V. SUPREMA DISTRIBUTIONS

1. Introduction.

We now turn our attention to the distribution of the supremum of a centered, Gaussian X over T, which, as always, we write either as $\|X\|_T$ or simply as $\|X\|$ when there is no danger of confusion. In particular, we shall be interested in the asymptotic behavior of $P\{\|X\| > \lambda\}$ as $\lambda \to \infty$.

As we have already seen, an upper bound to this probability comes from Borell's inequality, which tells us that for all $\lambda > 0$

$$P\{|\,\|X\| - E\|X\|\,| > \lambda\} \;<\; 2e^{\frac{1}{2}\lambda^2/\sigma_T^2}, \tag{5.1}$$

where $\sigma_T^2 = \sup_T EX_t^2$. This implies that for all $\lambda > E\|X\|$

$$P\{\|X\| > \lambda\} \;<\; 2e^{-\frac{1}{2}(\lambda^2 - 2\lambda E\|X\| + \|X\|^2)/\sigma_T^2}. \tag{5.2}$$

Since Borell's inequality was the key to finding sufficient conditions for Gaussian sample path continuity, that also turned out to be necessary, it seems reasonable to assume that it is close to sharp. Nevertheless, the aim of this chapter is to improve on (5.2) in two directions. Firstly, if we think of the right hand side of (5.2) as $f(\lambda)e^{-\frac{1}{2}\lambda^2/\sigma_T^2}$, then our aim will be to replace this by $\hat{f}(\lambda)e^{-\frac{1}{2}\lambda^2/\sigma_T^2}$, where $\hat{f}(\lambda)$ has a lower order of growth in λ than does the exponentially growing f. In many cases, it is possible to find a polynomial $\hat{f}$.

Having found such $\hat{f}$, we would also like to know if we have found the best possible, and so we shall also be interested in lower bounds for $P\{\|X\| > \lambda\}$. These almost always involve much more work than the upper bounds, so we shall generally suffice with statements without proofs. The situation is highly analagous to the continuity problem: sufficiency was easy, necessity was hard. What is rather interesting in the upper bound proofs, however, is that they proceed via a kind of "leap-froging", in which Borell's inequality is used to improve on itself!

Before we look at the general situation, however, it is worthwhile to look at an optimal situation, in which "almost everything" is known, so as to give us an idea as to what sort of results we can hope for in general.

5.1 THEOREM. *Let X be a centered, stationary Gaussian process on $\Re$, with covariance function R satisfying*

$$R(t) \;=\; 1 - C|t|^\alpha + o(|t|^\alpha), \qquad \text{as} \quad t \to 0, \tag{5.3}$$

where $\alpha \in (0,2]$ and C are positive constants. Then for each fixed $h > 0$ such that $\sup_{\epsilon \le t \le h} R(t) < 1$ for all $\epsilon > 0$,

$$\lim_{\lambda \to \infty} \frac{P\{\|X\|_{[0,h]} > \lambda\}}{\lambda^{2/\alpha}\Psi(\lambda)} \;=\; hC^{1/\alpha}H_\alpha, \tag{5.4}$$

where H_α is a finite constant depending only on α, and, as always,

$$\Psi(\lambda) = (2\pi)^{-1} \int_\lambda^\infty e^{-\frac{1}{2}x^2}\, dx, \tag{5.5}$$

is the probability that a standard normal variable exceeds λ.

This result dates back to Pickands (1969a,b), and you can find a full and detailed proof in Leadbetter, Lindgren and Rootzén (1983). The fact that the proof is 16 pages long indicates that this is not an easy result, and we should not be too hopeful about obtaining extensions for processes on general parameter spaces or for non-stationary processes. (There is an extension to random fields on $\Re^d$, however. See Adler (1981) and references therein.)

Indeed, even in the case treated in Theorem 5.1, the result is not quite as strong as it at first seems. The problem lies in the fact that except for $\alpha = 1$ and $\alpha = 2$ (see Exercises 1.1 and 1.2) it is not known how to calculate the constant H_α in (5.4). (It *is* known that if Y is the non-stationary process on $\Re$ with mean $-|t|^\alpha$ and covariance function $|s|^\alpha + |t|^\alpha - |t-s|^\alpha$ then

$$H_\alpha = \lim_{T\to\infty} T^{-1} \int_{-\infty}^0 e^{-x} P\Big\{ \sup_{0\le t\le T} Y_t > -x \Big\}\, dx,$$

but this does not seem to very instructive, nor of much help in computing H_α.) Aldous (1989), has some non-rigorous bounds on the values of H_α for general α. Most of these arguments are based on his "Poisson clumping heuristic", but they also use Slepian's inequality to interpolate H_α between its known, exact values. For details, turn to Aldous' monograph, which also makes interesting reading for a number of other reasons.

Perhaps the main message of Theorem 5.1 is that although there is clearly very strong structure to asymptotic extrema distributions, and one should be able to do much better than (5.1) and (5.2), we should not spend too much time worrying about constants, since it is unlikely that we shall ever be able to calculate them.

If we now return to the general situation, there are basically three different cases that require study. The first, and in many ways easiest, is when X is non-homogeneous, and there exists a unique point $t_o \in T$ such that $EX_{t_o}^2 = \sup_T EX_t^2$. It turns out that in remarkably many situations

$$\lim_{\lambda\to\infty} \frac{P\{\|X\| > \lambda\}}{\Psi(\lambda/\sigma_T)} = 1. \tag{5.6}$$

In Section 5.3 we shall give a characterization of these cases, due to Talagrand (1988a).

The second case, which is also not too difficult, is when X is either homogeneous or has constant variance. In the purely homogeneous case, and often when X has constant variance, it is sometimes possible to identify a function $M(\lambda)$, which is closely related to the metric entropy function, and a constant K such that

$$(5.7) \qquad K^{-1}M(\lambda/\sigma)\Psi(\lambda/\sigma) \;\le\; P\{\|X\| > \lambda\} \;\le\; KM(\lambda/\sigma)\Psi(\lambda/\sigma),$$

for all sufficiently large λ, where σ^2 is now the constant variance $\sigma^2 = EX_t^2$. (Weber, (1988), has results in this direction, that are of the precise form (5.7) only when $T = [0,1]$ and the covariance function satisfies some side conditions. Patrik Albin has just informed me that he has some very general results – not yet written up – of this form.)

The last case is probably the most interesting, particularly from the point of view of applications to empirical processes, but also seems to be the hardest. It is certainly the case where results are least tight (i.e. the known upper and lower bounds do not always agree). This is the case in which X does not have a constant variance (and, *a fortiori*, is not stationary) and achieves its maximal variance over a comparatively large set. Typical examples are the pinned Brownian sheet on $[0,1]^2$, which achieves its maximal variance of $1/4$ along the line $st(1-st) = 1/4$, or set indexed versions of this process. We shall treat this case in depth in Section 5.4.

We start, however, with some easy calculations.

2. Some Easy Bounds.

As we have noted at least once already, it is a remarkable and very convenient fact that Borell's inequality, together with a simple application of entropy concepts, can be used to improve on Borell's inequality.

To see this, let X be continuous and let $\epsilon > 0$ be arbitrary. In general, ϵ may depend on λ. For $t \in T$, set

$$(5.8) \qquad \mu(t,\epsilon) \;=\; E \sup_{s\in B(t,\epsilon)} X_s,$$

and

$$(5.9) \qquad \mu(\epsilon) \;=\; \sup_{t\in T} \mu(t,\epsilon).$$

Since $N(\epsilon)$ d-balls of radius ϵ cover T, it is an immediate consequence of Borell's inequality that, for $\lambda > \mu(\epsilon)$,

$$(5.10) \qquad P\{\|X\| > \lambda\} \;\le\; 2N(\epsilon)\, e^{-\frac{1}{2}(\lambda-\mu(\epsilon))^2/\sigma_T^2},$$

where, as usual, $\sigma_T^2 = \sup_T EX_t^2$. The following result shows how, with very little work, (5.10) leads to almost the best bounds for $P\{\|X\| > \lambda\}$.

5.2 THEOREM. *If $N(\epsilon) \le K\epsilon^{-\alpha}$, then for all λ sufficiently large,*

$$P\{\|X\| > \lambda\} \le C_\alpha \lambda^{\alpha+1+\eta} \Psi(\lambda/\sigma_T). \tag{5.11}$$

for every $\eta > 0$, where K and $C_\alpha = C(\alpha, K)$ are finite constants.

REMARK: It is possible to replace the condition that λ be sufficiently large by the seemingly weaker requirement that $\lambda > 0$, simply by choosing, if necessary, a larger C_α. We shall do this is in the future. Nevertheless, you should keep in mind the fact that both Theorem 5.2 and those following are really sharpest for $\lambda \to \infty$. (We shall see, however, that in a least one example infinity is reached very quickly, and asymptotic theory is actually good for very moderate values of λ.)

PROOF: To apply (5.10) we must first compute $\mu(\epsilon, t)$. By Corollary 4.15

$$\begin{aligned} \mu(t,\epsilon) &\le C\int_0^\epsilon \left(\log N(\epsilon)\right)^{\frac{1}{2}} d\epsilon \\ &\le C\int_0^\epsilon \left(\log(1/\epsilon)\right)^{\frac{1}{2}} d\epsilon \\ &\le C\epsilon\sqrt{\log(1/\epsilon)}, \end{aligned} \tag{5.12}$$

where $C = C(\alpha)$ may change from line to line. Set $\epsilon = \epsilon(\lambda) = \lambda^{-1}$, choose λ large enough so that $\lambda > C\lambda^{-1}\sqrt{\log\lambda}$ for the C of (5.12), and substitute into (5.10) to obtain

$$\begin{aligned} P\{\|X\| > \lambda\} &\le C_1 \lambda^\alpha e^{-\frac{1}{2}\left(\lambda - C_2\lambda^{-1}\sqrt{\log\lambda}\right)^2/\sigma_T^2} \\ &\le C_3 \lambda^{\alpha+1} e^{C_4\sqrt{\log\lambda}}\Psi(\lambda/\sigma_T), \end{aligned} \tag{5.13}$$

(Recall – (2.1) – that $\Psi(\lambda) \sim \lambda^{-1}e^{-\frac{1}{2}\lambda^2}$.) Since for $\eta > 0$ and λ large enough $e^{C\sqrt{\log\lambda}} < \lambda^\eta$, this completes the proof. ■

The bound in (5.11) can be improved slightly, to $C(\lambda\log\lambda)^{\alpha+1}\Psi(\lambda/\sigma_T)$, by choosing $\epsilon = \epsilon(\lambda)$ in the above proof to satisfy $\lambda^{-1} = \epsilon(\log(1/\epsilon))^{1/2}$. The gain, however, is rather small, in view of the fact that the following is true:

5.3 THEOREM. *Under the conditions of Theorem 5.2, there exists a finite, positive C_α such that for all $\lambda > 0$*

$$\begin{aligned} C_\alpha^{-1}\lambda^\alpha\left(\log\lambda\right)^{-\alpha/2}\Psi(\lambda/\sigma_T) &\le P\{\|X\| > \lambda\} \\ &\le C_\alpha \lambda^\alpha \Psi(\lambda/\sigma_T). \end{aligned} \tag{5.14}$$

Theorem (5.3) is a special case of Theorem 5.8 of Section 4 below. What is important to note is that the simple proof we applied to prove Theorem 5.2 does not give the sharper upper bound of Theorem 5.3. While this is not surprising, what *is* surprising, however, is the fact that an almost identical, easy, proof *does* give the sharpest upper bound in the following result:

5.4 THEOREM. *If* $N(\epsilon) < a \exp(b\epsilon^{-\alpha})$, $0 < \alpha < 2$, *then for all* $\lambda > 0$,

$$P\{\|X\| > \lambda\} \le C_1 \exp\left(C_2 \lambda^{2\alpha/(2+\alpha)}\right) \Psi(\lambda/\sigma_T), \tag{5.15}$$

where a, b, C_1, C_2 *are finite, positive constants.*

REMARK: Note that you cannot set $\alpha = 0$ in this result to recover either Theorem 5.2 or Theorem 5.3. That the upper bound given here is, under mild side conditions, also a lower bound, is shown in Samorodnitsky (1987b, 1990).

PROOF: The proof is identical to that of the previous theorem, with the exception that (5.12) must be replaced by

$$\begin{aligned} \mu(t,\epsilon) &\le K \int_0^\epsilon \left(\log\left(a \exp(b\epsilon^{-\alpha})\right)\right)^{\frac{1}{2}} d\epsilon \\ &< K\epsilon^{1-\alpha/2}, \end{aligned}$$

and the relation between ϵ and λ given by $\epsilon = \lambda^{-1/(1+\alpha/2)}$. The details of the proof are left to you. ■

3. Processes with a Unique Point of Maximal Variance.

Neither Borell's inequality, nor the improvement made on it in the previous section, made any particular use of the fact that the variance of X_t may vary on T. In this section, we make our first step in this direction, with the following elegant result of Talagrand (1988a).

5.5 THEOREM. *Conditions (5.16) and (a)+(b) below are equivalent.*

$$\lim_{\lambda \to \infty} \frac{P\{\|X\| > \lambda\}}{\Psi(\lambda/\sigma_T)} = 1. \tag{5.16}$$

(a) *There exists a unique* $t_o \in T$ *such that* $EX_{t_o}^2 = \sigma_T^2 = \sup_T EX_t^2$.
(b) *If, for* $h > O$, *we define*

$$T_h = \{t \in T \colon E(X_t X_{t_o}) \ge \sigma_T^2 - h^2\}$$

then

$$\lim_{h \to 0} h^{-1} E\|X - X_{t_o}\|_{T_h} = 0. \tag{5.17}$$

In view of the results of Chapter 4, condition (5.17) is quite easy to check. Here is an example, treated initially by Berman (1985).

5.6 COROLLARY. *Let X_t, $t \ge 0$ be centered Gaussian with stationary increments and $X_0 = 0$. Assume*

$$p^2(t) = E(X_{t+s} - X_s)^2 = EX_t^2$$

is convex, and that

$$\lim_{t\to 0} \frac{p^2(t)}{t} = 0. \tag{5.18}$$

Then

$$\lim_{\lambda\to\infty} \frac{P\{\|X\|_{[0,T]} > \lambda\}}{\Psi(\lambda/\sigma_T)} = 1, \tag{5.19}$$

where $\sigma_T^2 = EX_T^2$.

PROOF: Since X has stationary increments, condition (a) of the theorem is clearly satisfied for each $T > 0$ with $t_o = T$. Consider (b). Since

$$EX_T X_t = \tfrac{1}{2}\big(p^2(T) + p^2(t) - p^2(T-t)\big),$$

it follows that for $t \in T_h$ we have $p^2(t) \ge p^2(T) - 2h^2 = \sigma_T^2 - 2h^2$.

Since $p^2(t)$ is convex, it has a left derivative at each point, and so (draw a picture) for $t \in T_h$ we have $t \ge T - Kh^2$ for some finite K. Thus, since X has stationary increments, to show that (5.17) holds it suffices to show that

$$\lim_{h\to 0} h^{-1} E \sup_{0\le t\le h^2} |X_t| = 0. \tag{5.20}$$

But this follows immediately from an entropy bound such as Corollary 4.15, along with the condition (5.18). ■

PROOF OF THEOREM 5.5: We shall follow our by now standard practice of only proving the theorem in one direction. In this case, however, we are proving it in the harder, albeit more interesting direction; i.e. that (a)+(b) implies (5.16). For the other half of the proof, see Exercise 3.1.

To save a little on notation, set $\sigma_T = 1$. With t_o the unique point of maximal variance, start by setting

$$a(t) = E(X_t X_{t_o}), \qquad Y_t = X_t - a(t)X_{t_o}.$$

A simple covariance calculation shows that Y_t is independent of X_{t_o}, and since

$$Y_t - (X_t - X_{t_o}) = \big(1 - a(t)\big)X_{t_o},$$

we have, from the definition of T_h, that

$$E\|Y_t - (X_t - X_{t_o})\|_{T_h} \;\leq\; h^2 E|X_{t_o}|.$$

Thus, if (5.17) holds, we also have

$$\lim_{h\to 0} h^{-1} E\|Y\|_{T_h} \;=\; 0 \tag{5.21}$$

Thus we need only show that, under condition (a), (5.21) implies (5.16).

The idea of the proof is as follows: Since the maximal variance is at t_o, it is most likely that if the process achieves a high supremum it will do so close to t_o. Thus we divide the parameter space up into sets close, closer, and even closer, etc. to t_o, treating each one separately, and noting that in each one there is a maximal variance, strictly less than 1, that, by Borell's inequality, governs the distribution of the supremum. The details are as follows:

Take a $\eta > 0$ with $\eta < 1/16$, and $\alpha \in (0, 1/8)$ such that

$$E\|Y\|_{T_h} \;\leq\; \eta h, \qquad \text{for all } h \leq \alpha. \tag{5.22}$$

For $n \geq 0$, define the following subsets of T:

$$A_n \;=\; T_{2^{-n}\alpha} \;=\; \left\{t\colon a(t) \geq 1 - 2^{-2n}\alpha^2\right\}, \qquad B_n \;=\; A_n \setminus A_{n+1}.$$

(Draw a picture for $T = [0,1]$, $t_o = 1$, to see what is happenning.)

Note that the A_n are monotonic non-increasing, the B_n disjoint and

$$1 - 2^{-2n}\alpha^2 \;\leq\; a(t) \;\leq\; 1 - 2^{-2(n+1)}\alpha^2, \qquad \text{for all } t \in B_n. \tag{5.23}$$

Since t_o is the unique point of maximal variance

$$\sup\{EX_t^2\colon\, a(t) \leq 1 - \alpha^2\} \;<\; 1.$$

It follows from Borell's inequality that

$$\lim_{\lambda\to\infty} \frac{P\left\{\sup\{X_t\colon\, a(t) \leq 1 - \alpha^2\} > \lambda\right\}}{\Psi(\lambda)} \;=\; 0.$$

Thus to prove the theorem we need only show that

$$\limsup_{\lambda\to\infty} \frac{P\{\|X\|_{A_0} > \lambda\}}{\Psi(\lambda)} \;\leq\; 1 + K\eta^2 \tag{5.24}$$

for some universal constant K. (The corresponding lower bound needed to establish (5.16) is, of course, trivial.) To do this, we use the same type of conditioning argument as in the previous two proofs. Using (5.23) we have that for $x > 0$

$$
\begin{aligned}
P\{\|X\|_{B_n} > \lambda \mid X_{t_0} = x\} &= P\{\|Y_t + a(t)x\|_{B_n} \geq \lambda\} \\
&\leq P\{\|Y\|_{B_n} > \lambda - x(1 - 2^{-2(n+1)}\alpha^2)\}.
\end{aligned}
\tag{5.25}
$$

To bound this, we need to bound $\sup_{B_n} EY_t^2 \leq \sup_{A_n} EY_t^2$. But

$$
\begin{aligned}
\sup_{A_n} (EY_t^2)^{\frac{1}{2}} &= \sup_{A_n} \sqrt{\pi/2}\,\{E|Y_t|\} \\
&\leq \sqrt{\pi/2}\,E\{\sup_{A_n} |Y_t|\} \\
&\leq \sqrt{\pi/2}\,\{E|Y_{t_0}| + 2E\|Y\|_{A_n}\} \\
&\leq \sqrt{2\pi}\,\eta\alpha\, 2^{-n} \\
&\leq \eta\alpha\, 2^{-n+2},
\end{aligned}
$$

where the first inequality is a property of normal distributions, the second obvious, the third a consequence of Lemma 3.1, the fourth a consequence of (5.22) and the fact that $Y_{t_0} = 0$, and the final trivial.

Applying this, along with (5.22) and Borell's inequality to (5.25), we find that for $x > 0$

$$
\begin{aligned}
&P\{\|X\|_{B_n} > \lambda | X_{t_0} = x\} \\
&\leq 2\psi\left(\frac{\lambda - x(1 - 2^{-2(n+1)}\alpha^2) - \eta\alpha 2^{-n}}{\eta\alpha 2^{-n+2}}\right) \\
&= 2\psi\left(\frac{\lambda - x(1 - 2^{-2(n+1)}\alpha^2)}{\eta\alpha 2^{-n+2}} - \tfrac{1}{4}\right),
\end{aligned}
\tag{5.26}
$$

where $\psi(x) = \sqrt{2\pi}\phi(x)$ and, as usual, $\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$. Set

$$
\begin{aligned}
I_1(\lambda) &= \int_0^\lambda P\{\|X\|_{A_0} > \lambda \mid X_{t_0} = x\}\,\phi(x)\,dx, \\
I_2(\lambda) &= \int_{-\infty}^0 P\{\|X\|_{A_0} > \lambda \mid X_{t_0} = x\}\,\phi(x)\,dx,
\end{aligned}
$$

so that

$$
\begin{aligned}
P\{\|X\|_{A_0} > \lambda\} &= P\{X_{t_0} > \lambda\} + P\{\|X\|_{A_0} > \lambda,\, X_{t_0} < \lambda\} \\
&= \Psi(\lambda) + I_1(\lambda) + I_2(\lambda).
\end{aligned}
\tag{5.27}
$$

Treating the easier term first, take $x \leq 0$ and note that then

$$\begin{aligned} P\{\|X\|_{A_o} > \lambda \mid X_{t_o} = x\} &= P\{\sup_{A_o}(Y_t + a(t)x) \geq \lambda\} \\ &\leq P\{\|Y\|_{t_o} \geq \lambda\}. \end{aligned}$$

Thus $I_2(\lambda) \leq P\{\|Y\|_{A_o} \geq \lambda\}$. Since $\sup_{A_o} EY_t^2 < 1$, Borell's inequality gives us that $\lim_{\lambda\to\infty} I_2(\lambda)/\Psi(\lambda) = 0$. Thus it remains only to treat $I_1(\lambda)$.

Since $A_0 = \cup_{n\geq 0} B_n$, it follows from (5.26) that

$$\begin{aligned} I_1(\lambda) &\leq \sum_{n=0}^{\infty} \int_0^{\lambda} P\{\|X\|_{B_n} > \lambda \mid X_{t_o} = x\}\, \phi(x)\, dx \\ &\leq \sum_{n=0}^{\infty} \int_0^{\lambda} 2\psi\left(\frac{\lambda - x(1 - 2^{-2(n+1)}\alpha^2)}{\eta\alpha 2^{-n+2}} - \tfrac{1}{4}\right) \phi(x)\, dx. \end{aligned}$$

Make the change of variable $x = \lambda - z$, set

$$\gamma_n = \frac{2^{-(n+4)}\alpha}{\eta}, \qquad \delta_n = \frac{1 - 2^{-2(n+1)}\alpha^2}{\eta\alpha 2^{-n+2}},$$

to obtain

$$I_1(\lambda) \leq 2\sum_{n=0}^{\infty} \int_0^{\infty} \psi\big(\gamma_n\lambda + z\delta_n - \tfrac{1}{4}\big)\, \phi\big(\lambda - z\big)\, dz. \tag{5.28}$$

Note that for all $\lambda > 1$, and all z there exists a finite K such that

$$\psi(\lambda - z) \leq \psi(\lambda)\, e^{\lambda z} \leq K\lambda\, \Psi(\lambda)\, e^{\lambda z}.$$

Applying this to (5.28) we find

$$I_1(\lambda) \leq 2K\, \Psi(\lambda) \sum_{n=0}^{\infty} \int_{n=0}^{\infty} \lambda\psi\big(\gamma_n\lambda + z\delta_n - \tfrac{1}{4}\big)\, e^{\lambda z}\, dz. \tag{5.29}$$

But, since $\alpha^2 < 1/8$, $\eta < 1/16$, we have

$$\gamma_n\delta_n = 2^{-6}\left(\frac{1 - 2^{-2(n+1)}\alpha^2}{\eta^2}\right) \geq \frac{2^{-7}}{\eta^2} \geq 2, \tag{5.30}$$

so that returning to (5.29) we find that, for appropriate constants K,

$$\begin{aligned} I_1(\lambda) &\leq K\, \Psi(\lambda) \sum_{n=0}^{\infty} \int_0^{\infty} \lambda \exp\left(- \tfrac{1}{2}(\gamma_n\lambda - \tfrac{1}{4})^2 - \tfrac{1}{2}(\delta_n z - \tfrac{1}{4})^2\right) dz \\ &\leq K\, \Psi(\lambda) \sum_{n=0}^{\infty} \lambda\, \delta_n^{-1} \exp\left(- \tfrac{1}{2}(\gamma_n\lambda - \tfrac{1}{4})^2\right). \end{aligned}$$

But, applying (5.30) again yields

$$I_1(\lambda) \ \le \ \eta^2 K\,\Psi(\lambda) \sum_{n=0}^{\infty} \lambda\gamma_n \exp\Big(-\tfrac{1}{2}(\gamma_n\lambda - \tfrac{1}{4})^2\Big). \tag{5.31}$$

Since $\gamma_{n+1} = \frac{1}{2}\gamma_n$, it is immediate that for large enough λ the sum is bounded by a universal constant.

Thus, $\lim_{\lambda\to\infty} I_1(\lambda)/\Psi(\lambda) \le \eta^2 K$, as required, and since η was arbitrary, the proof is complete. ■

If we are prepared to assume a little more, it is possible to do better than Theorem 5.5. In the notation of Theorem 5.5, set

$$L(h) \ = \ E \sup_{t\in T_h} \big(X_t - a(t)X_{t_o}\big).$$

Recall that condition (5.17) of the Theorem is equivalent (c.f. (5.21)) to

$$\lim_{h\to 0} h^{-1}L(h) \ = \ 0,$$

i.e. $L(h) = o(h)$. Dobric, Marcus and Weber (1989) have shown that if we assume a little more about $L(h)$ we can improve Theorem 5.5 as follows.

5.7 THEOREM. *Suppose there exists a unique $t_o \in T$ such that $EX^2_{t_o} = 1 = \sup_T EX_t^2$, and two functions ω_1, ω_2, concave for $h \in [0, \bar{h}]$ for some $\bar{h} > 0$ with $\omega_i(0) = 0$. Define*

$$h_i(\lambda) \ = \ \sup\Big\{h\colon \frac{\omega_i(h)}{h^2} = \lambda\Big\}, \qquad i = 1,2.$$

Then if

$$\omega_1(h) \ \le \ L(h) \ \le \ \omega_2(h)$$

and $\sup_{t\in T_h} E[X_t - X_{t_o}E(X_tX_{t_o})]^2 < (2-\epsilon)h^2$ for $h \in [0,\bar{h}]$ for some $\epsilon > 0$, there exist constants C_1 and C_2 such that for all λ large enough

$$e^{C_1\lambda\omega_1(h_1(\lambda))} \ \le \ \frac{P\{\|X\| > \lambda\}}{\Psi(\lambda)} \ \le \ e^{C_2\lambda\omega_2(h_2(\lambda))}.$$

If, for example,

$$\limsup_{h\to 0} \frac{L(h)}{\omega_1(h)} \ \ge \ 1,$$

then, for all $\epsilon > 0$,

$$\limsup_{\lambda\to\infty} \frac{P\{\|X\| > \lambda\}}{\Psi(\lambda)\exp(k_1(1-\epsilon)\lambda\omega_1(h_1(\lambda)))} \ \ge \ 1.$$

You can find a proof and applications of this result in Dobric, Marcus and Weber (1988). I brought it here to show you two things: Firstly, one *can* do better than Theorem 5.5. Secondly, in terms of the simplicity of the result, one pays a lot the improvement.

4. General Bounds.

At the time of writing, this is the only remaining area in the general theory of Gaussian extremes (over bounded T) that is not yet fully "tidied up" under the influence of Borell's inequality. Thus, we shall suffice with a statement of the main result, without proof.

Nevertheless, the result and proof of the previous section give us a strong indication of how the proof must work. There, we treated processes with a unique maximal variance, and the proof relied on studying just how fast the variance decayed as one moved away from that point. In the general case, there will be a subset $T_{\max} \subset T$ where the maximal variance is achieved, and one has to study two things: the "size" of $T_{\max}$ (e.g. as measured in terms of metric entropy) and how rapidly EX_t^2 decays as we move out of $T_{\max}$.

To quantify this, define the following, for $\delta, \delta_1, \delta_2, \epsilon > 0$:

$$T_\delta^+ = \{t \in T : EX_t^2 > \delta\} \qquad T_\delta^- = \{t \in T : EX_t^2 \leq \delta\}$$

$$N^+(\delta, \epsilon) = N(T_\delta^+, \epsilon), \qquad N^-(\delta, \epsilon) = N(T_\delta^-, \epsilon)$$

$$N(\delta_1, \delta_2, \epsilon) = N(T_{\delta_1}^+ \cap T_{\delta_2}^-, \epsilon) \quad 0 < \delta_1 < \delta_2,$$

where $N(T', \epsilon)$ is the minimum number of d-balls of radius ϵ required to cover $T' \subseteq T$.

It is the function $N(\delta_1, \delta_2, \epsilon)$ that is crucial in measuring the non-stationary of X on T. Here is the main result, due to Samorodnitsky (1987a, 1990) and Adler and Samorodnitsky (1987).

5.8 THEOREM. *Suppose that for finite, positive a, α, β, the following conditions are satisfied for small enough, positive, δ and ϵ:*

$$a^{-1}\epsilon^{-\alpha}\delta^{\beta} \leq N^+(\sigma_T - \delta, \epsilon) \leq a\epsilon^{-\alpha}\delta^{\beta}. \tag{5.32}$$

Then, if $\alpha \geq 2\beta$, there exists a finite, positive K such that for all $\lambda > 0$

$$\begin{aligned} K^{-1}\lambda^{\alpha-2\beta}(\log\lambda)^{-\alpha/2}\Psi(\lambda/\sigma_T) &\leq P\{\|X\| > \lambda\} \\ &\leq K\lambda^{\alpha-2\beta}(\log\lambda)^{\beta}\Psi(\lambda/\sigma_T). \end{aligned} \tag{5.33}$$

If $\alpha < 2\beta$, then

$$K^{-1}\Psi(\lambda/\sigma_T) \leq P\{\|X\| > \lambda\} \leq K\Psi(\lambda/\sigma_T). \tag{5.34}$$

Under the additional assumption

$$N(\delta_1, \delta_2, \epsilon) \leq C\epsilon^{-\alpha}(\delta_2 - \delta_1)^{\beta}, \tag{5.35}$$

the upper bound in (5.33) holds without the logarithmic term.

This result, of course, treats only polynomial entropy functions of the type appearing in Theorems 5.2 and 5.3. (Note, in fact, that setting $\beta = 0$ above in fact gives Theorem 5.3.) The reason that we have not added the exponential entropy case to theorem 5.8 is due to the fact, already noted, that the easy bound of Theorem 5.4 is actually the best possible.

We shall conclude this section with two applications of Theorem 5.8. The first, which treats a non-stationary process on $[0,1]$, is designed to show how sample roughness and non-stationarity interact to determine the distribution of $\|X\|$. (You can get a less interesting result for stationary processes here by setting the parameter β to zero.) The second shows how to use the general theory to do an empirical process problem.

5.9 EXAMPLE. *Let X be stationary, centered on $[0,1]$ with covariance function R_X satisfying*

$$a_0 t^{\alpha} \leq 1 - R_X(t) \leq a_1 t^{\alpha}, \tag{5.36}$$

for all $t \in [0,\gamma]$ and a_0, a_1, α and γ positive. Let $\sigma(t)$ be positive, continuous, and monotonic increasing on $[0,1]$ such that

$$b_0 |t-s|^{\beta} \leq |\sigma(t) - \sigma(s)| \leq b_1 |t-s|^{\beta} \tag{5.37}$$

for all $s,t \in [0,1]$ and b_0, b_1 and β positive, and define a new, nonstationary, process Y by

$$Y(t) = \sigma(t)X(t), \qquad t \in [0,1]. \tag{5.38}$$

Finally, let σ, without a parameter, denote $\sigma(1)$. Then, if $0 < \beta \leq \alpha$, we have that for sufficiently large λ there exists a finite $C > 0$ such that

$$C^{-1}\lambda^{-1}e^{-\frac{1}{2}\lambda^2/\sigma^2} \leq P\{\sup_{t\in[0,1]} Y(t) > \lambda\} \leq C\lambda^{-1}e^{-\frac{1}{2}\lambda^2/\sigma^2}. \tag{5.39}$$

If, however, $0 < \alpha < \beta$, then

$$\begin{aligned} C^{-1}\lambda^{-1-\frac{2}{\beta}+\frac{2}{\alpha}}(\log\lambda)^{-\frac{1}{\alpha}}e^{-\frac{1}{2}\lambda^2/\sigma^2} &\leq P\{\sup_{t\in[0,1]} Y(t) > \lambda\} \\ &\leq C\lambda^{-1-\frac{2}{\beta}+\frac{2}{\alpha}}e^{-\frac{1}{2}\lambda^2/\sigma^2}. \end{aligned} \tag{5.40}$$

PROOF: Clearly everything hinges on finding good upper and lower bounds for $N(\delta_1, \delta_2, \epsilon) = N(T^+_{\delta_1} \cap T^-_{\delta_2}, \epsilon)$, $0 < \delta_1 < \delta_2$, so that we can establish (5.32) and (5.35). We shall show how to get an upper bound. That the same expression also serves as a lower bound is left as an exercise.

Note first that for $\sigma(0) \leq \delta_1 \leq \delta_2$,

$$\begin{aligned} T^+_{\delta_1} \cap T^-_{\delta_2} &= \{t \in [0,1] : \delta_1 \leq \sigma_t^2 < \delta_2\} \\ &= I(\delta_1, \delta_2) \end{aligned}$$

Note that, in view of (5.37), the length of the interval $I(\delta_1, \delta_2)$ lies between $((\delta_2 - \delta_1)/b_2)^{1/2}$ and $((\delta_2 - \delta_1)/b_1)^{1/2}$. Furthermore, in view of the definition (5.38),

$$\begin{aligned} E(Y_t - Y_s)^2 &= (\sigma_t - \sigma_s)^2 + 2\sigma_t\sigma_s(1 - R_X(t-s)) \\ &\leq b_1^2 |t-s|^{2\beta} + 2\sigma^2 a_1 |t-s|^\alpha, \end{aligned} \tag{5.41}$$

the inequality holding for all $|t - s| < \gamma$, by (5.38).

Moving now to the canonical metric on $[0,1]$ in order to calculate entropies, it follows that, if $0 < \alpha < 2\beta$, then (5.41) implies

$$d(s,t) \leq C\,|t-s|^{\alpha/2}.$$

Combining this with the comments above on the length of $I(\delta_1, \delta_2)$, we have that

$$N(\delta_1, \delta_2, \epsilon) \leq C\epsilon^{-2/\alpha}(\delta_2 - \delta_1)^{1/2}. \tag{5.42}$$

This is enough to establish the upper bounds in both (5.32) and (5.35) and so yield, by the Theorem, and Exercise 4.1, (5.39) and (5.40) for the case $\alpha < 2\beta$.

The case $\alpha \geq 2\beta$ is similarly treated, and leads to (5.39). This completes the proof. ∎

Our second example deals with the set indexed, pinned, Brownian sheet $\mathring{W}$ on the collection

$$\mathcal{A}_k = \{[s,t] : s,t \in [0,1]^k\}$$

of intervals in the k-dimensional unit square. Recall that if we take a general reference (probability) measure ν on $[0,1]^k$ then $\mathring{W}$ is the centered Gaussian process with covariance function

$$E\mathring{W}(A)\mathring{W}(B) = \nu(A \cap B) - \nu(A)\nu(B).$$

The result is

5.10 EXAMPLE. *Assume that the measure ν has a bounded density f that is everywhere positive. Then, with the above notation, there exists a finite $C > 0$ such that for large enough λ*

$$\begin{aligned} C^{-1}\lambda^{2(2k-1)}(\log\lambda)^{-2k}e^{-2\lambda^2} &\leq P\{\sup_{A\in\mathcal{A}_k} \mathring{W}(A) > \lambda\} \\ &\leq C\lambda^{2(2k-1)}e^{-2\lambda^2}. \end{aligned} \tag{5.43}$$

REMARK: The condition on ν can be relaxed to demanding that ν has a strictly positive density on some interval $[s-\epsilon, t+\epsilon]$ $(s,t,\epsilon \in [0,1]^k)$ for which $\nu([s,t]) \geq \frac{1}{2}$, or, indeed, even further to those used by Adler and Brown (1986) in obtaining corresponding bounds (but without the logarithmic term) for $\mathring{W}$ over the index set $\{[0,t]: t \in [0,1]^k\}$. See Samorodnitsky (1987a, 1990) for details.

PROOF: A detailed proof involves an unentertaining amount of algebra, so we shall suffice with only an outline.

For $0 \leq \delta_1 < \delta_2 \leq \frac{1}{2}$,

$$\begin{aligned} T^+_{\delta_1} \cap T^-_{\delta_2} &= \{A \in \mathcal{A}_k : \delta_1 \leq \nu(A) - \nu^2(A) \leq \delta_2\} \\ &= I^-_{\delta_1,\delta_2} \cup I^+_{\delta_1,\delta_2}, \end{aligned}$$

where

$$\begin{aligned} I^-_{\delta_1,\delta_2} &:= \{A \in \mathcal{A}_k : \tfrac{1}{2} - (\tfrac{1}{4}-\delta_1)^{\frac{1}{2}} \leq \nu(A) \leq \tfrac{1}{2} - (\tfrac{1}{4}-\delta_2)^{\frac{1}{2}}\} \\ I^+_{\delta_1,\delta_2} &:= \{A \in \mathcal{A}_k : \tfrac{1}{2} + (\tfrac{1}{4}-\delta_1)^{\frac{1}{2}} \leq \nu(A) \leq \tfrac{1}{2} + (\tfrac{1}{4}-\delta_2)^{\frac{1}{2}}\} \end{aligned}$$

Consider approximating the sets in I^- in terms of the canonical distance, in order to find an upper bound for the entropy function. (I^+ is similar. The lower bound is left to you. It is a little harder.) Choose as approximating sets k-dimensional intervals whose endpoints sit on the points of the lattice

$$L^k_\epsilon = \{t \in [0,1]^k : t_i = n_i\epsilon^2,\ n_i \in (0,1,\ldots,[\epsilon^{-2}]),\ i = 1,\ldots,k\}.$$

What now has to be shown is that one needs $O(\epsilon^{-4k}(\delta_2-\delta_1)^{\frac{1}{2}})$ such intervals to approximate every set in $I^-_{\delta_1,\delta_2}$ in the canonical metric. This, along with a similar bound for $I^-_{\delta_1,\delta_2}$ yields that

$$N(\delta_1,\delta_2,\epsilon) \leq C\epsilon^{-4k}(\delta_2-\delta_1)^{\frac{1}{2}},$$

which, in view of Theorem 5.9, is all we need in order to establish (5.43). ■

Once one has worked one of these examples in detail, it is relatively easy to work others. For example, for the Brownian sheet (based on Lebesgue measure) on $\{\{\text{half planes}\} \cap [0,1]^2\}$ a result similar to (5.43) holds, but with the power of λ reduced to 2 and the power of the logarithm equal to -2. (Samorodnitsky (1987a, 1990).) As I noted way back in the Preface, the monograph by Piterbarg (1988), which I received only recently, has a wealth of information related to the results of this subsection. His results also cover the more classical "extremal theory" of Gaussian processes, that arises when T is replaced by a sequence $\{T_n\}_{n \geq 1}$ of subsets of Euclidean space that grow towards all of $\Re^d$ as $n \to \infty$.

5. The Brownian Sheet on the Unit Square.

After all the general theory of this chapter it is rather natural to ask whether one could not do better, given a particular process, calculating $P\{\|X\| > \lambda\}$ for this process alone. Perhaps, then, it would be possible to identify the elusive constants in the general bounds, or, better still, calculate the precise distribution of the supremum. Sometimes, but rarely, this is possible. For example, there are only six different stationary Gaussian processes – i.e. six different covariance functions – on $\Re$ for which the precise distribution of the supremum over finite intervals is known.

One is the simple cosine process of Exercise 1.1, and another the process with triangular covariance function (Exercise 1.2). The most important is the Ornstein-Uhlenbeck process with covariance $\exp(|t|)$, and the other three are somehow related to either this or the triangular covariance process. In all of these six cases, however, the derivation of the distribution of the maximum is based not on general Gaussian techniques but either on the simplistic nature of the process (as in the cosine process) or on Markov methods. (The Ornstein-Uhlenbeck process is, of course, Markovian. The triangular covariance process is pseudo-Markovian. See Adler (1981) for definitions and references.) Thus these six "success stories" are not really part of the Gaussian framework.

Within the Gaussian framework, we noted already at the beginning of this chapter that pre-entropy results identify the constants appearing in the asymptotic bounds when $T = [0,1]$ and $R(s-t) \sim |t-s|^\alpha$ as $|t-s| \to 0$. However, even there it is not known how to obtain their precise *numerical* values. In general, even establishing the fact that such constants exist is an impossible task.

Nevertheless, in this section I want to show you how to calculate the constants for one very special process, the pinned Brownian sheet on $[0,1]^2$. There are two reasons for this. Firstly, the result has importance in bivariate Kolmogorov-Smirnov testing. Secondly, it is my favourite result in the entire theory of Gaussian random fields (i.e. processes defined on $\Re^k, k > 1$). The result and proof is due to Goodman (1975). A related upper bound is due to Cabaña and Wschebor (1981). It is harder to obtain and not as neat. A higher dimensional version, and a linkage of the result to Kolmogorov-Smirnov tests for empirical processes is given in Adler and Brown (1986). The only other example that I know of for which a similar argument works is in Adler (1984), where the random field on $\Re^2$ with covariance function

$$R(s,t) = \prod_{i=1}^{2} \max\left(0, 1 - |t_i - s_i|\right)$$

is treated. (It is probably worthwhile reiterating at this point that one of the reasons that special results of this type are of interest is that information on a specific process, along with Slepian's inequality, yields information on many others. See, for example, applications of this idea in Orsingher (1987) and Orsingher and Bassan (1988).)

Finally, it is important to note that the reasons that we can get results for the Brownian sheet are more related to its Markovian nature than its Gaussian properties. Nevertheless, it is a lovely calculation, particularly when one recalls its history, and the fact that it grew out of the theory of Banach space valued processes. It is rare indeed that such abstract theory can be used to carry out a calculation that provides useful numbers.

Once again, recall that the Brownian sheet on $[0,1]^2$ is the centered Gaussian process with covariance

$$EW_{st}W_{s't'} = (s \wedge s') \cdot (t \wedge t') \tag{5.44}$$

and the pinned sheet, a version of which is given by

$$\mathring{W}_{st} = W_{st} - stW_{11}. \tag{5.45}$$

has covariance function

$$E\{\mathring{W}_{st}\mathring{W}_{uv}\} = su(1-tv), \qquad (s,t),(u,v) \in [0,1]^2. \tag{5.46}$$

5.11 THEOREM. *If $\mathring{W}$ is the pinned Brownian sheet on $[0,1]^2$, then for all $\lambda > 0$*

$$P\{\sup_{[0,1]^2} \mathring{W}_{st} > \lambda\} > (1+2\lambda^2)e^{-2\lambda^2}. \tag{5.47}$$

The lower bound is, from the point of view of applications, the important one. For $\lambda > 1.5$, it can be shown that it underestimates the true probability by no more than .01, and that it actually gets tighter as $\lambda \to \infty$. (See Theorem 5.12 and the table following it.)

PROOF: The trick is to turn W_{st} from a real-valued, two-parameter process, to a Banach space valued, single parameter, Markov process, by, for each $t \in [0,1]$ letting W_t be the element of $C[0,1]$ defined pointwise by

$$W_t(s) = W_{st}, \qquad 0 \le s \le 1. \tag{5.48}$$

The proof will then rely on the fact that

$$\sup_{st} W_{st} = \sup_t(\sup_s W_t(s)). \tag{5.49}$$

We need to collect a few facts, however, before we can exploit this equivalence fully.

The first is that W_t is, in fact, a $C[0,1]$ valued Markov process. Since this statement is beyond the general interest of these notes, we shall assume it is correct. (If you are uncomfortable with such processes, you can take $0 \le s_1 \le ... \le s_k \le 1$, and work with the $\Re^k$-valued process $W_t^{(k)} = (W(s_1,t)...,W(s_k,t))$ in all that follows and, ultimately, send $k \to \infty$. There is no difficulty checking the Markovian nature of $W^{(k)}$).

The second fact is that one way to obtain $\mathring{W}$ from W is simply to condition W on taking the value zero at the point (1,1). This follows by comparing covariance functions.

Thirdly, since W_t is a well defined Markov process, it makes sense to condition W_t, $0 \le t < 1$ on W_1. A surprising fact, which is what makes the whole proof work, is that $W_t(s)$, conditioned on $W_1(s)$, is independent of $W_1(u)$ for all $u \ne s$. To see this, note that for any $0 \le s, u \le 1$, $0 \le t \le 1$

$$\begin{aligned} E\{(W_t(s) - tW_1(s))W_1(u)\} &= E\{(W_{st} - tW_{s1})W_{u1}\} \\ &= (s \wedge u)(t \wedge 1) - t(s \wedge u)(1 \wedge 1) \\ &= 0. \end{aligned} \tag{5.50}$$

Thus $E\{W_t(s) \mid W_1\} = tW_1(s)$, and our claim is established.

(It is worthwhile reiterating that (5.50) is a *very* special relationship, that will not work, for example, if the Gaussian white noise that W is based on is defined relative to any measure other than Lebesgue.)

The fourth, and final, fact, is that the process W_1, conditional on $W_1(1) = 0$, is a standard Brownian bridge (i.e. the one-dimensional Brownian sheet).

We can now start the calculation:

$$\begin{aligned} P\{\sup_{st} \mathring{W}_{st} < \lambda\} &= P\{\sup_{st} W(s,t) < \lambda \mid W(1,1) = 0\} \\ &\le \inf_s P\{\sup_t W_t(s)) < \lambda \mid W(1,1) = 0\} \\ &= \int_{\Omega^*} \inf_s P\{\sup_t W_t(s) < \lambda \mid W_1 = w\}\, dP^*(\omega), \end{aligned}$$

where P^* is the probability measure on $\Omega^* = \{f \in C[0,1]: f(0) = f(1) = 0\}$ induced by the standard Brownian bridge. (If the last equality bothers you – and it should – see Goodman's original proof for technical elucidation.)

By the third fact established above, we now have that

$$P\{\sup_{st} \mathring{W}_{st} < \lambda\} \le \int_{\Omega^*} \inf_s P\{\sup_t W_t(s) < \lambda \mid W_1(s) = \omega(s)\}\, dP^*(\omega). \tag{5.51}$$

But for each s, the real valued process $\{W_t(s)\}_{t\ge 0}$ is just a scaled Brownian motion, and the conditional probability that $W_t(s) < \lambda$ for all $0 \le t \le 1$, given the value of $W_1(s)$, is well known, as a property of Brownian motion. In fact (e.g. Feller (1971), page 341) we have

$$P\{\sup_t W_t(s) \; < \; \lambda \mid W_1(s) = w(s)\} \; = \; 1 - \exp\Big(- 2\lambda \frac{(\lambda - w(s))}{s}\Big).$$

Substituting into (5.51) now gives

$$P\{\sup_{st} \mathring{W}_{st} < \lambda\} \; \le \; \int_{\Omega^*} \Big\{1 - \exp\Big(2\lambda \sup_s \frac{(w(s) - \lambda)^-}{s}\Big)\Big\}\, dP^*(w). \tag{5.52}$$

However, the distribution of $Z = \sup_s\{(w(s) - \lambda)/s\}$ when ω is a Brownian bridge is also known, and has density

$$p(z) \; = \; \begin{cases} \lambda e^{-2\lambda(z+\lambda)} & z > -\lambda, \\ 0 & z < -\lambda. \end{cases}$$

Substituting into (5.52) gives

$$\begin{aligned} P\{\sup_{st} \mathring{W}_{st} < \lambda\} \; &\le \; \int_{-\lambda}^{0} \big(1 - e^{2\lambda z}\big)\, 2\lambda\, e^{-2(z+\lambda)\lambda}\, dz \\ &\le 1 - \big(1 + 2\lambda^2\big)\, e^{-2\lambda^2}. \end{aligned}$$

Consequently,

$$P\{\sup_{st} \mathring{W}_{st} > \lambda\} \; \ge \; \big(1 + 2\lambda^2\big)\, e^{-2\lambda^2},$$

as required. ■

While it is true that Theorem 5.11 is one of my favourite results, I should point out that one can actually do considerably better. Furthermore, whereas the proof given above relies on Markov and Banach space techniques, the better result is a truly Gaussian one. That is, the proof of the result below proceeds by noting that the largest contribution to the probability of the pinned Brownian sheet ever crossing a high level comes from that part of $[0,1]^2$ where the marginal probability of being above the level is highest; i.e. where the variance is highest. This region can then be carefully broken up into smaller pieces, and each one studied independently. This, of course, is precisely the technique that we have employed throughout this chapter. However, when carefully applied to a particular process, it is possible to keep track of all the constants involved, and Hogan and Seigmund (1986) (to whom you should turn for a proof) have thus managed to prove the following:

5.12 THEOREM. *If $\mathring{W}$ is the pinned Brownian sheet on $[0,1]^2$, then as $\lambda \to \infty$*

$$P\{\sup_{[0,1]^2} \mathring{W}_{st} > \lambda\} \sim 4\log 2\, \lambda^2 e^{-2\lambda^2}. \tag{5.53}$$

Note that in one sense, Theorem 5.12 is a little weaker than Theorem 5.11, as the latter gives an inequality for *all* $\lambda > 0$. It is interesting, nevertheless, to compare the two numerically, and so note that "∞" is a lot closer to zero than one might initially expect.

The following table does this, giving $P\{\sup_{[0,1]^2} \mathring{W}_{st} > \lambda\}$ for a range of λ, according to formulae (5.47), (5.53) and an approximation based on 400 simulations of a discrete approximation to a pinned Brownian sheet on a 100×100 grid on $[0,1]^2$ (graciously provided by Ron Pyke).

PINNED BROWNIAN SHEET TAIL PROBABILITIES			
λ	Lower bound	Asymptotic	Simulation
0.25	.9928	.1529	.9883
0.50	.9098	.4204	.9226
0.75	.6899	.5063	.7133
1.00	.4060	.3752	.4191
1.25	.1812	.1903	.1737
1.50	.0611	.0693	.0503
1.75	.0156	.0186	.0116
2.00	.0030	.0037	.0000
2.25	.0004	.0006	.0000

If you have bothered to read these figures then you are enough of a statistician to know that the last digit in the simulation results is totally spurious, and that very little faith should be had in even the third.

A little thought also shows that at the extreme tail the simulation figures should give lower probabilities than the true ones, since an approximation to $\mathring{W}$ on a finite grid will have a stochastically smaller supremum than $\mathring{W}$ itself.

Nevertheless, the overall agreement, at least for $\lambda \geq 1.25$ is surprisingly good, so that one is rather tempted to claim that, with an error of only 1%, $\infty \approx 2.5\sigma$! (Since $\sigma_T^2 = 1/4$.) In fact, experience, based both on numerical approximations and simulation indicates that this approximation is good for a wide variety of extremal results. ■

There is, of course, one thing wrong with both of the above Theorems: both refer only to the Brownian sheet based on the uniform measure on $[0,1]^2$. With only a little work, and an application of Slepian's inequality, one can do a lot better.

With a slight change of notation, so that the proof below will be easier to read, let F be a continuous distribution function on $\Re^2$. Write $\mathring{W}_F$ to denote the pinned Brownian sheet based on the measure with distribution function F, so that for $s, t \in \Re^2$

$$E\mathring{W}_F(s)\mathring{W}_F(t) = F(s \wedge t) - F(s)F(t), \tag{5.54}$$

where $s \wedge t$ is the coordinatewise minimum $(s_1 \wedge t_1, s_2 \wedge t_2)$.

Furthermore, let G denote the degenerate distribution, uniform on the negatively sloped diagonal $t_1 + t_2 = 1$ in $[0,1]^2$, so that

$$G(t) = (t_1 + t_2 - 1)^+, \qquad t \in [0,1]^2. \tag{5.55}$$

Here is a rather useful and somewhat surprising result:

5.13 THEOREM. *Let F be any continuous distribution function on $\Re^2$. For any $\lambda > 0$*

$$P\{\sup_{t \in \Re^2} \mathring{W}_F(t) > \lambda\} \leq P\{\sup_{t \in [0,1]^2} \mathring{W}_G(t) > \lambda\}. \tag{5.56}$$

Furthermore

$$P\{\sup_{t \in [0,1]^2} \mathring{W}_G(t) > \lambda\} \leq \sum_{n=1}^{\infty} (8n^2\lambda^2 - 2)\, e^{-2n^2\lambda^2}. \tag{5.57}$$

The value in this result lies in its applications to bivariate Kolmogorov-Smirnov tests. You can find more details in Adler and Brown (1986) from where the result comes, as well as in Adler, Brown and Lu (1990).

PROOF: We shall assume, without loss of generality, that F is concentrated on $[0,1]^2$ with uniform marginals. (Why is this true? Also, since this *is* true, why is it not true that we can also assume F to be uniform on $[0,1]^2$? Why *can* we assume this in one dimension?)

To save on space, write I^2 for $[0,1]^2$, and let m be the mapping from I^2 to the upper triangle $\{t \in I^2 : t_1 + t_2 \geq 1\}$, defined by

$$G(m(t)) = G(m_1(t), m_2(t)) = F(t), \qquad t \in I^2, \tag{5.58}$$

$$m_2(t) - m_1(t) = t_2 - t_1, \qquad t \in I^2. \tag{5.59}$$

We must check that m is well defined. But this is easy, for by (5.59) we have that we are mapping lines of slope one onto themselves. At the lower left end of each such line $F = G = 0$, and at the upper right end, as each line leaves I^2, the marginal uniformity of F and G ensures that they are again equal. Since both F and G are nondecreasing along such lines, and G is continuous and strictly increasing when it is not zero, it is easy to arrange (5.58) in a unique fashion.

Now consider the processes $\mathring{W}_F$ and $\mathring{W}_G$. We shall compare their suprema via Slepian's inequality, for which we need to compare variances and covariances. Note firstly that, for $t \in I^2$,

$$E\big(\mathring{W}_F(t)\big)^2 = E\big(\mathring{W}_G(m(t)\big)^2, \tag{5.60}$$

a simple consequence of (5.58) and (5.54). We want to show that

$$F(s \wedge t) \geq G\big(m(s) \wedge m(t)\big), \tag{5.61}$$

from which it would follow that

(5.62)
$$\begin{aligned} E\mathring{W}_F(s)\mathring{W}_F(t) &= F(s \wedge t) - F(s)F(t) \\ &\geq G\big(m(s) \wedge m(t)\big) - G\big(m(s)\big)G\big(m(t)\big) \\ &= E\mathring{W}_G\big(m(s)\big)\mathring{W}_G\big(m(t)\big). \end{aligned}$$

This will be enough to prove the first part of the Theorem, viz (5.56), since (5.60) and (5.62) are precisely the ingredients for Slepian's inequality.

Therefore, consider (5.61), and assume that $s < t$ in the sense of coordinatewise ordering. (i.e. $s_1 < t_1$ and $t_1 < t_2$.) Then $F(s \wedge t) = F(s) = G(m(s)) \geq G(m(s) \wedge m(t))$, and so (5.61) holds. The case $t < s$ is clearly identical, so consider now the case $s_1 > t_1$ and $s_2 < t_2$. (The remaining case is handled identically.) Then $s \wedge t = (t_1, s_2)$. Write $w = (m_1(t), m_2(s))$. There are three possible cases to consider: $m(s) \geq w \geq m(t)$, $m(t) \geq w \geq m(s)$, and $w = m(s) \wedge m(t)$. We shall look at the third case only, but you should check that our reasoning is valid for all three. Note (draw a picture!) that

$$\begin{aligned} F(s \wedge t) &= F(t_1, s_2) \\ &\geq \big[F(s) - (s_1 - t_1)\big] \vee \big[F(t) - (y_2 - s_2)\big] \\ &\qquad\qquad\qquad\qquad\qquad \text{(by marginal uniformity)} \\ &\geq \tfrac{1}{2}\big\{F(s) + F(t) - \big[(s_1 - s_2) - (t_1 - t_2)\big]\big\} \\ &= \tfrac{1}{2}\big\{G\big(m(s)\big) + G\big(m(t)\big) \\ &\qquad - \big[\big(m_1(s) - m_2(s)\big) - \big(m_1(t) - m_2(t)\big)\big]\big\} \quad \text{by (5.59), (5.60)} \\ &\geq m_2(s) + m_1(t) - 1 \quad \text{by (5.55).} \end{aligned}$$

Hence, if $m_2(s) + m_1(t) - 1 \geq 0$, then the above yields

$$F(s \wedge t) \geq G\big(m_1(t), m_2(s)\big). \tag{5.63}$$

On the other hand, if $m_2(s) + m_1(t) - 1 < 0$, then $G(m_1(t), m_2(s)) = 0$ and so (5.63) is trivially true. Thus, in general,

$$F(s \wedge t) \geq G\big(m_1(t), m_2(s)\big) = G(w) = G\big(m(s) \wedge m(t)\big).$$

From this we immediately obtain (5.61) and so the proof of (5.56).

To complete our proof we must establish (5.57). To this end, we define a two-parameter process $X(t)$ on I^2 by setting

$$X(t_1, t_2) = \begin{cases} \mathring{W}(t_1) - \mathring{W}(1 - t_2) & t_1 + t_2 - 1 \geq 0 \\ 0 & t_1 + t_2 - 1 \geq 0, \end{cases}$$

where $\mathring{W}$ is the usual single parameter Brownian bridge $\mathring{W}$ on $[0, 1]$. Comparison of covariance functions shows that X is a version of $\mathring{W}_G$. Thus

$$\begin{aligned} P\{\sup_{t \in I^2} \mathring{W}_G(t) > \lambda\} &= P\big\{\sup\big[\mathring{W}(t_1) - \mathring{W}(t_2) : t_1 + t_2 - 1 \geq 0\big] > \lambda\big\} \\ &= P\big\{\sup\big[\mathring{W}(t) - \mathring{W}(s) : 0 \leq s < t \leq 1\big] > \lambda\big\} \\ &\leq P\big\{\sup\big[\mathring{W}(t) - \mathring{W}(s) : s, t \in [0, 1]\big] > \lambda\big\} \\ &= P\Big\{\Big[\sup_{[0,1]} \big(\mathring{W}(t)\big)^+ + \sup_{[0,1]} \big(\mathring{W}(t)\big)^-\Big] > \lambda\Big\}. \end{aligned} \tag{5.64}$$

What we have gained here is that the supremum probability for the two-parameter $\mathring{W}_G$ has been expressed in terms of the single parameter process $\mathring{W}$, and, fortunately for us, Kac, Kiefer and Wolfowitz (1955, equation (4.6)) have shown, via clever use of the reflection principle – which in turn relies on the Markov type properties of the Brownian bridge – that the last probability above is precisely that given by the right hand side of (5.57). This completes our proof. ∎

REMARK: For the statistician who is still with us, and who might be interested in good, numerical, bounds, we should note that the one inequality in (5.64) is far from sharp. While it retains the right order of magnitude, a little thought shows that it "costs", roughly, a factor of two. Comparison of the upper bound of Theorem 5.13 and the lower bound of Theorem 5.11 (albeit for a specific process) seems to bear this out, as do simulations in Adler, Brown and Lu (1990), to where, once again, I refer you for details on how to use all of this.

Finally, it is interesting to ask what happens in treating the Brownian sheet in dimensions greater than two. Do similar results hold? Unfortunately, the answer seems to be negative, since there is no distribution in dimensions ≥ 3 which plays the same "extremal" role that G does in two dimensions.

You can look in Adler and Brown (1986) to see what happens in higher dimensions, but all you will find there are special cases of the very general theorems of this chapter. In fact, it was the results obtained there, for particular processes, that got me interested in the general formulation. Since I now seem to have got back to where I started, the cycle is complete, and the time has come to stop writing.

6. Exercises.

SECTION 5.1:

1.1 One of the useful consequences of Theorem 5.1 is that the asymptotic distribution of the maximum of stationary Gaussian processes on $\Re$ depends on the covariance only through the constants C and α in the expansion (5.3). Consequently, if we can find the value of H_α in (5.4) for a particular process X with covariance function satisfying (5.3) for a specific α, we have H_α for all such processes.

(i) Let A be a Rayleigh random variable, with density $f(a) = a\exp(-a^2/2)$, $a \geq 0$, and ϕ uniform on $[0, 2\pi)$, independent of A. Fix $\omega > 0$ and define the "cosine process with angular frequency ω" as

$$X(t) = A\cos(\omega t - \phi).$$

Show that X is centered, Gaussian, and stationary. Find its covariance function, and show that it satisfies (5.3) with $\alpha = 2$.

(ii) From elementary geometric considerations, show that for $0 < T < \pi/\omega$ and $\lambda > 0$

$$P\Big\{\sup_{0\leq t\leq T} X(t) > \lambda\Big\} = \Psi(\lambda) + \frac{\omega T}{2\pi}\exp\big(-\tfrac{1}{2}\lambda^2\big).$$

(iii) Apply (i), (ii), and Theorem 5.1 to show that $H_2 = 1/\sqrt{\pi}$.

1.2 Another specific case in which the distribution of the maximum can be explicitly calculated is the Gaussian process with triangular covariance

$$R(t) = 1 - |t|, \qquad |t| \leq 1.$$

The fact that this process satisfies a pseudo Markov property allowed Slepian (1961) to derive the distribution of the maximum. You might try going over his calculation, and, *en passant*, showing that $H_1 = 1$.

SECTION 5.3:

3.1 We shall now prove the necessity in Theorem 5.5
(i) Show that if (U, V) are centered, jointly Gaussian variables, with $EU^2 = EV^2 = 1$ and $EUV < 1$, then

$$\lim_{\lambda \to \infty} \frac{P\{\max(U,V) \geq \lambda\}}{\Psi(\lambda)} = 2.$$

Hence conclude that if (5.16) holds, then so must condition (a) following it. (That is, if $\|X\|$ behaves like a single Gaussian variable, then X achieves its maximal variance at only one point.)
(ii) Show that for $0 \leq \lambda \leq \nu$,

$$\Psi(\lambda) \geq \Psi(\nu) + \nu(\nu - \lambda)\Psi(\nu).$$

(iii) Assume that (5.16) is in force. By (i), this implies that condition (a) of Theorem 5.5 is also in force. Show now that (5.16) is equivalent to (5.17).
(iv) Complete the proof of necessity by showing that (5.17) implies (5.21). An outline of the main steps is as follows: Choose $\epsilon > 0$ small enough so that

$$P\{\|X\| > \lambda\} \leq (1 + \epsilon^2)\Psi(\lambda),$$

for large enough λ. With $a(t)$ and $Y(t)$ as in the proof of Theorem 5.5, and $T' = \{t \in T \colon a(t) \geq 0\}$, deduce that

$$\int_{-\infty}^{\lambda} \psi(y)\phi(y)\, dy \leq \epsilon^2 \Psi(\lambda),$$

where $\psi(y) = P\{\sup_{t \in T'}(Y(t) + a(t)y) > \lambda\}$, and ϕ, as usual, is a standard Gaussian density.

Now complete the proof. You will need both (ii) and the result of Exercise 1.1 of Chapter 3.

SECTION 5.4:

4.1 Complete the proof of Example 5.8 by showing that the upper bound (5.42) for $N(\delta_1, \delta_2, \epsilon)$ also serves, albeit with a different constant, as a lower bound.

4.2 Fill in all the missing details in the derivation of the upper bound for the entropy in the proof of Example 5.9. If you are feeling truly courageous, try proving that the upper bound also serves as a lower bound, as usual, with a different constant.