BOUNDS ON THE SUPREMA OF GAUSSIAN PROCESSES, AND OMEGA RESULTS FOR THE SUM OF A RANDOM MULTIPLICATIVE FUNCTION

ADAM J HARPER

Abstract. We prove new lower bounds for the upper tail probabilities of suprema of Gaussian processes. Unlike many existing bounds, our results are not asymptotic, but supply strong information when one is only a little into the upper tail. We present an extended application to a Gaussian version of a random process studied by Halász. This leads to much improved lower bound results for the sum of a random multiplicative function. We further illustrate our methods by improving lower bounds for some classical constants from extreme value theory, the Pickands constants H_α, as $\alpha \to 0$.

1. Introduction

Let $\mathcal{T}$ be a non-empty set, $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, and for each $t \in \mathcal{T}$ let $Z(t)$ be a random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that for any finite subset $\{t_1, t_2, ..., t_n\} \subseteq \mathcal{T}$, the random variable $(Z(t_1), ..., Z(t_n))$ has an n-variate normal distribution. We will then say, a little loosely, that $\{Z(t)\}_{t\in\mathcal{T}}$ is a *Gaussian process* with parameter set $\mathcal{T}$. We refer the reader to the book of Lifshits [9] for a general introduction to the theory of Gaussian processes.

In this paper we will be concerned with $\sup_{t\in\mathcal{T}} Z(t)$, and in particular with giving lower bounds for the probability that it is quite large. Results of this type have many applications, and the author's interest in them stems from a number-theoretic problem that will be described later. For overviews of results in this area we refer to two important books, by Leadbetter, Lindgren and Rootzén [7] and by Piterbarg [14].

Suppose that $\mathcal{T}$ is a finite set, so that $\sup_{t\in\mathcal{T}} Z(t)$ is certainly a genuine random variable, and

$$\mathbb{P}(\sup_{t\in\mathcal{T}} Z(t) > u)$$

is the probability that a multivariate normal random vector takes values in a certain subset of $\mathbb{R}^{\#\mathcal{T}}$. We will also be interested in processes with infinite index sets, but will study these by looking at suitably chosen finite subsets of points t. Unless the mean vector and covariance matrix of $\{Z(t)\}_{t\in\mathcal{T}}$ have special forms it is typically very difficult

Date: 30th November 2010.

2000 *Mathematics Subject Classification.* Primary 60G15; Secondary 60G70, 11N64.

The author is supported by a studentship from the Engineering and Physical Sciences Research Council of the United Kingdom.

to compute the tail probability exactly. However, it turns out that as $u \to \infty$, any correlations amongst the $Z(t)$ that are not perfect ± 1 correlations have an increasingly negligible effect on the tail behaviour. (This follows e.g. by the method of comparison, which we discuss more below). In fact something like this remains valid when $\#\mathcal{T} \to \infty$ along with u, *provided* that the "new" $Z(t)$ that are introduced are not too highly correlated. Readers familiar with e.g. Berman's theorem (as expounded in chapter 4 of Leadbetter, Lindgren and Rootzén [7]) should find this reasoning familiar.

Piterbarg [14] describes several asymptotic methods for analysing Gaussian processes, including the method of comparison, Pickands' method of double sums, and Rice-type methods based on calculation of moments. He presents a quantitative version of the method of comparison, in which results are formulated for fixed u; but these results involve unspecified constants that appear to depend on $\{Z(t)\}_{t \in \mathcal{T}}$, so one must wait for u to be sufficiently large (in an unspecifed sense) before they come into play[1]. No quantitative formulation of the method of double sums or the method of moments (for lower bounds) is attempted, and the general philosophy of these methods, that one need not analyse correlations of $\{Z(t)\}_{t \in \mathcal{T}}$ except for extremely large correlations, seems unsuited to obtaining results for fixed u.

One can also use metric entropy/capacity methods, such as Sudakov's minoration, to bound $\mathbb{E} \sup_{t \in \mathcal{T}} Z(t)$. See section 14 of Lifshits [9]. Together with suitable concentration inequalities, such as that of Borell/Sudakov–Tsyrelson, this yields explicit lower bounds on $\mathbb{P}(\sup_{t \in \mathcal{T}} Z(t) > u)$ for fixed u. However, the lower bounds that one obtains for $\mathbb{E} \sup_{t \in \mathcal{T}} Z(t)$ are typically off from the truth by a multiplicative factor, and then the lower bounds for $\mathbb{P}(\sup_{t \in \mathcal{T}} Z(t) > u)$ are very far from the truth for moderately sized u.

In this paper we develop an alternative approach to lower bounding the upper tail probability. The ingredients are an initial conditioning step, followed by a comparison (in the sense of the method of comparison) with a "model" Gaussian process that can be explicitly analysed. The resulting bounds are clean and can be non-trivial for moderately sized u, which is very important in our two applications.

We begin with the following straightforward result.

Proposition 1 (Conditioning Step). *Let $\{Z(t_i)\}_{1 \leq i \leq n}$ be jointly multivariate normal random variables. Write $r_{i,j} := \mathbb{E} Z(t_i) Z(t_j)$, and suppose that:*

- (centralisation and normalisation) $\mathbb{E} Z(t_i) = 0$ *and* $\mathbb{E} Z(t_i)^2 = 1$ *for all* $1 \leq i \leq n$*;*

[1]We present some normal comparison results in §3, and these may give an idea of the necessary size of u. When the author tried to study our §6 example by these methods, he could only show that the supremum there is larger than about $\log\log x/\sqrt{2}$ with high probability (by studying points t with spacing $1/\sqrt{\log x}$). Our Corollary 2 establishes that this supremum is larger than about $\log\log x$ with high probability.

- (no repeated variables) $|r_{i,j}| < 1$ *whenever* $i \neq j$.

Then for any $u \geq 0$ *and any* $H \geq 0$,

$$\mathbb{P}(\max_{1\leq i\leq n} Z(t_i) > u) \geq \frac{He^{-(u+H)^2/2}}{\sqrt{2\pi}} \sum_{m=1}^{n} \inf_{0\leq h\leq H} P(m,h),$$

where $P(m,h)$ *is*

$$\mathbb{P}\left(V_j \leq \frac{u - r_{j,m}(u+h)}{\sqrt{1-r_{j,m}^2}} \; \forall j \leq m-1 \right),$$

and the $V_j = V_{j,m}$ *are centralised, normalised, jointly multivariate normal random variables with correlations*

$$\frac{r_{j,k} - r_{j,m}r_{k,m}}{\sqrt{(1-r_{j,m}^2)(1-r_{k,m}^2)}}.$$

We give the short proof of Proposition 1 in §2. The author had a more involved proof of (a result like) Proposition 1, based on a "reversal of roles" in the normal comparison procedure. Since we will need some normal comparison results later, we present these in §3 and give a very brief description of the reversal of roles approach as well.

We now turn to the problem of what we will be able to say about $P(m,h)$. If the correlation structure of $\{Z(t_i)\}_{1\leq i\leq m}$ is arbitrary, the answer may be essentially nothing, in which case our attempt to give lower bounds will be at an end. However, under some conditions on the correlation structure we can be more optimistic, and to illustrate this we formulate the following result.

Proposition 2 (Comparison Step)**.** *Let* $u \geq 0$, *and suppose* h *is sufficiently small that the upper bounds on the* V_j *in the definition of* $P(m,h)$ *are non-negative. Suppose there exist numbers* $c_j = c_j(m,h), d_j = d_j(m,h) > 0$ *such that:*

(i) c_j/d_j *is a non-decreasing sequence,* $1 \leq j \leq m-1$;
(ii) $c_{\min\{j,k\}}d_{\max\{j,k\}}$ *is a strict lower bound for* $r_{j,k} - r_{j,m}r_{k,m}$, *for each pair* $1 \leq j,k \leq m-1$.

Then for any $\delta \geq 0$,

$$P(m,h) \geq \int_{-B(\delta)}^{B(\delta)} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \cdot \prod_{j=1}^{m-1} \Phi\left(\frac{(1-\delta)(u - r_{j,m}(u+h))}{\sqrt{1-r_{j,m}^2 - c_j d_j}} \right),$$

where $B(\delta) = \delta\sqrt{\frac{d_{m-1}}{c_{m-1}}} \min_{1\leq j\leq m-1} \frac{u-r_{j,m}(u+h)}{d_j}$, *and* Φ *denotes the standard normal cumulative distribution function.*

We will prove Proposition 2 in §4, by explicitly constructing a collection of Gaussian random variables with the lower bound correlation structure suggested by the c_j, d_j, and applying a Brownian motion maximal inequality to analyse those. The reader might think of this procedure as pulling out some of the dependence among the V_j, to be analysed non-trivially using the maximal inequality. By doing this we gain the subtracted terms $c_j d_j$ in the product, which will be very important, at the fairly small cost of introducing the factor involving $B(\delta)$ (and the multiplier $(1-\delta)$).

The reader might wonder where the numbers c_j, d_j are to come from, and whether the lower bound obtained will not be hopelessly small in situations of interest. In fact we can quickly deduce the following from Propositions 1 and 2.

Theorem 1. *Let $\{Z(t_i)\}_{1\leq i\leq n}$ be as in Proposition 1, and suppose further that the sequence is* stationary, *i.e. that $r_{j,k} = r(|j-k|)$ for some function r. Let $u \geq 1$, and suppose that:*

- *$r(m)$ is a decreasing non-negative function;*
- *$r(1)(1+2u^{-2})$ is at most 1.*

Then $\mathbb{P}(\max_{1\leq i\leq n} Z(t_i) > u)$ is

$$\gg n\frac{e^{-u^2/2}}{u}\min\left\{1, \sqrt{\frac{1-r(1)}{u^2 r(1)}}\right\}\prod_{j=1}^{n-1}\Phi\left(u\sqrt{1-r(j)}\left(1+O\left(\frac{1}{u^2(1-r(j))}\right)\right)\right),$$

where the implicit constants are absolute (in particular, not depending on $\{Z(t_i)\}_{1\leq i\leq n}$) and could be given explicit values.

Theorem 1 follows by choosing $H = u^{-1}$, $\delta = \min\{u^{-2}, \sqrt{r(1)/u^2(1-r(1))}\}$, $c_j = r_{j,m} = r(|m-j|)$, $d_j = 1 - r_{j,m} = 1 - r(|m-j|)$ in the preceding propositions. In this case if we did not have $c_j d_j$ in the denominators in Proposition 2, then $\sqrt{1-r(j)}$ would need to be replaced by $\sqrt{(1-r(j))/(1+r(j))}$ in the product. We do not actually use the theorem in this paper, as our examples require slightly different parameter choices. However, a reader familiar with classical limit theory for suprema of stationary processes (see e.g. chapter 4 of Leadbetter, Lindgren and Rootzén [7]) may find it instructive to compare with those results. We may not expect to obtain precisely sharp bounds from Theorem 1, because of the factor $\min\{1, \sqrt{(1-r(1))/u^2 r(1)}\}$, but it will supply good bounds provided u is large enough that the product term is $\gg 1$. For given $r(j)$ this may be a much weaker requirement on u than in proofs of the classical results, which rely on normal comparison inequalities.

We now move on to our two examples, which we hope will illustrate the usefulness of Propositions 1 and 2. In the theory of Gaussian processes, much attention has been paid to (mean zero, variance one) stationary processes whose covariance function satisfies

$$r(t) = 1 - C|t|^\alpha + o(|t|^\alpha) \quad \text{as } t \to 0,$$

where $C > 0$ and $0 < \alpha \le 2$. In particular, a 1969 theorem of Pickands [12] describes the asymptotic behaviour of suprema of such processes: if $h > 0$ is fixed, and if $\sup_{\epsilon \le t \le h} r(t) < 1$ for all $\epsilon > 0$, then

$$\lim_{u \to \infty} e^{u^2/2} u^{1-2/\alpha} \mathbb{P}(\sup_{0 \le t \le h} Z(t) > u) = \frac{hC^{1/\alpha} H_\alpha}{\sqrt{2\pi}},$$

where H_α is the so-called *Pickands constant*. In a second paper [13], Pickands used a result like this to determine the limiting distribution, as $T \to \infty$, of a scaled version of $\sup_{0 \le t \le T} Z(t)$. The scaling in that theorem thus involves H_α. See e.g. the paper of Shao [17] for further discussion of the role of H_α.

It appears that not very much is known about the size of H_α. Burnecki and Michna [1] describe as "mathematical folklore" the conjecture that $H_\alpha = 1/\Gamma(1/\alpha)$, but this is only known for $\alpha = 1, 2$. Bounds are available more generally, for example Shao [17] used a representation of H_α in terms of a non-stationary process, and various techniques from Gaussian process theory, to show that

$$(\frac{\alpha}{4})^{1/\alpha}(1 - e^{-1/\alpha}(1 + \frac{1}{\alpha})) \le H_\alpha \le \alpha^{1/\alpha}(2.41\sqrt{8.8 - \alpha \log(0.4 + 2.5/\alpha)} + 0.77\sqrt{\alpha})^{2/\alpha}$$

when $0 < \alpha < 1$, and other bounds when $1 \le \alpha \le 2$. Dębicki and Kisowski [2] subsequently improved the upper bound on the range $1 < \alpha < 2$. Dębicki, Michna and Rolski [3] proved that

$$\frac{\alpha}{8\Gamma(1/\alpha)}(\frac{1}{4})^{1/\alpha} \le H_\alpha, \quad 0 < \alpha \le 2,$$

and in a 2009 preprint Michna [10] improved this by a multiplicative factor of 2. Note that, since $\Gamma(1/\alpha) \sim \sqrt{2\pi\alpha}(1/e\alpha)^{1/\alpha}$ as $\alpha \to 0$, this is a much stronger bound than that of Shao [17] under that limit process.

Applying our methods, in §5 we improve the lower bound results as $\alpha \to 0$.

Corollary 1. *Uniformly for* $0 < \alpha \le 2$, $H_\alpha \gg \sqrt{\alpha}(e\alpha/2)^{1/\alpha}$.

This differs from the conjectured value $1/\Gamma(1/\alpha)$ by (essentially) a factor $\alpha 2^{-1/\alpha}$, as $\alpha \to 0$. One could work carefully here to obtain a completely explicit result, and thereby obtain non-trivial bounds for α away from zero as well.

For our main example, we give a detailed study of the following process:

$$\sum_{p \le x} g_p \frac{\cos(t \log p)}{p^{1/2 + 1/\log x}}, \qquad t \in \mathbb{R},$$

where the summation is restricted to prime numbers p, g_p are independent standard normal random variables, and x is a further (large) parameter[2].

[2]Roughly speaking, we are interested in the supremum of this process over $1 \le t \le 2$, obtaining probability bounds that are a function of x. As x is increased, the supremum becomes stochastically

The motivation for studying this is its connection with a number-theoretic problem of Wintner [18]. Let ϵ_p be a sequence of independent Rademacher random variables, (so that $\mathbb{P}(\epsilon_p = 1) = \mathbb{P}(\epsilon_p = -1) = 1/2$), and construct a "random multiplicative function" from these, as follows:

$$f(n) := \begin{cases} \prod_{p|n} \epsilon_p & \text{if } n \text{ is squarefree} \\ 0 & \text{otherwise} \end{cases}$$

We also set $M(x) := \sum_{n \leq x} f(n)$. One can view $f(n)$ as a heuristic model for some deterministic functions occurring in number theory, such as the Möbius function, or simply as an interesting object in its own right. In our first appendix, we reproduce an argument of Halász [5] showing that lower bound information about the supremum of a certain Rademacher process can be translated into lower bound information about $|M(x)|$. Exploiting this connection, Halász [5] established that there exists a constant $B > 0$ such that, almost surely,

$$M(x) \neq O(\sqrt{x} e^{-B\sqrt{\log\log x \log\log\log x}}) \quad \text{as } x \to \infty.$$

In §6, we use Propositions 1 and 2 to prove results like the following:

Corollary 2. *As* $x \to \infty$,

$$\mathbb{P}(\sup_{1 \leq t \leq 2(\log\log x)^2} \sum_{p \leq x} g_p \frac{\cos(t \log p)}{p^{1/2+1/\log x}} \leq \log\log x - \log\log\log x + O((\log\log\log x)^{3/4}))$$

is $O((\log\log\log x)^{-1/2})$.

The exact nature of the bound $O((\log\log\log x)^{-1/2})$ is not the important feature here: this could be improved a bit by fairly small changes to the proof, but any bound that is $o(1)$ would suffice for us. The crucial feature is the level $\log\log x - \log\log\log x + O((\log\log\log x)^{3/4}))$ that we know the supremum will typically exceed: standard methods[3] show that the supremum is at most $\log\log x + \log\log\log x$ (say) with probability $1 - o(1)$, so Corollary 2 is very precise in this respect. This precision is crucial if one wishes to deduce things about $|M(x)|$: indeed it is the size of the second order subtracted term $\log\log\log x$, together with the size of the interval over which the supremum is taken, that determines what can be said.

Together with a suitable version of the multivariate central limit theorem, given in our second appendix, Corollary 2 allows for a substantial improvement of Halász's [5] result about $M(x)$. However, it is possible to do better still.

larger: this is because the variance increases for fixed t, but also because the correlation at nearby values of t decreases. Ultimately, for reasons explained in §6.3, we allow the range of t to increase slowly with x as well.

[3]The process is "almost" stationary, as explained in §6.1, and a simple adaptation of Rice's formula yields upper bounds for its supremum. See e.g. the book of Leadbetter, Lindgren and Rootzén [7] for discussion of Rice's formula.

Corollary 3. *Let $A > 2.5$, and let $M(x)$ be the summatory function of a Rademacher random multiplicative function, as above. It almost surely holds that*

$$M(x) \neq O(\sqrt{x}(\log\log x)^{-A}).$$

Corollary 2 implies Corollary 3 with the restriction $A > 3$, and this is proved in §6. To prove Corollary 3 for all $A > 2.5$, an argument by contradiction is needed to slightly sharpen the result of Proposition 2. This argument is rather fiddly, and for example does not result in a direct analogue of Corollary 2. It is given in §7.

It seems extremely likely that, almost surely, $M(x) \neq O(\sqrt{x})$, and perhaps that $M(x)$ almost surely has fluctuations of order $\sqrt{x \log\log x}$ (by analogy with Kolmogorov's Law of the Iterated Logarithm). Indeed, $M(x)$ might well exhibit even larger fluctuations, since its probability distribution may have rather heavy tails : see, for example, the author's article [6]. However, an argument like our own, ultimately based on studying a certain average of $M(x)$, seems unable to detect these large but rare fluctuations.

We presented Proposition 2 in its current form, involving parameters c_j, d_j, δ, because this seems both easy to appreciate, and to lead to good results in our applications. However, as mentioned above, to prove the full version of Corollary 3 it is necessary to slightly strengthen Proposition 2. Such a strengthening may also be possible in the context of Corollary 1: some of the initial steps of the §7 argument transfer to that situation, but it is not clear whether the whole argument goes through (except that it does not trivially do so).

The author also believes that there will be other Gaussian processes to which Propositions 1 and 2 could usefully be applied, and hopes that the reader might have some examples at hand.

2. Proof of Proposition 1

In view of the decomposition

$$\mathbb{P}(\max_{1 \leq i \leq n} Z(t_i) > u) = \sum_{m=1}^{n} \mathbb{P}(Z(t_m) > u, Z(t_j) \leq u \; \forall j \leq m-1),$$

it will suffice to show that, for any $1 \leq m \leq n$ and any $H \geq 0$,

$$\mathbb{P}(Z(t_1), ..., Z(t_{m-1}) \leq u, Z(t_m) > u) \geq \frac{He^{-(u+H)^2/2}}{\sqrt{2\pi}} \inf_{0 \leq h \leq H} P(m,h).$$

It is well known (and easy to check, by computing correlations) that $Z(t_m)$ is independent of the collection of random variables

$$Z(t_j) - r_{j,m} Z(t_m), \quad 1 \leq j \leq m-1.$$

These have mean zero and correlations

$$r_{j,k} - r_{j,m}r_{k,m}, \quad 1 \leq j,k \leq m-1,$$

and, in particular, none of them are degenerate (by assumption in Proposition 1). Thus $\mathbb{P}(Z(t_1),...,Z(t_{m-1}) \leq u, Z(t_m) > u)$ is at least

$$\int_u^{u+H} \mathbb{P}(Z(t_j) - r_{j,m}Z(t_m) \leq u - r_{j,m}x \ \forall 1 \leq j \leq m-1) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx,$$

from which the proposition follows.

Q.E.D.

In our applications, it will turn out that

$$\mathbb{P}(Z(t_1),...,Z(t_{m-1}) \leq u, Z(t_m) > u+H)$$

decreases very rapidly as H increases. Indeed, we will always choose H so that its effect in $P(m,H)$ is negligibly small, and therefore only really need to understand $P(m,0)$. This is the point of introducing the initial decomposition of $\mathbb{P}(\max_{1 \leq i \leq n} Z(t_i) > u)$, rather than trying to understand $\mathbb{P}(Z(t_i) \leq u \ \forall 1 \leq i \leq n)$ directly by conditioning.

3. Normal comparison results

3.1. **Classical comparison results.** In this subsection we present the equality underlying normal comparison results, and state some fairly classical consequences of this. We will use these in a few places, and hopefully they will also supply an unfamiliar reader with some idea of how the method of comparison (as it is referred to by Piterbarg [14]) is traditionally employed. Our treatment largely follows Li and Shao [8], although we would also like to draw attention to a 1954 paper of Plackett [15], which contains a similar presentation of the basic comparison result[4].

If $\tilde{a}, \tilde{b} \in \mathbb{R}^n$, write $\tilde{a} \leq \tilde{b}$ to mean that every component of $\tilde{a}$ is at most the corresponding component of $\tilde{b}$. We have the following identity, which is the key part of the proofs of various normal comparison results:

Exact Formula 1 (Following Li and Shao, and others)**.** *Let $\tilde{X} = (X_1,...,X_n)$ and $\tilde{W} = (W_1,...,W_n)$ be centralised and normalised n-variate normal vectors, with non-singular covariance matrices $\mathrm{Var}(\tilde{X}) = (\mathrm{Cov}(X_i,X_j))_{1 \leq i,j \leq n} = (r_{i,j}^{(1)})$ and $\mathrm{Var}(\tilde{W}) = (r_{i,j}^{(0)})$ respectively. Let $\tilde{u} \in \mathbb{R}^n$. Then*

$$\mathbb{P}(\tilde{X} \leq \tilde{u}) - \mathbb{P}(\tilde{W} \leq \tilde{u}) =$$

[4]Plackett was interested in the numerical approximation of multivariate normal probabilities, but some later comparison results are readily obtained from his paper. Unfortunately this work does not seem to be very widely known.

$$= \sum_{1\le i<j\le n} (r_{i,j}^{(1)} - r_{i,j}^{(0)}) \int_0^1 \phi(u_i, u_j; r_{i,j}^{(h)}) \mathbb{P}(\tilde{Z}^{(h)} \le \tilde{u} | Z_i^{(h)} = u_i, Z_j^{(h)} = u_j) dh,$$

where $\tilde{Z}^{(h)} = (Z_1^{(h)}, ..., Z_n^{(h)})$ *is multivariate normal with covariance matrix*

$$(r_{i,j}^{(h)}) := h\mathrm{Var}(\tilde{X}) + (1-h)\mathrm{Var}(\tilde{W}),$$

and $\phi(x, y; r)$ *denotes the standard bivariate normal density with correlation* r*, viz.*

$$\frac{1}{2\pi\sqrt{1-r^2}} e^{-(x^2-2rxy+y^2)/2(1-r^2)}.$$

To prove the formula one writes

$$\mathbb{P}(\tilde{X} \le \tilde{u}) - \mathbb{P}(\tilde{W} \le u) = \int_0^1 \frac{d}{dh} \mathbb{P}(\tilde{Z}^{(h)} \le \tilde{u}) dh,$$

observing that

$$\frac{d}{dh}\mathbb{P}(\tilde{Z}^{(h)} \le \tilde{u}) = \sum_{1\le i<j\le n} \frac{\partial}{\partial r_{i,j}^{(h)}} \mathbb{P}(\tilde{Z}^{(h)} \le \tilde{u}) \frac{\partial r_{i,j}^{(h)}}{\partial h} = \sum_{1\le i<j\le n} (r_{i,j}^{(1)} - r_{i,j}^{(0)}) \int_{-\infty}^{\tilde{u}} \frac{\partial^2 f_h}{\partial y_i \partial y_j} d\tilde{y}.$$

Here f_h is the density function of $\tilde{Z}^{(h)}$, the range of integration has its obvious meaning, and the second equality uses the fact that

$$\frac{\partial f_h}{\partial r_{i,j}^{(h)}} = \frac{\partial^2 f_h}{\partial y_i \partial y_j},$$

which follows by expressing the multivariate normal density in terms of its characteristic function.

Exact Formula 1 provides rigorous support for the intuitive idea that distributions with “nearby” covariance matrices may have like behaviour. The inequalities that we derive next may express this in a more striking way: they are a composite of results of Li and Shao [8] and of Leadbetter, Lindgren and Rootzén [7], although in most respects are unchanged from bounds of Slepian (1962), Berman (1964, 1971), and Cramér (1967).

Comparison Inequality 1 (Following Leadbetter, Lindgren and Rootzén, and Li and Shao)**.** *If* $\tilde{X}, \tilde{W}, \tilde{u}$ *are as in Exact Formula 1, and* $\boldsymbol{1}$ *denotes the indicator function, then each of the following is an upper bound for* $\mathbb{P}(\tilde{X} \le \tilde{u}) - \mathbb{P}(\tilde{W} \le \tilde{u})$*.*

(i)

$$\frac{1}{2\pi} \sum_{1\le i<j\le n} \boldsymbol{1}_{r_{i,j}^{(1)} > r_{i,j}^{(0)}} \int_{r_{i,j}^{(0)}}^{r_{i,j}^{(1)}} \frac{1}{\sqrt{1-t^2}} e^{-(u_i^2+u_j^2)/2(1+|t|)} dt$$

(ii)

$$\frac{1}{2\pi} \sum_{1\le i<j\le n} \boldsymbol{1}_{r_{i,j}^{(1)} > r_{i,j}^{(0)}} (\arcsin(r_{i,j}^{(1)}) - \arcsin(r_{i,j}^{(0)})) e^{-(u_i^2+u_j^2)/2(1+\max\{|r_{i,j}^{(1)}|, |r_{i,j}^{(0)}|\})}$$

(iii)

$$\frac{2}{\pi}\sum_{1\le i<j\le n}\mathbf{1}_{r_{i,j}^{(1)}>r_{i,j}^{(0)}}\frac{(1+\max\{|r_{i,j}^{(1)}|,|r_{i,j}^{(0)}|\})^{3/2}}{(u_i^2+u_j^2)\sqrt{1-\max\{|r_{i,j}^{(1)}|,|r_{i,j}^{(0)}|\}}}e^{-(u_i^2+u_j^2)/2(1+\max\{|r_{i,j}^{(1)}|,|r_{i,j}^{(0)}|\})}$$

To obtain the first bound, we overestimate the conditional probability in Exact Formula 1 trivially by 1 and insert the definition of $\phi(u_i,u_j;r_{i,j}^{(h)})$, observing that

$$\begin{aligned}\int_0^1 \frac{e^{-(u_i^2-2r_{i,j}^{(h)}u_iu_j+u_j^2)/2(1-(r_{i,j}^{(h)})^2)}}{\sqrt{1-(r_{i,j}^{(h)})^2}}dh &\le \int_0^1 \frac{1}{\sqrt{1-(r_{i,j}^{(h)})^2}}e^{-(u_i^2+u_j^2)/2(1+|r_{i,j}^{(h)}|)}dh \\ &= \frac{1}{r_{i,j}^{(1)}-r_{i,j}^{(0)}}\int_{r_{i,j}^{(0)}}^{r_{i,j}^{(1)}}\frac{1}{\sqrt{1-t^2}}e^{-(u_i^2+u_j^2)/2(1+|t|)}dt.\end{aligned}$$

For the second bound, we overestimate the exponential by $e^{-(u_i^2+u_j^2)/2(1+\max\{|r_{i,j}^{(1)}|,|r_{i,j}^{(0)}|\})}$, and then evaluate the integral over t. Alternatively, by making a substitution $x=\sqrt{(1-t)/(1+t)}$ we find that for any $0\le a\le b<1$, and any $K\ge 0$,

$$\begin{aligned}\int_a^b \frac{1}{\sqrt{1-t^2}}e^{-K/(1+t)}dt &= 2e^{-K/2}\int_{\sqrt{(1-b)/(1+b)}}^{\sqrt{(1-a)/(1+a)}}\frac{1}{1+x^2}e^{-Kx^2/2}dx \\ &\le \frac{(1+b)^{3/2}}{\sqrt{1-b}K}e^{-K/2}\int_{\sqrt{(1-b)/(1+b)}}^{\sqrt{(1-a)/(1+a)}}Kxe^{-Kx^2/2}dx.\end{aligned}$$

Since this integral is at most $e^{-K(1-b)/2(1+b)}$, the third bound follows directly.

As Leadbetter, Lindgren and Rootzén [7] point out, the assumption that $\tilde{X}$ and $\tilde{W}$ are non-singular is not necessary for the above bounds, as one may pass to that case by making arbitrarily small changes to the entries of the covariance matrices, and the first bound (from which we derived the others) is a continuous function of those entries.

Typically, one would apply Comparison Inequality 1 by observing that the covariance matrix of $\tilde{X}$ "looks rather like" the covariance matrix of a well understood multivariate normal distribution, e.g. that it looks like the identity matrix. See the paper of Li and Shao [8] for some examples. If the entries of the covariance matrices are sufficiently close together, or if one can afford to choose the entries of $\tilde{u}$ very large, then Comparison Inequality 1 can supply strong information.

We finish with a well known qualitative consequence of Comparison Inequality 1.

Comparison Inequality 2. *Let $\tilde{X}=(X_1,...,X_n)$ and $\tilde{W}=(W_1,...,W_n)$ be centralised and normalised n-variate normal vectors, with covariance matrices $\mathrm{Var}(\tilde{X})=(r_{i,j}^{(1)})$ and $\mathrm{Var}(\tilde{W})=(r_{i,j}^{(0)})$ respectively. Let $\tilde{u}\in\mathbb{R}^n$. If $r_{i,j}^{(1)}\le r_{i,j}^{(0)}$ for each $1\le i,j\le n$, then*

$$\mathbb{P}(\tilde{X}\le\tilde{u})\le\mathbb{P}(\tilde{W}\le\tilde{u}).$$

The special case of this result where $\tilde{u} = (u, u, ..., u)$, for some $u \in \mathbb{R}$, is usually referred to as Slepian's lemma.

3.2. **Reversal of roles.** As promised in the introduction, we now give a very brief description of the reversal of roles argument that originally served in place of Proposition 1. For the applications in this paper, Proposition 1 entirely supersedes such an argument, but it is possible that it may be useful in other contexts.

We aim to give an estimate for

$$\mathbb{P}(Z(t_m) > u, Z(t_j) \leq u \,\forall j \leq m-1),$$

under the conditions of Proposition 1. Our idea is to apply the methodology of Exact Formula 1, but viewing the sum of integrals that arises as a main term for subsequent analysis, and the subtracted probability as an error term. Thus we do not choose $\tilde{W}$ to have a standard distribution, but so that this subtracted probability is zero.

More concretely, we let $A_1, ..., A_m$ be a collection of $N(0,1)$ random variables, all independent of one another and of the $Z(t_i)$. Let $\epsilon > 0$, and define

$$X_i = W_i := \frac{Z(t_i) + \epsilon A_i}{\sqrt{1+\epsilon^2}}, \;\; 1 \leq i \leq m-1; \quad X_m := \frac{Z(t_m) + \epsilon A_m}{\sqrt{1+\epsilon^2}}; \quad W_m := \frac{Z(t_{m-1}) + \epsilon A_m}{\sqrt{1+\epsilon^2}}.$$

Precisely analogously to Exact Formula 1, and adopting the same notation $r_{i,j}^{(h)}$ as there, we find that

$$\begin{aligned}
&\mathbb{P}(X_1, ..., X_{m-1} \leq u, X_m > u) - \mathbb{P}(W_1, ..., W_{m-1} \leq u, W_m > u) \\
&= - \sum_{1 \leq i \leq m-1} (r_{i,m}^{(1)} - r_{i,m}^{(0)}) \int_0^1 \phi(u, u; r_{i,m}^{(h)}) \mathbb{P}(\tilde{Z}^{(h)} \leq \tilde{u} | Z_i^{(h)} = u, Z_m^{(h)} = u) dh \\
&= - \sum_{1 \leq i \leq m-1} \frac{r_{i,m} - r_{i,m-1}}{1+\epsilon^2} \int_0^1 \phi(u, u; \frac{h r_{i,m} + (1-h) r_{i,m-1}}{1+\epsilon^2}) P(i, h, \epsilon) dh,
\end{aligned}$$

say. We need the ϵ perturbations here to ensure that we work with non-singular multivariate normal distributions. However, at the end of the argument we can let $\epsilon \to 0$, whereby we will have compared $\mathbb{P}(Z(t_m) > u, Z(t_j) \leq u \,\forall j \leq m-1)$ with

$$\mathbb{P}(Z(t_1), ..., Z(t_{m-1}) \leq u, Z(t_{m-1}) > u) = 0.$$

It is less straightforward to analyse $P(i, h, \epsilon)$ for $1 \leq i \leq m-2$ than to analyse $P(m-1, h, \epsilon)$, and to give lower bounds one can replace those probabilities by $\mathbf{1}_{r_{i,m} > r_{i,m-1}}$. In our examples these other terms give a lower order contribution, but this need not always be so. However, to analyse $P(m-1, h, \epsilon)$ one can note (as did Li and Shao [8]) that for any $1 \leq i \leq m-1$, the collection of random variables

$$Y_j^{(h)} := Z_j^{(h)} - \left(\frac{r_{j,i}^{(h)} - r_{i,m}^{(h)} r_{j,m}^{(h)}}{1 - (r_{i,m}^{(h)})^2} \right) Z_i^{(h)} - \left(\frac{r_{j,m}^{(h)} - r_{i,m}^{(h)} r_{j,i}^{(h)}}{1 - (r_{i,m}^{(h)})^2} \right) Z_m^{(h)}$$

$$= Z_j^{(1)} - \left(\frac{r_{j,i}^{(1)} - r_{i,m}^{(h)} r_{j,m}^{(h)}}{1-(r_{i,m}^{(h)})^2}\right) Z_i^{(1)} - \left(\frac{r_{j,m}^{(h)} - r_{i,m}^{(h)} r_{j,i}^{(1)}}{1-(r_{i,m}^{(h)})^2}\right) Z_m^{(h)}, \quad 1 \leq j \leq m-1, j \neq i$$

is independent of $\{Z_i^{(h)}, Z_m^{(h)}\}$. In our examples this leads, after some slightly fiddly manipulations, to a probability estimate much like Proposition 1. (For, in our examples, $Z(t_{m-1})$ and $Z(t_m)$ are always very highly correlated, and so $P(m-1,h,\epsilon)$ is essentially the same as the simple conditional probability in the proof of Proposition 1.)

4. Proof of Proposition 2

In view of Comparison Inequality 2, and assumption (ii) in the statement of Proposition 2, we may proceed on the assumption that for $1 \leq j,k \leq m-1$ and $j \neq k$, $\mathbb{E}V_jV_k$ is equal to

$$\frac{c_{\min\{j,k\}} d_{\max\{j,k\}}}{\sqrt{(1-r_{j,m}^2)(1-r_{k,m}^2)}}.$$

The key to the proof is the explicit construction of such random variables from a collection of independent normal random variables.

Let $Y_1, ..., Y_n, Z_1, ..., Z_n$ be independent standard normal random variables, and for $1 \leq i \leq n$ let α_i, β_i be real numbers satisfying

$$\beta_i^2 \sum_{j \leq i} \alpha_j^2 < 1.$$

Then the random variables

$$X_i := \beta_i \sum_{j \leq i} \alpha_j Y_j + \sqrt{1 - \beta_i^2 \sum_{j \leq i} \alpha_j^2} Z_i$$

are again jointly multivariate normal, have zero means and unit variances, and satisfy

$$\mathbb{E}X_iX_j = \beta_i\beta_j \sum_{k \leq \min\{i,j\}} \alpha_k^2, \quad i \neq j.$$

We also note that if $u_1, ..., u_n$ are any real numbers, if $\beta_i > 0 \ \forall 1 \leq i \leq n$, and if $\delta \in \mathbb{R}$, then

$$\begin{aligned} \mathbb{P}(X_i \leq u_i \ \forall 1 \leq i \leq n) &= \mathbb{P}\left(Z_i \leq \frac{u_i - \beta_i \sum_{j \leq i} \alpha_j Y_j}{\sqrt{1-\beta_i^2 \sum_{j \leq i} \alpha_j^2}} \ \forall 1 \leq i \leq n\right) \\ &\geq \mathbb{P}(\sum_{j \leq i} \alpha_j Y_j \leq \frac{\delta u_i}{\beta_i} \ \forall 1 \leq i \leq n) \prod_{i=1}^n \Phi\left(\frac{u_i(1-\delta)}{\sqrt{1-\beta_i^2 \sum_{j \leq i} \alpha_j^2}}\right). \end{aligned}$$

We now set $n = m-1$, and define real numbers α_i, β_i by

$$\beta_i := \frac{d_i}{\sqrt{1-r_{i,m}^2}}, \quad \sum_{j \leq i} \alpha_j^2 := \frac{c_i}{d_i}, \quad 1 \leq i \leq m-1.$$

The conditions on c_i, d_i in Proposition 2 ensure that we can define α_i, β_i in this way, and that they satisfy the various hypotheses above. The reader may also check that the X_i have the correlation structure that we wanted, and that the product term in the previous paragraph is as in Proposition 2 (when u_i is taken as $(u - r_{i,m}(u+h))/\sqrt{1-r_{i,m}^2}$). It remains to give a suitable lower bound for $\mathbb{P}(\sum_{j \leq i} \alpha_j Y_j \leq \frac{\delta u_i}{\beta_i} \ \forall 1 \leq i \leq m-1)$.

It should not come as a surprise that the behaviour of partial sums of independent normal random variables is rather well understood. For example, writing $\{W_t\}_{t \geq 0}$ for the standard Brownian motion, (see e.g. chapter 5 of Lifshits [9] for much discussion of this process), one has the following neat result, which we quote from chapter 13.4 of Grimmett and Stirzaker [4]: if $t \geq 0$, then

$$\max_{0 \leq s \leq t} W_s \stackrel{d}{=} |W_t| \stackrel{d}{=} |N(0,t)|.$$

This is useful to us because $(\sum_{j \leq i} \alpha_j Y_j)_{1 \leq i \leq m-1} \stackrel{d}{=} (W_{\sum_{j \leq i} \alpha_j^2})_{1 \leq i \leq m-1}$, so that

$$\mathbb{P}(\sum_{j \leq i} \alpha_j Y_j \leq \frac{\delta u_i}{\beta_i} \ \forall 1 \leq i \leq m-1) \geq \Phi(B) - \Phi(-B),$$

where

$$B = \frac{\delta}{\sqrt{\sum_{j \leq m-1} \alpha_j^2}} \min_{1 \leq i \leq m-1} \frac{u_i}{\beta_i} = \delta \sqrt{\frac{d_{m-1}}{c_{m-1}}} \min_{1 \leq i \leq m-1} \frac{u - r_{i,m}(u+h)}{d_i},$$

as claimed in Proposition 2.

Q.E.D.

The proof just given divided naturally into two parts: first we constructed the X_j to explicitly model the V_j, allowing us to extract some of their dependence in the manageable form of the Y_j; and then we analysed the Y_j using a result about Brownian motion. Both of these steps could conceivably be improved, potentially leading to a better lower bound for $P(m,h)$.

In the analysis of the Y_j, we used a fact about the probability that a Brownian motion remains below a constant level for a period of "time" t. We could have used results about the probability that it remains below, for example, a sloping line, allowing some flexibility in the upper bounds that we ask for. However, in our applications these probabilities are never particularly small, and the author doubts that a more complicated approach would be advantageous in many situations.

It appears to the author that the modelling part of the argument is weaker. Thus, in our examples, our lower bound $c_{\min\{j,k\}}d_{\max\{j,k\}}$ for $r_{j,k} - r_{j,m}r_{k,m}$ is not very tight when j and k are close together. An alternative way to think about this is to note that we can replace the independent Z_j in our construction by any standard normal A_j with

$$\mathbb{E}A_jA_k \leq \frac{r_{j,k} - r_{j,m}r_{k,m} - c_{\min\{j,k\}}d_{\max\{j,k\}}}{\sqrt{(1-r_{j,m}^2 - c_jd_j)(1-r_{k,m}^2 - c_kd_k)}}.$$

The correlation bound here looks complicated, but this may be somewhat illusory; for example, if we were able to make the choices $c_j = r_{j,m}$, $d_j = 1 - r_{j,m}$, as for certain stationary processes, we would want

$$\mathbb{E}A_jA_k \leq \frac{r_{j,k} - r_{\min\{j,k\},m}}{\sqrt{(1-r_{j,m})(1-r_{k,m})}}.$$

These quantities are not likely to be easier to work with than the correlations $r_{j,k}$ of our original random variables. However, to prove Proposition 2 we need upper bounds for upper tail probabilities (which then lower bound the probability that none of the A_j are too big), and these may be easier to come by than lower bounds, for example by using Rice's formula as part of a first moment argument. Another approach to improving Proposition 2 along these lines is worked out in §7.

5. Application to estimating Pickands' constants

Suppose that $t_1 < t_2 < ... < t_M$ is a set of equally spaced real numbers. Suppose, moreover, that $\{Z(t_i)\}_{1\leq i\leq M}$ is a mean zero, variance one, stationary Gaussian process, with decreasing covariance function $r(t)$, $t \geq 0$. If $a > 0$, then (the proof of) Proposition 1 implies that

$$\begin{aligned}\mathbb{P}(\max_{1\leq i\leq M} Z(t_i) > u) &\geq M\mathbb{P}(Z(t_M) > u, Z(t_j) \leq u\ \forall j < M) \\ &\geq \frac{Me^{-u^2/2}}{\sqrt{2\pi}u} \cdot ae^{-a-a^2/2u^2} \inf_{0\leq h\leq a/u} P(M,h).\end{aligned}$$

In a paper from 1996, Shao [17] considers a (mean zero, variance one) stationary Gaussian process indexed by the half-line $[0,\infty)$, with covariance function

$$r(t) = \frac{1}{2}(e^{\alpha t/2} + e^{-\alpha t/2} - (e^{t/2} - e^{-t/2})^{\alpha}), \quad t \geq 0.$$

Such a process exists for each fixed $0 < \alpha < 2$. As $t \to 0$, we see (as did Shao [17]) that $r(t) = 1 - t^\alpha/2 + O(t^2)$. We also note that, for $t > 0$,

$$\begin{aligned} r'(t) &= \frac{\alpha}{4}(e^{\alpha t/2} - e^{-\alpha t/2} - (e^{t/2} + e^{-t/2})(e^{t/2} - e^{-t/2})^{\alpha-1}) \\ &= \frac{\alpha}{4}(e^{\alpha t/2} - e^{-\alpha t/2} - e^{\alpha t/2}(1 + e^{-t})(1 - e^{-t})^{\alpha-1}) \\ &\leq \frac{\alpha}{4}(e^{\alpha t/2} - e^{-\alpha t/2} - e^{\alpha t/2}(1 - e^{-2t})) \\ &< 0. \end{aligned}$$

In proving Corollary 1, we shall assume that α is smaller than a certain fixed positive constant, which could be found explicitly if desired and should certainly be less than 1. There is no loss in doing this, because $H_\alpha \gg 1$ for α larger than such a constant. To prove the corollary, we will study Shao's stationary process at the sample points $t_i = i/M$. Making the simple choice $a = 1$ in the above discussion, we have

$$H_\alpha \gg 2^{1/\alpha} \lim_{u \to \infty} (M u^{-2/\alpha} \inf_{0 \leq h \leq 1/u} P(M, h)),$$

and we will investigate the largest value of M, depending on u and α, for which we can show that $\inf_{0 \leq h \leq 1/u} P(M, h) \gg 1$. Note that a large value of M corresponds to a close packing of sample points in the interval $[0, 1]$. The reader should also note that there is nothing intrinsically asymptotic about most of our calculations, although we are interested in letting $u \to \infty$ to compare with Pickands' theorem.

We want to apply Proposition 2, and can do so with the natural choices

$$c_j = r((M-j)/M), \quad d_j = 1 - r((M-j)/M), \quad 1 \leq j \leq M-1,$$

since $r(t)$ is decreasing and positive. Thus $P(M, h)$ is at least

$$\begin{aligned} &(\Phi(B) - \Phi(-B)) \prod_{j=1}^{M-1} \Phi\left((1 + O(\frac{1}{u^2(1 - r(j/M))}))u(1-\delta)\sqrt{1 - r(j/M)}\right) \\ &= (\Phi(B) - \Phi(-B)) \prod_{j=1}^{M-1} \Phi\left((1 + O(\frac{M^\alpha}{u^2 j^\alpha}))(1-\delta)\sqrt{u^2(j^\alpha/2M^\alpha + O(j^2/M^2))}\right), \end{aligned}$$

where $B = B(\delta)$ is as in Proposition 2, and δ will be chosen later in terms of α. Together with the known and conjectured bounds for Pickands' constants, this suggests taking $M = [(bu^2\alpha/2)^{1/\alpha}]$, where now we investigate how large b may be chosen. For definiteness in our calculations, we declare that we shall certainly have $1 \leq b \leq 10$ (and of course our conclusion will be that taking b as $e/2$ is permissible).

First we note that the part of the product over $j > M^{1/4}$ is $1 + o(1)$ as $u \to \infty$. For, since $r(t)$ is decreasing and α, δ are small, each of those terms is at least

$$\Phi((1 + O(\alpha))u(1-\delta)\sqrt{1 - r(M^{-3/4})}) \geq \Phi((1/2)\sqrt{u^2(M^{-3\alpha/4}/2 + O(M^{-3/2}))}).$$

If u is large enough, this is

$$\geq 1 - e^{-(1/8)u^2(M^{-3\alpha/4}/2+O(M^{-3/2}))} \geq 1 - e^{-\sqrt{u}},$$

and clearly $(1 - e^{-\sqrt{u}})^M$ is $1 + o(1)$ as $u \to \infty$ with α fixed.

When $j \leq M^{1/4}$, provided that u is large enough in terms of $\alpha \leq 1$ we see

$$u^2 j^2/M^2 \leq u^2 M^{-3/2} = O(u^{-1}\alpha^{-3/2}), \quad \text{and} \quad M^\alpha/u^2 j^\alpha = O(\alpha/j^\alpha).$$

Since for $x \geq 2$, say, we have $\Phi(x) \geq 1 - x^{-1}e^{-x^2/2} \geq e^{-2x^{-1}e^{-x^2/2}}$, the part of the product over $j \leq M^{1/4}$ is at least $e^{-f(b,\delta,\alpha,u)}$, where

$$f(b,\delta,\alpha,u) = O\Big(\sum_{j \leq M^{1/4}} e^{-(1-\delta)^2 j^\alpha/2b\alpha}\sqrt{b\alpha}/(1-\delta)j^{\alpha/2}\Big).$$

Note that, since we assume that α and δ are small, the arguments of Φ in the product are all at least 2. Now

$$\begin{aligned}
\sum_{j \leq M^{1/4}} e^{-(1-\delta)^2 j^\alpha/2b\alpha} &\leq \int_0^{M^{1/4}} e^{-(1-\delta)^2 t^\alpha/2b\alpha} dt \\
&= \frac{2b}{(1-\delta)^2} \int_0^{(1-\delta)^2 M^{\alpha/4}/2b\alpha} \left(\frac{2b\alpha y}{(1-\delta)^2}\right)^{1/\alpha - 1} e^{-y} dy \\
&\leq \left(\frac{2b}{(1-\delta)^2}\right)^{1/\alpha} \alpha^{1/\alpha-1}\Gamma(1/\alpha).
\end{aligned}$$

In view of Stirling's formula, the right hand side is asymptotic to $\sqrt{2\pi/\alpha}(2b/e(1-\delta)^2)^{1/\alpha}$ as $\alpha \to 0$, so is at most $4\alpha^{-1/2}(2b/e(1-\delta)^2)^{1/\alpha}$, say, when α is small.

Finally, observe that

$$B(\delta) = \delta\sqrt{(1 - r(1/M))/r(1/M)}u(1 + O(1/u^2(1 - r(1/M)))) \gg \delta/\sqrt{\alpha},$$

provided that u is large enough in terms of α. If we make the choice $\delta = \alpha$, then b can be chosen as large as $e/2$ whilst still ensuring that $f(b,\delta,\alpha,u) = O(1)$. Corollary 1 follows from making these choices.

6. Application to a number-theoretic process

6.1. **Preliminary calculations.** Before we can apply Propositions 1 and 2 to our second example, we must reduce to studying a finite set of sample points t, and determine the covariance structure of the corresponding random variables. As might be expected, variants of some of these calculations appear in Halász's paper [5], although we need to be more precise in several places.

It is useful initially to ignore the contribution from "very small" primes to our random sums. Let y be a parameter, later to be chosen as a suitable function of x. It is immediate that if $s, t \in \mathbb{R}$, then

$$\begin{aligned}\mathbb{E}(\sum_{y \le p \le x} g_p \frac{\cos(t \log p)}{p^{1/2+1/\log x}} \cdot \sum_{y \le p \le x} g_p \frac{\cos(s \log p)}{p^{1/2+1/\log x}}) &= \sum_{y \le p \le x} \frac{\cos(t \log p)\cos(s \log p)}{p^{1+2/\log x}} \\ &= \frac{1}{2} \sum_{y \le p \le x} \frac{\cos((t+s)\log p) + \cos((t-s)\log p)}{p^{1+2/\log x}}.\end{aligned}$$

For $t \in \mathbb{R}$, we let $Z_y(t)$ denote the normalised random variable

$$\frac{\sum_{y \le p \le x} g_p \cos(t \log p)/p^{1/2+1/\log x}}{\sqrt{\sum_{y \le p \le x} \cos^2(t \log p)/p^{1+2/\log x}}} = \frac{\sum_{y \le p \le x} g_p \cos(t \log p)/p^{1/2+1/\log x}}{\sqrt{\sum_{y \le p \le x} 1/2p^{1+2/\log x} + \sum_{y \le p \le x} \cos(2t \log p)/2p^{1+2/\log x}}}.$$

By a strong form of the prime number theorem (see e.g. chapter 6 of Montgomery and Vaughan [11]) we have

$$\pi(z) := \#\{p \le z : p \text{ is prime}\} = \int_2^z \frac{du}{\log u} + O(ze^{-d\sqrt{\log z}}), \quad z \ge 2,$$

where $d > 0$ is a certain constant. Then if $z \le x$,

$$\begin{aligned}\sum_{p \le z} \frac{1}{p^{1+2/\log x}} = \int_2^z \frac{1}{u^{1+2/\log x}} d\pi(u) &= \int_2^z \frac{u^{-2/\log x}}{u \log u} du + c(x) + O(e^{-d\sqrt{\log z}}) \\ &= \log\log z + O(1),\end{aligned}$$

where $c(x)$ depends on x only. Moreover, if $\alpha \neq 0$,

$$\begin{aligned}\sum_{y \le p \le x} \frac{\cos(\alpha \log p)}{p^{1+2/\log x}} &= \int_y^x \frac{\cos(\alpha \log u) u^{-2/\log x}}{u \log u} du + O((1+|\alpha|)e^{-d\sqrt{\log y}}) \\ &= \int_{\log y}^{\log x} \frac{\cos(\alpha u)}{u} du + \int_{\log y}^{\log x} \frac{\cos(\alpha u)}{u}(e^{-2u/\log x} - 1) du + \\ &\quad + O((1+|\alpha|)e^{-d\sqrt{\log y}}) \\ &= \int_{\alpha \log y}^{\alpha \log x} \frac{\cos u}{u} du + O(\frac{1}{\alpha \log x}) + O((1+|\alpha|)e^{-d\sqrt{\log y}}),\end{aligned}$$

where the third equality follows using integration by parts, since $\frac{d}{du}((e^{-2u/\log x} - 1)/u) = O(1/\log^2 x)$ for $\log y \le u \le \log x$. We deduce that if $s, t \ge 1$, and $s \neq t$, then

$$\mathbb{E} Z_y(t) Z_y(s) = \frac{\int_{|t-s|\log y}^{|t-s|\log x} \frac{\cos u}{u} du + O(\frac{1}{(t+s)\log y}) + O(\frac{1}{|t-s|\log x}) + O((t+s)e^{-d\sqrt{\log y}})}{\int_y^x \frac{du}{u^{1+2/\log x} \log u} + O(\frac{1}{\log y}) + O((t+s)e^{-d\sqrt{\log y}})}.$$

We now set out the specific situation to which our Gaussian process results will be applied. Let $E \ge 1$ be a further parameter, (to be chosen later as a function of x), and for $n \in \mathbb{N} \cup \{0\}$ and $M \le \log x/E$ introduce the sets

$$\mathcal{T}_n = \mathcal{T}_{n,x,E,M} := \{2n + 1 + iE/\log x : 1 \le i \le M\} \subseteq [2n+1, 2n+2].$$

We seek lower bound information about $\max_{0 \leq n \leq B} \sup_{t \in \mathcal{T}_n} Z_y(t)$, for certain B.

At this point the reader may be rather appalled by the number of parameters around, so we hasten to point out that most of these will "select themselves" in due course, and can essentially be ignored. The sets $\mathcal{T}_n$ are sufficiently separated that the behaviour of $Z_y(t)$ on different sets is roughly independent (see §6.3). Moreover, up to error terms the correlation $\mathbb{E}Z_y(t)Z_y(s)$ depends on s, t through $|t-s|$ only (i.e. our process is approximately stationary). Thus we focus on understanding $\sup_{t \in \mathcal{T}_0} Z_y(t)$, and defer thinking about larger values of n until we put our results together in §6.3.

The parameter E controls the spacing of sample points within their blocks $\mathcal{T}_n$, and in §6.2 it will be chosen as small as possible such that we obtain good probability lower bounds from Proposition 2. We declare for now that we shall certainly have $E \leq e^{\sqrt{\log\log x}}$, say. We would like to take M as large as possible, but to simplify our calculations we choose $M = [\log x / KE \log y]$, where K is an absolute constant that forces $\mathbb{E}Z_y(t)Z_y(s) \geq 1/\log\log x$, say, for $t, s \in \mathcal{T}_0$ (see below). The parameter y is present to get rid of "beginning of series" effects, in particular ensuring that we have enough independence of $Z_y(t)$ over different blocks $\mathcal{T}_n$. It will be selected in §6.3, but we declare for now that we shall certainly have $\log x \leq y \leq e^{(\log\log x)^{100}}$.

In the above set-up, if $s, t \in \mathcal{T}_0$ are distinct then

$$\begin{aligned}
\mathbb{E}Z_y(t)Z_y(s) &= \frac{\int_{|t-s|\log y}^{\log x} \frac{\cos u}{u} du + O(\frac{1}{|t-s|\log x})}{\int_y^x \frac{du}{u^{1+2/\log x}\log u}} + O\left(\frac{1}{\log y \log\log x}\right) \\
&= \frac{\int_{|t-s|\log y}^{1} \frac{\cos u}{u} du}{\int_y^x \frac{du}{u^{1+2/\log x}\log u}} + O\left(\frac{1}{\log\log x}\right) \\
&= \frac{\log(1/|t-s|\log y)}{\log\log x - \log\log y} + O\left(\frac{1}{\log\log x}\right).
\end{aligned}$$

6.2. **Implementation of Propositions 1 and 2.** We order the points of $\mathcal{T}_0$ in the obvious and natural way, writing $t_i = 1 + iE/\log x$, $1 \leq i \leq M$. We aim to show that the maximum of our original random sum is about $\log\log x$, and the standard deviations that we normalised by are about $\sqrt{(\log\log x - \log\log y)/2}$, so we take $u = \sqrt{2(\log\log x - \log\log y)}$. Then

$$u(1 - r_{m-1,m}) = \Theta(\log E/\sqrt{\log\log x - \log\log y}) = \Theta(\log E/u),$$

so we can safely make the canonical choice $H = 1/u$ in Proposition 1.

We now seek to apply Proposition 2 to give a lower bound for $P(m,h)$, where $1 \leq m \leq M$ and $h \leq H$. Let $j < k \leq m-1$. If $|j-k| \leq \log^{1/3} x$ then

$$\begin{aligned} r_{j,k} = 1 - \frac{\log(|j-k|E)}{\log\log x - \log\log y} + O\left(\frac{1}{\log\log x}\right) &\geq \max\{1/2, r_{j,m}\} + O\left(\frac{1}{\log\log x}\right) \\ &\geq r_{j,m} + O\left(\frac{r_{j,m}}{\log\log x}\right). \end{aligned}$$

In fact this is also true when $|j-k| > \log^{1/3} x$: since $\int_{\alpha \log y}^{\log x} (\cos u/u) du$ is a decreasing function of $0 < \alpha \leq 1/\log y$, we have

$$\begin{aligned} r_{j,k} &= \frac{\int_{|j-k|E \log y/\log x}^{\log x} \frac{\cos u}{u} du + O(\frac{1}{|j-k|E})}{\int_y^x \frac{du}{u^{1+2/\log x} \log u}} + O\left(\frac{1}{\log y \log\log x}\right) \\ &\geq r_{j,m} + O\left(\frac{1}{\log y \log\log x}\right), \end{aligned}$$

and $r_{j,m} \geq 1/\log\log x \geq 1/\log y$ because $m \leq M$. This means that it is legitimate to choose

$$c_j = 1 - \frac{\log(m-j)E + O(1)}{\log\log x - \log\log y}, \quad d_j = \frac{\log(m-j)E + O(1)}{\log\log x - \log\log y}, \quad 1 \leq j,k \leq m-1$$

in Proposition 2. Setting $\delta = 1/\log\log x$, to match the size of the other "big Oh" term, we discover that

$$B(1/\log\log x) = \Theta\left(\frac{u\sqrt{\log E}}{(\log\log x)^{3/2}}\right) = \Theta\left(\frac{\sqrt{\log E}}{\log\log x}\right).$$

It follows from all of this that

$$\begin{aligned} P(m,h) &\gg \frac{\sqrt{\log E}}{\log\log x} \prod_{j=1}^{m-1} \Phi\left((1 + O(\frac{1}{\log(m-j)E}))\sqrt{2\log(m-j)E}\right) \\ &\gg \frac{\sqrt{\log E}}{\log\log x} e^{-\Theta(\sum_{j=1}^{m-1} \frac{1}{(m-j)E\sqrt{\log(m-j)E}})}, \end{aligned}$$

provided always that E is larger than an absolute constant. Making the choice $E = \sqrt{\log\log x}$, the exponential becomes $\Theta(1)$, and we find $P(m,h) \gg \sqrt{\log\log\log x}/\log\log x$.

Plugging this lower bound into Proposition 1, it follows immediately that

$$\begin{aligned} \mathbb{P}(\sup_{t \in \mathcal{T}_0} Z_y(t) > \sqrt{2(\log\log x - \log\log y)}) &\gg \frac{M\sqrt{\log\log\log x}e^{-u^2/2}}{u \log\log x} \\ &\gg \frac{\sqrt{\log\log\log x}}{(\log\log x)^2}. \end{aligned}$$

6.3. **Exploitation of the lower bound.** The lower bound obtained at the end of §6.2 is useful raw information about $\{Z_y(t)\}_{t \in \mathcal{T}_0}$. However, in order to deduce results about the summatory function $M(x)$ of a random multiplicative function, (as described in the introduction), we need to be able to say that the supremum is large with probability close to 1.

To do this, our idea is to "sample the supremum several times independently": since the probability that the supremum is large is not too small, if we sample a few times it is very likely that we will obtain a large value. Although we do not have lots of independent copies of $\{Z_y(t)\}$, we can achieve something like this by considering $\{Z_y(t)\}_{t\in\mathcal{T}_n}$ for different n. If $Be^{-d\sqrt{\log y}} \leq \frac{1}{\log y}$, say, then for distinct $1 \leq s,t \leq 2B+2$ we have

$$\mathbb{E}Z_y(t)Z_y(s) = \frac{\int_{|t-s|\log y}^{\log x} \frac{\cos u}{u} du + O(\frac{1}{|t-s|\log x})}{\int_y^x \frac{du}{u^{1+2/\log x}\log u}} + O\left(\frac{1}{\log y \log\log x}\right),$$

as at the end of §6.1. For such s,t with $|s-t| \geq 1$, the calculations in §6.1 supply a more precise result, namely that

$$\mathbb{E}Z_y(t)Z_y(s) = O\left(\frac{1}{|t-s|\log y \log\log x} + \frac{(t+s)e^{-d\sqrt{\log y}}}{\log\log x}\right).$$

Thus, by the second part of Comparison Inequality 1,

$$|\mathbb{P}(\max_{0\leq n\leq B}\sup_{t\in\mathcal{T}_n} Z_y(t) \leq \sqrt{2(\log\log x - \log\log y)}) - \prod_{0\leq n\leq B}\mathbb{P}(\sup_{t\in\mathcal{T}_n} Z_y(t) \leq \sqrt{2(\log\log x - \log\log y)})|$$

$$\begin{aligned}
&\ll \frac{\log^2 y}{\log^2 x}\sum_{0\leq i<j\leq B}\sum_{1\leq k,l\leq M}|\mathbb{E}Z_y(2i+1+kE/\log x)Z_y(2j+1+lE/\log x)| \\
&\ll \frac{\log^2 y M^2}{\log^2 x \log\log x}\sum_{0\leq i<j\leq B}\left(\frac{1}{|i-j|\log y} + (i+j)e^{-d\sqrt{\log y}}\right) \\
&\ll \frac{1}{(\log\log x)^2}\left(\frac{B\log B}{\log y} + B^3 e^{-d\sqrt{\log y}}\right).
\end{aligned}$$

We have seen that, at the level of precision required in §6.2, the correlation structure of $\{Z_y(t)\}_{t\in\mathcal{T}_n}$ is the same for each $0 \leq n \leq B$. Thus our calculations concerning $\sup_{t\in\mathcal{T}_0} Z_y(t)$ go through for $\sup_{t\in\mathcal{T}_n} Z_y(t)$ as well, and $\mathbb{P}(\max_{0\leq n\leq B}\sup_{t\in\mathcal{T}_n} Z_y(t) \leq \sqrt{2(\log\log x - \log\log y)})$ is

$$\ll \frac{1}{(\log\log x)^2}\left(\frac{B\log B}{\log y} + B^3 e^{-d\sqrt{\log y}}\right) + e^{-\Theta((B+1)\sqrt{\log\log\log x}/(\log\log x)^2)}.$$

The right hand side is $O(e^{-\Theta(\sqrt{\log\log\log x})})$ if we take $B = (\log\log x)^2$, and $y \geq \log x$.

For our application to $M(x)$, we need a version of the above result in which $Z_y(t)$ is replaced by

$$\frac{\sum_{y\leq p\leq x} f(p)\cos(t\log p)/p^{1/2+1/\log x}}{\sqrt{\sum_{y\leq p\leq x}\cos^2(t\log p)/p^{1+2/\log x}}},$$

with $f(p)$ independent Rademacher random variables. This can be achieved using a multivariate central limit theorem, as explained in our second appendix, at the cost of replacing $\sqrt{2(\log\log x - \log\log y)}$ by $\sqrt{2(\log\log x - \log\log y) - 1}$. In the application of the central limit theorem, we need y to be at least a certain power of $\log x$, say $y = \log^8 x$. This choice is also permissible for all of the preceding calculations.

Finally note that for fixed $t \in \mathbb{R}$,

$$\mathbb{E}(\sum_{p<y} \frac{g_p \cos(t \log p)}{p^{1/2+1/\log x}})^2 = \mathbb{E}(\sum_{p<y} \frac{f(p) \cos(t \log p)}{p^{1/2+1/\log x}})^2 = O(\log\log y) = O(\log\log\log x) \quad \text{as } x \to \infty,$$

as in §6.1. Applying Chebychev's inequality to this estimate, it follows that

$$\mathbb{P}(|\sum_{p<y} \frac{g_p \cos(t \log p)}{p^{1/2+1/\log x}}| > (\log\log\log x)^{3/4}) = O((\log\log\log x)^{-1/2}),$$

similarly if the g_p are replaced by Rademacher random variables $f(p)$. The behaviour of these sums is independent of the behaviour of the sums over $y \leq p \leq x$, so that Corollary 2 quickly follows from our bounds.

As discussed in our first appendix, the tail sum $\sum_{p>x} f(p) \cos(t \log p)/p^{1/2+1/\log x}$ is almost surely convergent, and in fact it converges in square mean, so that

$$\mathbb{E}(\sum_{p>x} \frac{f(p) \cos(t \log p)}{p^{1/2+1/\log x}})^2 \leq \sum_{p>x} \frac{1}{p^{1+2/\log x}} = O(\int_x^\infty \frac{du}{u^{1+2/\log x} \log u}) = O(1).$$

We temporarily set $d(x) := \inf_{1 \leq t \leq 2(\log\log x)^2} \sqrt{\mathbb{E}(\sum_{y \leq p \leq x} f(p) \cos(t \log p)/p^{1/2+1/\log x})^2}$, so that $d(x) = \sqrt{(\log\log x - \log\log y)/2 + O(1)}$ by the calculations in §6.1. Applying Chebychev's inequality again, together with the Rademacher version of our estimate for $Z_y(t)$, we find that

$$\mathbb{P}(\frac{1}{d(x)} \sup_{1 \leq t \leq 2(\log\log x)^2} \sum_p \frac{f(p) \cos(t \log p)}{p^{1/2+1/\log x}} \leq \sqrt{2(\log\log x - \log\log y) - 1} + \frac{(\log\log\log x)^{3/4}}{d(x)})$$

is $O((\log\log\log x)^{-1/2})$. Then applying the First Borel–Cantelli Lemma, one quickly deduces that: for any fixed $A > 3$, there almost surely exists a sequence (x_k), tending to infinity, with

$$\sup_{1 \leq t \leq 2(\log\log x_k)^2} \sum_p \frac{f(p) \cos(t \log p)}{p^{1/2+1/\log x_k}} - 2\log\log\log x_k \geq \log\log x_k - A \log\log\log x_k.$$

By the argument in our first appendix, (and specifically by Supplementary Lemma 1 from that appendix), this implies Corollary 3 for $A > 3$.

7. Refinement of Proposition 2 for the random multiplicative functions application

As discussed at the end of §4, Proposition 2 may be refined in that the product term can be replaced by any lower bound for

$$\mathbb{P}\left(A_j \leq \frac{(1-\delta)(u-r_{j,m}(u+h))}{\sqrt{1-r_{j,m}^2-c_jd_j}} \forall 1 \leq j \leq m-1\right),$$

for *any* standard normal random variables A_j satisfying

$$\mathbb{E}A_jA_k \leq \frac{r_{j,k}-r_{j,m}r_{k,m}-c_{\min\{j,k\}}d_{\max\{j,k\}}}{\sqrt{(1-r_{j,m}^2-c_jd_j)(1-r_{k,m}^2-c_kd_k)}}.$$

It will be convenient to write $U(j,k)$ for this upper bound on the permissible correlations. By assumption about the numbers c_j, d_j, we always have $U(j,k) \geq 0$.

In our application to random multiplicative functions, $U(j,k)$ is at least

$$\frac{r_{j,k}-1}{\sqrt{(1-r_{j,m}^2-c_jd_j)(1-r_{k,m}^2-c_kd_k)}}+\frac{1-r_{j,m}}{\sqrt{(1-r_{j,m}^2-c_jd_j)(1-r_{k,m}^2-c_kd_k)}}$$

$$=\frac{-\log|j-k|E+O(1)}{\sqrt{(\log|j-m|E+O(1))(\log|k-m|E+O(1))}}+\frac{\log|j-m|E+O(1)}{\sqrt{(\log|j-m|E+O(1))(\log|k-m|E+O(1))}},$$

for $1 \leq j < k \leq m-1$. It seems sensible to consider intervals $L^i/E < |m-j|, |m-k| \leq L^{i+1}/E$, (with $L \leq 2$ a parameter to be chosen), on which we see

$$U(j,k) \geq 1-\frac{\log(|j-k|E)}{i\log L}+O\left(\frac{1}{i\log L}\right).$$

In the random multiplicative functions example, the upper bound $(1-\delta)(u-r_{j,m}(u+h))/\sqrt{1-r_{j,m}^2-c_jd_j}$ for the A_j on such an interval is at least $(1+O(1/i\log L))\sqrt{2i\log L}$. So considering intervals independently, we can replace the product in Proposition 2 by

$$\prod_{\substack{i=0,\\ L^i\geq E/2}}^{[\log(Em)/\log L]} \mathbb{P}(A_j \leq (1-\frac{c}{i\log L})\sqrt{2i\log L}\ \ \forall L^i/E < |m-j| \leq L^{i+1}/E),$$

where c is an absolute constant, and A_j are any standard normal random variables whose correlations are bounded as described.

The crucial point is that on each interval, and up to the "big Oh" term, the bound on $U(j,k)$ corresponds to a stationary correlation structure that we can hope to understand. Indeed, it is essentially a re-scaled version of the original correlation structure of our random multiplicative functions process.

Using these ideas, we shall establish the following result:

Proposition 3. *If E is a sufficiently large constant, then the following is true. Let $\{Z_y(t)\}_{t \in \mathcal{T}_0} = \{Z_y^x(t)\}_{t \in \mathcal{T}_0}$ be the Gaussian process described in §6.1, for such a choice of E. Let $\epsilon(x)$ be any function tending to zero as $x \to \infty$. Then for some sequence of x, tending to infinity, we have*

$$\mathbb{P}(\sup_{t \in \mathcal{T}_0} Z_y(t) > \sqrt{2(\log\log x - \log\log y)}) \geq \frac{\epsilon(x)\sqrt{\log E}}{E(\log\log x)^{3/2}}.$$

Here we include a superscript x to explicitly record that $Z_y(t) = Z_y^x(t)$ depends on x, and we remind the reader that we had $y = \log^8 x$.

Recall from §6.1 that

$$\mathbb{E}Z_y^x(1+\frac{jE}{\log x})Z_y^x(1+\frac{kE}{\log x}) = 1 - \frac{\log(|j-k|E)+O(1)}{\log\log x - \log\log y}, \quad 1 \leq j,k \leq \frac{\log x}{KE\log y}, \quad j \neq k.$$

When L^i is larger than an absolute constant, we can choose $x(i) \in \mathbb{R}$ such that

$$\sqrt{2(\log\log x(i) - \log\log y(i))} = (1 - \frac{c}{i\log L})\sqrt{2i\log L}.$$

Here we write $y(i) = y(x(i)) = \log^8 x(i)$. Then if $C \in \mathbb{N}$ is larger than an absolute constant, and we take $L = 1 + 1/KC^3$, we have

$$\mathbb{E}Z_{y(i)}^{x(i)}(1+\frac{jCE}{\log x(i)})Z_{y(i)}^{x(i)}(1+\frac{kCE}{\log x(i)}) = 1 - \frac{\log(|j-k|E)+\log C+O(1)}{(1-c/(i\log L))^2 i\log L} \leq U(j,k),$$

where $U(j,k)$ denotes the bound for interval i. This only makes sense if $jC, kC \leq \log x(i)/KE\log y(i)$, but that will hold for example if $j,k \leq L^i/KEC^2$. Then if i is sufficiently large that

$$\left[\frac{L^{i+1}}{E}\right] - \left[\frac{L^i}{E}\right] \leq \left[\frac{L^i}{KEC^2}\right],$$

we can say $\mathbb{P}(A_j \leq (1-\frac{c}{i\log L})\sqrt{2i\log L} \;\; \forall L^i/E < |m-j| \leq L^{i+1}/E)$ is at least

$$\mathbb{P}(\sup_{t \in \mathcal{T}_0} Z_{y(i)}^{x(i)}(t) \leq \sqrt{2(\log\log x(i) - \log\log y(i))}).$$

Notice that, for our fixed choice of L, the various requirements for i to be "sufficiently large" will all be satisfied if $i \geq i_E + D$, where i_E is least for which $L^i \geq E/2$, and D is an absolute constant. Thus the product term in Proposition 2 may be replaced by

$$\prod_{j=1}^{[L^D]} \Phi\left((1+O(\frac{1}{\log jE}))\sqrt{2\log jE}\right) \prod_{i=i_E+D}^{[\frac{\log(Em)}{\log L}]} \mathbb{P}(\sup_{t \in \mathcal{T}_0} Z_{y(i)}^{x(i)}(t) \leq \sqrt{2(\log\log x(i) - \log\log y(i))}).$$

We also note that, obviously, $x(i)$ tends to infinity with i.

Now suppose that the proposition failed, so for all sufficiently large x the tail probability was smaller than required. Then for all i from some point onwards we would have

$$\mathbb{P}(\sup_{t\in\mathcal{T}_0} Z_{y(i)}^{x(i)}(t) \leq \sqrt{2(\log\log x(i) - \log\log y(i))}) \geq 1 - \frac{1}{(\log\log x(i))^{3/2}} \geq 1 - O\left(\frac{1}{(i\log L)^{3/2}}\right),$$

so the product term in Proposition 2 could be replaced by a positive constant. But then the argument of §6.2 would supply that

$$\mathbb{P}(\sup_{t\in\mathcal{T}_0} Z_y(t) > \sqrt{2(\log\log x - \log\log y)}) \gg \frac{\sqrt{\log E}}{E(\log\log x)^{3/2}},$$

which is a contradiction for x sufficiently large.

Q.E.D.

Armed with Proposition 3, we can repeat the argument of §6.3 with E chosen to be a large constant (rather than $\sqrt{\log\log x}$), and B then chosen as $(\log\log x)^{3/2}\log\log\log x$, say (rather than $(\log\log x)^2$). The reader should note that there is a subtlety involved, as this requires lower bounds for

$$\mathbb{P}(\sup_{t\in\mathcal{T}_n} Z_y(t) > \sqrt{2(\log\log x - \log\log y)}), \quad 0 \leq n \leq B,$$

whilst Proposition 3 concerns $\sup_{t\in\mathcal{T}_0} Z_y(t)$ only. However, modifying the choice of E and K by some multiplicative constants in the definition of $\mathcal{T}_n$, $n \neq 0$, so that E is larger but EK remains the same, we can arrange using Comparison Inequality 2 that

$$\mathbb{P}(\sup_{t\in\mathcal{T}_n} Z_y(t) > \sqrt{2(\log\log x - \log\log y)}) \geq \mathbb{P}(\sup_{t\in\mathcal{T}_0} Z_y(t) > \sqrt{2(\log\log x - \log\log y)}).$$

We also only have probability bounds for a sequence of x tending to infinity, rather than all x, but we do not require that in §6.3. Corollary 3 follows from these considerations.

Appendix A. Random multiplicative functions and Rademacher processes

In this appendix we sketch the connection between the summatory function $M(x)$ of a random multiplicative function (as defined in the introduction), and a certain Rademacher random process. The argument we give is essentially that of Halász [5].

In view of Wintner's [18] result that for each $\epsilon > 0$, $M(x) = O(x^{1/2+\epsilon})$ almost surely, we know that the Dirichlet series

$$F(s) := \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$$

is almost surely convergent in the half plane $\Re(s) > 1/2$, and then satisfies

$$F(s) = s\int_1^{\infty} \frac{M(z)}{z^{s+1}}dz.$$

On the other hand, writing $\zeta(s) := \sum_n 1/n^s, \Re(s) > 1$ for the Riemann zeta function, we have the Euler product identity

$$\begin{aligned} F(s) = \prod_p (1 + \frac{f(p)}{p^s}) &= e^{\sum_p f(p)/p^s - \sum_p 1/2p^{2s} + \sum_{k\geq 3}\sum_p (-1)^{k+1} f(p)^k/kp^{ks}} \\ &= e^{\sum_p f(p)/p^s - \log\zeta(2s)/2 + \sum_{k\geq 2}\sum_p 1/2kp^{2ks} + \sum_{k\geq 3}\sum_p (-1)^{k+1} f(p)^k/kp^{ks}}. \end{aligned}$$

This is certainly valid when $\Re(s) > 1$, and almost surely extends to $\Re(s) > 1/2$ in view of Kolmogorov's Three Series theorem[5] and the identity theorem of complex analysis. Thus in the domain $1/2 < \sigma < 1, 1 \leq t \leq 2$, say, we almost surely have

$$e^{\sum_p f(p)\cos(t\log p)/p^{\sigma}} \ll \int_1^{\infty} \frac{|M(z)|}{z^{\sigma+1}}dz \leq \sup_{z\geq 1} \frac{|M(z)|}{\sqrt{z(\sigma - 1/2)}} + \sup_{z\geq z_0} \frac{|M(z)|}{\sqrt{z}(\sigma-1/2)},$$

where the second inequality follows by splitting the integral at $z_0 := e^{1/\sqrt{\sigma-1/2}}$. Taking $\sigma = 1/2 + 1/\log x$, where $x \geq 2$ is a parameter, we find that

$$e^{\sum_p f(p)\cos(t\log p)/p^{1/2+1/\log x}} \ll \sqrt{\log x}\sup_{z\geq 1}\frac{|M(z)|}{\sqrt{z}} + \log x \sup_{z\geq e^{\sqrt{\log x}}} \frac{|M(z)|}{\sqrt{z}}, \quad 1 \leq t \leq 2.$$

For the proof of Corollary 3, we need a version of the preceding inequality that is valid for a larger range of t. Using the estimate $|\log\zeta(\sigma+it)| \leq \log\log|t| + O(1), \sigma \geq 1, |t| \geq 2$, which is contained in e.g. Theorem 6.7 of Montgomery and Vaughan [11], we can say that for $t \geq 1$,

$$e^{\sum_p f(p)\cos(t\log p)/p^{1/2+1/\log x} - \log t - \log\log(t+2)/2} \ll \sqrt{\log x}\sup_{z\geq 1}\frac{|M(z)|}{\sqrt{z}} + \log x \sup_{z\geq e^{\sqrt{\log x}}} \frac{|M(z)|}{\sqrt{z}}.$$

This immediately implies the following result:

Supplementary Lemma 1. *Let $g(z)$ be a decreasing function. If, with positive probability, we have $M(z) = O(\sqrt{z}g(z))$ as $z \to \infty$, then with positive probability we have*

$$\sup_{t\geq 1} e^{\sum_p f(p)\cos(t\log p)/p^{1/2+1/\log x} - \log t - \log\log(t+2)/2} = O(g(1)\sqrt{\log x} + g(e^{\sqrt{\log x}})\log x)$$

for all $x \geq 2$.

Since Halász's paper [5] seems to be difficult to get hold of, it is perhaps worthwhile to briefly discuss Halász's [5] own use of the foregoing argument. He shows that there

[5]This implies that $\sum_p f(p)/p^s$ converges almost surely when $\Re(s) > 1/2$. We then use the standard fact, proved using partial summation, that a Dirichlet series is a holomorphic function strictly to the right of its abscissa of convergence.

almost surely exist sequences of real numbers x_k, tending to infinity, and of sets $S_k \subseteq [1,2]$, of measure $> 1/\log x_k$, and of sets $B_k \subseteq [1,2]$, of measure $\leq 1/\log x_k$, such that

$$\sum_{p \leq x_k} f(p) \frac{\cos(t \log p)}{\sqrt{p}} \geq \log\log x_k - \sqrt{29 \log\log x_k \log\log\log x_k} \quad \forall t \in S_k,$$

$$\sum_{p \leq x_k} f(p) \frac{\cos(t \log p)}{\sqrt{p}} - \sum_{p} f(p) \frac{\cos(t \log p)}{p^{1/2+1/\log x_k}} = O(\sqrt{\log\log x_k}) \quad \forall t \in [1,2] \backslash B_k.$$

In particular, there almost surely exists a sequence x_k such that

$$\sup_{t \in [1,2]} \sum_{p} f(p) \frac{\cos(t \log p)}{p^{1/2+1/\log x_k}} \geq \log\log x_k - \sqrt{29 \log\log x_k \log\log\log x_k} - O(\sqrt{\log\log x_k}),$$

which is enough to imply the omega result for $M(x)$ attributed to Halász in the introduction.

Very roughly, Halász [5] investigates the process $\sum_{p \leq x} f(p) \frac{\cos(t \log p)}{\sqrt{p}}, t \in [1,2]$ by estimating moments of the counting function

$$\int_1^2 \mathbf{1}_{\sum_{p \leq x} f(p) \cos(t \log p)/\sqrt{p} \geq M} dt,$$

where M is a parameter. However, the details are rather complicated, as it is actually necessary to split the sum over p into several ranges, and then reduce the range of integration to progressively smaller random subsets of $[1,2]$. This splitting is, in a sense, quite natural, as the parts of the sum taken over large primes are less correlated at nearby values of t (see §6.1). On the other hand, the splitting causes an accumulation of error terms in the analysis, one from each range of summation. The iterative approach is also highly reliant on being presented with the process as a random sum over p, whereas (at least if the $f(p)$ were independent Gaussians) one might just as well be given a description of the process only in terms of its covariance structure.

Appendix B. A multivariate central limit theorem

In this appendix we discuss a multivariate central limit theorem of Reinert and Röllin [16]. We view this as a "universality result", which sometimes lets us transfer conclusions about suprema of Gaussian processes to conclusions about the suprema of corresponding Rademacher processes. Reinert and Röllin's [16] approach is based on Stein's method of exchangeable pairs.

Suppose that $\mathcal{T}$ is a finite set, and that $\alpha_i(t) \in \mathbb{R}$ for $1 \leq i \leq n$ and $t \in \mathcal{T}$. Suppose also that $(\epsilon_i)_{i=1}^n$ is a sequence of independent Rademacher random variables, and that $(g_i)_{i=1}^n$ is a sequence of independent standard normal random variables. We wish to approximate the (joint) distribution of $\{X_t\}_{t\in\mathcal{T}}$ by that of $\{Y_t\}_{t\in\mathcal{T}}$, where

$$X_t := \sum_{i=1}^n \alpha_i(t)\epsilon_i, \qquad Y_t := \sum_{i=1}^n \alpha_i(t) g_i.$$

In the typical way, we construct random variables X_t' such that $((X_t)_{t\in\mathcal{T}}, (X_t')_{t\in\mathcal{T}})$ is an exchangeable pair of vectors. (That is, such that the law of this tuple is the same as the law of $((X_t')_{t\in\mathcal{T}}, (X_t)_{t\in\mathcal{T}})$.) Let I be a random variable having the discrete uniform distribution on $\{1, 2, ..., n\}$, independently of everything else; and let $(\epsilon_i')_{i=1}^n$ be an independent copy of $(\epsilon_i)_{i=1}^n$. We define X_t' as follows: conditional on the event $\{I = i\}$, set

$$X_t' = X_t - \alpha_i(t)\epsilon_i + \alpha_i(t)\epsilon_i', \quad t \in \mathcal{T}.$$

The reader may check that the exchangeability property does then hold, together with the following regression property:

$$\mathbb{E}(X_t' - X_t | (X_s)_{s\in\mathcal{T}}) = -\frac{1}{n} X_t.$$

With a view to applying Theorem 2.1 of Reinert and Röllin [16], we calculate two further quantities:

$$\mathbb{E}((X_t' - X_t)(X_s' - X_s) | (X_u)_{u\in\mathcal{T}}) = \frac{1}{n}\sum_{i=1}^n \alpha_i(t)\alpha_i(s)\mathbb{E}((\epsilon_i' - \epsilon_i)^2 | (X_u)_{u\in\mathcal{T}}) = \frac{2}{n}\sum_{i=1}^n \alpha_i(t)\alpha_i(s);$$

$$\mathbb{E}|(X_t' - X_t)(X_s' - X_s)(X_u' - X_u)| = \frac{1}{n}\sum_{i=1}^n |\alpha_i(t)\alpha_i(s)\alpha_i(u)| \mathbb{E}|\epsilon_i' - \epsilon_i|^3 = \frac{4}{n}\sum_{i=1}^n |\alpha_i(t)\alpha_i(s)\alpha_i(u)|.$$

The reader should notice that, whilst we did not use the fact that the ϵ_i are Rademacher random variables up until this point, in the first calculation it allows us to conclude that the left hand side is deterministic. This means that one of the error terms in Reinert and Röllin's [16] theorem is identically zero; indeed, if $h : \mathbb{R}^{\#\mathcal{T}} \to \mathbb{R}$ is a three times differentiable function, and if the covariance matrix of $(X_t)_{t\in\mathcal{T}}$ is non-singular, their theorem implies that

$$|\mathbb{E}h((X_t)_{t\in\mathcal{T}}) - \mathbb{E}h((Y_t)_{t\in\mathcal{T}})| \le \frac{1}{3} \sup_{s,t,u\in\mathcal{T}, \tilde{x}\in\mathbb{R}^{\#\mathcal{T}}} \left|\frac{\partial^3 h(\tilde{x})}{\partial x_s \partial x_t \partial x_u}\right| \sum_{s,t,u\in\mathcal{T}} \sum_{i=1}^n |\alpha_i(s)\alpha_i(t)\alpha_i(u)|.$$

The condition that the covariance matrix should be non-singular is evidently unnecessary here (at least if h is bounded, say), since we can ensure this by introducing $\#\mathcal{T}$ dummy random variables whose coefficients $\alpha_i(t)$ have absolute value at most δ, and then let $\delta \to 0$.

Specialising to our random multiplicative functions application, we would like to choose h to be the indicator function of a box in $\mathbb{R}^{\#\mathcal{T}}$, but this would not satisfy the three times differentiability condition. Reinert and Röllin devote a section of their paper [16] to this "unsmoothing" problem, but the results they obtain are rather involved, and in this case we can easily overcome the difficulty directly. Let $s : \mathbb{R} \to [0,1]$ be a three times differentiable function satisfying

$$s(z) = \begin{cases} 1 & \text{if } z \leq \sqrt{2(\log\log x - \log\log y) - 1}, \\ 0 & \text{if } z \geq \sqrt{2(\log\log x - \log\log y)}. \end{cases}$$

The interval on which $s(z)$ must transition from 1 to 0 has length $\Theta(1/\sqrt{\log\log x})$, so we can find such a function whose derivatives satisfy $|s^{(r)}(z)| = O((\log\log x)^{r/2})$, $0 \leq r \leq 3$, $z \in \mathbb{R}$. Setting $h((x_t)_{t\in\mathcal{T}}) = \prod_{t\in\mathcal{T}} s(x_t)$, we conclude that

$$\mathbb{P}(\max_{t\in\mathcal{T}} X_t \leq \sqrt{2(\log\log x - \log\log y) - 1})$$

$$\leq \mathbb{P}(\max_{t\in\mathcal{T}} Y_t \leq \sqrt{2(\log\log x - \log\log y)}) + O((\log\log x)^{3/2}(\#\mathcal{T})^3 \sum_{i=1}^{n} \max_{t\in\mathcal{T}} |\alpha_i(t)|^3).$$

The reader may check that in the random multiplicative functions case, the error term on the right has order at most

$$(\#\mathcal{T})^3 \sum_{y\leq p\leq x} \frac{1}{p^{3/2}} \ll \frac{(\#\mathcal{T})^3}{\sqrt{y}\log y}.$$

We have $\#\mathcal{T} = (B+1)M \ll (\log\log x)^2 \log x$, so this is $o(1)$ as $x \to \infty$ provided that y is at least $\log^8 x$, say. The multivariate central limit theorem has supplied an extremely good bound, presumably because any individual ϵ_p (or g_p) has a very tiny impact on the random multiplicative function processes.

Acknowledgements. The author would like to thank his PhD supervisor, Ben Green, for introducing him to Halász's work on random multiplicative functions, and for reading a draft of this paper. He would also like to thank Nathanael Berestycki and Richard Nickl for discussions on these topics.

References

[1] K. Burnecki, Z. Michna. Simulation of Pickands constants. *Probability and Mathematical Statistics*, **22**, no. 1, pp 193-199. 2002

[2] K. Dębicki, P. Kisowski. A note on upper estimates for Pickands constants. *Statistics and Probability Letters*, **78**, pp 2046-2051. 2008

[3] K. Dębicki, Z. Michna, T. Rolski. Simulation of the asymptotic constant in some fluid models. *Stochastic Models*, **19**, pp 407-423. 2003

[4] G. R. Grimmett, D. R. Stirzaker. *Probability and Random Processes.* Third edition, published by Oxford University Press. 2001

[5] G. Halász. On Random Multiplicative Functions. In *Proceedings of the Hubert Delange colloquium*, (Orsay, 1982), pp 74-96. Publ. Math. Orsay, Univ. Paris XI, 1983.

[6] A. J. Harper. On the limit distributions of some sums of a random multiplicative function. *Preprint.* 2009

[7] M. Leadbetter, G. Lindgren, H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Published in the Springer Series in Statistics, 1983.

[8] W. Li, Q.M. Shao. A normal comparison inequality and its applications. *Probab. Theory Relat. Fields*, **122**, pp 494-508. 2002

[9] M. Lifshits. *Gaussian Random Functions.* Published in the Mathematics and Its Applications series, by Kluwer Academic Publishers, 1995.

[10] Z. Michna. Remarks on Pickands theorem. *Preprint.* 2009

[11] H. L. Montgomery, R. C. Vaughan. *Multiplicative Number Theory I: Classical Theory.* First edition, published by Cambridge University Press. 2007

[12] J. Pickands III. Upcrossing probabilities for stationary Gaussian processes. *Trans. Amer. Math. Soc.*, **145**, pp 51-73. 1969

[13] J. Pickands III. Asymptotic properties of the maximum in a stationary Gaussian process. *Trans. Amer. Math. Soc.*, **145**, pp 75-86. 1969

[14] V. Piterbarg. *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* Volume 148 of the American Mathematical Society Translations of Mathematical Monographs, 1996.

[15] R. L. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, **41**, pp 351-360. 1954

[16] G. Reinert, A. Röllin. Multivariate Normal Approximation with Stein's Method of Exchangeable Pairs under a General Linearity Condition. *Annals of Probability*, **37**, no. 6, pp 2150-2173. 2009

[17] Q.M. Shao. Bounds and estimators of a basic constant in extreme value theory of Gaussian processes. *Statistica Sinica*, **6**, pp 245-257. 1996

[18] A. Wintner. Random factorizations and Riemann's hypothesis. *Duke Mathematical Journal*, **11**, no. 2, pp 267-275. 1944

Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WA, England

E-mail address: A.J.Harper@dpmms.cam.ac.uk