle)(1 - \langle v, v' \rangle)$$

which is non-negative for all $(u,v), (u',v') \in T$. Therefore, by the Sudakov-Fernique inequality we get $\mathbf{E}[X^*] \le \mathbf{E}[Y^*]$. Clearly $Y^* \le \|\xi\| + \|\eta\|$ and $\mathbf{E}[\|\xi\|] \le \sqrt{\mathbf{E}[\|\xi\|^2]} = \sqrt{m}$ and $\mathbf{E}[\|\eta\|] \le \sqrt{\mathbf{E}[\|\eta\|^2]} = \sqrt{n}$. But X^* is precisely s_m. Therefore $\mathbf{E}[s_m] \le \sqrt{n} + \sqrt{m}$.

Next observe that $s_1 = \min_u \max_v X_{u,v}$. We have already seen that

$$\mathbf{E}[|Y_{u,v} - Y_{u',v'}|^2] \ge \mathbf{E}[|X_{u,v} - X_{u',v'}|^2] \text{ for all } u, v, u', v',$$

$$\mathbf{E}[|Y_{u,v} - Y_{u,v'}|^2] = \mathbf{E}[|X_{u,v} - X_{u,v'}|^2] \text{ for all } u, v, v'.$$

For the second, observe that $\langle u, u' \rangle = 1$ when $u = u'$. Gordon's inequality applies to give $\mathbf{E}[s_1] \ge \mathbf{E}[\min_u \max_v Y_{u,v}]$.■ As the last step in the proof, observe that picking $v = \eta / \|\eta\|$ and $u = -\xi/\|\xi\|$ achieves the $\min_u \max_v Y_{u,v}$ and gives $\mathbf{E}[\min_u \max_v Y_{u,v}] = \mathbf{E}[\|\eta\|] - \mathbf{E}[\|\xi\|]$. Since $\|\eta\|^2 \sim \chi_{n-1}^2$,

$$\mathbf{E}[\|\eta\|] = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty \sqrt{x} e^{-x} x^{\frac{n}{2}-1} dx = \frac{\sqrt{2}\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}$$

and similarly $\mathbf{E}[\|\xi\|] = \frac{\sqrt{2}\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})}$. Thus the theorem is proved if we show that $\mathbf{E}[\|\eta\|] - \mathbf{E}[\|\xi\|] \ge \sqrt{n} - \sqrt{m}$. Deduce this from Exercise **??**. ■

Exercise 2. Show that $\nu \to \frac{\sqrt{2}\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} - \sqrt{\nu}$ is increasing for $\nu \ge 1$.

Location of individual singular values of a Gaussian matrix: Let $A_{m,n}$ be a real symmetric matrix such that $a_{i,j}$, $i \le j$ are i.i.d. $N(0,1)$ (it is okay to allow the diagonals to have variance 2 to make it exactly a GOE matrix). Let $\lambda_{n,1} < \ldots < \lambda_{n,n}$ be the eigenvalues of $A_n/\sqrt{n}$ (normalized so that the empirical distribution of eigenvalues converges to the semicircle distribution as n tends to infinity)

Theorem 3. *There exist deterministic numbers $t_{n,k}$ such that $\mathbf{P}\{|\lambda_{n,k} - t_{n,k}| \ge u\} \le Ce^{-cnu^2}$ for all $k \le n$.*

---

[10]This material is taken from the paper *Local operator theory, random matrices and Banach spaces*, by Davidson and Szarek. Roman Vershynin has several lecture notes that cover this and much more.

Proof. Recall the min-max representation

$$\lambda_{n,n-k+1} = \frac{1}{\sqrt{n}} \min_{\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}} \max_{\mathbf{u}:\mathbf{u}\perp\mathbf{v}_j} \mathbf{u}^t A\mathbf{u}.$$

From this (since $A \to \mathbf{u}^t A\mathbf{u}$ is linear for each $\mathbf{u}$) it follows that the function $(a_{i,j})_{i\le j\le n} \to \lambda_{n,n-k+1}$ is $\text{Lip}(2/\sqrt{n})$. By the Gaussian concentration inequality, if $t_{n,n-k+1}$ is a median of $\lambda_{n,n-k+1}$ then $\mathbf{P}\{|\lambda_{n,n-k+1} - t_{n,n-k+1}| \ge u\} \le 2\overline{\Phi}(u/2\sqrt{n}) \le 2e^{-nu^2/8}$. ■

Remark 4. The well-known Wigner's semicircle law says that the histogram of eigenvalues is close to the semi-circle density $c\sqrt{4-x^2}$. This does not imply a quantitative estimate for the location of individual eigenvalues. In contrast, the above theorem shows that each eigenvalue is concentrated in a window of length essentially $1/\sqrt{n}$. However the actual facts (proved by harder methods) are that eigenvalues are concentraetd in even smaller windows (of length $1/n$ if k is away from 1 and n and of length $n^{-2/3}$ if k is close to 1 or n).