CORRECTION

The following 7 pages replace pages 42–47 on my lecture notes *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, (1990), IMS Lecture Notes-Monograph Series, **12**.

They "enlarge" on the comment on the third line from the bottom of page 46, which is concerned with the proof of Borell's inequality, where I said "To complete the proof note simply that...". The word "simply" is simply out of place, and, in fact, Lemma 2.2 is false as stated in the Notes. All the details are in the corrected version following[1].

This is also a good place to point out that Borell's inequality is due not only to Borell, but, as Michel Talagrand pointed out to me, was also established independently in

Cirelśon, B. S., Ibragimov, I. A. and Sudakov, V. N. Norms of Gaussian sample functions. *Proceedings of the Third Japan-USSR Symposium on Probability Theory (Tashkent, 1975)*, *Lecture Notes in Math.*, **550**, Springer, Berlin, (1976), 20–41.

In fact, their proof is more or less the one I used, building on the ideas of Pisier, who also was unaware of this paper.

ADVERTISEMENT

A fully corrected version of the entire Lecture Notes, along with a totally rewritten edition of my 1981 *Geometry of Random Fields*, is currently under preparation, and should appear in 2001 as *Random Fields and their Geometry*, as a Birkhaüser publication.

[1]The first person to notice this problem was a Stanford student of David Siegmund, and the corrected version is basically due to Amir Dembo.

II. TWO BASIC RESULTS

Despite the title of this Chapter, there are probably *three* basic results in the theory of Gaussian processes, that make this theory both manageable and special. The first is the existence theorem that to any positive semi-definite function R there corresponds a centered Gaussian process with covariance function R; an important, but not particularly exciting result.

The second is that the supremum of a Gaussian process behaves much like a single Gaussian variable with variance equal to the largest variance achieved by the entire process. In the way that we shall present it, this is Borell's inequality, and is the key to all results about Gaussian continuity, boundedness, and suprema.

The third is that if two centered processes have identical variances (i.e. $EX_t^2 = EY_t^2$ for all $t \in T$), but one process is more "correlated" than the other (i.e. if $EX_tX_s \geq EY_tY_s$ for all $s, t \in T$) then the more correlated process has the stochastically smaller maximum, in the sense that $P\{\sup X_t > \lambda\} \leq P\{\sup Y_t > \lambda\}$ for all $\lambda > 0$. This is Slepian's inequality, and without this result many of the most basic results in the theory of Gaussian processes would have no proof.

Both Borell's and Slepian's inequality are very special in that analagous results for non-Gaussian processes are extremely rare. (We shall see some exceptions to this rule later). The fact that even for Gaussian processes the $\sup X_t$ in Slepian's inequality cannot be replaced by as simple a variant as $\sup |X_t|$ is also indicative how very lucky we are that a result of this kind holds at all.

1. Borell's Inequality.

Let X be a centered Gaussian random variable with variance σ^2. Then choosing

$$\Psi(\lambda) \;=\; (2\pi)^{-\frac{1}{2}} \int_{\lambda}^{\infty} e^{-\frac{1}{2}x^2}\,dx,$$

to denote the standard Gaussian distribution function, straightforward approximations give that for all $\lambda > 0$

$$(2.1)\qquad \begin{aligned} (1-\sigma^2\lambda^{-2})(\sigma/\sqrt{2\pi})\lambda^{-1}e^{-\frac{1}{2}\lambda^2/\sigma^2} \;&\leq P\{X > \lambda\} \\ &= \Psi(\lambda/\sigma) \\ &\leq (\sigma/\sqrt{2\pi})\lambda^{-1}e^{-\frac{1}{2}\lambda^2/\sigma^2} \end{aligned}$$

One immediate consequence of (2.1) is that

$$(2.2)\qquad \lim_{\lambda\to\infty} \lambda^{-2}\,\log P\{X > \lambda\} \;=\; -(2\sigma^2)^{-1}.$$

There is a classical result of Landau and Shepp (1970) and Marcus and Shepp (1971) that gives a result closely related to (2.2), but for the supremum of a general centered Gaussian process. If we assume that $\{X_t\}_{t\in T}$ has bounded sample paths with probability one, then they showed that

$$\lim_{\lambda\to\infty} \lambda^{-2} \log P\{\sup_{t\in T} X_t > \lambda\} \;=\; -(2\sigma_T^2)^{-1}, \tag{2.3}$$

where

$$\sigma_T^2 \;:=\; \sup_{t\in T} EX_t^2$$

is a notation that will remain with us for the remainder of these notes. An immediate consequence of (2.3) is that for all $\epsilon > 0$ and large enough λ

$$P\{\sup_{t\in T} X_t > \lambda\} \;\le\; e^{\epsilon\lambda^2 - \frac{1}{2}\lambda^2/\sigma_T^2}. \tag{2.4}$$

Since $\epsilon > 0$ is arbitrary, comparing (2.4) and (2.1) we reach the conclusion described above that the supremum of a centered, bounded Gaussian process behaves much like a single Gaussian variable with a suitably chosen variance.

In Chapter 5 we shall investigate (2.4) in considerable detail, and show how to close the gap between (2.4) and (2.1) (i.e. between λ^{-1} and $e^{\epsilon\lambda^2}$).

Most proofs of results like (2.3) rely on geometrical arguments and the so-called Brunn-Minkowski inequality for Gauss space (k-dimensional Euclidean space with a k-dimensional Gaussian measure). The strongest form is due to Borell (1975) in a highly abstract setting and with a difficult proof. Maurey and Pisier (Pisier (1986)) recently found a very short proof of a version of Borell's inequality, which avoids the need to appeal to areas outside of probability theory. This is, in essence, the proof that we shall give. It has the advantage of being more self-contained for a probabilistic audience, and the disadvantage that it cannot reach all the cases that proofs based on isoperimetric inequalities can. Nevertheless, it is my favourite application of Itô's formula, for who would have expected to be able to use stochastic analysis to prove results in Gaussian processes? (By the way – the stochastic analysis/Gaussian process interface is now a two way street. See Chapter 6 for details on this.) The result is:

2.1 Theorem. *Let $\{X_t\}_{t\in T}$ be a centered Gaussian process with sample paths bounded a.s. Let $\|X\| = \sup_{t\in T} X_t$. Then $E\|X\| < \infty$, and for all $\lambda > 0$*

$$P\{|\,\|X\| - E\|X\|\,| \;>\; \lambda\} \;\le\; 2e^{-\frac{1}{2}\lambda^2/\sigma_T^2}. \tag{2.5}$$

An immediate consequence of (2.5) is that for all $\lambda > E\|X\|$,

$$P\{\|X\| \;>\; \lambda\} \;\le\; 2e^{-\frac{1}{2}(\lambda - E\|X\|)^2/\sigma_T^2}. \tag{2.6}$$

Thus (2.3) and (2.4) are easily seen to be consequences of Borell's inequality.

Indeed, a far stronger result is true, for (2.4) can be replaced by

$$P\{\sup_{t\in T} X_t > \lambda\} \;\le\; e^{C\lambda - \frac{1}{2}\lambda^2/\sigma_T^2}, \tag{2.7}$$

where C is a constant depending on $E\|X\|$.

Of course, the sharper forms (2.5) and (2.6) will only be useful if we can manage to calculate $E\|X\|$. This, in fact, is one of the main tasks facing us, and we shall see that this single expectation is the key to a Pandora's box of other results.

Theorem 2.1 is true in much more generality than we have indicated, and can also be formulated somewhat differently. Borell's original result, for example, used the median of $\|X\|$ instead of the mean $E\|X\|$ in (2.5). In this formulation, the process X can be allowed to take values in a quite general Banach space, and $\|\ \|$ is then the norm of the Banach space. (In fact, since the passage from the inequality with the median to that with the mean, or *vice versa*, is far from immediate, it is really not quite precise to refer to (2.5) as "Borell's" inequality. Nevertheless, we shall not let a minor point like this change our nomenclature.)

Similar results, involving Banach space valued processes, using both the natural norm and its expectation, are also available, but with a constant other than $\frac{1}{2}$ in the exponent in (2.5). (For details see Pisier (1986, 1989).)

Throughout these notes you should always remember that $\|\ \| \equiv \sup$ is not a true norm, and that very often one needs bounds on the tail of $\sup_t |X_t|$ rather than $\|X\| = \sup_t X_t$. However, a symmetry argument immediately gives one that

$$P\big\{\sup_t |X_t| > \lambda\big\} \;\le\; 2\,P\big\{\sup_t X_t > \lambda\big\},$$

so that Borell's inequality helps out here as well.

For more on the relation between stochastic analysis and isoperimetric inequalities, see, for example, Ledoux (1988) and Pisier (1986, 1989). ■

The following lemma forms the main step in the proof of Borell's inequality, and is also of considerable independent interest. (As usual, we shall also denote the usual Euclidean norm by $\|\ \|$, hopefully without creating too much confusion.)

2.2 LEMMA. *Let X be a k-dimensional vector of centered, unit variance, independent, Gaussian variables. If $f : \Re^k \to \Re$ is Lifshitz, with Lifshitz constant σ – i.e. $|f(x) - f(y)| \le \sigma\|x - y\|$ for all $x, y \in \Re^k$ – then for all $\lambda > 0$*

$$P\{|f(X) - Ef(X)| > \lambda\} \;\le\; 2e^{-\frac{1}{2}\lambda^2/\sigma^2}, \tag{2.8}$$

PROOF: To start, assume that f has derivatives of up to second order, which certainly implies that it is Lifshitz, as required by the Lemma..

The main part of the proof is a little out of place in a book about Gaussian processes, since we are going to need two excursions into stochastic analysis, which seems, at first sight, to have nothing to do with the problem at hand. However, if $\{B_t\}_{t\geq 0} = \{B_t^1, ..., B_t^k\}_{t\geq 0}$ is a k-dimensional Brownian motion – i.e. the B^i are i.i.d. standard, real-valued Brownian motions – then B_1 and X are identically distributed.

The first excursion starts by noting that if $G\colon \Re^k \to \Re^k$ is continuous and coordinatewise bounded, and if $\langle\ ,\ \rangle$ is the usual Euclidean inner product, then we have that

$$\exp\Big\{ \int_0^t \langle G(B_s),\, dB_s\rangle \;-\; \tfrac{1}{2}\int_0^t \|G(B_s)\|^2\, dt\Big\}$$

is an (exponential) martingale with initial value, and so constant mean, 1. (For information on exponential martingales, or, indeed, on any of the stochastic analysis arguments that follow, Karatzas and Shreve (1988) is a very accessible reference. In this case, the requisite result is on page 199.) Taking expectations, and setting

$$\alpha \;=\; \sup_{x\in\Re^d} \|G(x)\|,$$

we obtain that for all real θ

$$E\Big\{\exp\Big(\theta\int_0^1 \langle G(B_s),\, dB_s\rangle\Big)\Big\} \;\leq\; e^{\frac{1}{2}\theta^2\alpha^2}.$$

A standard Chebycheff type argument gives us that

$$\begin{aligned} &P\Big\{\Big|\int_0^1 \langle G(B_s),\, dB_s\rangle\Big| \;>\; \lambda\Big\} \\ &\qquad = \; P\Big\{\int_0^1 \langle G(B_s),\, dB_s\rangle \;>\; \lambda\Big\} + P\Big\{\int_0^1 \langle G(B_s),\, dB_s\rangle \;<\; -\lambda\Big\} \\ &\qquad \leq \; 2e^{-\theta\lambda}\, E\Big\{\exp\Big(\theta\int_0^1 \langle G(B_s),\, dB_s\rangle\Big)\Big\} \qquad (2.9) \\ &\qquad \leq \; 2e^{-\theta\lambda}\, e^{\frac{1}{2}\theta^2\alpha^2} \\ &\qquad = \; 2e^{-\frac{1}{2}\lambda^2/\alpha^2}, \end{aligned}$$

the factor of two in the first inequality coming from symmetry considerations and the last inequality being a consequence of setting $\theta = \lambda/\alpha^2$.

Our second excursion involves Itô's formula for real valued functions of vector valued Brownian motion. (Karatzas and Shreve (1988), page 153.) The form we shall need states that for a sufficiently smooth $F = F(x,t)\colon \Re^k \times \Re_+ \to \Re$,

$$\begin{aligned} F(B_t,t) - F(B_s,s) \;&=\; \int_s^t \langle \nabla_x F(B_u,u), dB_u\rangle \\ &\quad + \int_s^t \big(\tfrac{1}{2}\Delta_{xx}F(B_u,u) + F_t(B_u,u)\big)\, du, \qquad (2.10) \end{aligned}$$

where ∇_x and Δ_{xx} denote derivatives of $F(x,t)$ with respect to x, and $F_t(x,t) = \partial F(x,t)/\partial t$. We shall also need $\{P_t\}_{t\geq 0}$, the Markov semi-group associated with B, determined by the fact that for smooth $g\colon \Re^k \to \Re$

$$\begin{aligned}(P_t g)(x) &= E^x g(B_t) \\ &= (2\pi t)^{-k/2} \int_{\Re^k} g(y) e^{-\frac{1}{2}\|x-y\|^2/t}\, dy,\end{aligned}$$

where E^x denotes expectation with respect to the Brownian motion B starting at the point $x \in \Re^k$ at time zero.

We can now put the above two parts together to prove our Lemma.

Setting $F(x,t) = (P_{1-t}f)(x)$, the conditions of the lemma imply that F is sufficiently smooth for Itô's formula to hold. With $t = 1,\ s = 0$, (2.10) yields

$$f(B_1) - Ef(B_1) = \int_0^1 \langle \nabla(P_{1-u}f)(B_u), dB_u\rangle. \tag{2.11}$$

The last expression in the Itô formula disappears due to the specific form of the semi-group P_t. If you are not familiar with this (it is the heat equation that makes everything work) you should do the algebra to convince yourself that everything works as claimed.

Since P_t is a *contraction* semi-group, the fact that $|f(x)-f(y)| < \sigma\|x-y\|$ immediately implies that $P_t f$ satisfies the same inequality for every $t \geq 0$, and so $\|\nabla P_t f(x)\| \leq \sigma$ for almost every x. It then easily follows from (2.9) that

$$P\{|f(B_1) - Ef(B_1)| > \lambda\} \leq 2e^{-\frac{1}{2}\lambda^2/\sigma^2}.$$

To finish the proof, we need to lift the differentiability we imposed on f. Any standard approximation procedure (such as convolution with C^∞ functions) will work, and so this part of the proof is left to you. ∎

Proof of Theorem 2.1: We have two things to prove. Firstly, Theorem 2.1 will follow immediately from Lemma 2.2 in the case of finite T and X having i.i.d. components once we show that sup(.), or max(.) in this case, was Lifshitz. We shall show this, and lift the i.i.d. restriction, in one step. The second part of the proof involves lifting the result from finite to general T.

To start, let $V = E\{X' \cdot X\}$ be the covariance matrix of X so that, in this case, we have

$$\sigma_T^2 = \sup_{1\leq i\leq k} V(i,i) = \sup_{1\leq i\leq k} EX_i^2.$$

Let A be such that $A' \cdot A = V$, so that $X \stackrel{\mathcal{L}}{=} AB_1$, and $\max_i X_i \stackrel{\mathcal{L}}{=} \max_i (AB_1)_i$.

Now let $e_i \in \Re^k$ denote the vector with 1 in position i and zeroes elsewhere, and consider the function $f(x) = \max_i (Ax)_i$.

$$\begin{aligned} \big|\max_i(Ax)_i - \max_i(Ay)_i\big| &= \big|\max_i(e_iAx) - \max_i(e_iAy)\big| \\ &\le \max_i \big|e_iA(x-y)\big| \\ &\le \max_i \big|e_iA\big| \cdot \|x-y\|, \end{aligned}$$

where the first inequality is elementary and the second is Cauchy-Schwartz.

But

$$|e_iA|^2 = e_i'A'Ae_i = e_i'Ve_i = V(i,i),$$

so that

$$\big|\max_i(Ax)_i - \max_i(Ay)_i\big| \le \sigma_T\|x-y\|,$$

which, in view of Lemma 2.2 and the equivalence in law of $\max_i X_i$ and $\max_i(AB_1)_i$, establishes the Theorem for finite T.

We now turn to lifting the result from finite to general T. This is, almost, an easy exercise in approximation. For each $n > 0$ let T_n be a finite subset of T such that $T_n \subset T_{n+1}$ and T_n increases to a dense subset of T. By separability,

$$\sup_{t\in T_n} X_t \overset{\text{a.s.}}{\to} \sup_{t\in T} X_t,$$

and, since the convergence is monotone, we also have that

$$E\sup_{t\in T_n} X_t \to E\sup_{t\in T} X_t.$$

Since $\sigma^2_{T_n} \to \sigma^2_T < \infty$, (again monotonely) this would be enough to prove the general version of Borell's inequality from the finite T version if only we knew that the one worrisome term, $E\sup_T X_t$, were definitely finite, as claimed in the statement of the Theorem. Thus if we now that the assumed a.s. finiteness of $\|X\|$ implies also the finiteness of its mean, we shall have a complete proof to both parts of the Theorem.

We proceed by contradiction. Thus, assume $E\|X\| = \infty$, and choose $\lambda_o > 0$ such that

$$e^{-\lambda_o^2/\sigma_T^2} \le \tfrac{1}{4}, \qquad P\big\{\sup_{t\in T} X_t < \lambda_o\big\} \ge \tfrac{3}{4}.$$

Now choose $n \ge 1$ such that $E\|X\|_{T_n} > 2\lambda_o$, possible since $E\|X\|_{T_n} \to E\|X\|_T = \infty$. Borell's inequality on the finite space T_n then gives

$$\begin{aligned} \tfrac{1}{2} \ge 2\,e^{-\lambda_o^2/\sigma_T^2} &\ge 2\,e^{-\lambda_o^2/\sigma_{T_n}^2} \\ &\ge P\big\{\big|\,\|X\|_{T_n} - E\|X\|_{T_n}\big| > \lambda_o\big\} \\ &\ge P\big\{E\|X\|_{T_n} - \|X\|_T > \lambda_o\big\} \\ &\ge P\big\{\|X\|_T < \lambda_o\big\} \\ &\ge \tfrac{3}{4}. \end{aligned}$$

This provides the required contradiction, and so we are done. ■

I cannot overemphasise how important a result Borell's inequality is. For example, an almost immediate consequence of Borell's inequality is that the a.s. finiteness of $\|X\|$ implies that it also has all regular, and some exponential, moments. (c.f. Theorem 3.2.) In later chapters, especially Chapter 5, we shall see how one can apply Borell's inequality a number of times, with almost no other tools, to obtain even sharper bounds on tail probabilities for suprema.

Now, however, we turn to the second, and equally central, result about Gaussian processes.