GAPS BETWEEN LOGS

JENS MARKLOF AND ANDREAS STRÖMBERGSSON

Abstract. We calculate the limiting gap distribution for the fractional parts of $\log n$, where n runs through all positive integers. By rescaling the sequence, the proof quickly reduces to an argument used by Barra and Gaspard in the context of level spacing statistics for quantum graphs. The key ingredient is Weyl equidistribution of irrational translations on multi-dimensional tori. Our results extend to logarithms with arbitrary base; we deduce explicit formulas when the base is transcendental or the rth root of an integer. If the base is close to one, the gap distribution is close to the exponential distribution.

§1. The gap distribution is a popular measure to quantify the degree of randomness in a given deterministic sequence. Rudnick and Zaharescu have shown [12] that the gap distribution of the fractional parts of $2^n\alpha$ $(n = 1, \dots, N)$, for almost all α, converges in the limit $N \to \infty$ to an exponential distribution—the gap distribution of a Poisson point process. (2^n may be replaced by any lacunary sequence.) The same is expected to hold for the fractional parts of $n^k\alpha$, for any integer $k \geq 2$ and α of bounded type, but this remains unproven [10, 11, 8, 4]. Numerical experiments also suggest that the fractional parts of n^β may have an exponential gap distribution provided $\beta \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$. In the case $\beta = \frac{1}{2}$, Elkies and McMullen [3] have calculated an explicit formula for the limiting gap distribution, which is evidently not exponential. Remarkably, their distribution coincides with the limiting gap distribution of directions of points in the affine lattice $\mathbb{Z}^2 + \boldsymbol{a}$ for any fixed vector $\boldsymbol{a} \notin \mathbb{Q}^2$ [9]. Both of these findings follow from the equidistribution of translates of certain curves on the homogeneous space $\mathrm{ASL}(2, \mathbb{Z}) \backslash \mathrm{ASL}(2, \mathbb{R})$. In the present paper we show that also the fractional parts of $\log_b n$ have a non-exponential limiting gap distribution, which we calculate explicitly. Our derivation reduces quickly to an argument used by Barra and Gaspard [1] in the context of spectral statistics of quantum graphs. The key ingredient here is Weyl equidistribution on multi-dimensional tori, similar in spirit to [5, Chap. 3].

§2. To state our main results, let us denote by ξ_n the fractional parts of $\log_b n$. We denote by $\{\xi_{n,N}\}_{n=1,\dots,N}$ the ordered set of the first N elements of ξ_n, so that

$$0 \leq \xi_{1,N} \leq \xi_{2,N} \leq \ldots \leq \xi_{N,N} < 1. \tag{1}$$

For purely notational reasons it will be convenient to set $\xi_{N+1,N} := \xi_{1,N} + 1$, and also consider the gap between the first and last element mod 1; this additional gap has of course no effect on the existence and form of the limiting gap distribution.

For given N, the gap distribution $P_N(s)$ is defined as the probability density on $\mathbb{R}_{\geq 0}$,

$$P_N(s) = \frac{1}{N} \sum_{n=1}^{N} \delta(s - N(\xi_{n+1,N} - \xi_{n,N})). \tag{2}$$

Date: 6 January 2013.

2010 *Mathematics Subject Classification.* 11K06, 11J71.

J.M. is supported by a Royal Society Wolfson Research Merit Award and ERC Advanced Grant HFAKT. A.S. is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

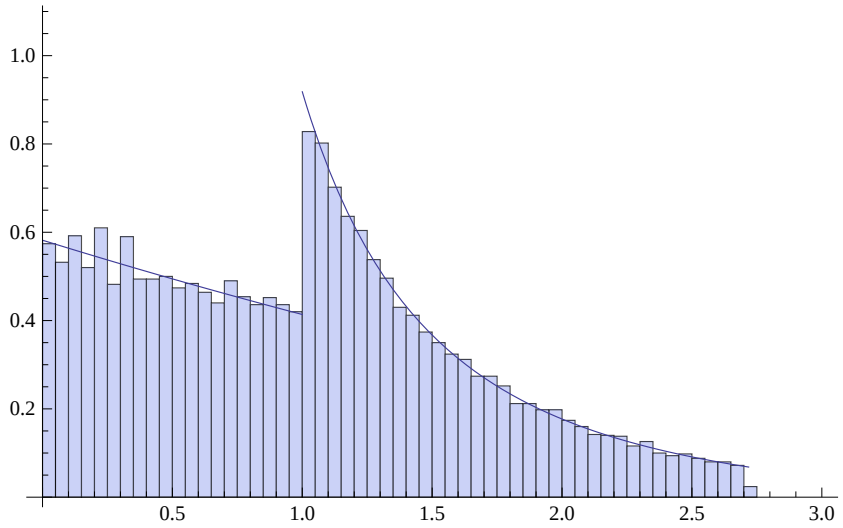

FIGURE 1. The distribution of gaps between the fractional parts of $\log n$, where $n = 1, \dots, 10^4$. The piecewise continuous curve is the limit distribution $P(s)$ of Theorem 1 for $b = \mathrm{e}$.

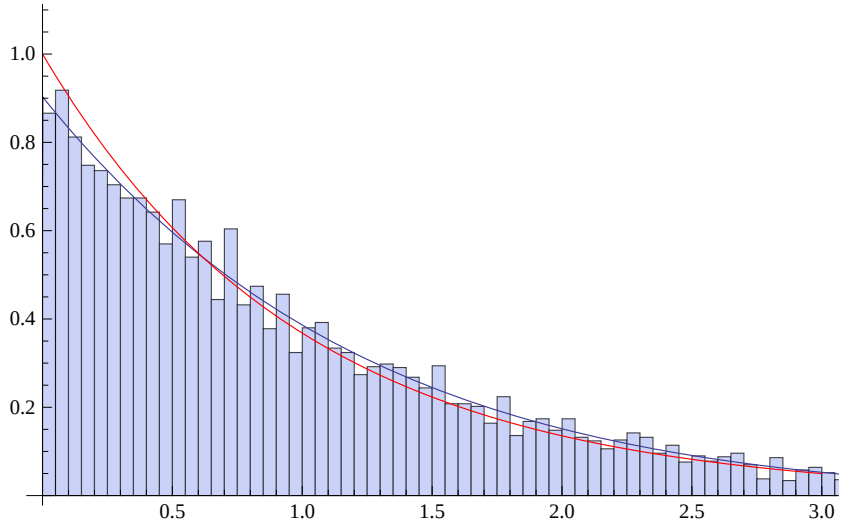

FIGURE 2. The distribution of gaps between the fractional parts of $\log_b n$ with base $b = \mathrm{e}^{1/5} = 1.221402758\ldots$, where $n = 1, \dots, 10^4$. The blue curve is the limit distribution $P(s)$ of Theorem 1, the red curve is the exponential distribution e^{-s}.

We denote by $\mathcal{T}$ the set of transcendental numbers $b > 1$; $b = \mathrm{e}$ is a natural example, cf. Figure 1. The technique we present here also works for algebraic b, but in general leads to more intricate limit distributions. In §10 we discuss the simple case of integer base b, and in §11 the case when the base is the rth root of an integer.

Theorem 1. *Let $b \in \mathcal{T}$. For any bounded continuous $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_0^\infty f(s)\, P_N(s)\, ds = \int_0^\infty f(s)\, P(s)\, ds, \tag{3}$$

where

$$P(s) = \begin{cases} \frac{1}{\log b}\, s^{-2}(F(s\log b) - F(sb^{-1}\log b)) & \text{if } 0 < s < \frac{1}{\log b} \\ -\frac{1}{\log b}\, s^{-2} F(sb^{-1}\log b) & \text{if } \frac{1}{\log b} < s < \frac{b}{\log b} \\ 0 & \text{if } s > \frac{b}{\log b}, \end{cases} \tag{4}$$

$$F(a) = a\frac{\partial}{\partial a}(a; b^{-1}) - (a; b^{-1}) \tag{5}$$

and

$$(a; q) = \prod_{n=0}^{\infty}(1 - aq^n) \qquad \textit{(q-Pochhammer symbol).} \tag{6}$$

The derivative of the Pochhammer symbol is explicitly

$$\frac{\partial}{\partial a}(a; q) = -(a; q)\sum_{j=0}^{\infty}\frac{1}{q^{-j} - a}. \tag{7}$$

Note that $P(s)$ has jump discontinuities at $s = \frac{1}{\log b}$ and $\frac{b}{\log b}$; however both the left and right limits exist at these points, and are finite. We also note that $\lim_{s\to 0^+} P(s) = \frac{\log b}{b-1}$.

A more elementary formulation, which is equivalent to the above, is that for every $s \geq 0$

$$\lim_{N\to\infty}\frac{1}{N}\big|\{n \leq N : N(\xi_{n+1,N} - \xi_{n,N}) > s\}\big| = \int_s^{\infty} P(s')\, ds'. \tag{8}$$

We will show in the final paragraph §12 of this paper that, in the limit $b \to 1$, $P(s)$ converges to the exponential distribution e^{-s}, cf. Figure 2.

§3. To gain a better insight into the limiting gap distribution $P(s)$ for fixed b, let us recall that the fractional parts of $\log_b n$ are *not* uniformly distributed mod 1 [6]. However, the sequence of η_n $(n = 1, \ldots, N)$ given by the fractional parts of $\log_b(n/N) = \log_b n - \log_b N$ has a limit density mod 1, and evidently the same gap distribution (to see this, recall that we have added the gap between the first and last element of the sequence mod 1, and the ordering in the sequence remains the same). It is an easy exercise to show that for any interval $A \subset [0, 1]$,

$$\lim_{N\to\infty}\frac{1}{N}\big|\{n \leq N : \eta_n \in A\}\big| = \int_A \rho(x)\, dx \tag{9}$$

with density

$$\rho(x) = \frac{\log b}{b-1}\, b^x. \tag{10}$$

§4. A more refined statement of Theorem 1 is to consider the joint distribution of η_n and the subsequent gap; define the corresponding probability density on $[0, 1] \times \mathbb{R}_{\geq 0}$ by

$$P_N(x, s) = \frac{1}{N}\sum_{n=1}^{N}\delta(x - \eta_{n,N})\ \delta(s - N(\eta_{n+1,N} - \eta_{n,N})), \tag{11}$$

where $\{\eta_{n,N}\}_{n=1,\ldots,N}$ is again the ordered set of the first N elements of η_n. The joint distribution allows us to compare the statistics of gaps between elements near one point x of the unit interval with the gap statistics at a second point x'.

The explicit formula for the limit distribution will involve the density

$$R(a,b) = (b^{-1};b^{-1})\,\delta(a-1) + \begin{cases} \dfrac{\partial^2}{\partial a^2}(a;b^{-1}) & \text{if } 0<a<1 \\ 0 & \text{if } a>1. \end{cases} \tag{12}$$

The second derivative of the Pochhammer symbol is

$$\frac{\partial^2}{\partial a^2}(a;q) = (a;q) \sum_{\substack{j,k=0 \\ j\neq k}}^{\infty} \frac{1}{(q^{-j}-a)(q^{-k}-a)}. \tag{13}$$

Theorem 2. *Let $b \in \mathcal{T}$. For any bounded continuous $f : [0,1] \times \mathbb{R}_{\geq 0} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_0^1 \int_0^\infty f(x,s)\, P_N(x,s)\, ds\, dx = \int_0^1 \int_0^\infty f(x,s)\, P(x,s)\, ds\, dx, \tag{14}$$

where

$$P(x,s) = (b^{x-1}\log b)^2 R(sb^{x-1}\log b, b). \tag{15}$$

§5. To unravel the previous statements further, let us consider a strictly increasing function $[0,1] \to [0,1]$ which maps the sequence η_j to a new sequence $\tilde\eta_j$ which now is uniformly distributed mod 1. This function is given by (cf. [7])

$$x \mapsto \int_0^x \rho(y)\, dy = \frac{b^x - 1}{b-1}, \tag{16}$$

and thus

$$\tilde\eta_n = \frac{b^{\eta_n}-1}{b-1}, \qquad \tilde\eta_{n,N} = \frac{b^{\eta_{n,N}}-1}{b-1}. \tag{17}$$

We have, explicitly, for $n \in [Nb^{-k}, Nb^{-k+1})$ and $k \in \mathbb{Z}_{\geq 0}$,

$$\tilde\eta_n = \frac{b^k n - N}{N(b-1)}. \tag{18}$$

We will see in §8 that the knowledge of the joint distribution $P_N(x,s)$ for the original sequence and the analogue for the rescaled and ordered sequence,

$$\tilde P_N(x,s) = \frac{1}{N}\sum_{n=1}^{N} \delta(x-\tilde\eta_{n,N})\, \delta(s - N(\tilde\eta_{n+1,N} - \tilde\eta_{n,N})) \tag{19}$$

are equivalent:

Theorem 3. *Let $b \in \mathcal{T}$. For any bounded continuous $f : [0,1] \times \mathbb{R}_{\geq 0} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_0^1 \int_0^\infty f(x,s)\, \tilde P_N(x,s)\, ds\, dx = \int_0^1 \int_0^\infty f(x,s)\, \tilde P(s)\, ds\, dx, \tag{20}$$

where

$$\tilde P(s) = (1-b^{-1})^2 R\big((1-b^{-1})s, b\big). \tag{21}$$

Note that the limit distribution is independent of x. This is not a consequence of $\tilde\eta_j$ being uniformly distributed mod 1, but reflects the fact that the gap distribution is the same if we restrict the sequence to an arbitrary subinterval of $[0,1]$.

We will first prove Theorem 3 in §6 and §7, and then retrace our steps by subsequently showing in §8 that Theorem 3 implies Theorem 2, where

$$P(x,s) = \tilde P(\rho(x)s)\rho(x)^2 = (b^{x-1}\log b)^2 R(sb^{x-1}\log b, b). \tag{22}$$

We obtain Theorem 1 from Theorem 2 by taking an x-independent test function.

§6. We now turn to the proof of Theorem 3 and study a more general setting. Consider a finite sequence $\omega_1, \dots, \omega_J$ of positive real numbers (the "frequencies"). With each ω_j we associate a real number β_j. We are interested in the statistical properties of the multiset

$$\mathcal{S} = \biguplus_{j=1}^{J} (\beta_j + \omega_j^{-1}\mathbb{Z}). \tag{23}$$

Here $\uplus$ denotes multiset sum, i.e. a union where we keep track of the multiplicity of each element (this is only relevant if the sets $\beta_j + \omega_j^{-1}\mathbb{Z}$, $j = 1, \dots, J$ are not mutually disjoint). The number of elements of $\mathcal{S}$ that fall into the interval $I = [t - \frac{L}{2}, t + \frac{L}{2}]$ is

$$N(t, L) = \sum_{j=1}^{J} N_j(t, L), \qquad N_j(t, L) = \big|I \cap (\beta_j + \omega_j^{-1}\mathbb{Z})\big|. \tag{24}$$

Evidently $N_j(t, L)$ is periodic in t with period ω_j^{-1}. Furthermore

$$\omega_j L - 1 < N_j(t, L) \le \omega_j L + 1. \tag{25}$$

Therefore

$$\lim_{L\to\infty} \frac{N(t, L)}{L} = \sum_{j=1}^{J} \omega_j, \tag{26}$$

which means that the asymptotic density of $\mathcal{S}$ in $\mathbb{R}$ equals $\sum_j \omega_j$.

For any $x \in \mathbb{R}$ we write $|x|_{\mathbb{Z}} \in [0, \frac{1}{2}]$ for the distance between x and the nearest integer. We also set

$$n(x, L) = \big|[x - \tfrac{L}{2}, x + \tfrac{L}{2}] \cap \mathbb{Z}\big|. \tag{27}$$

The verification of the following lemma is a simple exercise.

Lemma 1. *Given $x \in \mathbb{R}$, $L \ge 0$ and $k \in \mathbb{Z}_{\ge 0}$, we have $n(x, L) = k$ if and only if $|x - \frac{1}{2}k|_{\mathbb{Z}}$ is both $\ge \frac{1}{2}(k - L)$ and $> \frac{1}{2}(L - k)$.*

In particular, if x is picked at random with respect to the uniform probability measure $\mathbb{P}_0$ on $[0, 1]$, then

$$\mathbb{P}_0(n(x, L) = k) = E_1(k, L) := \begin{cases} 1 - |k - L| & \text{if } L - 1 < k < L + 1 \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

The following observation is due to Barra and Gaspard [1].

Theorem 4. *Assume that the frequencies $\omega_1, \dots, \omega_J$ are linearly independent over $\mathbb{Q}$. Then, for $-\infty < a < b < \infty$ and $k \in \mathbb{Z}_{\ge 0}$,*

$$\lim_{T\to\infty} \frac{\operatorname{meas}(\{t \in [aT, bT] : N(t, L) = k\})}{(b - a)T} = E(k, L) \tag{29}$$

where

$$E(k, L) = \sum_{k_1 + \dots + k_J = k} \prod_{j=1}^{J} E_1\big(k_j, \omega_j L\big). \tag{30}$$

In particular,

$$E(0,L)=\begin{cases}\prod_{j=1}^{J}\big(1-\omega_j L\big) & \text{if } L<\max(\omega_1,\dots,\omega_J)^{-1}\\ 0 & \text{otherwise.}\end{cases} \tag{31}$$

The convergence in (29) *is uniform over all choices of* $\beta_1,\dots,\beta_J\in\mathbb{R}$ *(but keeping* $\omega_1,\dots,\omega_J$ *fixed).*

Proof. Decomposing $[aT,bT]$ into intervals of length ω_1^{-1} and using $N(t,L)=\sum_{j=1}^{J}N_j(t,L)$, we see that $\mathrm{meas}(\{t\in[aT,bT]: N(t,L)=k\})$ equals

$$\sum_{k_1+\dots+k_J=k}\sum_{n=n_a}^{n_b-1}\omega_1^{-1}\,\mathrm{meas}(\{s\in[0,1): N_j(\omega_1^{-1}(s+n),L)=k_j \text{ for } j=1,\dots,J\})+E, \tag{32}$$

where $n_a=\lfloor\omega_1 aT\rfloor$, $n_b=\lfloor\omega_1 bT\rfloor$, and E is a real number whose absolute value is bounded above by the measure of the symmetric difference between $[aT,bT]$ and $[n_a\omega_1^{-1},n_b\omega_1^{-1}]$; thus $|E|\le 2\omega_1^{-1}$. Let us define $\alpha_{j,n}\in\mathbb{R}/\mathbb{Z}$ through the relation $\beta_j+\omega_j^{-1}\mathbb{Z}=n\omega_1^{-1}+\omega_j^{-1}(\alpha_{j,n}+\mathbb{Z})$. Then $N_j(\omega_1^{-1}(s+n),L)=k_j$ holds if and only if $\big|[\omega_1^{-1}s-\frac{L}{2},\omega_1^{-1}s+\frac{L}{2}]\cap\omega_j^{-1}(\alpha_{j,n}+\mathbb{Z})\big|=k_j$, and by Lemma 1 this holds if and only if $2\big|\frac{\omega_j}{\omega_1}s-\alpha_{j,n}-\frac12 k_j\big|_{\mathbb{Z}}$ is both $\ge k_j-\omega_j L$ and $>\omega_j L-k_j$. It follows that $\mathrm{meas}(\{t\in[aT,bT]: N(t,L)=k\})$ equals

$$\sum_{k_1+\dots+k_J=k}\sum_{n=n_a}^{n_b-1}\omega_1^{-1}f_{k_1,\dots,k_J}(\alpha_{1,n},\alpha_{2,n},\dots,\alpha_{J,n})+E, \tag{33}$$

where $f_{k_1,\dots,k_J}:(\mathbb{R}/\mathbb{Z})^J\to[0,1]$ is defined by

$$f_{k_1,\dots,k_J}(\alpha_1,\alpha_2,\dots,\alpha_J)=\int_0^1\prod_{j=1}^{J}I\Big(2\Big|\frac{\omega_j}{\omega_1}s-\alpha_j-\frac12 k_j\Big|_{\mathbb{Z}}>|k_j-\omega_j L|\Big)\,ds. \tag{34}$$

Note that each function $f_{k_1,\dots,k_J}$ is continuous on $(\mathbb{R}/\mathbb{Z})^J$. It follows from our definition of $\alpha_{j,n}$ that $\alpha_{j,n+1}=\alpha_{j,n}-\omega_j\omega_1^{-1} \mod\mathbb{Z}$. In particular $\alpha_{1,n}$ is independent of n; in fact $\alpha_{1,n}=\alpha_1:=\omega_1\beta_1$ in $\mathbb{R}/\mathbb{Z}$ for all n. On the other hand our assumption implies that the numbers $1,\omega_2\omega_1^{-1},\omega_3\omega_1^{-1},\dots,\omega_J\omega_1^{-1}$ are linearly independent over $\mathbb{Q}$, and hence by Weyl equidistribution we have

$$\frac{\sum_{n=n_a}^{n_b-1}f_{k_1,\dots,k_J}(\alpha_{1,n},\alpha_{2,n},\dots,\alpha_{J,n})}{n_b-n_a}\to\int_{(\mathbb{R}/\mathbb{Z})^{J-1}}f_{k_1,\dots,k_J}(\alpha_1,\alpha_2,\dots,\alpha_J)\,d\alpha_2\dots d\alpha_J \tag{35}$$

as $n_b-n_a\to\infty$. Note that the convergence here is uniform over all choices of $\beta_1,\dots,\beta_J\in\mathbb{R}$. The right hand side equals

$$\int_{(\mathbb{R}/\mathbb{Z})^{J-1}}\int_0^1\prod_{j=1}^{J}I\Big(2\Big|\frac{\omega_j}{\omega_1}s-\alpha_j-\frac12 k_j\Big|_{\mathbb{Z}}>|k_j-\omega_j L|\Big)\,ds, \tag{36}$$

and changing order of integration and then substituting $\alpha_j:=x_j+\frac{\omega_j}{\omega_1}s-\frac12 k_j$ ($x_j\in\mathbb{R}/\mathbb{Z}$) we see that the expression factors as $\prod_{j=1}^{J}E_1(k_j,\omega_j L)$. Finally noticing also that $n_b-n_a\sim\omega_1(b-a)T$ as $T\to\infty$, we obtain (29). □

We now turn to the gap distribution. We order the elements of $\mathcal{S}$ and label them as

$$\dots\le\lambda_{-2}\le\lambda_{-1}\le\lambda_0\le\lambda_1\le\lambda_2\le\dots. \tag{37}$$

Theorem 5. *Assume that the frequencies ω_j are linearly independent over $\mathbb{Q}$, and that ω_1 is the largest among $\omega_1, \dots, \omega_J$. Then, for $-\infty < a < b < \infty$, and any $s \geq 0$, $s \neq \omega_1^{-1}$,*

$$\lim_{T\to\infty} \frac{1}{(b-a)T} |\{j \in \mathbb{Z} : \lambda_j \in [aT, bT] \text{ and } \lambda_{j+1} - \lambda_j > s\}| = \int_s^\infty P_\omega(s')ds' \tag{38}$$

where

$$P_\omega(s) = \omega_1 \prod_{j=2}^{J} \Big(1 - \frac{\omega_j}{\omega_1}\Big)\delta(s - \omega_1^{-1}) + \begin{cases} \displaystyle\sum_{\substack{h,i=1 \\ h\neq i}}^{J} \omega_h\omega_i \Big[\prod_{\substack{j=1 \\ j\neq h,i}}^{J} \big(1 - \omega_j s\big)\Big] & \text{if } 0 < s < \omega_1^{-1} \\ 0 & \text{if } s > \omega_1^{-1}. \end{cases} \tag{39}$$

The convergence in (38) *is uniform over all choices of $\beta_1, \dots, \beta_J \in \mathbb{R}$ (but keeping $\omega_1, \dots, \omega_J$ fixed).*

Proof. This is a corollary of Theorem 4, cf. e.g. Theorem 2.2 in [7]. The limiting gap density is given by the formula

$$P_\omega(s) = \frac{d^2E(0,s)}{ds^2}. \tag{40}$$

□

§7. *Proof of Theorem 3.* It suffices to prove that (20) holds when f is the characteristic function of a box $[0, a] \times [0, A]$ with $A \neq (1 - b^{-1})^{-1}$, i.e. to prove that for any fixed $a \in [0, 1]$ and $A \geq 0$, $A \neq (1 - b^{-1})^{-1}$, we have

$$G_{a,A}(N) \to a \int_0^A \tilde{P}(s)\, ds \tag{41}$$

as $N \to \infty$, where

$$G_{a,A}(N) := \frac{1}{N} \Big| \Big\{ n \in \{1, \dots, N\} : \tilde{\eta}_{n,N} \in [0, a],\ N(\tilde{\eta}_{n+1,N} - \tilde{\eta}_{n,N}) \in [0, A] \Big\} \Big| \tag{42}$$

(cf., e.g., [2, Example 2.3]).

Set $\omega_j = b^{-j}(b-1)$ and $\beta_j^{(N)} = \beta^{(N)} = -\frac{N}{b-1}$ for $j = 1, 2, \dots$. For any given $J, N \in \mathbb{Z}_{\geq 1}$ we note that by (18), $\{N\tilde{\eta}_1, \dots, N\tilde{\eta}_N\}$ equals the multiset sum of $(\beta^{(N)} + \omega_j^{-1}\mathbb{Z}) \cap [0, N)$ for $j = 1, 2, \dots, J$ and a remaining multiset of cardinality $\lceil Nb^{-J} \rceil$ consisting of $N\tilde{\eta}_n$ for $n = N$ and all $n < Nb^{-J}$. Hence if we order the elements of $\mathcal{S}_N := \uplus_{j=1}^J (\beta^{(N)} + \omega_j^{-1}\mathbb{Z})$ as $\dots \leq \lambda_{-2} \leq \lambda_{-1} \leq \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots$, then

$$G_{a,A}(N) = \frac{1}{N} \Big| \Big\{ j \in \mathbb{Z} : \lambda_j \in [0, aN],\ \lambda_{j+1} - \lambda_j \in [0, A] \Big\} \Big| + O\big(N^{-1} + b^{-J}\big). \tag{43}$$

Since b is transcendental, the frequencies $\omega_1, \dots, \omega_J$ are linearly independent over $\mathbb{Q}$; thus Theorem 5 applies, and we obtain, for any fixed $J \in \mathbb{Z}_{\geq 1}$,

$$\begin{aligned} \limsup_{N\to\infty} G_{a,A}(N) &\leq a \int_0^A P_{\omega,J}(s')\, ds' + O(b^{-J}); \\ \liminf_{N\to\infty} G_{a,A}(N) &\geq a \int_0^A P_{\omega,J}(s')\, ds' - O(b^{-J}), \end{aligned} \tag{44}$$

where $P_{\omega,J}(s) = 0$ if $s > \omega_1^{-1}$ while for $s < \omega_1^{-1}$ we have

$$P_{\omega,J}(s) = \omega_1 \prod_{j=2}^{J} \Big(1 - \frac{\omega_j}{\omega_1}\Big)\delta(s - \omega_1^{-1}) + \frac{d^2}{ds^2} \prod_{j=1}^{J} \big(1 - \omega_j s\big). \tag{45}$$

We now compute that

$$\int_0^A P_{\omega,J}(s')\,ds' = \sum_{j=1}^{J} \omega_j - \begin{cases} 0 & \text{if } A > \omega_1^{-1} \\ \displaystyle\sum_{j=1}^{J} \omega_j \prod_{\substack{i=1 \\ (i\neq j)}}^{J} (1 - \omega_i A) & \text{if } A < \omega_1^{-1}. \end{cases} \tag{46}$$

Hence, letting $J \to \infty$ in (44) and using $\omega_j = b^{-j}(b-1)$, we conclude

(47)

$$\lim_{N\to\infty} G_{a,A}(N) = a + a \left\{ \begin{matrix} 0 & \text{if } A > \frac{b}{b-1} \\ \Big(\frac{\partial}{\partial s}\big((1-b^{-1})s; b^{-1}\big)\Big)_{|s=A} & \text{if } A < \frac{b}{b-1} \end{matrix} \right\} = a \int_0^A \tilde{P}(s)\,ds,$$

where the last equality follows by a direct computation using (12) and (21). Hence (41) holds, and Theorem 3 is proved. □

§8. Let us now prove that Theorem 3 implies Theorem 2. (An entirely analogous argument also shows that Theorem 2 implies Theorem 3, so that the statements of these two theorems are in fact equivalent.) Given a bounded continuous function $f : [0,1] \times \mathbb{R}_{\geq 0} \to \mathbb{R}$, we wish to prove that (14) holds. Since the sequence of probability measures $P_N(x,s)\,ds\,dx$ is tight (cf., e.g., Lemma 2.1 in [7]), a familiar approximation argument shows that without loss of generality we may assume that f has compact support, i.e. there is some $B > 0$ such that $f(x,s) = 0$ whenever $s \geq B$.

Let $T : [0,1] \to [0,1]$ be the map in (16) and define $\mathfrak{T} : [0,1] \times \mathbb{R}_{\geq 0} \to [0,1] \times \mathbb{R}_{\geq 0}$ through $\mathfrak{T}(x,s) = (T(x), T'(x)s)$. This is a bijection, with inverse given by

$$\mathfrak{T}^{-1}(x,s) = \Big(T^{-1}(x), \frac{s}{T'(T^{-1}(x))}\Big). \tag{48}$$

Now apply Theorem 3 with the test function $f \circ \mathfrak{T}^{-1}$. Then in (20) we have

$$\int_0^1 \int_0^\infty f \circ \mathfrak{T}^{-1}(x,s)\,\tilde{P}_N(x,s)\,ds\,dx = \frac{1}{N} \sum_{n=1}^{N} f(\eta_{n,N}, s_{n,N}), \tag{49}$$

where $s_{n,N} := N \cdot \dfrac{T(\eta_{n+1,N}) - T(\eta_{n,N})}{T'(\eta_{n,N})}$. Let $\mathcal{F}_N$ be the set of those $n \in \{1,\dots,N\}$ for which $\eta_{n+1,N} - \eta_{n,N} \leq BbN^{-1}$. Then since $T'(x) = \rho(x)$ is continuous and bounded away from zero on $[0,1]$ we have $s_{n,N} = N(\eta_{n+1,N} - \eta_{n,N}) + o(1)$ uniformly over all $n \in \mathcal{F}_N$ as $N \to \infty$, and hence, since f is uniformly continuous,

$$f(\eta_{n,N}, s_{n,N}) = f(\eta_{n,N}, N(\eta_{n+1,N} - \eta_{n,N})) + o(1), \tag{50}$$

uniformly over all $n \in \mathcal{F}_N$. On the other hand if $n \notin \mathcal{F}_N$ then $N(\eta_{n+1,N} - \eta_{n,N}) > Bb > B$ and also $s_{n,N} > Bb\frac{\inf T'}{\sup T'} = B$, so that both sides of (50) vanish. Hence (50) in fact holds uniformly over *all* $n \in \{1,\dots,N\}$ as $N \to \infty$, and it follows that the limit in the left hand side of (14) exists, and equals the limit in (20) (with $f \circ \mathfrak{T}^{-1}$). Finally substituting $(x,s) = \mathfrak{T}(x',s')$ in the right hand side of (20) we see (via (22)) that this limit equals the right hand side of (14), and the proof is complete.

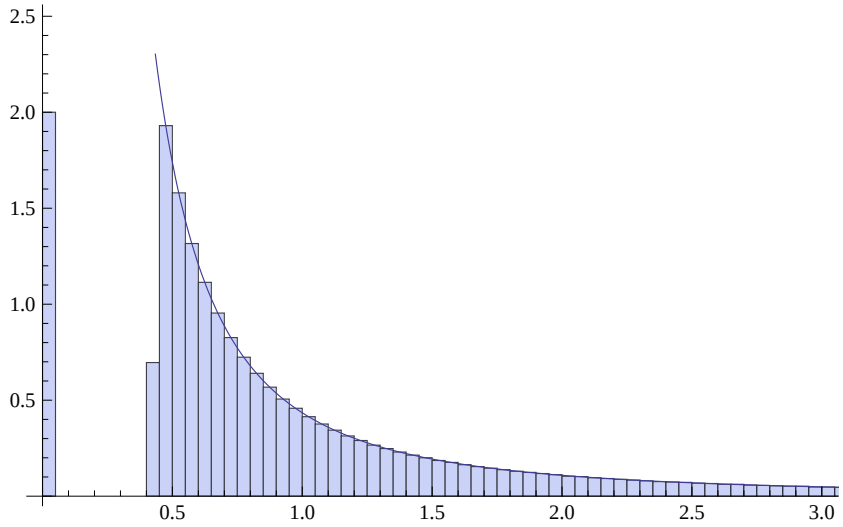

FIGURE 3. The distribution of gaps between the fractional parts of $\log_b n$ with base $b = 10$, where $n = 1, \ldots, 10^4$. The piecewise continuous curve is the limit distribution $P(s)$ in (56).

§9. Theorem 1 is now a direct consequence of Theorem 2. As to the explicit formula for the gap distribution,

$$\begin{aligned} P(s) &= \int_0^1 P(x,s)\,dx \\ &= \int_0^1 (b^{x-1}\log b)^2 R(sb^{x-1}\log b, b)\,dx. \end{aligned} \tag{51}$$

With the variable substitution $a = sb^{x-1}\log b$, $da = sb^{x-1}(\log b)^2\,dx = (\log b)\,a\,dx$ we have

$$P(s) = \frac{1}{s^2\log b}\int_{sb^{-1}\log b}^{s\log b} R(a,b)\,a\,da. \tag{52}$$

Recall the definition of $R(a,b)$ in (12). The integral over the first term in (12) yields $\frac{(b^{-1};b^{-1})}{s^2\log b}$ if $sb^{-1}\log b < 1 < s\log b$ and 0 otherwise. For the second term, we use integration by parts,

$$\begin{aligned} \int_{a_0}^{a_1} a\frac{\partial^2}{\partial a^2}(a;b^{-1})\,da &= \left[a\frac{\partial}{\partial a}(a;b^{-1})\right]_{a_0}^{a_1} - \int_{a_0}^{a_1}\frac{\partial}{\partial a}(a;b^{-1})\,da \\ &= \left[a\frac{\partial}{\partial a}(a;b^{-1}) - (a;b^{-1})\right]_{a_0}^{a_1}, \end{aligned} \tag{53}$$

where $a_0 = \min\{1, sb^{-1}\log b\}$ and $a_1 = \min\{1, s\log b\}$. This yields (4).

§10. As we have noted, the technique presented here works also for algebraic b. Let us consider briefly the simplest case of $b > 1$ being an integer. In this case, the multiset sum $\mathcal{S}_N = \uplus_{j=1}^{J}(\beta^{(N)} + \omega_j^{-1}\mathbb{Z})$ considered in the proof of Theorem 3 has all its points lying in $\beta^{(N)} + \omega_1^{-1}\mathbb{Z}$, i.e. the only gap lengths appearing are 0 and $\omega_1^{-1} = (1-b^{-1})^{-1}$; furthermore $\mathcal{S}_N$ is periodic with period $\omega_J^{-1} = b^{J-1}\omega_1^{-1}$. One finds that the average multiplicity of $\beta^{(N)} + \omega_1^{-1}k$ in $\mathcal{S}_N$ for varying $k \in \mathbb{Z}$ equals $\frac{1-b^{-J}}{1-b^{-1}}$. Hence following the proof of Theorem 3 and taking the limit $J \to \infty$, we conclude that for $b > 1$ an integer,

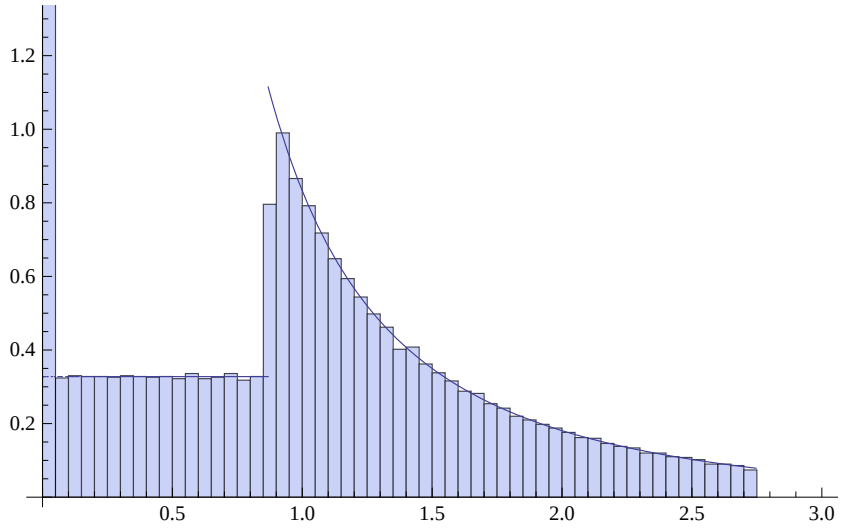

FIGURE 4. The distribution of gaps between the fractional parts of $\log_b n$ with base $b = \sqrt{10}$, where $n = 1, \dots, 10^4$. The piecewise continuous curve is the limit distribution $P(s)$ in (62) with $r = 2$.

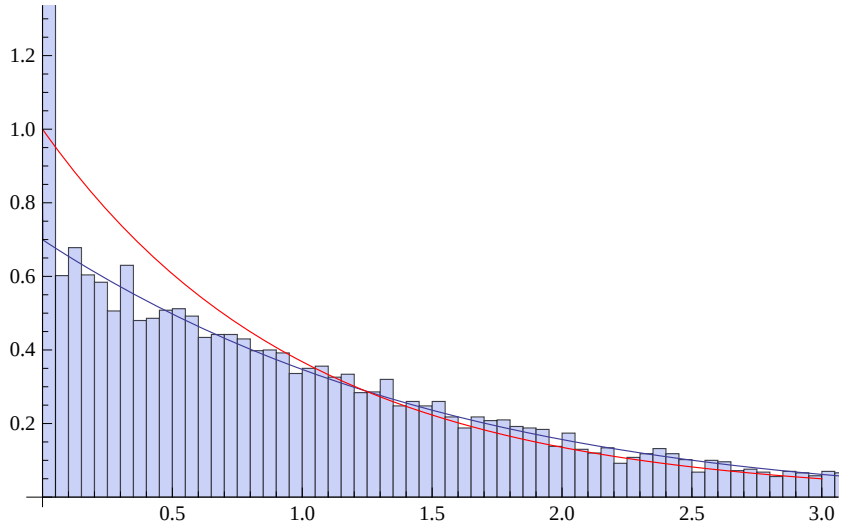

FIGURE 5. The distribution of gaps between the fractional parts of $10 \log_{10} n$ ($= \log_b n$ with base $b = 10^{1/10} = 1.258925411\ldots$), where $n = 1, \dots, 10^4$. The blue curve is the limit distribution $P(s)$ in (62) with $r = 10$. The red curve is the exponential distribution e^{-s}.

we have a limit result as in (20) in Theorem 3, but with the limit density being given simply by

$$\tilde{P}(s) = b^{-1}\delta(s) + (1 - b^{-1})\delta\big(s - (1 - b^{-1})^{-1}\big). \tag{54}$$

As to the limit of the unscaled sequence η_n, the first equality in (22) yields

$$P(x, s) = \frac{\log b}{b - 1}\, b^{x-1}\delta(s) + (\log b) b^{x-1}\delta\bigg(s - \frac{b^{1-x}}{\log b}\bigg), \tag{55}$$

and therefore

$$P(s) = b^{-1}\delta(s) + \begin{cases} \frac{1}{\log b}\, s^{-2} & \text{if } \frac{1}{\log b} < s < \frac{b}{\log b} \\ 0 & \text{otherwise.} \end{cases} \tag{56}$$

See Figure 3 for an illustration in the case $b = 10$.

§11. The findings of the previous paragraph can be generalised to bases of the form $b = m^{1/r}$, with m and r positive integers such that $m^{1/p}$ is not an integer for any prime divisor p of r. In this case, the multiset sum $\mathcal{S}_N = \uplus_{j=1}^{J}(\beta^{(N)} + \omega_j^{-1}\mathbb{Z})$ considered in the proof of Theorem 3 has all points in the set $\cup_{j=1}^{r}(\beta^{(N)} + \omega_j^{-1}\mathbb{Z})$. Our assumption implies that the polynomial $X^r - m$ is irreducible over $\mathbb{Q}$ (since any monic polynomial of degree $d < r$ dividing $X^r - m$ must have zeroth coefficient of absolute value $m^{d/r}$, which is irrational). Hence $\omega_1, \ldots, \omega_r$ are linearly independent over $\mathbb{Q}$, and we now see by the same argument as in the proof of Theorem 4 that the limit (29) exists in the case $k = 0$ and is given by

$$E(0, L) = \begin{cases} \prod_{j=1}^{r} \left(1 - b^{-j}(b-1)L\right) & \text{if } L < (1 - b^{-1})^{-1} \\ 0 & \text{otherwise.} \end{cases} \tag{57}$$

Using the finite q-Pochhammer symbol

$$(a; q)_r = \prod_{n=0}^{r-1}(1 - aq^n), \tag{58}$$

we can write this as

$$E(0, L) = \begin{cases} ((1 - b^{-1})L; b^{-1})_r & \text{if } L < (1 - b^{-1})^{-1} \\ 0 & \text{otherwise.} \end{cases} \tag{59}$$

We note that for $L < (1 - b^{-1})^{-1}$

$$-\frac{d}{dL}E(0, L) = (1 - b^{-1})E(0, L)\sum_{j=0}^{r-1}\frac{1}{b^j - (1 - b^{-1})L} \tag{60}$$

and thus

$$-\frac{d}{dL}E(0, L)\Big|_{L=0} = (1 - b^{-1})\sum_{j=0}^{r-1}\frac{1}{b^j} = 1 - b^{-r}. \tag{61}$$

The fact that this value is less than 1 is due to the non-trivial multiplicity of values in our sequence, which results in a positive density of zero gaps. The statements of Theorems 1–3 therefore hold with the following limit distributions. The limiting gap distribution of Theorem 1 is

$$P(s) = b^{-r}\,\delta(s) + \begin{cases} \frac{1}{\log b}\, s^{-2}(F_r(s\log b) - F_r(sb^{-1}\log b)) & \text{if } 0 < s < \frac{1}{\log b} \\ -\frac{1}{\log b}\, s^{-2}F_r(sb^{-1}\log b) & \text{if } \frac{1}{\log b} < s < \frac{b}{\log b} \\ 0 & \text{if } s > \frac{b}{\log b}, \end{cases} \tag{62}$$

with

$$F_r(a) = a\frac{\partial}{\partial a}(a; b^{-1})_r - (a; b^{-1})_r. \tag{63}$$

The joint limiting gap distribution of Theorem 2 is

$$P(x, s) = \frac{\log b}{b-1}\, b^{x-r}\,\delta(s) + (b^{x-1}\log b)^2 R_r(sb^{x-1}\log b, b) \tag{64}$$

and the rescaled gap distribution of Theorem 3 is

$$\tilde{P}(s) = b^{-r}\,\delta(s) + (1-b^{-1})^2 R_r\big((1-b^{-1})s, b\big), \tag{65}$$

with

$$R_r(a,b) = (b^{-1};b^{-1})_{r-1}\,\delta(a-1) + \begin{cases} \dfrac{\partial^2}{\partial a^2}(a;b^{-1})_r & \text{if } 0<a<1 \\ 0 & \text{if } a>1. \end{cases} \tag{66}$$

See Figures 4 and 5 for illustrations in the cases $b = 10^{1/2}$ and $b = 10^{1/10}$.

§12. We conclude this investigation with a few remarks on the family of limit distributions which we have obtained for different values of $b \in \mathcal{T}$. It is helpful to introduce the random point process in $\mathbb{R}$ given by the sequence

$$\varphi_N^{(b)} = \{N(\tilde{\eta}_n + t + m) : 1 \le n \le N,\ m \in \mathbb{Z}\} \tag{67}$$

where t is uniformly distributed in $[0,1]$. This process is clearly stationary and has intensity one. By using the same strategy as in the proof of Theorem 3, one can deduce from Theorem 4 a limit law for the one-dimensional distribution of $\varphi_N^{(b)}$: For every closed interval $A \subset \mathbb{R}$, $k \in \mathbb{Z}_{\ge 0}$,

$$\lim_{N\to\infty} \mathbb{P}(|\varphi_N^{(b)} \cap A| = k) = E^{(b)}(k, \mathrm{meas}(A)), \tag{68}$$

where

$$E^{(b)}(k,L) = \lim_{J\to\infty} \sum_{k_1+\ldots+k_J=k} \prod_{j=1}^{J} E_1\big(k_j, (b-1)b^{-j}L\big), \tag{69}$$

with $E_1(k,L)$ as in (28). In the case $k=0$, relation (68) in fact follows directly from Theorem 3 (exploiting again Theorem 2.2 in [7]), where

$$\tilde{P}^{(b)}(s) = \frac{d^2 E^{(b)}(0,s)}{ds^2}. \tag{70}$$

Although (68) requires $b \in \mathcal{T}$, the family of distributions (69) is well defined for any real $b \in (1,\infty)$. Let us briefly analyse the limiting cases $b \to \infty$ and $b \to 1$.

For $b \to \infty$, $E^{(b)}(k,L)$ converges to the statistics of the point process given by the randomly shifted integer lattice $\mathbb{Z} + t$, with t uniformly distributed in $[0,1]$. That is, $E^{(b)}(k,L) \to E_1(k,L)$ as in (28). The rescaled gap distribution satisfies

$$\tilde{P}^{(b)}(s) \to R(s,\infty) = \delta(s-1). \tag{71}$$

The limit of the raw $P^{(b)}(s)$ distribution is slightly more involved since the asymptotic density $\rho^{(b)}(x)$ in (10) converges to $\delta(x-1)$. We therefore cannot expect $P^{(b)}(s)$ to converge to a non-trivial limit without further rescaling. Let us extend $\rho^{(b)}(x)$ to a probability density on $\mathbb{R}_{\ge 0}$ by setting $\rho^{(b)}(x) = 0$ for $x > 1$. We have the scaling limit

$$\frac{b-1}{\log b}\rho^{(b)}\left(\frac{x}{\log b}\right) \to \mathrm{e}^x, \tag{72}$$

which, however, is not a probability density on $\mathbb{R}_{\ge 0}$. Working directly from (4), we see that for s fixed,

$$\frac{1}{\log b}\, P^{(b)}\left(\frac{s}{\log b}\right) \to \begin{cases} 0 & \text{if } 0<s<1 \\ s^{-2} & \text{if } s>1. \end{cases} \tag{73}$$

In the limit $b \to 1$, $E^{(b)}(k, L)$ converges to the Poisson distribution. That is,

$$E^{(b)}(k, L) \to \frac{L^k}{k!}\mathrm{e}^{-L}, \qquad \tilde{P}^{(b)}(s) \to \mathrm{e}^{-s}. \tag{74}$$

To see this, note that for L fixed and b sufficiently close to 1,

$$\log E^{(b)}(0, L) = \sum_{j=1}^{\infty} \log(1 - (b-1)b^{-j}L) = -L + O(b-1), \tag{75}$$

which proves the claim for $k = 0$ and the gap distribution $\tilde{P}(s)$ (via Theorem 2.2 in [7]). Again, for L fixed and b sufficiently close to 1, we have for the remaining cases $k \geq 1$,

$$\begin{aligned} E^{(b)}(k, L) &= E^{(b)}(0, L) \sum_{j_1<\ldots<j_k} \frac{E_1(1, (b-1)b^{-j_1}L)\cdots E_1(1, (b-1)b^{-j_k}L)}{E_1(0, (b-1)b^{-j_1}L)\cdots E_1(0, (b-1)b^{-j_k}L)} \\ &= (b-1)^k L^k E^{(b)}(0, L) \sum_{j_1<\ldots<j_k} \frac{b^{-(j_1+\ldots+j_k)}}{(1-(b-1)b^{-j_1}L)\cdots(1-(b-1)b^{-j_k}L)} \\ &= L^k E^{(b)}(0, L)\left(\frac{1}{k!} + O(b-1)\right). \end{aligned} \tag{76}$$

As to the limit of the unscaled sequence η_n, we have for the density

$$\rho^{(b)}(x) \to 1, \qquad x \in [0, 1], \tag{77}$$

i.e., we have uniform distribution mod one. In view of (22), $P^{(b)}(x, s) \to \mathrm{e}^{-s}$ and thus, for the raw gap distribution, $P^{(b)}(s) \to \mathrm{e}^{-s}$.

The above discussion can be readily adapted to bases of the form $b = m^{1/r}$ considered in §11. In particular, we again observe that both the raw and rescaled limiting gap distributions (62) and (65) converge to the exponential distribution e^{-s} when $r \to \infty$ for fixed m, cf. Fig. 5.

Acknowledgments. We thank Jon Keating and Zeév Rudnick for their comments on the first draft of this paper.

References

[1] F. Barra and P. Gaspard, On the level spacing distribution in quantum graphs. J. Statist. Phys. 101 (2000) 283-319.
[2] P. Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1999.
[3] N.D. Elkies and C.T. McMullen, Gaps in $\sqrt{n}$ mod 1 and ergodic theory. Duke Math. J. 123 (2004) 95–139.
[4] D.R. Heath-Brown, Pair correlation for fractional parts of αn^2. Math. Proc. Cambridge Philos. Soc. 148 (2010) 385-407.
[5] M. Kac, Statistical independence in probability, analysis and number theory. The Carus Mathematical Monographs, No. 12. Published by the Mathematical Association of America. Distributed by John Wiley and Sons Inc., New York 1959.
[6] L. Kuipers and H. Niederreiter, Uniform distribution of sequences. Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974.
[7] J. Marklof, Distribution modulo one and Ratner's theorem. Equidistribution in number theory, an introduction, 217-244, NATO Sci. Ser. II Math. Phys. Chem., 237, Springer, Dordrecht, 2007.
[8] J. Marklof and A. Strömbergsson, Equidistribution of Kronecker sequences along closed horocycles. Geom. Funct. Anal. 13 (2003) 1239-1280.
[9] J. Marklof and A. Strömbergsson, The distribution of free path lengths in the periodic Lorentz gas and related lattice point problems. Ann. of Math. (2) 172 (2010) 1949-2033.
[10] Z. Rudnick and P. Sarnak, The pair correlation function of fractional parts of polynomials, Comm. Math. Phys. 194 (1998) 61-70.

[11] Z. Rudnick, P. Sarnak and A. Zaharescu, The distribution of spacings between the fractional parts of $n^2\alpha$. Invent. Math. 145 (2001) 37-57.
[12] Z. Rudnick and A. Zaharescu, The distribution of spacings between fractional parts of lacunary sequences. Forum Math. 14 (2002) 691-712.

School of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.
j.marklof@bristol.ac.uk

Department of Mathematics, Box 480, Uppsala University, SE-75106 Uppsala, Sweden
astrombe@math.uu.se