Printed: 8 December 2015

Chapter 6

Gaussian processes

SECTION 6.1 *states three beautiful facts about multivariate normal distributions: the Sudakov inequality; the Fernique comparison inequality; and the concentration inequality for Lipschitz functionals, with the Borell inequality as a special case.*
SECTION 6.2 *sketches a proof of the Fernique inequality, then shows how it implies the Sudakov inequality.*
SECTION 6.3 *presents four different proofs for slightly differnt version of the Lipschitz concentration inequality. The proofs use techniques that have proven themselves most useful for the study of Gaussian processes.*

6.1 Introduction

Gaussian::S:intro

This chapter has two aims:

(i) to describe the technical tools that are needed (in Chapter 7) to establish the various equivalences, for centered Gaussian processes, between the finiteness of $\mathbb{P}\sup_{t\in T} X_t$ and the existence of majorizing measures, as described in Section 4.6;

(ii) to describe some surprising properties of Gaussian processes that have been the starting point for a flourishing literature on the concentration of measure phenomenon, as discussed in Chapters 11 and 12.

Happily the two aims overlap.

An essential ingredient for Talgrand's majoring measure argument is an inequality usually attributed to Sudakov (but consult the references in Section 6.5 for a more complete account of the history).

Gaussian::Sudakov <1> **Theorem.** *("Sudakov's minoration") Let $Y := (Y_1, Y_2, \dots, Y_n)$ have a centered (zero means) multivariate normal distribution, with $\mathbb{P}|Y_j - Y_k|^2 \geq \delta^2$ for all $j \neq k$. Then $(4\pi)^{1/2}\mathbb{P}\max_{i\leq n} Y_i \geq \delta\sqrt{\log_2 n}$.*

Remark. The lower bound is sharp within a constant, in the following sense. If $\mathbb{P}|Y_j - Y_k|^2 \leq \delta^2$ for all $j \neq k$ then $\mathbb{P}\max_i Y_i = \mathbb{P}Y_1 + \mathbb{P}\max_i(Y_i - Y_1) = \mathbb{P}\max_i(Y_i - Y_1)$ and

$$
\begin{aligned}
&\exp\left(\mathbb{P}\max_i(Y_i - Y_1)/2\delta\right)^2 \\
&\quad\leq \mathbb{P}\max_i \exp\left((Y_i - Y_1)^2/4\delta^2\right) \qquad \text{by Jensen} \\
&\quad\leq n\mathbb{P}\exp\left(W^2\right) \qquad \text{with } W \sim N\left(0, \tfrac{1}{4}\right).
\end{aligned}
$$

Thus $\mathbb{P}\max_i Y_i$ is bounded above by $2\delta\sqrt{\log(\sqrt{2}n)}$.

The minoration can be proved (Section 6.2) by using a comparison theorem due to Fernique (1975, page 18).

fernique.thm <2> **Fernique's comparison inequality.** *Suppose X and Y both have centered (zero means) multivariate normal distributions, with*

$$\mathbb{P}|X_i - X_j|^2 \leq \mathbb{P}|Y_i - Y_j|^2 \qquad \text{for all } i, j.$$

Then

$$\mathbb{P}f(\max_i X_i - \min_i X_i) \leq \mathbb{P}f(\max_i Y_i - \min_i Y_i)$$

for each increasing, convex function f on $\mathbb{R}^+$.

Section 6.2 sketches the proof of this inequality. The method of proof illustrates an important technique: construct a path between X and Y along which the expected value of interest increases.

The other ingredient in the majorizing measure argument is a concentration inequality for the supremum of a Gaussian process. To avoid measurability issues, assume the index set is at worst countably infinite.

Gaussian::Borell.subg <3> **Borell's inequality.** *Suppose $\{Y_t : t \in T\}$ is Gaussian process with T finite or countably infinite. Assume both $m := \mathbb{P}\sup_{t\in T} Y_t < \infty$ and $\sigma^2 := \sup_{t\in T}\text{var}(Y_t) < \infty$. Then*

$$\mathbb{P}\{|\sup_{t\in T} Y_t - m| \geq \sigma u\} \leq 2\exp(-u^2/2) \qquad \text{for all } u \geq 0.$$

Consequently, $\|\sup_{t\in T} Y_t - m\|_{\Psi_2} \leq C_{Bor}\sigma$, with C_{Bor} a universal constant.

In special cases (such as independent $N(0,1)$-distributed variables, as shown by the Problems to Chapter 4) one can get tighter bounds, but Borell's inequality has the great virtue of being impervious to the effects of possible dependence between the Y_t.

Theorem <3> can be deduced from a more basic fact about the $N(0, I_n)$ distribution on $\mathbb{R}^n$.

For vectors in $\mathbb{R}^n$ write $|\cdot|$ for the usual ℓ^2 distance: $|x|^2 = \sum_i x_i^2$.

Gaussian::Lipschitz.fnal <4> **Theorem.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz function, with $\|f\|_{\text{Lip}} \le \kappa$. That is, $|f(x) - f(y)| \le \kappa|x - y|$ for all $x, y \in \mathbb{R}^n$. Then, for a universal constant C,*

$$\gamma_n\{f(x) \ge \gamma_n f + \kappa u\} \le e^{-u^2/(2C)} \qquad \text{for all } u \ge 0.$$

where γ_n denotes the $N(0, I_n)$ distribution.

Remark. Notice that the dimension n does not appear explicitly in the upper bound, although it might enter implicitly through κ for some functionals.

This Theorem provides a good illustration of several different arguments that have been developed for Gaussian processes. Section 6.3 contains four different proofs of the Theorem. The easiest method (Pisier-Maurey, subsection 6.3.1) gives the concentration bound with $C = \pi^2/4$. The smart path method (subsection 6.3.2) improves the constant to 2. The stochastic calculus method (subsection 6.3.3) improves the constant to 1. The deepest method (subsection 6.3.4), based on the Gaussian isoperimetric inequality, again gives the constant 1 but with centering at the median of $f(x)$. Together the four methods offer a mini-course in Gaussian tricks.

Remark. The constant $C = 1$ is the best possible in general. If u is a unit vector the linear function $f(x) = u'x$ is Lipschitz with $\kappa = 1$. Under γ_n the function $f(x)$ has a $N(0,1)$ distribution, whose tails decrease like $\exp(-u^2/2)$.

Let me show you how Theorem <4> implies the analog of the Borell inequality with the $u^2/2$ in the exponent replaced by $u^2/(2C)$ for whichever constant C you feel comfortable to use. (Different C's just lead to different values for C_{Bor}, but have no important effect on the arguments in Chapter 7.)

Suppose $T = \mathbb{N}$. Define $M_n = \max_{i \le n} Y_i$. For each fixed n we can think of each Y_i as a linear functional, $Y_i(x) = \mu_i + a_i'x$, on $\mathbb{R}^n$ equipped with γ_n, with $A = [a_1, \ldots, a_n]$ an $n \times n$ matrix with $A'A$ equal to the variance matrix of $(Y_1, \ldots, Y_n)$. That gives $|a_i|_2^2 = \text{var}(Y_i) \le \sigma^2$.

The functional $f(x) := \max_{i\le n} Y_i(x)$ is Lipschitz:

$$\begin{aligned}|f(x)-f(z)| &= |\max_{i\le n}(\mu_i + a_i'x) - \max_{i\le n}(\mu_i + a_i'z)| \\ &\le \max_{i\le n}|(\mu_i + a_i'x) - (\mu_i + a_i'z)| \\ &\le \max_{i\le n}|a_i|\,|x-z| \qquad \text{by Cauchy-Schwarz} \\ &\le \sigma|x-z|.\end{aligned}$$

Theorem <4> gives

$$\mathbb{P}\{M_n \ge \mathbb{P}M_n + \sigma u\} \le e^{-u^2/(2C)},$$

which implies

$$\mathbb{P}\{M_n > r\} \le e^{-u^2/(2C)} \qquad \text{for } r > m + \sigma u \text{ and each } n.$$

In the limit, as $n\to\infty$, we get a one-sided analog of Theorem <3>. Repeat the argument with f replaced by $-f$ to deduce the two-sided bound.

6.2 The Fernique inequality

Gaussian::S:Fernique

The following sketch of Fernique's argument summarizes the more detailed exposition by Pollard (2001, Section 12.3).

First a smoothing argument shows that the function f could be assumed to be infinitely differentiable with second derivative having compact support, which sidesteps integrability questions and allows uninhibited appeals to integration-by-parts.

Suppose $X \sim N(0,V_0)$ and $Y \sim N(0,V_1)$. The main idea is to interpolate between X and Y along a path $X(\theta) = \sqrt{1-\theta} + \sqrt{\theta}Y$, for $0\le\theta\le 1$. The random vector $X(\theta)$ has a $N(0,V_\theta)$ distribution, where

$$V_\theta = (1-\theta)V_0 + \theta V_1 = V_0 + \theta D$$

By Fourier inversion, the $N(0,V_\theta)$ distribution has density

$$g_\theta(x) = (2\pi)^{-n}\int_{\mathbb{R}^n} \exp\left(-ix't - \tfrac{1}{2}t'V_\theta t\right).$$

Differentiation under the integral sign leads to the identity

$$\frac{\partial g_\theta(x)}{\partial\theta} = \frac{1}{2}\sum_{j=1}^n\sum_{k=1}^n D_{j,k}\frac{\partial^2 g_\theta(x)}{\partial x_j \partial x_k}.$$

It remains to show that the function

$$H(\theta) := \mathbb{P} f\left(\max_i X_i(\theta) - \min_i X_i(\theta)\right) = \int_{\mathbb{R}^n} f\left(\max_i x_i - \min_i x_i\right) g_\theta(x)\,dx$$

is increasing in θ, or that

$$H'(\theta) = \tfrac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} D_{j,k}\int_{\mathbb{R}^n} f\left(\max_i x_i - \min_i x_i\right)\frac{\partial^2 g_\theta(x)}{\partial x_j \partial x_k}\,dx$$

is nonnegative.

Split the range of integration according to which x_i is the maximum and which x_i is the minimum. On each region integration-by-parts leads to a representation

$$H'(\theta) = \tfrac{1}{2}\sum\nolimits_{j,k}\{j<k\}\left(D_{j,j} - 2D_{j,k} + D_{k,k}\right)\left(A_{j,k} + B_{j,k}\right),$$

where $A_{j,k}$ is an $n-1$-dimensional integral of the nonnegative function $f' g_\theta$ over a boundary set and $B_{j,k}$ is an n-dimension integral of the nonnegative function $f'' g_\theta$. And the coefficient $(D_{j,j} - 2D_{j,k} + D_{k,k})$ is also nonengative because it equals

$$\mathbb{P}|Y_j - Y_k|^2 - \mathbb{P}|X_j - X_k|^2 \ge 0 \qquad \text{by assumption.}$$

Done.

The Sudakov's minoration follows directly from the Fernique inequality with f chosen as the identity function. Without loss of generality suppose n equals 2^k, a power of 2, so that the index set can be identified with $\mathcal{S} := \{-1,+1\}^k$. Construct the process $\{X_{\mathfrak{s}} : \mathfrak{s} \in \mathcal{S}\}$ from a set $Z_1, \dots, Z_k$ of independendent $N(0,1)$'s,

$$X_{\mathfrak{s}} := \tfrac{1}{2}\delta k^{-1/2}\sum\nolimits_{j=1}^{k} \mathfrak{s}_j Z_j$$

for which $\mathbb{P}|X_{\mathfrak{s}} - X_{\mathfrak{s}'}|^2 = \frac{1}{4}\delta^2 k^{-1}\sum_j(\mathfrak{s}_j - \mathfrak{s}'_j)^2 \le \delta^2$.

By Fernique's inequality,

$$\mathbb{P}\left(\max_{\mathfrak{s}} Y_{\mathfrak{s}} - \min_{\mathfrak{s}} Y_{\mathfrak{s}}\right) \ge \mathbb{P}\left(\max_{\mathfrak{s}} X_{\mathfrak{s}} - \min_{\mathfrak{s}} X_{\mathfrak{s}}\right).$$

Symmetry of the multivariate normal implies that $\max_{\mathfrak{s}} Y_{\mathfrak{s}}$ has the same distribution as $\max_{\mathfrak{s}}(-Y_\alpha) = -\min_{\mathfrak{s}} Y_\alpha$, and similarly for the X's. The last inequality implies

$$\begin{aligned}\mathbb{P}\max_{\mathfrak{s}} Y_{\mathfrak{s}} &\geq \mathbb{P}\max_{\mathfrak{s}} X_{\mathfrak{s}}\\ &= \tfrac{1}{2}\delta k^{-1/2}\mathbb{P}\left(\max_{\mathfrak{s}} \sum\nolimits_{j=1}^{k} \mathfrak{s}_j Z_j\right)\\ &= \tfrac{1}{2}\delta k^{-1/2}\mathbb{P}\sum\nolimits_{j=1}^{k} |Z_j| = \tfrac{1}{2}\delta k^{1/2}\mathbb{P}|Z_1|.\end{aligned}$$

6.3 Concentration of Lipschitz functionals

Gaussian::S:Lipschitz

As promised, here are four different methods for proving versions of Theorem <4>. The aim is to show, for various choices of the constant C, that

$$\mathbb{P}\{f(X) \geq \mathbb{P}f(X) + \kappa u\} \leq \exp\left(-u^2/(2C)\right) \qquad \text{for all } u \geq 0,$$

if X has distribution γ_n and f has Lipschitz constant κ. For the first three methods the bound follows from the usual subgaussian moment generating function control,

\E@ Lip.mgf <5>

$$\mathbb{P}e^{\lambda(f(X)-\gamma_n f)} \leq e^{C\lambda^2\kappa^2/2} \qquad \text{for all } \lambda \geq 0.$$

That is,

$$\mathbb{P}\{f(X) \geq \gamma_n f + \kappa u\} \leq \inf\nolimits_{\lambda\geq 0}\mathbb{P}\exp\left(-\lambda\kappa u + C\lambda^2\kappa^2/2\right)$$

with the inimum achieved at $\lambda = u/(C\kappa)$.

As shown in Problem [1], a smoothing argument reduces to the case where f is infinitely differentiable, with $|\nabla f(x)| \leq \kappa$ everywhere. That is, if $f_i(x)$ denotes $\partial f(x)/\partial x_i$ then

\E@ gradf <6>

$$|\nabla f(x)|^2 = \sum\nolimits_{i\leq n} f_i(x)^2 \leq \kappa^2 \qquad \text{for all } x \in \mathbb{R}^n.$$

6.3.1 The Pisier-Maurey approach

Gaussian::PisierMaurey

The simplest bound for the left-hand side of <5> comes from Jensen's inequality,

$$e^{\lambda\left(f(X(\omega))-\gamma_n^y f(y)\right)} \leq \gamma_n^y e^{\lambda(f(X(\omega))-f(y))} \qquad \text{for each } \omega.$$

Equivalently, if Y is another $N(0, I_n)$-distributed random vector that is independent of X then

\E@ Lip.sym <7>

$$\mathbb{P}e^{\lambda(f(X)-\gamma_n f)} \leq \mathbb{P}e^{\lambda(f(X)-f(Y))}.$$

> **Remark.** Notice that $\text{var}(f(X) - f(Y)) = 2\text{var}(f(X))$. This "symmetrization" approach inevitably leads to at least a doubling of the constant C.

We could bound $f(X) - f(Y)$ in <7> by $\kappa|X - Y|$ but that would introduce an explicit dependence on n in the upper bound, because the distribution of $|X-Y|$ depends on n. Instead we need to exploit cancellations due to independent along a one-dimensional path from $Y = X_0$ to $X = X_1$,

$$X_\theta = X\sin(\pi\theta/2) + Y\cos(\pi\theta/2) \qquad \text{for } 0 \le \theta \le 1,$$

with derivative

$$\frac{\partial X_\theta}{\partial\theta} = \frac{\pi}{2}\left(X\cos(\pi\theta/2) - Y\sin(\pi\theta/2)\right) =: \frac{\pi}{2}Z_\theta.$$

Note that both X_θ and Z_θ have distribution γ_n, and X_θ is independent of Z_θ because $\text{cov}(X_\theta, Z_\theta) = 0$ for each θ. Moreover

$$f(X) - f(Y) = \int_0^1 \frac{\partial f(X_\theta)}{\partial\theta}\,d\theta = \int_0^1 \frac{\pi}{2}Z_\theta\cdot\nabla f(X_\theta)\,d\theta.$$

By Jensen's inequality for Lebesgue measure on $[0,1]$,

$$\exp\left(\lambda\left(f(X) - f(Y)\right)\right) \le \int_0^1 \exp\left(\frac{\lambda\pi}{2}Z_\theta\cdot\nabla f(X_\theta)\right)d\theta.$$

Take expectations with respect to $\mathbb{P}$, first conditioning on X_θ and using the fact that $\nabla f(X_\theta)$ is independent of Z_θ and $\mathbb{P}\exp(Z_\theta\cdot t) = \exp(|t|^2/2)$ for each fixed t in $\mathbb{R}^n$, to deduce that

$$\begin{aligned}\mathbb{P}\exp\left(\lambda\left(f(X) - f(Y)\right)\right) &\le \int_0^1 \mathbb{P}\exp\left(\frac{\lambda^2\pi^2}{8}|\nabla f(X_\theta)|^2\right)d\theta \\ &\le \exp\left(\lambda^2\kappa^2\pi^2/8\right) \qquad \text{by <6>.}\end{aligned}$$

We have inequality <5> with $C = \pi^2/4$.

6.3.2 The smart path method

Gaussian::smart

This refinement of the path method comes from Talagrand (2003, Section 1.3)]. It improves the constant C by creating a path through Gaussian random $2n$-vectors built from independent random $N(0, I_n)$-distributed vectors X, Y, and Z,

$$\begin{aligned} W_0 &= (Z, Z) \quad \text{AND} \quad W_1 = (X, Y) \\ W_\theta &= \alpha_\theta W_0 + \beta_\theta W_1 = (\alpha_\theta Z + \beta_\theta X, \alpha_\theta Z + \beta_\theta Y) \\ X_\theta &= \frac{\partial W_\theta}{\partial \theta} = \dot{\alpha}_\theta W_0 + \dot{\beta}_\theta W_1 \end{aligned}$$

where $\alpha_\theta = \sqrt{1-\theta}$ and $\beta_\theta = \sqrt{\theta}$ for $0 \le \theta \le 1$ and the dots denote derivatives (to avoid confusion with primes for transpose). The random vector W_θ has a $N(0, V_\theta)$ distribution with

$$V_\theta = \alpha_\theta^2 \text{var}(W_0) + \beta_\theta^2 \text{var}(W_1) = I_{2n} + (1-\theta)\begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$$

The random vector (X_θ, W_θ) also has a multivariate normal distribution with

$$2\text{cov}(X_\theta, W_\theta) = 2\alpha_\theta\dot{\alpha}_\theta V_0 + 2\beta_\theta\dot{\beta}_\theta V_1 = D := -\begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}.$$

The functional $G : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, defined by $G(x, y) = e^{\lambda(f(x)-f(y))}$, is evaluated along the W_θ path to create a function

$$H(\theta) := \mathbb{P}G(W_\theta) \qquad \text{for } 0 \le \theta \le 1$$

with $H(0) = 1$ and $H(1) = \mathbb{P}e^{\lambda(f(X)-f(Y))} \ge \mathbb{P}e^{\lambda(f(X)-\gamma_n f)}$.

Write G_i for the partial derivative of G with respect to its ith argument, that is

$$G_i(x, y) = \begin{cases} \lambda f_i(x) G(x, y) & \text{if } 1 \le i \le n \\ -\lambda f_{i-n}(y) G(x, y) & \text{if } n+1 \le i \le 2n. \end{cases}$$

Then

$$H'(\theta) = \mathbb{P}\frac{\partial G(W_\theta)}{\partial \theta} = \sum\nolimits_{i=1}^{2n} \mathbb{P}X_{\theta,i} G_i(W_\theta).$$

An integration-by-parts (Problem [2]) gives

$$\mathbb{P}X_{\theta,i} G_i(W_\theta) = \sum\nolimits_{j=1}^{2n} \tau_{i,j} \mathbb{P}G_{i,j}(W_\theta)$$

where

$$\tau_{i,j} = \text{cov}(X_{\theta,i}, W_{\theta,j}) = \tfrac{1}{2} D_{i,j} = \begin{cases} -\frac{1}{2} & \text{if } |i-j| = n \\ 0 & \text{otherwise} \end{cases}.$$

Many terms disappear in the double sum:

$$\begin{aligned}
&\sum\nolimits_{1=1}^{2n}\sum\nolimits_{j=1}^{2n}\tau_{i,j}G_{i,j}(x,y)\\
&\quad= -\tfrac{1}{2}\sum\nolimits_{1=1}^{n} 2\lambda^2(-1)f_i(x)f_i(y)G(x,y)\\
&\quad\le \lambda^2 G(x,y)\sqrt{\left(\sum_{1=1}^{n} f_i(x)^2\right)\left(\sum_{1=1}^{n} f_i(y)^2\right)} \qquad \text{by Cauchy-Schwarz}\\
&\quad\le \lambda^2 G(x,y)\kappa^2 \qquad \text{by <6>.}
\end{aligned}$$

Thus $H'(\theta) \le \lambda^2\kappa^2 \mathbb{P}G(W_\theta) = \lambda^2\kappa^2 H(\theta)$, that is,

$$\frac{d\log H(\theta)}{d\theta} \le \lambda^2\kappa^2$$

which integrates to give

$$\mathbb{P}e^{\lambda(f(X)-\gamma_n f)} \le H(1) \le e^{\lambda^2\kappa^2}.$$

Theorem <4> holds with $C = 2$.

Gaussian::stoch.calc

6.3.3 The stochastic calculus (Brownian motion) method

This proof creates a diffent sort of path, from $\gamma_n f$ to $f(X)$, using a stochastic integral with respect to an n-dimensional Brownian motion,

$$B_t = (X_{1,t}, \dots, X_{n,t}) \qquad \text{for } 0 \le t \le 1.$$

That is, the X_i processes are independent Brownian motions on $[0,1]$.

The key idea is that the process

$$M_t = \mathbb{P}_{\mathcal{F}_t} f(B_1) \qquad \text{for } 0 \le t \le 1$$

is a martingale with $M_1 = f(B_1)$ and $M_0 = \gamma_n f$. (Here $\mathbb{P}_{\mathcal{F}_t}$ denotes the conditional expectation with respect to the sigma-field $\mathcal{F}_t$ generated by $\{B_s : 0 \le s \le t\}$.) By the Markov property of Brownian motion and the independence of its increments, the martingale has an explicit representation, $M_t = F(B_t, t)$, where

\E@ F.def <8>

$$F(x,t) = \mathbb{P}f(x + (B_1 - B_t)) = \int_{\mathbb{R}^n} f(x + z\sqrt{1-t}\,)\phi_n(z)\,dz$$

and $\phi_n(z_1, \dots, z_n) = (2\pi)^{-n/2}\exp(-|z|^2/2)$, the $N(0, I_n)$ density.

Remark. The random variable M_t is only defined up to an almost sure equivalence. The representation $F(B_t, t)$ gives a version of the process with continuous sample paths.

At the risk of some notational confusion, write F_i for $\partial F/\partial x_i$ and $F_{i,j}$ for $\partial^2 F/\partial x_i \partial x_j$ and F_t for $\partial F/\partial t$. Before I present the very slick stochastic calculus proof via the Itô formula, let me give you a more heuristic argument.

The trick with stochastic integration is to carry Taylor expansions of functions of B_t out to second order, using the fact that for small $\delta > 0$

$$\Delta B_t =: (\Delta X_{1,t}, \dots, \Delta X_{n,t}) := B_{t+\delta} - B_t$$

has a $N(0, \delta I_n)$ distribution independent of $\mathcal{F}_t$. For example,

\E@ M.incr1 <9>

$$\begin{aligned}\Delta M_t &:= M_{t+\delta} - M_t \\ &\approx F_t(B_t,t)\delta + \sum\nolimits_{i=1}^n F_i(B_t,t)\Delta X_{i,t} \\ &\quad + \tfrac{1}{2}\sum\nolimits_{i=1}^n \sum\nolimits_{j=1}^n F_{i,j}(B_t,t)\Delta X_{i,t}\Delta X_{j,t}\end{aligned}$$

For $i \neq j$ the term $A_{i,j} := F_{i,j}(B_t,t)\Delta X_{i,t}\Delta X_{j,t}$ has $\mathbb{P}_{\mathcal{F}_t} A_{i,j} = 0$ and $\mathbb{P}_{\mathcal{F}_t} A_{i,j}^2 = F_{i,j}(B_t,t)\delta^2$. All those cross-product terms can be absorbed into the error of approximation. However, for $i = j$ we have $\mathbb{P}_{\mathcal{F}_t} A_{i,i} = F_{i,i}(B_t,t)\delta$ and $\mathbb{P}_{\mathcal{F}_t}\left(A_{i,i} - F_{i,i}(B_t,t)\delta\right)^2 = F_{i,i}(B_t,t)^2 O(\delta^2)$. The deviations of $A_{i,i}$ from $\mathbb{P}_{\mathcal{F}_t} A_{i,i}$ can be ignored but the conditional expectation itself makes an important contribution. That is,

$$\Delta M_t \approx \sum\nolimits_{i=1}^n F_i(B_t,t)\Delta X_{i,t} + \delta\left(F_t(B_t,t) + \tfrac{1}{2}\sum\nolimits_{i=1}^n F_{i,i}(B_t,t)\right).$$

The martingale properties of B and M now kill another contribution:

$$0 = \mathbb{P}_{\mathcal{F}_t}\Delta M_t \approx \delta\left(F_t(B_t,t) + \tfrac{1}{2}\sum\nolimits_{i=1}^n F_{i,i}(B_t,t)\right),$$

which corresponds to the fact (Problem [3]) that F is a solution to the heat equation,

$$\frac{\partial F(x,t)}{\partial t} + \tfrac{1}{2}\sum\nolimits_i \frac{\partial^2 F(x,t)}{\partial x_i^2} = 0 \qquad \text{for } 0 < t < 1,$$

with boundary conditions $F(x,1) = f(x)$ and $F(x,0) = \int f(x+z)\phi(z)\,dz$. The approximation simplifies even more,

\E@ M.incr2 <10>

$$\Delta M_t \approx \sum\nolimits_{i=1}^n F_i(B_t,t)\Delta X_{i,t}.$$

Similar reasoning gives

\E@ M.qv <11>
$$\mathbb{P}_{\mathcal{F}_t}(\Delta M_t)^2 \approx \sum\nolimits_i F_i(B_t,t)^2 \mathbb{P}_{\mathcal{F}_t}\Delta X_{i,t}^2 = \delta|\nabla_x F(B_t,t)|^2,$$

which suggests that the process

$$Z_t := \exp\left(\lambda M_t - \tfrac{1}{2}\lambda^2\int_0^t |\nabla_x F(B_s,s)|^2 ds\right)$$

should also be a martingale:

$$\begin{aligned}\mathbb{P}_{\mathcal{F}_t} Z_{t+\delta} &\approx Z_t \mathbb{P}_{\mathcal{F}_t}\exp\left(\lambda\Delta M_t - \tfrac{1}{2}\lambda^2\delta|\nabla_x F(B_t,t)|^2\right)\\ &\approx Z_t\mathbb{P}_{\mathcal{F}_t}\left(1+\lambda\Delta M_t - \tfrac{1}{2}\lambda^2\delta|\nabla_x F(B_t,t)|^2 + \tfrac{1}{2}(\lambda\Delta M_t)^2\right)\\ &\approx Z_t.\end{aligned}$$

Here the higher-order terms in $\delta|\nabla_x F(B_t,t)|^2$, which are of order δ^2, have been absorbed into the error of approximation.

Assuming (correctly—see below) the validity of this martingale assertion we now have

\E@ Zmg.equality <12>
$$\mathbb{P}e^{\lambda M_0} = \mathbb{P}Z_0 = \mathbb{P}Z_1 = \mathbb{P}\exp\left(\lambda f(B_1) - \tfrac{1}{2}\lambda^2\int_0^1 |\nabla_x F(B_s,s)|^2 ds\right).$$

Finally we come to place where the Lipschitz property of f plays a role. The gradient of F inherits that that property:

$$\begin{aligned}|\nabla_x F(x,t)|^2 &= \left|\int \nabla_x f(x+z\sqrt{1-t}\,)\phi_n(z)\,dz\right|^2\\ &\le \int |\nabla_x f(x+z\sqrt{1-t}\,)|^2\phi_n(z)\,dz \le \kappa^2.\end{aligned}$$

Thus inequality <12> tells us that

$$\mathbb{P}e^{\lambda f(B_1)} \le e^{\lambda^2\kappa^2/2}\mathbb{P}Z_1 \le e^{\lambda\gamma_n f + \lambda^2\kappa^2/2},$$

which is inequality <5> with $C = 1$.

I would be sympathetic if you had reservations about all these approximations. A rigorous derivation uses the versatile theorems of stochastic calculus, as expounded by Chung and Williams (2014, Section 5.4) = C&W. The argument is very clean. The process M is an $\mathcal{L}^2$ martingale with continuous sample paths. By the Itô formula,

$$\begin{aligned}M_t - M_0 &= F(B_t,t) - F(B_0,0)\\ &= \sum\nolimits_i \int_0^t F_i(B_s,s)\,dX_{i,s}\\ &\quad + \int_0^t F_t(B_s,s)\,ds + \tfrac{1}{2}\sum\nolimits_{i=1}^n\sum\nolimits_{j=1}^n \int_0^t F_{i,j}(B_s,s)\,d[X_i,X_j]_s,\end{aligned}$$

an $\mathcal{L}^2$ martingale plus a process with sample paths of bounded variation. The process $M_t - M_0 - \sum_i \int_0^t F_i(B_s,s)\,dX_{i,s}$ is an $\mathcal{L}^2$ martingale whose sample paths are both continuous and of bounded variation, which forces it to be the zero process [C&W Corollary 4.5]. That is, even without the benefit of Problem [3] we know that the bounded variation contribution is zero, leaving

$$M_t = M_0 + \sum\nolimits_i \int_0^t F_i(B_s,s)\,dX_{i,s}.$$

The quadratic variation process [cf. C&W Theorem 5.7] is given by

$$\begin{aligned}[M]_t &= \sum\nolimits_{i,j} \int_0^t F_i(B_s,s)F_j(B_s,s)\,d[X_i,X_j]_s \\ &= \sum\nolimits_i \int_0^t F_i(B_s,s)^2 ds \le \kappa^2 t.\end{aligned}$$

The process $Z_t = \exp(\lambda M_t - \frac{1}{2}\lambda^2[M]_t)$ is a local martingale, that is, for some sequence of stopping times $\tau_j \uparrow \infty$, the process $Z(t\wedge\tau_j)$ is a martingale [C&W Theorem 6.2]. For each j,

$$\begin{aligned}\mathbb{P}Z_0 = \mathbb{P}Z(1\wedge\tau_j) &= \mathbb{P}\exp\left(\lambda M_{1\wedge\tau_j} - \tfrac{1}{2}\lambda^2[M]_{1\wedge\tau_j}\right) \\ &\ge \mathbb{P}\exp\left(\lambda M_{1\wedge\tau_j} - \tfrac{1}{2}\lambda^2\kappa^2\right).\end{aligned}$$

Complete the argument by an appeal to Fatou's Lemma as $j\to\infty$,

$$\mathbb{P}e^{\lambda f(B_1)} = \mathbb{P}e^{\lambda M_1} \le e^{\lambda\gamma_n f + \kappa^2\lambda^2/2},$$

which again is inequality <5> with $C = 1$.

6.3.4 The Gaussian isoperimetric inequality

Gaussian::iso

For each subset A of $\mathbb{R}^n$ define $d(z,A) = \inf\{|z-y| : y\in A\}$ and $A^\delta = \{z\in\mathbb{R}^n : d(a,A)\le\delta\}$. Write Φ for the $N(0,1)$ distribution function and $\bar\Phi$ for $1-\Phi$. That is, if Z is $N(0,1)$ distributed then $\mathbb{P}\{Z\le x\} = \Phi(x)$ and $\mathbb{P}\{Z > x\} = \bar\Phi(x)$.

The most stunning fact about γ_n—the so-called isoperimetric inequality—was established independently by Borell (1975) and Sudakov and Tsirel'son (1978).

Gaussian::gisop <13> **Gaussian isoperimetric inequality.** *If A is a Borel subset of $\mathbb{R}^n$ with $\gamma_n A = \Phi(\alpha)$ then $\gamma_n A^\delta \le \bar\Phi(\alpha+\delta)$ for each $\delta\ge 0$. The upper bound is achieved when A is any closed halfspace with Gaussian measure $\Phi(\alpha)$.*

For an exposition of a proof due to Ehrhard (1983a,b) see Pollard (2001, Section 12.5). The inequality can be rewritten more compactly as

$$\gamma_k(A^\delta) \geq \Phi\left(\Phi^{-1}(\gamma_k A) + \delta\right),$$

slightly disguising the fact that equality is achieved by halfspaces but leading towards a functional form of the inequality that was developed by Bobkov (1996, 1997). For elegant reformulations of the functional approach see Ledoux (1998) and Barthe and Maurey (2000).

It is the reduction from an n-dimensional problem, with n arbitrarily large, to a one-dimensional calculation for the lower bound that makes the isoperimetric inequality so powerful, as shown by the inequalities in the next Example.

Recall that a median of a (real valued) random variable X is any constant m for which $\mathbb{P}\{X \geq m\} \geq 1/2$ and $\mathbb{P}\{X \leq m\} \geq 1/2$. Such an m always exists, but it need not be unique.

Gaussian::gauss.conc <14> **Example.** Suppose f is a Lipschitz function on $\mathbb{R}^n$ with $\|f\|_{\text{Lip}} \leq \kappa$. Under γ_n, the random variable $f(z)$ has at least one median, a number M for which

$$\gamma_n\{f(z) \leq M\} \geq \tfrac{1}{2} \qquad \text{AND} \qquad \gamma_n\{f(z) \geq M\} \geq \tfrac{1}{2}.$$

Define $A = \{z \in \mathbb{R}^n : f(z) \leq M\}$ so that $\gamma_n A \geq 1/2 = \Phi(0)$. If $d(x, A) < u$ then there exist a point $z \in A$ with $d(z, x) < u$. From the Lipschitz property and the fact that $f(x) \leq M$ we then get

$$f(x) < f(z) + \kappa u < M + \kappa u$$

Conversely, if $f(x) \geq M + \kappa u$ then $d(x, A) \geq u$. It follows that

$$\gamma_n\{f(x) \geq M + \kappa u\} \leq \gamma_n\{d(x, A) \geq u\} \leq \bar{\Phi}(0 + u) \leq \tfrac{1}{2}\exp(-u^2/2).$$

The companion lower bound follows by analogous argument for deviations from the set $\{z \in \mathbb{R}^n : f(z) \geq M\}$. Together the two bounds give a concentration property for f,

\E@ med.conc <15>
$$\gamma_n\{z : |f(z) - M| \geq \kappa y\} \leq \exp(-y^2/2),$$

where M is a median for f under γ_n.

For many purposes it is more convenient to center the functional at its expected value $\mu = \gamma_n f$. Inequality <15> implies

$$|\mu - M| \leq \gamma_n|f - M| \leq \kappa \int_0^\infty \gamma_n\{|f(z) - M| \geq \kappa y\}\, dy = C\kappa,$$

where $C = 1/(2\sqrt{2\pi})$. Thus

$$\gamma_n\{|f-\mu| \ge \kappa(C+y)\} \le \gamma_n\{|f-M| \ge \kappa y\} \le \exp(-y^2/2),$$

which implies a concentration inequality around μ.

□

6.4 Problems

Gaussian::S:Ga.problems

Gaussian::P:smooth.Lip

[1] Suppose f is a real valued function on $\mathbb{R}^n$ with $\|f\|_{\text{Lip}} = \kappa$. Let ψ be an infinitely differentiable, nonegative function on $\mathbb{R}^n$ with compact support and $\int \psi(z)\,dz = 1$. For each $\sigma > 0$, define $f_\sigma(x) := \int f(x+\sigma z)\psi(z)\,dz$.

(i) Show that f_σ has continuous partial derivatives of all orders and

$$|f_\sigma(x) - f_\sigma(y)| \le \int |f(x+\sigma z) - f(y+\sigma z)|\psi(z)\,dz \le \kappa|x-y|$$

That is, $\|f_\sigma\|_{\text{Lip}} \le \kappa$.

(ii) Use the inequality

$$|f(x+z) - f(x) - z\cdot\nabla f(x)| \le \left|\int_0^1 z\cdot\left[\nabla f(x+\theta z) - \nabla f(x)\right]\,d\theta\right|$$

and the continuity of ∇f as $z \to 0$ to deduce that $|\nabla f(x)| \le \kappa$ everywhere.

(iii) Show that

$$\sup_x |f_\sigma(x) - f(x)| \le \sup_x \int |f(x+\sigma z) - f(x)|\psi(z)\,dz \le \kappa\sigma\int |z|\psi(z)\,dz.$$

Deduce that f_σ converges uniformly to f as σ tends to zero.

(iv) Suppose $\mathbb{P}\{f_\sigma(X) - \mathbb{P}f_\sigma(X) \ge t\} \le H(t)$ for all $t > 0$, uniformly in σ, where H is a continuous function. Deduce that $\mathbb{P}\{f(X) - \mathbb{P}f(X) \ge t\} \le H(t)$ for all $t > 0$. Hint: Choose σ small enough that $\sup_x |f_\sigma(x) - f(x)| \le \delta$.

Gaussian::P:gauss.ibp

[2] Suppose $Z \sim \gamma_1 = N(0,1)$ and $(X, W_1, \dots, W_m)$ has multivariate normal distribution with $\mathbb{P}X = 0$. Suppose $G : \mathbb{R}^m \to \mathbb{R}$ has partial derivatives G_i for which $\mathbb{P}|G_i(W_1,\dots,W_m)| < \infty$. Show that

$$\mathbb{P}XG(W_1,\dots,W_m) = \sum\nolimits_{i\le m} \tau_i \mathbb{P}G_i(W_1,\dots,W_m) \qquad \text{where } \tau_i := \text{cov}(X, W_i)$$

by following these steps.

(i) For each absolutely continuous $g : \mathbb{R} \to \mathbb{R}$ for which $\mathbb{P}|g'(Z)| < \infty$ show that $\mathbb{P}Zg(Z) = \mathbb{P}g'(Z)$. Hint: The function $\phi(u)g(u)$ has almost sure derivative $-u\phi(u)g(u) + \phi(u)g'(u)$.

(ii) Without loss of generality suppose $X \sim N(0,1)$. Define $W_i := Y_i - \tau_i X$ for $i = 1,\dots,m$. By calculating covariances show that W is independent of X. Invoke part (i) with $g(x) = G(W + x\tau)$ for a fixed realization of W to deduce that

$$\gamma_1^x xG(W + x\tau) = \sum\nolimits_{i=1}^m \tau_i \gamma_1^x G_i(W + x\tau),$$

then take expected values over W.

Gaussian::P:heat

[3] For the function F defined by equation <8>, show that

$$\begin{aligned} \frac{\partial F(x,t)}{\partial t} &= \int_{\mathbb{R}^n} \frac{\partial f(x + z\sqrt{1-t}\,)\phi_n(z)}{\partial t}\,dz \\ &= -\tfrac{1}{2}(1-t)^{-1/2}\sum\nolimits_i \int f_i(x + z\sqrt{1-t}\,)z_i\phi_n(z)\,dz \\ &= -\tfrac{1}{2}\sum\nolimits_i \int f_{i,i}(x + z\sqrt{1-t}\,)\phi_n(z)\,dz \\ &= -\tfrac{1}{2}\sum\nolimits_i \frac{\partial^2 F(x,t)}{\partial^2 x_i}. \end{aligned}$$

6.5 Notes

Gaussian::S:Notes

For a proof of Sudakov's inequality <1> (also known as the Sudakov minoration) see Ledoux and Talagrand (1991, page 79). They used the Slepian-Gordon inequalities, whose proof they borrowed from Kahane (1986). See also their Notes on pages 87–88 for more about the history and where credit is due. The Notes to Dudley (1999, Chapter 2) indicate that credit is also due to Chevet (1970).

The method in subsection 6.3.1 is a special case of Theorem 2.2 of Pisier (1985, page 176), who commented that "The proof below is a simplification, due to Maurey, of my original proof which used an expansion in Hermite polynomials". He aslo (page 180) sketched a stochastic calculus proof of the sharper result for $C = 1$, with the comment "B. Maurey found a proof of theorem 2.1 with the best constant $K = 1/2$ [that is, $C = 1$]. His proof uses stochastic integrals and apparently does not extend to the setting of theorem 2.2." Ledoux (2001, page 45) attributed the stochastic calculus proof of the to Cirel'son et al. (1976), who in turn attributed the result to (the 1974 Russian version of) Sudakov and Tsirel'son (1978). See Adler (1990, page 43) for a most readable exposition.

See the concise and informative book by Ledoux (2001) for more about concentration inequalities.

References

Adler90gauss — Adler, R. J. (1990). *An introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Volume 12 of *Lecture Notes–Monograph series*. Hayward, CA: Institute of Mathematical Statistics.

BartheMaurey2000AIHP — Barthe, F. and B. Maurey (2000). Some remarks on isoperimetry of Gaussian type. *Annales de l'Institut Henri Poincaré, Probability and Statistics 36*(4), 419–434.

Bobkov1996JFA — Bobkov, S. (1996). A functional form of the isoperimetric inequality for the Gaussian measure. *Journal of Functional Analysis 135*, 39–49.

Bobkov1997AnnProb — Bobkov, S. G. (1997). An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Annals of Probability 25*(1), 206–214.

Borell:75 — Borell, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Inventiones Mathematicae 30*, 207–216.

Chevet1970ASUCF — Chevet, S. (1970). Mesures de Radon sur R^n et mesures cylindriques. *Annals scientifiques de l'Université de Clermont-Ferrand 2, série Mathématiques 43*(6), 91–158.

ChungWilliams2014book — Chung, K. L. and R. J. Williams (2014). *Introduction to Stochastic Integration* (Second ed.). Boston: Birkhäuser.

nIbragimovSudakov1976norms — Cirel'son, B., I. Ibragimov, and V. Sudakov (1976). Norms of Gaussian sample functions. In *Proceedings of the Third Japan—USSR Symposium on Probability Theory*, Volume 550 of *Springer Lecture Notes in Mathematics*, pp. 20–41. Springer.

Dudley1999UCLT — Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

Ehrhard83MathScand — Ehrhard, A. (1983a). Symétrisation dans l'espace de Gauss. *Mathematica Scandinavica 53*, 281–301.

Ehrhard83slnm Ehrhard, A. (1983b). Un principe de symétrisation dans les espaces de Gauss. *Springer Lecture Notes in Mathematics 990*, 92–101.

Fernique75StFlour Fernique, X. (1975). Regularité des trajectoires des fonctions aléatoires gaussiennes. *Springer Lecture Notes in Mathematics 480*, 1–97. Ecole d'Eté de Probabilités de Saint-Flour IV—1974.

Kahane1986IsraelJM Kahane, J.-P. (1986). Une inégalité du type de slepian et gordon sur les processus gaussiens. *Israel Journal of Mathematics 55*(1), 109–110.

Ledoux1998PP Ledoux, M. (1998). A short proof of the Gaussian isoperimetric inequality. *Progress in Probability 43*, 229–232.

Ledoux01conc Ledoux, M. (2001). *The Concentration of Measure Phenomenon*, Volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society.

LedouxTalagrand91book Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. New York: Springer.

Pisier85Varenna Pisier, G. (1985). Probabilistic methods in the geometry of Banach spaces. *Springer Lecture Notes in Mathematics 1206*, 167–241.

PollardUGMTP Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

SudakovTsirelson1978JSM Sudakov, V. N. and B. S. Tsirel'son (1978). Extremal properties of half-spaces for spherically invariant measures. *Journal of Soviet Mathematics 9*, 419–434. (Translated from Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo instituta im. V. A. Steklova AN SSSR, Vo!. 41, pp. 14-24, 1974.).

Talagrand03spin Talagrand, M. (2003). *Spin Glasses: a Challenge to Mathematicians*, Volume 46 of *Ergbnisse der Mathematik und ihrer Grenzgebiete*. New York: Springer-Verlag.