The Annals of Applied Probability
2013, Vol. 23, No. 2, 584–616
DOI: 10.1214/12-AAP847

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

BY ADAM J. HARPER[1]

University of Cambridge

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 nonempty 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 $\{t_1, t_2, \dots, t_n\} \subseteq \mathcal{T}$, the random variable $(Z(t_1), \dots, Z(t_n))$ 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 [12] 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 [10] and by Piterbarg [17].

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

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

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

Received January 2011; revised January 2012.

[1]Supported by a studentship from the Engineering and Physical Sciences Research Council of the United Kingdom.

MSC2010 subject classifications. Primary 60G15; secondary 60G70, 11N64.

Key words and phrases. Gaussian processes, bounds on tail probabilities, Pickands constants, random multiplicative functions.

mean vector and covariance matrix of $\{Z(t)\}_{t\in\mathcal{T}}$ have special forms, it is typically very difficult to compute the tail probability exactly. Nevertheless, existing results offer two broad options for lower bounding $\mathbb{P}(\sup_{t\in\mathcal{T}} Z(t) > u)$.

- One can use metric entropy/capacity methods, such as Sudakov's minoration, to bound $\mathbb{E}\sup_{t\in\mathcal{T}} Z(t)$ (see Lifshits [12], Section 14). 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.
- One can use techniques such as the method of comparison (which we discuss more below), Pickands' method of double sums or Rice-type methods (based on calculation of moments) to estimate the probability asymptotically as $u \to \infty$ (see Piterbarg's book [17]).

The methods listed can be powerful when attacking certain problems, but have some unfortunate limitations. 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. The asymptotic techniques ultimately rely on, among other things, the fact that as $u \to \infty$, any correlations among the $Z(t)$ that are not perfect ± 1 correlations have an increasingly negligible effect on the tail behavior. (Readers familiar with, e.g., Berman's theorem should find this reasoning familiar.) Unfortunately u may need to be extremely large before the techniques guarantee this effect to occur.

Piterbarg [17] does not formulate the method of double sums or the method of moments (for lower bounds) for fixed u, and the general philosophy of those methods, that one need not analyze correlations of $\{Z(t)\}_{t\in\mathcal{T}}$ except for extremely large correlations, seems unsuited to obtaining such results. His version of the method of comparison involves 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 unspecified sense, before it comes into play. (We present some normal comparison inequalities in Section 3, and when the author tried to study our Section 6 example using them, 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 shows that supremum is larger than about $\log\log x$ with high probability.)

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 analyzed. The resulting bounds are clean and can be nontrivial for moderately sized u. Indeed, in our number-theoretic application we will have nontrivial bounds for u just larger than $\mathbb{E}\sup_{t\in\mathcal{T}} Z(t)$ (and, in particular, will be able to identify the expectation up to second order terms). As our bounds are completely explicit, they also give information about the "mysterious" constants in some asymptotic results, and our other application is a new lower bound for the Pickands constants (defined later).

We begin with the following straightforward result.

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

- (*centralization and normalization*) $\mathbb{E}Z(t_i) = 0$ *and* $\mathbb{E}Z(t_i)^2 = 1$ *for all* $1 \le i \le n$;
- (*no repeated variables*) $|r_{i,j}| < 1$ *whenever* $i \ne j$.

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

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

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

$$\mathbb{P}\bigg(V_j \le \frac{u - r_{j,m}(u+h)}{\sqrt{1-r_{j,m}^2}}\ \forall j \le m-1\bigg),$$

and the $V_j = V_{j,m}$ *are centralized, normalized, 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 Section 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 Section 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 $\{Z(t_i)\}_{1\le i\le m}$ 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 show this we formulate the following result.

PROPOSITION 2 (Comparison step). *Let* $u \ge 0$, *and suppose* h *is sufficiently small that all the upper bounds* $(u - r_{j,m}(u+h))/\sqrt{1-r_{j,m}^2}$ *in the definition of* $P(m,h)$ *are nonnegative. 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 nondecreasing sequence*, $1 \le j \le m-1$;
(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* $1 \le j,k \le m-1$.

Then for any $\delta \ge 0$,

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

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

We will prove Proposition 2 in Section 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 analyze those. The reader might think of this procedure as pulling out some of the dependence among the V_j, to be analyzed nontrivially 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 may wonder where the c_j, d_j will 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* $\{Z(t_i)\}_{1\le i\le n}$ *be as in Proposition* 1, *and suppose further that the sequence is* stationary, *that is*, *that* $r_{j,k} = r(|j-k|)$ *for some function* r. *Let* $u \ge 1$, *and suppose that*:

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

Then

$$\mathbb{P}\Big(\max_{1\le i\le n} Z(t_i) > u\Big) \ge n\frac{e^{-u^2/2}}{40u}\min\left\{1, \sqrt{\frac{1-r(1)}{u^2 r(1)}}\right\}$$
$$\times \prod_{j=1}^{n-1}\Phi\left(u\sqrt{1-r(j)}\left(1+O\left(\frac{1}{u^2(1-r(j))}\right)\right)\right),$$

where the implicit constant in the "big Oh" notation is absolute [*in particular*, *not depending on* $\{Z(t_i)\}_{1\le i\le n}$], *and could be found explicitly*.

Theorem 1 follows by taking $H = u^{-1}$, $\delta = \min\{u^{-2}, \sqrt{r(1)/u^2(1-r(1))}\}$, $c_j = r_{j,m} = r(|m-j|)$, $d_j = 1 - r_{j,m} = 1 - r(|m-j|)$ 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., Leadbetter, Lindgren and Rootzén [10], Chapter 4) 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 at least $1/2$, say. 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. [The bound in Theorem 1 is seen to be good because, since we assumed that $r(m)$ is nonnegative, the tail probability cannot be larger than $1 - \Phi(u)^n = O(ne^{-u^2/2}/u)$. An unfamiliar reader may deduce this from Comparison Inequality 2 in Section 3.1.]

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

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

where $C > 0$ and $0 < \alpha \leq 2$. In particular, a 1969 theorem of Pickands [15] describes the asymptotic behavior of suprema of such processes; if $h > 0$ is fixed and if $\sup_{\varepsilon \leq t \leq h} r(t) < 1$ for all $\varepsilon > 0$, then

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

where H_{α} is the so-called *Pickands constant*. In a second paper [16], Pickands used a result like this to determine the limiting distribution, as $T \to \infty$, of a scaled version of $\sup_{0 \leq t \leq T} Z(t)$. The scaling in that theorem thus involves H_{α} (see, e.g., the paper of Shao [20] 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 to hold for $\alpha = 1, 2$. Bounds are available more generally; for example, Shao [20] used a representation of H_{α} in terms of a nonstationary process, and various techniques from Gaussian process theory, to show that

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

when $0 < \alpha < 1$, and other bounds when $1 \leq \alpha \leq 2$. Dębicki and Kisowski [3] subsequently improved the upper bound on the range $1 < \alpha < 2$. Dębicki, Michna and Rolski [4] proved that

$$\frac{\alpha}{8\Gamma(1/\alpha)}\Big(\frac{1}{4}\Big)^{1/\alpha} \leq H_{\alpha}, \qquad 0 < \alpha \leq 2,$$

and in a 2009 preprint Michna [13] 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 [20] under that limit process.

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

COROLLARY 1. *There is an absolute constant* $c > 0$, *which could be found explicitly, such that* $H_\alpha \ge c\sqrt{\alpha}(e\alpha/2)^{1/\alpha}$ *for all* $0 < \alpha \le 2$.

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.

The motivation for studying this is its connection with a number-theoretic problem of Wintner [21]. Let ε_p be a sequence of independent Rademacher random variables [so that $\mathbb{P}(\varepsilon_p = 1) = \mathbb{P}(\varepsilon_p = -1) = 1/2$] and construct a "random multiplicative function" from these, as

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

We also set $M(x) := \sum_{n \le x} f(n)$. One can view $f(n)$ as a heuristic model for some deterministic functions occurring in number theory, such as the Möbius function. There has been quite a lot of recent work on the behavior of $f(n)$ (e.g., by Chatterjee and Soundararajan [2], Harper [7], Hough [8] and Lau, Tenenbaum and Wu [9]). However, the best known lower bound result for $|M(x)|$ remains that of Halász [6], who proved in 1982 that there exists a constant $B > 0$ such that, almost surely,

$$M(x) \ne O\left(\sqrt{x} e^{-B\sqrt{\log\log x \log\log\log x}}\right) \qquad \text{as } x \to \infty.$$

His proof, discussed in Appendix A, shows that lower bound information about the supremum of a certain Rademacher process (essentially the process above, with the g_p replaced by independent Rademacher random variables) can be translated into lower bound information about $|M(x)|$.

In Section 6, we use Propositions 1 and 2 to prove results like the following.

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

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

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

We stress that Corollary 2 is *not* an asymptotic result for a single Gaussian process, but a statement about an infinite sequence of processes depending on the parameter x. As x grows, the variance of the random sum grows for each fixed t, but also the correlation at nearby values of t decreases. (The reader may wish to look back at these comments after he or she has read Section 6.1.) For each x, we now know the supremum will typically exceed the level $\log\log x - \log\log\log x + O((\log\log\log x)^{3/4})$; standard methods 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. (For each x the process is "almost" stationary, as explained in Section 6.1, and a simple adaptation of Rice's formula yields upper bounds for its supremum.) 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 Appendix B, Corollary 2 allows a substantial improvement of Halász's [6] result about $M(x)$. However, it is possible to do better still.

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*

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

Corollary 2 implies Corollary 3 with the restriction $A > 3$, and this is proved in Section 6. The proofs in Section 6 are a bit fiddly, mostly because we must handle error terms arising in estimates for prime number sums. A "repeated sampling" argument is used to deduce Corollary 2 in the form we need, and care is also needed to arrange that the multivariate central limit theorem applies. However, ultimately our results follow simply by substituting some correlation values and parameter choices into Propositions 1 and 2. To prove Corollary 3 for all $A > 2.5$, an argument by contradiction is needed to slightly sharpen the result of Proposition 2. This is given in Section 7.

It seems extremely likely that, almost surely, $M(x) \neq O(\sqrt{x})$, and perhaps $M(x)$ almost surely has fluctuations of order $\sqrt{x\log\log x}$ (by analogy with Kolmogorov's law of the iterated logarithm). In fact, $M(x)$ might well exhibit even larger fluctuations, since its probability distribution may have rather heavy tails (see, e.g., Harper's article [7]). However, an argument like our own, ultimately based on studying a certain average of $M(x)$ (see Appendix A for justification of this comment), 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. 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 Section 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

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

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

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

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

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

These have mean zero and correlations

$$r_{j,k} - r_{j,m}r_{k,m}, \qquad 1 \le j,k \le m-1,$$

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

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

from which the proposition follows. □

In our applications, it will turn out that

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

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 $\mathbb{P}(\max_{1\le i\le n} Z(t_i) > u)$, rather than trying to understand $\mathbb{P}(Z(t_i) \le u \ \forall 1 \le i \le n)$ 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 give an unfamiliar reader some idea of how the method of comparison, as it is referred to by Piterbarg [17], is traditionally employed. Our treatment largely follows Li and Shao [11], although we would also like to draw attention to a 1954 paper of Plackett [18] which contains a similar presentation of the basic comparison result. (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.)

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 $\tilde{X} = (X_1, \ldots, X_n)$ and $\tilde{W} = (W_1, \ldots, W_n)$ be centralized and normalized n-variate normal vectors, with covariance matrices* $\mathrm{Var}(\tilde{X}) = (\mathrm{Cov}(X_i, X_j))_{1 \leq i,j \leq n} = (r_{i,j}^{(1)})$ *and* $\mathrm{Var}(\tilde{W}) = (r_{i,j}^{(0)})$ *that are nonsingular. Let $\tilde{u} \in \mathbb{R}^n$. Then*

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

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

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

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

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

To prove the formula one writes

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

observing that

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

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 behavior. The inequalities that we derive next may express this in a more striking way; they are a composite of results of Li and Shao [11] and of Leadbetter, Lindgren and Rootzén [10], although in most respects are unchanged from bounds of Slepian, Berman and Cramér from the 1960s (see Leadbetter, Lindgren and Rootzén's book for the history and references).

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 $\mathbf{1}$ 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)

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

(ii)

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

(iii)

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

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

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

For bound (ii), 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 [10] point out, the assumption that $\tilde{X}$ and $\tilde{W}$ are nonsingular 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, for example, that it looks like the identity matrix (see the paper of Li and Shao [11] 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 $\tilde{X} = (X_1, \dots, X_n)$, $\tilde{W} = (W_1, \dots, W_n)$ be centralized and normalized 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 $\tilde{u} = (u, u, \dots, u)$, 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

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

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

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

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

$$\begin{aligned}
&\mathbb{P}(X_1, \dots, X_{m-1} \le u, X_m > u) - \mathbb{P}(W_1, \dots, W_{m-1} \le u, W_m > u)\\
&\quad = -\sum_{1 \le i \le m-1} \big(r_{i,m}^{(1)} - r_{i,m}^{(0)}\big) \int_0^1 \phi\big(u, u; r_{i,m}^{(h)}\big)\\
&\qquad\qquad\qquad\qquad \times \mathbb{P}\big(\tilde{Z}^{(h)} \le \tilde{u} \,|\, Z_i^{(h)} = u, Z_m^{(h)} = u\big)\, dh\\
&\quad = -\sum_{1 \le i \le m-1} \frac{r_{i,m} - r_{i,m-1}}{1+\varepsilon^2} \int_0^1 \phi\left(u, u; \frac{h r_{i,m} + (1-h) r_{i,m-1}}{1+\varepsilon^2}\right)\\
&\qquad\qquad\qquad\qquad \times P(i, h, \varepsilon)\, dh,
\end{aligned}$$

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

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

It is less straightforward to analyze $P(i, h, \varepsilon)$ for $1 \le i \le m-2$ than to analyze $P(m-1, h, \varepsilon)$, 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 analyze $P(m-1, h, \varepsilon)$ one can note (as did Li and Shao [11]) that for any $1 \le i \le m-1$, the collection of random variables

$$\begin{aligned} Y_j^{(h)} &:= Z_j^{(h)} - \left(\frac{r_{j,i}^{(h)} - r_{i,m}^{(h)} r_{j,m}^{(h)}}{1 - (r_{i,m}^{(h)})^2}\right) Z_i^{(h)} - \left(\frac{r_{j,m}^{(h)} - r_{i,m}^{(h)} r_{j,i}^{(h)}}{1 - (r_{i,m}^{(h)})^2}\right) Z_m^{(h)} \\ &= Z_j^{(1)} - \left(\frac{r_{j,i}^{(1)} - r_{i,m}^{(h)} r_{j,m}^{(h)}}{1 - (r_{i,m}^{(h)})^2}\right) Z_i^{(1)} - \left(\frac{r_{j,m}^{(h)} - r_{i,m}^{(h)} r_{j,i}^{(1)}}{1 - (r_{i,m}^{(h)})^2}\right) Z_m^{(h)}, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 1 \le j \le m-1,\ j \ne i, \end{aligned}$$

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. [Since, in our examples, $Z(t_{m-1})$ and $Z(t_m)$ are always very highly correlated, and so $P(m-1, h, \varepsilon)$ 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 $1 \le j, k \le m-1$ and $j \ne k$, $\mathbb{E} V_j V_k$ 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 $Y_1, \ldots, Y_n, Z_1, \ldots, Z_n$ be independent standard normal random variables, and for $1 \le i \le n$ let α_i, β_i be real numbers satisfying

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

Then the random variables

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

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

$$\mathbb{E} X_i X_j = \beta_i \beta_j \sum_{k \le \min\{i,j\}} \alpha_k^2, \qquad i \ne j.$$

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

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

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

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

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 $\mathbb{P}(\sum_{j \le i} \alpha_j Y_j \le \frac{\delta u_i}{\beta_i} \ \forall 1 \le i \le m-1)$.

It should not come as a surprise that the behavior of partial sums of independent normal random variables is rather well understood. For example, writing $\{W_t\}_{t \ge 0}$ for the standard Brownian motion (see, e.g., Lifshits [12], Chapter 5, for much discussion of this process), one has the following neat result, which we quote from Grimmett and Stirzaker [5], Chapter 13.4: if $t \ge 0$, then

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

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

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

where

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

as claimed in Proposition 2. $\square$

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 analyzed 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 modeling 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 Section 7.

5. Application to estimating Pickands' constants. Suppose that $t_1 < t_2 < \cdots < t_M$ 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

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

In a paper from 1996, Shao [20] considers a mean zero, variance one stationary Gaussian process indexed by the half-line $[0, \infty)$, with covariance function

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

Such a process exists for each fixed $0 < \alpha < 2$. As $t \to 0$, we see (as did Shao [20]) that $r(t) = 1 - t^{\alpha}/2 + O(t^2)$. We also note that, for $t > 0$,

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

From now on it will be convenient to employ Vinogradov's notation $\gg$, meaning "greater than, up to a multiplicative constant." Thus $p(\alpha) \gg q(\alpha)$ means the same as $q(\alpha) = O(p(\alpha))$. In proving Corollary 1, we shall assume that α is smaller than a certain positive constant less than 1; a suitable explicit value could be extracted from our calculations if desired. There is no loss in this because $H_\alpha \gg 1 \gg \sqrt{\alpha}(e\alpha/2)^{1/\alpha}$ for α larger than such a constant. To prove the corollary, we will study Shao's stationary process at the sample points $t_i = i/M$. Simply choosing $a = 1$ in the above discussion, and comparing with Pickands' theorem in the Introduction, we see

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

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

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

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

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

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 $1 \le b \le 10$ (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

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

If u, and therefore the argument of Φ, is large enough, this is

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

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

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

$$u^2 j^2/M^2 \le u^2 M^{-3/2} = O(u^{-1}\alpha^{-3/2})$$

and

$$M^\alpha/u^2 j^\alpha = O(\alpha/j^\alpha)$$

and so the terms in the product are

$$\Phi\big((1 + O(\alpha/j^\alpha))(1-\delta)\sqrt{j^\alpha/(b\alpha) + O(1/u\alpha^{3/2})}\big) \\ = \Phi\big((1 + O(\alpha/j^\alpha))(1-\delta)\sqrt{j^\alpha/(b\alpha)}\big).$$

Thus, since $\Phi(x) \ge 1 - x^{-1}e^{-x^2/2} \ge e^{-2x^{-1}e^{-x^2/2}}$ for $x \ge 2$, the part of the product over $j \le M^{1/4}$ is at least $e^{-f(b,\delta,\alpha,u)}$, where

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

(Since we assume that α and δ are small, the arguments of Φ are all at least 2.) Now

$$\sum_{j \le M^{1/4}} e^{-(1-\delta)^2 j^\alpha/2b\alpha} \le \int_0^{M^{1/4}} e^{-(1-\delta)^2 t^\alpha/2b\alpha}\,dt \\ = \frac{2b}{(1-\delta)^2}\int_0^{(1-\delta)^2 M^{\alpha/4}/2b\alpha}\bigg(\frac{2b\alpha y}{(1-\delta)^2}\bigg)^{1/\alpha - 1} e^{-y}\,dy \\ \le \bigg(\frac{2b}{(1-\delta)^2}\bigg)^{1/\alpha}\alpha^{1/\alpha-1}\Gamma(1/\alpha).$$

By Stirling's formula, the right-hand side is asymptotic to

$$\sqrt{2\pi/\alpha}\big(2b/e(1-\delta)^2\big)^{1/\alpha}$$

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

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

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 $e/2$ while still ensuring that $f(b, \delta, \alpha, u) = O(1)$. Corollary 1 follows from making these choices. $\square$

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 [6], but we must 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

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

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

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

By a strong form of the prime number theorem (see, e.g., Montgomery and Vaughan [14], Chapter 6) we have

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

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

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

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

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

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 $\log y \leq u \leq \log x$. We deduce that if $s, t \geq 1$, and $s \neq t$, then

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

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

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

We seek lower bound information on $\max_{0 \leq n \leq B} \sup_{t \in \mathcal{T}_n} Z_y(t)$, 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 behavior of $Z_y(t)$ on different sets is roughly independent (see Section 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 Section 6.3.

The parameter E controls the spacing of sample points within their blocks $\mathcal{T}_n$, and in Section 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 $E \le e^{\sqrt{\log\log x}}$, 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 $\mathbb{E}Z_y(t)Z_y(s) \ge 1/\log\log x$, 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 Section 6.3, but we declare for now that we shall certainly have $\log x \le y \le e^{(\log\log x)^{100}}$.

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

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

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

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

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 $1 \le m \le M$ and $h \le H$. Let $j < k \le m-1$. If $|j-k| \le \log^{1/3} x$ then

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

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 $0 < \alpha \le 1/\log y$, we have

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

and we always have $r_{j,m} \ge 1/\log\log x \ge 1/\log y$ because $|m-j| \le M$. This means that $r_{j,k} - r_{j,m} r_{k,m} \ge r_{j,m}(1 - r_{k,m} + O(1/\log\log x))$, so it is legitimate to choose

$$\begin{aligned} c_j &= 1 - \frac{\log((m-j)E) + O(1)}{\log\log x - \log\log y}, \\ d_j &= \frac{\log((m-j)E) + O(1)}{\log\log x - \log\log y}, \qquad 1 \le j \le m-1, \end{aligned}$$

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

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

It follows from all of this that, for $1 \le m \le M$ and $h \le H$,

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

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

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

6.3. *Exploitation of the lower bound.* The lower bound obtained at the end of Section 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 just sample a few times we will very likely 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 $Be^{-d\sqrt{\log y}} \le \frac{1}{\log y}$, say, then for distinct $1 \le s, t \le 2B+2$ we have

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

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

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

Thus, by the second bound in Comparison Inequality 1,

$$\begin{aligned}
&\left|\mathbb{P}\left(\max_{0\le n\le B}\sup_{t\in\mathcal{T}_n} Z_y(t) \le \sqrt{2(\log\log x - \log\log y)}\right)\right.\\
&\qquad\left. - \prod_{0\le n\le B}\mathbb{P}\left(\sup_{t\in\mathcal{T}_n} Z_y(t) \le \sqrt{2(\log\log x - \log\log y)}\right)\right|\\
&\quad\ll \frac{\log^2 y}{\log^2 x}\sum_{0\le i<j\le B}\sum_{1\le k,l\le M}\left|\mathbb{E}Z_y\left(2i+1+\frac{kE}{\log x}\right)Z_y\left(2j+1+\frac{lE}{\log x}\right)\right|\\
&\quad\ll \frac{\log^2 y M^2}{\log^2 x \log\log x}\sum_{0\le i<j\le B}\left(\frac{1}{|i-j|\log y} + (i+j)e^{-d\sqrt{\log y}}\right)\\
&\quad\ll \frac{1}{(\log\log x)^2}\left(\frac{B\log B}{\log y} + B^3 e^{-d\sqrt{\log y}}\right).
\end{aligned}$$

We noted above that, at the level of precision required in Section 6.2, the correlation structure of $\{Z_y(t)\}_{t\in\mathcal{T}_n}$ is the same for each $0 \le n \le B$. 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, so that $\mathbb{P}(\sup_{t\in\mathcal{T}_n} Z_y(t) \le \sqrt{2(\log\log x - \log\log y)}) \le e^{-\Theta(\sqrt{\log\log\log x}/(\log\log x)^2)}$ for each $0 \leq n \leq B$, and

$$\mathbb{P}\Big(\max_{0\leq n\leq B} \sup_{t\in\mathcal{T}_n} Z_y(t) \leq \sqrt{2(\log\log x - \log\log y)}\Big)$$
$$\ll \frac{1}{(\log\log x)^2}\Big(\frac{B\log B}{\log y} + B^3 e^{-d\sqrt{\log y}}\Big)$$
$$+ e^{-\Theta((B+1)\sqrt{\log\log\log x}/(\log\log x)^2)}.$$

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 probability estimate in which $\max_{0\leq n\leq B}\sup_{t\in\mathcal{T}_n} Z_y(t)$ is replaced by

$$\max_{0\leq n\leq B} \sup_{t\in\mathcal{T}_n} \frac{\sum_{y\leq p\leq x} f(p)\cos(t\log p)/p^{1/2+1/\log x}}{\sqrt{\sum_{y\leq p\leq x}\cos^2(t\log p)/p^{1+2/\log x}}}$$

with $f(p)$ independent Rademacher random variables. This can be achieved using a multivariate central limit theorem, as explained in Appendix B, if we replace the upper bound $\sqrt{2(\log\log x - \log\log y)}$ that we demand 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 $t \in \mathbb{R}$,

$$\mathbb{E}\Big(\sum_{p<y}\frac{g_p\cos(t\log p)}{p^{1/2+1/\log x}}\Big)^2 = \mathbb{E}\Big(\sum_{p<y}\frac{f(p)\cos(t\log p)}{p^{1/2+1/\log x}}\Big)^2 = O(\log\log y)$$
$$= O(\log\log\log x)$$

as $x \to \infty$, as in Section 6.1. Applying Chebyshev's inequality to this estimate,

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

also if the g_p are replaced by Rademacher random variables $f(p)$. These sums are independent of the sums over $y \leq p \leq x$, so temporarily setting $d(x) := \inf_{1\leq t\leq 2(\log\log x)^2}\sqrt{\mathbb{E}(\sum_{y\leq p\leq x} g_p\cos(t\log p)/p^{1/2+1/\log x})^2}$ we find

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

is $O((\log\log\log x)^{-1/2})$. Corollary 2 quickly follows since, by the calculations in Section 6.1, we have $d(x) = \sqrt{(\log\log x - \log\log y)/2 + O(1)}$. □

As noted in Appendix A, 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

$$\begin{aligned}\mathbb{E}\biggl(\sum_{p>x}\frac{f(p)\cos(t\log p)}{p^{1/2+1/\log x}}\biggr)^2 &\le \sum_{p>x}\frac{1}{p^{1+2/\log x}}\\ &= O\biggl(\int_x^\infty \frac{du}{u^{1+2/\log x}\log u}\biggr)\\ &= O(1).\end{aligned}$$

Applying Chebyshev's inequality again, together with the Rademacher version of our estimate for $Z_y(t)$, we have that

$$\begin{aligned}\mathbb{P}\biggl(&\sup_{1\le t\le 2(\log\log x)^2}\sum_p \frac{f(p)\cos(t\log p)}{p^{1/2+1/\log x}}\\ &\le \log\log x - \log\log y - O(1) - (\log\log\log x)^{3/4}\biggr)\end{aligned}$$

is $O((\log\log\log x)^{-1/2})$. Applying the first Borel–Cantelli lemma at a lacunary set of points x, one quickly deduces that for any fixed $A > 3$, there almost surely exists a sequence (x_k), tending to infinity, with

$$\begin{aligned}&\sup_{1\le t\le 2(\log\log x_k)^2}\sum_p \frac{f(p)\cos(t\log p)}{p^{1/2+1/\log x_k}} - 2\log\log\log x_k\\ &\quad\ge \log\log x_k - A\log\log\log x_k.\end{aligned}$$

By the argument in Appendix A (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 Section 4, Proposition 2 may be refined in that the product term can be replaced by any lower bound for

$$\mathbb{P}\biggl(A_j \le \frac{(1-\delta)(u - r_{j,m}(u+h))}{\sqrt{1 - r_{j,m}^2 - c_j d_j}}\ \forall 1 \le j \le m-1\biggr)$$

for *any* standard normal random variables A_j satisfying

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

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 $U(j,k) \ge 0$.

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

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

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, on such an interval the upper bound $(1-\delta)(u-r_{j,m}(u+h))/\sqrt{1-r_{j,m}^2-c_jd_j}$ that we demand for the A_j is at least $(1+O(1/i\log L))\sqrt{2i\log L}$. Thus, taking A_j on distinct intervals to be independent of one another (rather than *all* A_j necessarily being independent), we can replace the product in Proposition 2 by

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

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. In its statement 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 $y=\log^8 x$.

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 Section 6.1, for such a choice of E. Let $\varepsilon(x)$ be any function tending to zero as $x\to\infty$. Then for some sequence of x, tending to infinity, we have*

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

Recall from Section 6.1 that

$$\mathbb{E}Z_y^x\left(1+\frac{jE}{\log x}\right)Z_y^x\left(1+\frac{kE}{\log x}\right)=1-\frac{\log(|j-k|E)+O(1)}{\log\log x-\log\log y},$$
$$1\le j,k\le\frac{\log x}{KE\log y},\ j\ne k,$$

where K is an absolute constant in the definition of $\mathcal{T}_0$. Let us fix a large absolute constant $C\in\mathbb{N}$, and set $L=1+1/KC^3$. When L^i is large enough, we can choose $x(i)\in\mathbb{R}$ such that

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

Here we wrote $y(i)=y(x(i))=\log^8 x(i)$. Then we will have

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

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

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

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

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

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

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

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 onward we would have

$$\mathbb{P}\Big(\sup_{t\in\mathcal{T}_0} Z_{y(i)}^{x(i)}(t) \le \sqrt{2\big(\log\log x(i) - \log\log y(i)\big)}\Big) \ge 1 - \frac{1}{(\log\log x(i))^{3/2}}$$
$$\ge 1 - O\Big(\frac{1}{(i\log L)^{3/2}}\Big),$$

so [since $\prod_{i=2}^{\infty}(1-1/i^{3/2})$ is convergent] the product term in Proposition 2 could be replaced by a positive constant. But then the argument of Section 6.2 would supply that

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

which is a contradiction for x sufficiently large. □

Armed with Proposition 3, we can repeat the argument of Section 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

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

while 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 $\mathcal{T}_n$, $n \ne 0$, so that E is larger but EK remains the same, we can arrange using Comparison Inequality 2 that

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

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 Section 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 sum $M(x) = \sum_{n\le x} f(n)$ 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 [6].

In view of Wintner's [21] result that for each $\varepsilon > 0$, $M(x) = O(x^{1/2+\varepsilon})$ 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

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

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 and the identity theorem of complex analysis. [The three series theorem 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 such a Dirichlet series is a holomorphic function strictly to the right of its abscissa of convergence.]

Thus in the domain $1/2 < \sigma < 1$, $1 \leq t \leq 2$, say, we almost surely have

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

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

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

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, for example, Montgomery and Vaughan [14], Theorem 6.7, 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*

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

for all $x \geq 2$.

Since Halász's paper [6] seems to be difficult to get hold of, it is perhaps worthwhile to briefly discuss Halász's own use of the foregoing argument. He shows that there 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

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

and

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

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

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

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

Very roughly, Halász [6] investigates the process $\sum_{p\le x} f(p)\frac{\cos(t\log p)}{\sqrt{p}}$, $t\in[1,2]$, by estimating moments of the counting function

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

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 Section 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 [19]. 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 [19] 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 $1\le i\le n$ and $t\in\mathcal{T}$. Suppose also that $(\varepsilon_i)_{i=1}^n$ 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

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

In the usual way, we construct random variables X_t' so $((X_t)_{t\in\mathcal{T}},(X_t')_{t\in\mathcal{T}})$ is an exchangeable pair of vectors [i.e., so 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 $\{1,2,\dots,n\}$, independently of everything else and let $(\varepsilon_i')_{i=1}^n$ be an independent copy of $(\varepsilon_i)_{i=1}^n$. We define X_t' as follows: conditional on the event $\{I=i\}$, set

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

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

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

With a view to applying Theorem 2.1 of Reinert and Röllin [19], we calculate two further quantities:

$$\begin{aligned}
\mathbb{E}\big((X_t'-X_t)(X_s'-X_s)|(X_u)_{u\in\mathcal{T}}\big) &= \frac{1}{n}\sum_{i=1}^{n}\alpha_i(t)\alpha_i(s)\mathbb{E}\big((\varepsilon_i'-\varepsilon_i)^2|(X_u)_{u\in\mathcal{T}}\big)\\
&= \frac{2}{n}\sum_{i=1}^{n}\alpha_i(t)\alpha_i(s);\\
\mathbb{E}|(X_t'-X_t)(X_s'-X_s)(X_u'-X_u)| &= \frac{1}{n}\sum_{i=1}^{n}|\alpha_i(t)\alpha_i(s)\alpha_i(u)|\mathbb{E}|\varepsilon_i'-\varepsilon_i|^3\\
&= \frac{4}{n}\sum_{i=1}^{n}|\alpha_i(t)\alpha_i(s)\alpha_i(u)|.
\end{aligned}$$

The reader should notice that, while 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 [19] 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 nonsingular, their theorem implies that

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

The condition that the covariance matrix should be nonsingular 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 $\delta\to 0$.

Specializing 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 [19] 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

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

The interval on which $s(z)$ must drop from 1 to 0 has length $\Theta(1/\sqrt{\log\log x})$, so we can find such s with derivatives satisfying $|s^{(r)}(z)|=O((\log\log x)^{r/2})$, $0\le$

$r \leq 3$, $z \in \mathbb{R}$. Setting $h((x_t)_{t\in\mathcal{T}}) = \prod_{t\in\mathcal{T}} s(x_t)$, we conclude that

$$\begin{aligned}
&\mathbb{P}\Bigl(\max_{t\in\mathcal{T}} X_t \leq \sqrt{2(\log\log x - \log\log y) - 1}\Bigr) \\
&\quad\leq \mathbb{P}\Bigl(\max_{t\in\mathcal{T}} Y_t \leq \sqrt{2(\log\log x - \log\log y)}\Bigr) \\
&\qquad + O\Biggl((\log\log x)^{3/2}(\#\mathcal{T})^3 \sum_{i=1}^{n} \max_{t\in\mathcal{T}} |\alpha_i(t)|^3\Biggr).
\end{aligned}$$

The reader may check that in the random multiplicative functions case, the error term on the right-hand side 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.

Acknowledgments. The author would like to thank his Ph.D. 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 Nathanaël Berestycki and Richard Nickl for discussions on these topics and the anonymous referee for his or her comments.

REFERENCES

[1] BURNECKI, K. and MICHNA, Z. (2002). Simulation of Pickands constants. *Probab. Math. Statist.* **22** 193–199. MR1944151

[2] CHATTERJEE, S. and SOUNDARARAJAN, K. (2012). Random multiplicative functions in short intervals. *Int. Math. Res. Not.* **3** 479–492.

[3] DĘBICKI, K. and KISOWSKI, P. (2008). A note on upper estimates for Pickands constants. *Statist. Probab. Lett.* **78** 2046–2051. MR2458013

[4] DĘBICKI, K., MICHNA, Z. and ROLSKI, T. (2003). Simulation of the asymptotic constant in some fluid models. *Stoch. Models* **19** 407–423. MR1993949

[5] GRIMMETT, G. R. and STIRZAKER, D. R. (2001). *Probability and Random Processes*, 3rd ed. Oxford Univ. Press, New York. MR2059709

[6] HALÁSZ, G. (1983). On random multiplicative functions. In *Hubert Delange Colloquium* (*Orsay*, 1982). *Publications Mathématiques d'Orsay* **83** 74–96. Univ. Paris XI, Orsay. MR0728404

[7] HARPER, A. J. (2012). On the limit distributions of some sums of a random multiplicative function. *J. Reine Angew. Math.* To appear.

[8] HOUGH, B. (2011). Summation of a random multiplicative function on numbers having few prime factors. *Math. Proc. Cambridge Philos. Soc.* **150** 193–214. MR2770059

[9] LAU, Y. K., TENENBAUM, G. and WU, J. (2013). On mean values of random multiplicative functions. *Proc. Amer. Math. Soc.* **141** 409–420.

[10] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York. MR0691492

[11] LI, W. V. and SHAO, Q.-M. (2002). A normal comparison inequality and its applications. *Probab. Theory Related Fields* **122** 494–508. MR1902188

[12] LIFSHITS, M. A. (1995). *Gaussian Random Functions*. *Mathematics and Its Applications* **322**. Kluwer Academic, Dordrecht. MR1472736

[13] MICHNA, Z. (2009). Remarks on Pickands theorem. Preprint.

[14] MONTGOMERY, H. L. and VAUGHAN, R. C. (2007). *Multiplicative Number Theory. I. Classical Theory*. *Cambridge Studies in Advanced Mathematics* **97**. Cambridge Univ. Press, Cambridge. MR2378655

[15] PICKANDS, J. III (1969). Upcrossing probabilities for stationary Gaussian processes. *Trans. Amer. Math. Soc.* **145** 51–73. MR0250367

[16] PICKANDS, J. III (1969). Asymptotic properties of the maximum in a stationary Gaussian process. *Trans. Amer. Math. Soc.* **145** 75–86. MR0250368

[17] PITERBARG, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. *Translations of Mathematical Monographs* **148**. Amer. Math. Soc., Providence, RI. MR1361884

[18] PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biometrika* **41** 351–360. MR0065047

[19] REINERT, G. and RÖLLIN, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150–2173. MR2573554

[20] SHAO, Q.-M. (1996). Bounds and estimators of a basic constant in extreme value theory of Gaussian processes. *Statist. Sinica* **6** 245–257. MR1379060

[21] WINTNER, A. (1944). Random factorizations and Riemann's hypothesis. *Duke Math. J.* **11** 267–275. MR0010160

DEPARTMENT OF PURE MATHEMATICS
AND MATHEMATICAL STATISTICS
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD, CAMBRIDGE CB3 0WA
UNITED KINGDOM
E-MAIL: A.J.Harper@dpmms.cam.ac.uk