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 $\lambda > 0$

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

One immediate consequence of (2.1) is that

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

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

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

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 $f: \Re^k \to \Re$ have derivatives of up to second order, bounded pointwise by $Ae^{B\|x\|}$ for some $A, B < \infty$, and let X be a k-dimensional, centered, Gaussian variable with covariance matrix V. If $|f(x) - f(y)| \le \|x-y\|$ for all $x, y \in \Re^k$, then for all $\lambda > 0$*

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

where

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

REMARK: It may seem strange at first that the upper bound depends neither on the function f nor the covariance structure of the X_i (other than via the variances). If one rewrites the result in the form of (2.6), however, the fact that $Ef(X)$ now appears in the bound, and the fact that this expectation depends explicitly on f and implicitly on the full covariance structure of the X_i, shows that all is in heuristic order.

PROOF: Let $\{B_t\}_{t\geq 0} = \{B_t^1, ..., B_t^k\}_{t\geq 0}$ be a k-dimensional Brownian motion; i.e. the B^i are i.i.d. standard, real-valued Brownian motions.

We can link the vector X to B_t by noting that $X_1, \ldots, X_k$ is distributed exactly like $V^{\frac{1}{2}} B_1$, where $V^{\frac{1}{2}}$ satisfies $V = V^{\frac{1}{2}} \cdot (V^{\frac{1}{2}})'$, and exists by virtue of the positive semi-definiteness of V. Before we can fully utilise this fact, however, we need to make two excursions into stochastic analysis.

The first 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

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

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

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

A standard Chebycheff type argument gives us that

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

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 $F = F(x,t) : \Re^k \times \Re_+ \to \Re$,

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

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 $g: \Re^k \to \Re$

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

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.

Let $\hat{f}: \Re^k \to \Re^1$ satisfy the differentiability requirements of the f of the lemma, and assume $|\hat{f}(x) - \hat{f}(y)| \leq \sigma\|x-y\|$. Setting $F(x,t) = (P_{1-t}\hat{f})(x)$, 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

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

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 $|\hat{f}(x) - \hat{f}(y)| < \sigma\|x-y\|$ immediately implies that $P_t\hat{f}$ satisfies the same inequality for every $t \geq 0$, and so $\|\nabla P_t \hat{f}(x)\| \leq \sigma$ for almost every x. It then easily follows from (2.9) that

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

To complete the proof note simply that $f(X) \stackrel{\mathcal{L}}{=} f(V^{\frac{1}{2}}B_1)$, so that (2.7) follows from (2.11) with $\hat{f}(x) = f(V^{\frac{1}{2}}x)$. The function $\hat{f}$ then satisfies all the requirements placed on it. ∎

PROOF OF THEOREM 2.1: We have two things to prove. Firstly, Theorem 2.1 would follow immediately from Lemma 2.2 in the case of finite T if only sup(.) were a sufficiently smooth function. Unfortunately, it is not, being non-differentiable on the diagonal in $\Re^k$. Fortunately, however, it is approximable by smooth functions. Any standard approximation procedure (such as convolution with C^∞ functions) will work, and so this part of the proof is left to you.

The second part of the proof involves 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

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

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

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

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.

2. Slepian's Inequality.

There are a variety of different ways to present Slepian-like inequalities today. We choose the following formulation, (from Joag-Dev, Perlman and Pitt (1983)) which actually includes a number of interesting side results.

2.3 THEOREM. *Let $X_1, \dots, X_k$ be centered Gaussian variables with covariance matrix $R = (r_{ij})_{i,j=1}^k$, $r_{ij} = EX_iX_j$. Let $h\colon \Re^k \to \Re$ be C^2, and assume that, together with its derivatives, it satisfies a $O(\|x\|^N)$ growth condition at infinity for some finite N. Let*

$$(2.12) \qquad \mathcal{H}(R) = Eh(X_1, \dots, X_k).$$

and assume that for a pair (i,j), $1 \le i < j \le k$

$$(2.13) \qquad \frac{\partial^2 h(x)}{\partial x_i \partial x_j} \ge 0 \qquad \text{for all } x.$$

Then $\mathcal{H}(R)$ is an increasing function of r_{ij}.

Note that the theorem is actually true without the growth conditions on h, if one is prepared to attribute veracity to the result if $\mathcal{H}$ is identically infinite. The proof is reasonably straightforward, and, in its important details, goes back to Slepian (1962).

PROOF: We have to show that

$$\frac{\partial \mathcal{H}(R)}{\partial r_{ij}} \ge 0$$

whenever $\partial^2 h / \partial x_i \partial x_j \ge 0$.

To make our lives a little easier we assume that R is non-singular, so that it makes sense to write $\phi(x) = \phi_R(x)$ for the centered Gaussian density on $\Re^k$ with covariance matrix R. Then some algebra (dating back at least to Plackett (1954)) shows that

$$(2.14) \qquad \frac{\partial \phi}{\partial r_{ii}} = \frac{1}{2} \frac{\partial^2 \phi}{\partial x_i^2}, \qquad \frac{\partial \phi}{\partial r_{ij}} = \frac{\partial^2 \phi}{\partial x_i \partial x_j}, \qquad i \ne j.$$

(The algebra of this calculation is almost identical to that needed to justify (2.11) in the proof of Borell's inequality!) Apply this result and our assumptions on h to justify two integrations by parts to obtain

$$\frac{\partial \mathcal{H}(R)}{\partial r_{ij}} = \int_{\Re^k} h(x)\frac{\partial \phi(x)}{\partial r_{ij}}dx = \int_{\Re^k} \frac{\partial^2 h(x)}{\partial x_i \partial x_j}\phi(x)dx \geq 0.$$

(If the integrations by parts bother you, see Berman (1987).) This completes the proof for the case of non-singular R. The general case can be handled by approximating a singular R via a sequence of non-singular covariance matrices. ∎

Here are some of the consequences of Theorem 2.3:

2.4 COROLLARY (SLEPIAN'S INEQUALITY). *If X and Y are a.s. bounded, centered Gaussian processes on T such that $EX_t^2 = EY_t^2$ for all $t \in T$ and*

$$E(X_t - X_s)^2 \leq E(Y_t - Y_s)^2 \qquad \text{for all } s,t \in T, \tag{2.15}$$

then for all real λ

$$P\{\sup_{t\in T} X_t > \lambda\} \leq P\{\sup_{t\in T} Y_t > \lambda\} \tag{2.16}$$

PROOF: By separability, and the final argument in the proof of Borell's inequality, it suffices to prove (2.16) for T finite. Note that since $EX_t^2 = EY_t^2$ for all $t \in T$, (2.15) implies that $EX_sX_t \geq EY_sY_t$ for all $s,t \in T$. Let $h(x) = \prod_{i=1}^k f_i(x_i)$, where each f_i is a positive non-increasing, C^2 function satisfying the growth conditions placed on h in the statement of Theorem 2.3 and k is the number of points in T. Note that, for $i \neq j$

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} = f_i'(x_i)f_j'(x_j)\prod_{\substack{n\neq i\\ n\neq j}} f_n(x_n) \geq 0,$$

since both f_i' and f_j' are non-positive. It therefore follows from the theorem that

$$E\prod_{i=1}^k f_i(X_i) \geq E\prod_{i=1}^k f_i(Y_i).$$

Now take $\{f_i^{(n)}\}_{n=1}^\infty$ to be a sequence of positive, non-increasing, C^2 approximations to the indicator function of the interval $(-\infty,\lambda]$, to derive that

$$P\{\|X\| < \lambda\} \geq P\{\|Y\| < \lambda\}.$$

That is,

$$P\{\|X\| > \lambda\} \le P\{\|Y\| > \lambda\},$$

which is what we had to prove. ■

Slepian's inequality will form one the basic building blocks of the results of Chapter 4. In general, it is written with the condition $EX_sX_t \ge EY_sY_t$ rather than (2.15), since this more clearly indicates the difference between the dependence structures of the two processes.

It is immediate from symmetry considerations that a similar result holds for infima rather than suprema, as it does for the range of X_t, but, as we shall soon see by example, no such result holds for $\sup_t |X_t|$.

Another consequence of Slepian's inequality is

2.5 COROLLARY. *Under the conditions of Corollary 2.4*

$$E \sup_{t\in T} X_t \le E \sup_{t\in T} Y_t \tag{2.17}$$

PROOF:

$$\begin{aligned} E\|X\| &= \int_0^\infty P\{\|X\| > \lambda\}\, d\lambda - \int_{-\infty}^0 P\{\|X\| < \lambda\}\, d\lambda \\ &\le \int_0^\infty P\{\|Y\| > \lambda\}\, d\lambda - \int_{-\infty}^0 P\{\|Y\| < \lambda\}\, d\lambda \\ &= E\|Y\|. \end{aligned}$$

This proves the result. ■

The main problem with Slepian's inequality is that it requires that X_t and Y_t have identical variances. To compare processes with differing variances it turns out that we have to concentrate on inequalities of the form (2.17) rather that (2.16). Our first result in this regard is the almost trivial

2.6 OBSERVATION. *If X is an a.s. bounded, centered Gaussian process on T and Y a centered Gaussian variable (not necessarily independent of X), then*

$$E\Big(\sup_{t\in T}(X_t + Y)\Big) = E \sup_{t\in T} X_t \tag{2.18}$$

As trite as this observation is, it ceases to be valid if, for example, we investigate $\sup_t |X_t|$ rather than $\sup_t X_t$. It has surprisingly important implications. For example, it allows us to prove

2.7 LEMMA. *Let X and Y be two centered Gaussian random variables. Then*

$$E\max(X,Y) \;=\; \frac{\left(E(X-Y)^2\right)^{\frac{1}{2}}}{\sqrt{2\pi}}. \tag{2.19}$$

PROOF: Set $\sigma^2 = E(X-Y)^2$. By Observation 2.6

$$\begin{aligned} E\max(X,Y) &= E\max(X-Y, Y-Y) \\ &= E\max(X-Y, 0) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_0^\infty z\, e^{-z^2/2\sigma^2}\, dz \\ &= \frac{\sigma}{\sqrt{2\pi}}, \end{aligned}$$

as required. ■

A useful inequality that follows from this lemma is that for any a.s. bounded, centered X,

$$E\sup_{t\in T} X_t \;\geq\; \frac{\left(\sup_{s,t\in T} E(X_s - X_t)^2\right)^{\frac{1}{2}}}{\sqrt{2\pi}}. \tag{2.20}$$

What is more important, however, is that Observation 2.6 will allow us to prove

2.8 THEOREM. *If X and Y are a.s. bounded, centered Gaussian processes on T such that*

$$E(X_t - X_s)^2 \;\leq\; E(Y_t - Y_s)^2 \tag{2.21}$$

then

$$E\sup_T X_t \;\leq\; 2E\sup_T Y_t. \tag{2.22}$$

Note that the important difference between Lemma 2.5 and Theorem 2.8 is not so much the additional factor of 2 on the right hand side, but the fact that we no longer require that X and Y have identical variances.

PROOF: Fix a point $t_o \in T$ and set $\alpha^2 = \sup_{t\in T} E(Y_t - Y_{t_o})^2$. Furthermore, set

$$\widetilde{X}_t \;=\; X_t - X_{t_o}, \qquad \widetilde{Y}_t \;=\; Y_t - Y_{t_o},$$

and let η and η' be two standard Gaussian variables, independent of one another and the two processes X_t and Y_t. Now define

$$\begin{aligned}\widehat{X}_t &= \widetilde{X}_t + \eta\alpha,\\ \widehat{Y}_t &= \widetilde{Y}_t + \eta' g(t),\end{aligned}$$

where

$$g^2(t) := \alpha^2 - E(Y_t - Y_{t_o})^2 + E(X_t - X_{t_o})^2.$$

Note that $\alpha^2 \geq g^2(t) \geq 0$ for all $t \in T$, and that $E\widehat{X}_t^2 = E\widehat{Y}_t^2$ for all $t \in T$. Furthermore, since

$$\begin{aligned}(2.23) \qquad E\big(\widehat{Y}_t - \widehat{Y}_s\big)^2 &= E\big[\widetilde{Y}_t - \widetilde{Y}_s) + \eta'\big(g(t) - g(s)\big)\big]^2\\ &= E\big(\widetilde{Y}_t - \widetilde{Y}_s\big)^2 + \big(g(t) - g(s)\big)^2\\ &\geq E\big(\widetilde{Y}_t - \widetilde{Y}_s\big)^2,\end{aligned}$$

it follows that

$$\begin{aligned}E\big(\widehat{X}_t - \widehat{X}_s\big)^2 &= E\big(\widetilde{X}_t - \widetilde{X}_s\big)^2\\ &\leq E\big(\widetilde{Y}_t - \widetilde{Y}_s\big)^2\\ &\leq E\big(\widehat{Y}_t - \widehat{Y}_s\big)^2,\end{aligned}$$

the first inequality following from (2.21) and the second from (2.23).

Observation 2.6 gives us that

$$E\|X\| = E\|\widetilde{X}\| = E\|\widehat{X}\|,$$

and Corollary 2.5 gives us that

$$E\|\widehat{X}\| \leq E\|\widehat{Y}\|.$$

But,

$$\begin{aligned}E\|\widehat{Y}\| &= E\sup_{t\in T}\big(\widetilde{Y}_t + \eta' g(t)\big)\\ &\leq E\|\widetilde{Y}\| + E\sup_{t\in T}\eta' g(t)\\ &\leq E\|\widetilde{Y}\| + E\big\{\sup_{t\in T}\eta' g(t)1_{\eta'>0}\big\}\\ &= E\|\widetilde{Y}\| + \sup_{t\in T} g(t)\, E\{\eta' 1_{\eta'>0}\}\\ &\leq E\|\widetilde{Y}\| + \frac{\alpha}{\sqrt{2\pi}},\end{aligned}$$

the last line following from the definition of g and a standard Gaussian integration. However, by (2.20), the last line here is less that $2E\|\widetilde{Y}\| = 2E\|Y\|$. This concludes the proof. ■

There is, in fact, a stronger version of Theorem 2.8, in which the factor of 2 in (2.22) does not appear. The proof is somewhat more involved than that of Theorem 2.8, and relies on new calculations rather than a clever application of previous results.

2.9 THEOREM (SUDAKOV-FERNIQUE INEQUALITY). *Under the conditions of Theorem 2.8, (2.22) continues to hold if the factor of 2 is removed from the right hand side of the inequality.*

PROOF: Actually, I shall only show you how to start the proof, and shall leave the last part, which involves considerable calculus, up to you. You can find the details in Fernique (1975) or Jain and Marcus (1978).

The main trick of the proof lies in noting that it is sufficient to show that

$$E \sup_{s,t \in T} |X_t - X_s| \;\le\; E \sup_{s,t \in T} |Y_t - Y_s|.$$

The result we seek will then follow since, for example,

$$\begin{aligned} E \sup_{s,t \in T} |Y_t - Y_s| &= E \sup_{s,t \in T} (Y_t - Y_s) \\ &= E\big\{ \sup_{t \in T} Y_t + \sup_{t \in T}(-Y_t) \big\} \\ &= 2\,E\|Y\|. \end{aligned}$$

As usual, we assume that T is finite with k points, and use separability and the final argument in the proof of Borell's inequality to complete the proof. Let $\mathbf{X}$ and $\mathbf{Y}$ denote the k-vectors of values of X_t and Y_t on T. Put copies of $\mathbf{X}$ and $\mathbf{Y}$ onto the same probability space, and assume they are independent. For $\theta \in [0,1]$ set

$$\mathbf{Z}(\theta) \;=\; \sqrt{1-\theta}\,\mathbf{X} + \sqrt{\theta}\,\mathbf{Y},$$

and $\psi(\theta) = E(\sup_i (Z(\theta))_i)$. We need to show that $\psi(0) \le \psi(1)$, for which it is sufficient to show that $\psi'(\theta) \ge 0$ for all $\theta \in [0,1]$.

If $\mathbf{R}(\theta)$ is the covariance matrix of $\mathbf{Z}(\theta)$, then using Fourier transforms we have that the k-dimensional density of $\mathbf{Z}(\theta)$ is given by

$$p_\theta(\mathbf{z}) \;=\; (2\pi)^{-k} \int_{\Re^k} \exp\left\{ i\langle \mathbf{z}, \mathbf{x}\rangle \;-\; \tfrac{1}{2}\langle \mathbf{R}(\theta)\mathbf{x}, \mathbf{x}\rangle \right\} d\mathbf{x}.$$

Since

$$\psi(\theta) \;=\; \int_{\Re^k} \max\{z_1, \dots, z_k\}\, p_\theta(\mathbf{z})\, d\mathbf{z},$$

we have

$$\psi'(\theta) = \int_{\Re^k} \max\{z_1,\ldots,z_k\} \frac{\partial p_\theta(\mathbf{z})}{\partial\theta}\, d\mathbf{z}.$$

If you now start differentiating, keeping (2.14) in mind, you will find that

$$\psi'(\theta) = \tfrac{1}{2}\sum_{i,j} \frac{\partial r_{ij}(\theta)}{\partial\theta} \int_{\Re^k} \max\{z_1,\ldots,z_k\} \frac{\partial^2 p_\theta(\mathbf{z})}{\partial z_i \partial z_j}\, d\mathbf{z}.$$

What remains to complete the proof is to show that this expression is positive. This is left to you, with the help of the references given above. ■

The Sudakov-Fernique inequality is a good place to look to see how far one can generalize results of this kind. For example, the inequality cannot work for $\sup|X_t|$. To see this, let Z be a centered Gaussian variable, ξ a positive real number, and X a bounded Gaussian process. Let Y^ξ be the process defined by

$$Y_t^\xi = \xi Z + X_t, \qquad t \in T.$$

Then $E(Y_s^\xi - Y_t^\xi)^2 = E(X_s - X_t)^2$ for all $s,t \in T$, but $E\sup_{t\in T}|Y_t^\xi| \to \infty$ as $\xi \to \infty$, while $E\sup_{t\in T}|X_t|$ is bounded.

To show that the original Slepian's inequality also fails for $\sup_{t\in T}|X_t|$, take $T=\{1,2\}$, with X_1 and X_2 standard normal with correlation ρ. Writing $P_\rho(\lambda)$ for the probability under correlation ρ that $\max(X_1,X_2) > \lambda$ and Ψ as usual for the right hand tail of the standard normal distribution function we see that

$$P_{-1}(\lambda) = P_{-1}\{\max(X_1, X_2) > \lambda\} = P\{|X| > \lambda\} = 2\Psi(\lambda)$$

$$P_0(\lambda) = 2\Psi(\lambda) - \Psi^2(\lambda)$$

$$P_{+1}(\lambda) = P_1\{\max(X_1, X_2) > \lambda\} = P\{X_1 > \lambda\} = \Psi(\lambda)$$

Hence $P_{-1}(\lambda) \geq P_0(\lambda) \geq P_1(\lambda)$ as Slepian's inequality requires. But if $\hat{P}_\rho$ is the probability that $\max(|X_1|,|X_2|) > \lambda$, then $\hat{P}_{-1}(\lambda) = \hat{P}_{+1}(\lambda) = 2\Psi(\lambda)$, $\hat{P}_0(\lambda) = 4[\Psi(\lambda) - \Psi^2(\lambda)]$, so that for all $\lambda > 0$

$$\hat{P}_{-1}(\lambda) < \hat{P}_0(\lambda) \qquad \hat{P}_0(\lambda) > \hat{P}_1(\lambda),$$

and the monotonicity required by Slepian breaks down. ■

Here is a list of interesting variations of Slepian's inequality. You can skip them without harming your understanding of the following chapters, but it would be a shame.

ELLIPTICALLY CONTOURED DISTRIBUTIONS: If $p_\Sigma(x)$ is a density on $\Re^k$ of the form

$$p_\Sigma(x) = |\Sigma|^{-\frac{1}{2}} p(x'\Sigma^{-1}x),$$

where Σ is an invertible, positive definite matrix and p: $[0,\infty) \to \Re_+$ is assumed to satisfy $\int_0^\infty \lambda^{k-1} p(\lambda^2)\, d\lambda < \infty$, we say that p is *elliptically contoured.* A version of Slepian's inequality holds for such densities, for writing $\Sigma = (\sigma^{ij})_{1\le i,j\le n}$ and defining

$$\mathcal{H}(\Sigma) = \int_{\Re^k} h(x) p_\Sigma(x) dx,$$

with h satisfying the appropriate continuity and growth conditions (as in Theorem 2.3), then Joag-Dev *et al.* have shown that $\mathcal{H}(\Sigma)$ is an increasing function of σ^{ij}. The proof is almost identical to that we gave for Theorem 2.3.

GORDON'S INEQUALITY: An interesting extension of Slepian's inequality, which is a result about Gaussian maxima, is a result of Gordon's about the min-max of a rectangular array of Gaussian variables. Let $(X_{ij})_I$, $(Y_{ij})_I$, $I = \{(i,j): 1 \le i \le n, 1 \le j \le m\}$ be two collections of centered Gaussian variables satisfying the following three conditions:

$$\begin{aligned} EX_{ij}^2 &= EY_{ij}^2 && (i,j) \in I, \\ EX_{ij}X_{ik} &\le EY_{ij}Y_{ik} && (i,j),(i,k) \in I, \\ EX_{ij}X_{\ell k} &\ge EY_{ij}Y_{\ell k} && (i,j),(\ell,k) \in I,\ i \ne \ell. \end{aligned}$$

Then, for all real λ_{ij},

$$P\Big\{\bigcap_{i=1}^{n}\bigcup_{j=1}^{m}[X_{ij} > \lambda_{ij}]\Big\} \ge P\Big\{\bigcap_{i=1}^{n}\bigcup_{j=1}^{m}[Y_{ij} > \lambda_{ij}]\Big\}.$$

This implies, for example, that for any increasing function g on $\Re$

$$E\{\min_{1\le i\le n}\max_{1\le j\le m} g(X_{ij})\} \ge E\{\min_{1\le i\le n}\max_{1\le j\le m} g(Y_{ij})\}, \tag{2.24}$$

and that for all $\lambda > 0$

$$P\{\min_{1\le i\le n}\max_{1\le j\le m} X_{ij} \ge \lambda\} \ge P\{\min_{1\le i\le n}\max_{1\le j\le m} Y_{ij} > \lambda\}. \tag{2.25}$$

This result has extensions from minmax to minmaxmin..., etc. and to elliptically contoured distributions as well. It is a consequence of the following theorem, whose proof is due to Kahane (1986).

2.10 THEOREM. *Let* $\mathbf{X} = (X_i)$ *and* $\mathbf{Y} = (Y_i)$, $i = 1,\dots,k$ *be two collections of Gaussian variables, and* I *and* J *subsets of* $\{1,\dots,k\}^2$ *such that*

$$EX_iX_j \le EY_iY_j \qquad (i,j) \in I, \tag{2.26}$$

$$EX_iX_j \ge EY_iY_j \qquad (i,j) \in J, \tag{2.27}$$

$$EX_iX_j = EY_iY_j \qquad (i,j) \notin I \cup J. \tag{2.28}$$

Let $h\colon \Re^k \to \Re$ be C^2, and assume that, together with its derivatives, it satisfies a $O(\|x\|^N)$ growth condition at infinity for some finite N. Furthermore, assume

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} \ \geq\ 0 \qquad (i,j) \in I, \tag{2.29}$$

$$\frac{\partial^2 h(x)}{\partial x_i \partial x_j} \ \leq\ 0 \qquad (i,j) \in J, \tag{2.30}$$

then

$$Eh(\mathbf{X}) \ \leq\ Eh(\mathbf{Y}). \tag{2.31}$$

Note that Theorem 2.3, and so Slepian's inequality and all that it implied, are consequences of this result. Thus what follows is, in fact, an alternative proof of Slepian's inequality.

PROOF: As in the proof of Theorem 2.9, put copies of $\mathbf{X}$ and $\mathbf{Y}$ onto the same probability space, and assume they are independent. For $t \in [0,1]$ set

$$\mathbf{Z}(\theta) \ =\ \sqrt{1-\theta}\,\mathbf{X} + \sqrt{\theta}\,\mathbf{Y},$$

and $\psi(\theta) = Eh(\mathbf{Z}(\theta))$. We need to show that $\psi(0) \leq \psi(1)$, for which it is sufficient to show that $\psi'(\theta) \geq 0$ for all $\theta \in [0,1]$. Writing h_i for $\partial h(x)/\partial x_i$, and allowing ourselves the luxury of interchanging the order of expectation and differentiation, we have

$$\psi'(\theta) \ =\ \sum_{j=1}^{N} E\{h_j(\mathbf{Z}(\theta))\, Z_j'(\theta)\}. \tag{2.32}$$

Fix θ and j, and note that

$$E\{Z_i(\theta)\, Z_j'(\theta)\} \ =\ \tfrac{1}{2}E\{Y_jY_i \ -\ X_jX_i\}.$$

It thus follows from (2.26)–(2.28) that if $W_1,\ldots,W_N$ is a new sequence of Gaussian variables, independent of both the X_i and Y_i, then we can express the Z_i as follows:

$$Z_i(\theta) \ =\ \alpha_i Z_j'(\theta) + W_i, \tag{2.33}$$

where $\alpha_i \geq 0$ if $(i,j) \in I$, $\alpha_i \leq 0$ if $(i,j) \in J$, and $\alpha_i = 0$ if $(i,j) \notin I \cup J$.

With j still fixed, consider the behaviour of a typical summand in (2.33). Then

(2.34)

$$E\{h_j(\mathbf{Z}(\theta))Z_j'(\theta)\} = \int\cdots\int h_j(\alpha_1 z_j' + w_1, \ldots, \alpha_N z_j' + w_N)z_j' \times \phi_{Z_j'}(z_j')\phi_W(w_1, \ldots, w_N)\, dz_j' dw_1 \ldots dw_N,$$

where $\phi_{Z_j'}$ and ϕ_W are the obvious Gaussian densities.

To complete the argument, differentiate (2.34) with respect to each α_i, and note, that from the properties of the α_i (c.f. the line following (2.33)) and the inequalities (2.29) and (2.30), that the resulting expression is positive if $(i,j) \in I$, and negative if $(i,j) \in J$. That is, $E\{h_j(\mathbf{Z}(\theta))Z_j'(\theta)\}$ is an increasing function of α_i for $(i,j) \in I$, and decreasing for $(i,j) \in J$.

On the other hand, it is clear that $E\{h_j(\mathbf{Z}(\theta))Z_j'(\theta)\} = 0$ if all the $\alpha_i = 0$. Consequently, this expectation must always be non-negative. That is, $\psi'(\theta) \geq 0$, which is what we had to prove. ∎

As one might expect, Gordon's inequality can be extended in much the same way as Fernique's inequality can. For example, there is an analagous result for elliptically contoured distributions. You can find further results, as well as applications of Gordon's inequality, (which we shall return to in the following section) in Gordon (1987, 1988a,b). Note also from the proof of Theorem 2.10 that the restriction to twice differentiable h is unnecessarily restrictive. It would have been enough, for example, to assume that the first order derivatives h_j are all absolutely continuous with positive generalised derivatives.

INFINITELY DIVISIBLE PROCESSES: We consider only a simple example.

Let μ be a non-negative measure on $\Re^d$ and let N_μ be a Poisson point process on $\Re^d$ with intensity μ. Let $A_1, \ldots, A_n$ be Borel sets in $\Re^d$ and define $X_i = N_\mu(A_i)$. Let $Y_i = N_\nu(A)$ be similarly defined with respect to a Poisson process with intensity ν. If, for all $B \subseteq \{1, \ldots, n\}$, $B \neq \emptyset$,

$$\mu(\bigcap_{i\in B} A_i) \leq \nu(\bigcap_{i\in B} A_i), \qquad \mu(\bigcup_{i\in B} A_i) \geq \nu(\bigcup_{i\in B} A_i), \tag{2.35}$$

then for all $\lambda > 0$

$$P\{\max_{1\leq i\leq n} X_i > \lambda\} \geq P\{\max_{1\leq i\leq n} Y_i > \lambda\}. \tag{2.36}$$

That is, a form of Slepian's inequality holds. Since Poisson processes are the natural building blocks of infinitely divisible processes, it is not hard to see that a result of this form must extend to a far more general situation. For details, see Brown and Rinott (1988) and Ellis (1988).

The primary interest in (2.36) is that it allows one to build multidimensional Kolmogorov-Smirnov tests based on multivariate empirical distribution functions, which, unlike the examples discussed earlier, hold for finite sample size. You can find details (and tables!) in Adler, Brown and Lu (1988).

The fact that a Slepian-like inequality holds for these processes is somewhat surprising, since there is nothing Gaussian-like in their structure. Thus it is *not* surprising that whereas the proofs of the basic Gaussian result, the results for elliptically contained distributions, and the min-max result all look basically the same, the proof of (2.36) under (2.35) is completely different.

STABLE PROCESSES: We have already discussed these processes in the previous Chapter. It has long been part of the folklore of stable processes that no Slepian like inequality holds, and counter-examples have been constructed to demonstrate this: e.g. Fernique (1983), and Linde (1986). In a certain sense, however, this claim is not quite true. The examples given all show that ordering two stable processes in terms of the *covariation* function, does not imply stochastic ordering of the sample maxima. (The covariation function is a stable analogue of the covariance function for Gaussian processes, and is defined as

$$\operatorname{cov}(s,t) \;=\; \int \beta(s)|\beta(t)|^{\alpha-1}\operatorname{sgn}(\beta(t))\, m(d\beta),$$

where m is the spectral measure of (1.74).)

However, whereas covariance functions determine Gaussian processes, the same is not true of covariation functions for stable processes. Thus, it is not necessarily natural to try to base a stochastic majoration result on comparison of covariation functions. Gennady Samorodnitsky has recently shown me a Slepian type result for stable processes based on comparison of spectral measures. Hopefully it will be written up sometime soon.

One factor to note, however, is that one of the reasons that Slepian's inequality is so useful is that it is so very easy to state and apply. This is not the case with Samorodnitsky's result for stable processes. Thus it still seems that a "pure" Slepian inequality is basically a Gaussian result only, and we are very lucky to have it.

DIFFUSIONS AND THEIR ILK: There is a substantial body of results in the diffusion literature on stochastic domination problems. For example, let $X_1(t)$ and $X_2(t)$ be two diffusions on $[0,\infty)$ satisfying the stochastic differential equations

$$dX_i(t) \;=\; a_i\big(X_i(t)\big)\,dt \;+\; b\big(X_i(t)\big)\,dW_i(t), \tag{2.37}$$

where the $W_i(t)$ are standard Brownian motions. Note that the X_i have different drift, but the same "speed" function. If $a_1(x) \le a_2(x)$ for all x, then

$$P\{\|X_1\|_T > \lambda\} \;\le\; P\{\|X_2\|_T > \lambda\}$$

for all λ and all $T \subset [0,\infty)$. Indeed, much more than this Slepian like inequality is true, for one can build X_1 and X_2 on the same sample space in such a way that, with probability one, $X_1(t) \leq X_2(t)$ for all $t \in [0,\infty)$. The construction is easy. Just start the processes off in such a way that $X_1(0) \leq X_2(0)$, and use the same Brownian motion in (2.37) to generate each of the two processes. The fact that $a_1(X) \leq a_2(X)$ will then ensure that $X_1(t) \leq X_2(t)$ for all t.

In the light of this result, it is natural to ask the following question: Given two Gaussian processes satisfying the conditions of Slepian's inequality, can one construct them on the same sample space so that the maximum of one a.s. dominates that of the other? Simple examples with two-point parameter spaces show that this is not the case.

Once again, to convince you that in spite of what we have just indicated Slepian's inequality is a primarily Gaussian result, you should remember that the diffusion result is heavily restricted to processes on the real line only, and has no extension (as, in fact, can be said of diffusions themselves) to unstructured parameter spaces.

3. Applications in Banach Spaces.

One of the most fruitful areas in the theory of probability over the past decade has been the development of an active interface with the theory of Banach spaces. Much of this activity has centered around the general theory of Gaussian processes, and in this section I want to briefly describe two closely related results in the local theory of Banach spaces and their proof via Gaussian methods.

It is beyond the intended scope of these notes to enter into a detailed treatment of these results. Nevertheless, I do want you to see them, if only to show how useful the general theory of Gaussian processes is, even outside the usual settings of probability and stochastic processes. For details you should see the notes by Pisier (1986) and his brand new monograph Pisier (1989), as well as the monographs by Milman and Schechtman (1986) and Linde (1986). These cover not only results of the kind discussed below, but also a wide variety of other applications of Gaussian processes to Banach space theory.

We start by recalling some basic facts and definitions. Let E, F be Banach spaces and let $\lambda > 1$. Then E and F are called λ-isomorphic if there is an isomorphism $T\colon E \to F$ such that $\|T\| \cdot \|T^{-1}\| \leq \lambda$, where, as usual,

$$\|T\| = \sup_{\|x\|_E \leq 1} \|T(x)\|_F ,$$

and $\| \ \|_E$ and $\| \ \|_F$ and the norms on E and F. If E and F are λ-isomorphic we write $E \overset{\lambda}{\sim} F$.

This concept can be used to define a (Banach-Mazur) distance between E and F by

$$\rho(E,F) = \inf\{\lambda : E \overset{\lambda}{\sim} F\}.$$

The importance of this distance function is that it is a good measure of how close two Banach spaces are in terms of their local characteristics. To provide an example of this, let ℓ_p^n denote the Euclidean space $\Re^n$ equipped with the norm $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$. Here is a result dating back, in essence, to Dvoretzky (1961):

2.11 THEOREM. *Let B be a Banach space, and $\epsilon > 0$. Then there exists a $n = n(\epsilon)$ such that B has a subspace F with $F \overset{1+\epsilon}{\sim} \ell_2^n$.*

Here is a more precise result in a more specific setting.

2.12 THEOREM. *Let $\epsilon > 0$ and $n \geq 1$. If $N \geq \alpha \exp(\beta n \epsilon^{-2})$ is an integer (where $\alpha, \beta > 0$ are universal constants), then for every convex set $B \subset \Re^N$ which contains the origin in its interior, there is a subspace F of dimension n and a constant $a > 0$, such that*

$$aS^n \subseteq F \cap B \subseteq \left(\frac{1+\epsilon}{1-\epsilon}\right) aS^n, \tag{2.38}$$

where S^n is the unit ball in $\Re^n$.

These two results are really versions of the same basic truth, Theorem 2.12 being the sharper of the two because of the additional information it provides on n. The bound $\alpha \exp(\beta n \epsilon^{-2})$ that it gives can actually be shown to be sharp.

Neither of these results looks like it has anything to do with Gaussian processes, but here is the briefest of outlines of a proof that will show there is a very strong connection indeed.

Recall firstly the comment made after the statement of Theorem 2.1 that Borell's inequality (2.5) holds also for Banach space valued Gaussian processes in which the norm in (2.5) is the Banach norm.

To construct the subspace F of Theorem 2.11, let X be a B-valued Gaussian variable, take $n \geq 1$, and let $X_1, \ldots, X_n$ be independent copies of X. Let $M = E\|X\|$. Then some reasonably simple calculations, put together with the Banach space version of Borell's inequality, show that for the right values of n there is a subset Ω_0 of our generic probability space, with $P\{\Omega_0\} > 0$, such that for $\omega \in \Omega_0$ the following is true for all $x \in \Re^n$:

$$(1+\epsilon)^{-\frac{1}{2}} \Big(\sum_{i=1}^n |x_i|^2\Big)^{\frac{1}{2}} \leq M^{-1} \Big\| \sum_{i=1}^n x_i X_i(\omega) \Big\| \leq (1+\epsilon)^{\frac{1}{2}} \Big(\sum_{i=1}^n |x_i|^2\Big)^{\frac{1}{2}}.$$

Taking the span of $\{X_1(\omega), \ldots, X_n(\omega)\}$ for $\omega \in \Omega_0$ provides an example of F.

This proves Theorem 2.11. Details can be found in Pisier (1986, 1989). The proof of Theorem 2.12 follows the same principle; i.e. one uses random Gaussian mappings satisfying certain conditions with positive probability to provide proof that such mappings exist at all. The proof of Theorem 2.12, however, is based on a very clever use of Gordon's inequality (2.24). The proof is in Gordon (1988).

4. Exercises.

SECTION 2.1:

1.1 By expanding the normal probability density in a power series, or otherwise, establish the two sides of the basic inequality (2.1).

1.2 If you have not already done so, prove (2.11) by applying Itô's formula (2.10) to $F(x,t) = (P_{1-t}\hat{f})(x)$. Note how very neatly various complicated terms cancel out, and marvel in the beauty and speciality of the Gaussian density which makes this happen.

SECTION 2.2:

2.1 Show by example on a two point parameter space that the a.s. domination of suprema that holds for diffusions does not hold for Gaussian processes. Why does it not hold for Gaussian diffusions?