Annales Univ. Sci. Budapest., Sect. Comp. 21 (2004) 177-186

A PURELY PROBABILISTIC PROOF FOR A THEOREM OF BAKSTYS ON MULTIPLICATIVE ARITHMETICAL FUNCTIONS

J. Galambos (Philadelphia, PA, USA)
I. Simonelli (Commerce, TX, USA)

Abstract. We give a purely probabilistic proof for a theorem of Bakstys on multiplicative functions and, as a consequence, for the Erdős-Wintner theorem on additive functions. This is made possible by a recent result on products of independent random variables by Simonelli. Neither Simonelli nor us use any analytic tools in the proofs and in the reduction method to independent random variables no sieve methods are utilized.

1. Introduction

Let $g(m)$ be a multiplicative arithmetical function with $g(1) = 1$. For primes p we define $s_p = s_p(m)$ to be the exponent of p in the prime factorization of

$$m = \prod_p p^{s_p}. \tag{1}$$

Then

$$g(m) = \prod_p g(p^{s_p}). \tag{2}$$

Consider the probability space $S_N = (\Omega_N, \mathcal{A}_N, P_N)$, where $\Omega_N = \{1, 2, \cdots, N\}$, $\mathcal{A}_N$ is the collection of all subsets of Ω_N and P_N is the probability measure that assigns mass $1/N$ to each element in Ω_N. If A is a set $A = \{a_1, a_2, \cdots\}$ of positive integers a_j, we set $A_N = A \cap \Omega_N$. Then

$$\lim P_N(A_N) = d(A), \quad N \to +\infty,$$

whenever the limit above exists, is called the (natural) density of A. In the sequel we just put $P_N(A)$ for $P_N(A_N)$. We write

$$F_N(x) = P_N\left(\{m: \, m \in \Omega_N, \, g(m) \le x\}\right), \quad x \text{ real.}$$

The density $\lim F_N(x) = F(x)$, $N \to +\infty$, when it exists, is called the limiting distribution function of $g(m)$. The limiting distribution function $F(x)$ is called symmetric if $F(x) = 1 - F(-x)$ for all continuity points of $F(x)$. We deal with nonsymmetric limiting distribution functions in this paper. Note that, for integers $k_j \ge 0$ and for every $t \ge 1$,

$$P_N\left(\{m: \, m \in \Omega_N, \, s_{p_j}(m) \ge k_j, \, 1 \le j \le t\}\right) = \frac{1}{N}\left\lfloor \frac{N}{p_1^{k_1} p_2^{k_2} \cdots p_t^{k_t}} \right\rfloor, \tag{3}$$

where $\lfloor y \rfloor$ signifies the integer part of y. We once again drop m and $m \in \Omega_N$ from the notation and abbreviate the left hand side to $P_N(s_{p_j}(m) \ge k_j, 1 \le$ $\le j \le t)$. The right hand side of (3) expresses an almost stochastic independence of the functions $s_{p_j} = s_{p_j}(m)$. It is indeed our aim of analyzing the existence of $F(x)$ above through a probabilistic model in which the s_{p_j} are completely independent. For this, we introduce an abstract probability space $(\Omega, \mathcal{A}, P)$ on which there are random variables $e_{p_j} = e_{p_j}(\omega)$ which take the nonnegative integers $k_j \ge 0$ with distribution

$$P\left(e_{p_j} \ge k_j, \, 1 \le j \le t\right) = \frac{1}{p_1^{k_1} p_2^{k_2} \cdots p_t^{k_t}}.$$

Hence, the random variables e_{p_j}, $j \ge 1$, are stochastically independent. Simonelli (2001) obtained a complete solution to the existence of limiting distribution of the product

$$G_N(\omega) = \prod g(p^{e_p}). \tag{4}$$

Simonelli's argument is purely probabilistic, free of transforms of distributions. Hence, our proofs on $g(m)$ will be purely probabilistic if we reduce the distributional properties of the multiplicative function $g(m)$ of (2) to that of $G_N(\omega)$ of (4) without any number theoretic tools such as sieve methods.

2. The theorem of Bakstys and the Erdős-Wintner theorem

Our aim is to give a new proof for the following theorem of Bakstys (1968).

Theorem 1. *Let $g(m) \neq 0$ and assume*

$$\sum_{g(p)<0} \frac{1}{p} < +\infty.$$

Then $g(m)$ has a limiting distribution function $F(x)$, continuous at 0, if, and only if, $g^*(m) = |g(m)|$ does. Equivalently, $g(m)$ has a limiting distribution function continuous at 0 if, and only if, for arbitrary $0 < M < +\infty$, the three series

$$(5) \qquad (i) \sum{}^{*} \frac{1}{p}, \qquad (ii) \sum{}^{**} \frac{\log|g(p)|}{p}, \qquad (iii) \sum{}^{**} \frac{\log^2|g(p)|}{p}$$

converge, where summation in $\sum^*$ is over primes p such that $|\log|g(p)|| >$ $> M$, and in $\sum^{**}$, $|\log|g(p)|| < M$. The limiting distribution function $F(x)$ is symmetric if, and only if, $g(2^k) = -1$ for all k.

From (5) it is clear that $g(m)$ and $g^*(m)$ have limiting distribution functions at the same time.

The special case $g(m) > 0$, for all $m \geq 1$, allows one to take logarithm of $g(m)$. We have from (2)

$$(6) \qquad f(m) = \log g(m) = \sum_p \log g(p^{s_p}) = \sum_p f(p^{s_p}).$$

The property at (6) is referred to as $f(m)$'s being additive. Hence, Theorem 1 implies the celebrated Erdős-Wintner theorem (Erdős and Wintner (1939), or Elliott (1979), p. 187).

Theorem 2. *An additive arithmetical function $f(m)$ has a limiting distribution if, and only if, the three series at (5) converge (upon replacing $\log|g(p)|$ by $f(p)$).*

Theorem 2 resembles the three-series theorem of Kolmogorov from probability theory. If one rewrites (6) as

$$f(m) = \sum_p \epsilon_{p,s_p}(m) f(p^{s_p}),$$

where, with the notation at (1),

$$\epsilon_{p,s_p}(m) = \begin{cases} 1 & \text{if } p^{s_p} \| m, \\ 0 & \text{otherwise,} \end{cases}$$

an additive function $f(m)$ becomes an "infinite sum of random variables". It came up in the literature several times, and both P. Erdős and A. Rényi suggested to one of us (Galambos) in personal discussions to reduce Theorem 2 from the Kolmogorov three-series theorem rather than proceeding with a dependent model and extending the three series theorem to such model in order to prove Theorem 2 on probabilistic lines (see Galambos [4]). Such a solution is now given in the present paper via proving the more general Theorem 1.

When we use Chebyshev's inequality in one argument in the next section the variance of the logarithm of $|g(m)|$ enters, that can be estimated by (iii) of (5). In order to avoid new computations we refer there to the Kubilius-Turán inequality (see Kubilius [6] or Elliott [2], p.147).

3. Proof of Theorem 1

For the random variable $G_N(\omega)$ of (4), Simonelli (2001) proved the following theorem.

Theorem 3. *Let* $g(m) \neq 0$ *and assume*

$$\sum_{g(p)<0} \frac{1}{p} < +\infty.$$

(i) $G_N(\omega)$ *converges weakly to a random variable continuous at zero if, and only if,* $|G_N(\omega)|$ *does. The limiting distribution of* $G_N(\omega)$ *is symmetric if, and only if,* $g(2^k) = -1$ *for all* $k \geq 1$.

(ii) $G_N(\omega)$ *converges weakly to a random variable discontinuos at zero if, and only if,* $G_N(\omega)$ *converges to zero almost surely.*

We wish to stress that no analytic tools (transforms such as characteristic functions) are employed by Simonelli.

Proof of Theorem 1. First let us assume the validity of the three series in (5). This, Kolmogorov three series theorem, and Theorem 3 immediately imply that $G_N(\omega)$ converges weakly to a random variable continuous at 0. We claim that $G_N(\omega)$ and $g(m)$ have the same limiting distribution. Since for each T, $\prod_{p \leq T} g(p^{s_p(m)})$ converges weakly to $G_T(\omega)$, as $N \to +\infty$, to prove our claim it suffices to show that for any $\delta > 0$,

$$R = \lim_{T\to+\infty} \limsup_{N\to+\infty} P_N\Big(\Big|\prod_{p\leq N} g(p^{s_p(m)}) - \prod_{p\leq T} g(p^{s_p(m)})\Big| > \delta\Big) = 0.$$

Now, the following estimates

$$P_N\Big(\prod_{T<p\le N} g(p^{s_p(m)})<0\Big)\le\left(\sum_{p>T:g(p)<0}\frac{1}{p}+\sum_{p>T}\frac{1}{p^2}\right)\to 0\ \text{ as}\,T\to+\infty$$

and

$$P_N\Big(s_p(m)\ne\beta_p(m),\,p>T\Big)\le\sum_{p>T}\frac{1}{p^2}\to 0\ \text{ as }\ T\to+\infty,$$

where $\beta_p(m)=1$ if $p|m$ and 0 otherwise, give

$$R=\lim_{T\to+\infty}\limsup_{N\to+\infty}P_N\Big(\Big|\prod_{p\le T}g(p^{s_p(m)})\Big|\Big|\prod_{T<p\le N}g(p^{s_p(m)})-1\Big|>\delta\Big)\le$$
$$\le\lim_{T\to+\infty}\limsup_{N\to+\infty}P_N\Big(\Big|\prod_{p\le T}g(p^{s_p(m)})\Big|\Big|\prod_{T<p\le N}\Big|g(p^{\beta_p(m)})\Big|-1\Big|>\delta\Big).$$

This last estimate and the weak convergence of $G_T(\omega)$ to a proper distribution imply that $R=0$ holds whenever for arbitrary $\epsilon>0$

$$\text{(7)}\qquad\lim_{T\to+\infty}\limsup_{N\to+\infty}P_N\Big(\Big|\prod_{T<p\le N}\Big|g(p^{\beta_p(m)})\Big|-1\Big|>\epsilon\Big)=0.$$

The convergence of the first series in (5) and the inequality

$$P_N\Big(\Big|\prod_{T<p\le N}\Big|g(p^{\beta_p(m)})\Big|-1\Big|>\epsilon\Big)\le$$
$$\le\sum_{T<p\le N}^{*}\frac{1}{p}+P_N\Big(\Big|\prod_{T<p\le N}^{**}|g(p^{\beta_p(m)})|-1\Big|>\epsilon\Big),$$

where $\prod^{**}$ is over all primes p such that $|\log|g(p)||<M$, give that (7) is equivalent to

$$\lim_{T\to+\infty}\limsup_{N\to+\infty}P_N\Big(\Big|\prod_{T<p\le N}^{**}|g(p^{\beta_p(m)})|-1\Big|>\epsilon\Big)=\lim_{T\to+\infty}\limsup_{N\to+\infty}R_{N,T}=0.$$

Let $0<\epsilon_1<\epsilon$. The convergence of the second series in (5) implies we can find a T_o such that whenever $N>T>T_o$,

$$\Big|\prod_{T<p\le N}^{**}|g(p)|^{\frac{1}{p}}-1\Big|<\epsilon_1.$$

Hence,

$$\left\{ \Big| \prod_{T<p\le N}^{**} |g(p^{\beta_p(m)})| - 1 \Big| > \epsilon \right\} =$$

$$= \left\{ \prod_{T<p\le N}^{**} |g(p^{\beta_p(m)})| > 1+\epsilon \quad \text{or} \quad \prod_{T<p\le N}^{**} |g(p^{\beta_p(m)})| < 1-\epsilon \right\} \subseteq$$

$$\subseteq \left\{ \prod_{T<p\le N}^{**} \frac{|g(p^{\beta_p(m)})|}{|g(p)|^{\frac{1}{p}}} (1+\epsilon_1) > 1+\epsilon \quad \text{or} \right.$$

$$\left. \prod_{T<p\le N}^{**} \frac{|g(p^{\beta_p(m)})|}{|g(p)|^{\frac{1}{p}}} (1-\epsilon_1) < 1-\epsilon \right\} =$$

$$= \left\{ \ln \left(\prod_{T<p\le N}^{**} \frac{|g(p^{\beta_p(m)})|}{|g(p)|^{\frac{1}{p}}} \right) > \ln \left(\frac{1+\epsilon}{1+\epsilon_1} \right) \quad \text{or} \right.$$

$$\left. \ln \left(\prod_{T<p\le N}^{**} \frac{|g(p^{\beta_p(m)})|}{|g(p)|^{\frac{1}{p}}} \right) < \ln \left(\frac{1-\epsilon}{1-\epsilon_1} \right) \right\} \subseteq$$

$$\subseteq \left\{ \Big| \sum_{T<p\le N}^{**} \ln |g(p^{\beta_p(m)})| - \sum_{T<p\le N}^{**} \frac{\ln |g(p)|}{p} \Big| > \epsilon' \right\},$$

where $\epsilon' = \min \left\{ \ln \left(\frac{1+\epsilon}{1+\epsilon_1} \right), -\ln \left(\frac{1-\epsilon}{1-\epsilon_1} \right) \right\}$. Hence

$$R_{N,T} \le P_N \Big(\Big| \sum_{T<p\le N}^{**} \ln |g(p^{\beta_p(m)})| - \sum_{T<p\le N}^{**} \frac{\ln |g(p)|}{p} \Big| > \epsilon' \Big| \Big).$$

Now, by Chebyshev's inequality, after estimating the variance by Kubilius-Turán inequality, $P_N()$ above is at most

$$\frac{C}{\epsilon'^2} \sum_{p>T}^{**} \frac{\ln^2 |g(p)|}{p},$$

and by assumption this bound goes to 0 as $T \to +\infty$. Hence $R = 0$ and the weak convergence of $g(m)$ and $|g(m)|$ is established.

Next we assume $|g(m)|$ has a limiting distribution continuous at zero. Let $|G_{T_k}(\omega)|$ be an arbitrary subsequence of $|G_N(\omega)|$ that converges weakly, and denote by $H(x)$ its limiting distribution. We claim that $H(x)$ is a proper distribution continuous at zero. If this is indeed the case, the convergence of $|g(m)|$ and the first part of the proof would imply that $H(x)$ must coincide with the limiting distribution of $|g(m)|$, and since $|G_{T_k}(\omega)|$ was arbitrary this would give the convergence of $|G_N(\omega)|$. Kolmogorov three series theorem and the first part of our proof would further imply the convergence of the three series in (5) and of $g(m)$. So let us first assume that $H(x)$ is not continuous at zero. Then from Theorem 3 $|G_{T_k}(\omega)|$ converges to zero almost surely. We show that this is not possible since it would imply that the limiting distribution of $|g(m)|$ is also degenerate at 0, contradicting our initial assumption. Let $\beta_p(m)$ be as in the first part of the proof. To prove our claim it suffices to show that for arbitrary $\epsilon > 0$,

$$\lim_{k\to+\infty} \limsup_{n\to+\infty} P_{T_n}\left[\prod_{T_k<p\le T_n} |g(p^{\beta_p(m)})| > \epsilon\right] = 0. \tag{8}$$

In fact the convergence of $|g(m)|$ and (8) immediately give that for arbitrary k

$$\lim_{N\to+\infty} P_N\left[\prod_{p\le N} |g(p^{s_p})| > \epsilon\right] = \lim_{n\to+\infty} P_{T_n}\left[\prod_{p\le T_n} |g(p^{s_p})| > \epsilon\right] \le$$

$$\le \limsup_{n\to+\infty}\left(P_{T_n}\left[\prod_{T_k<p\le T_n} |g(p^{\beta_p(m)})| > \epsilon\right] + \right.$$

$$\left. + P_{T_n}\left[\prod_{p\le T_k} |g(p^{s_p(m)})| \ge 1\right] + \sum_{p>T_k}\frac{1}{p^2}\right),$$

and the right hand side of the above inequality goes to zero as $k \to +\infty$. By Markov inequality ([5], p.50)

$$\limsup_{n\to+\infty} P_{T_n}\left[\prod_{T_k<p\le T_n} |g(p^{\beta_p(m)})| > \epsilon\right] \le \frac{1}{\epsilon}\limsup_{n\to+\infty} E_{T_n}\left[\prod_{T_k<p\le T_n} |g(p^{\beta_p(m)})|\right],$$

and by Schwarz's inequality ([5], p.50 and 419) the above expectation is bounded by

$$\left(E_{T_n}\left[\prod_{T_k<p\le T_{k+t}} |g(p^{\beta_p(m)})|^2\right] E_{T_n}\left[\prod_{T_{k+t}<p\le T_n} |g(p^{\beta_p(m)})|^2\right]\right)^{\frac{1}{2}}. \tag{9}$$

Since the first expectation in (9) goes to zero as n and then t go to infinity, the validity of (8) will follow if we can show that

$$\alpha = \lim_{t\to+\infty} \limsup_{n\to+\infty} E_{T_n}\left[\prod_{T_{k+t}<p\leq T_n} |g(p^{\beta_p(m)})|^2\right]$$

is bounded. Let t_n be the number of primes p, $T_{k+t} < p \leq T_n$. By applying Hölder inequality $t_n - 1$ times one obtains that α is at most

$$\lim_{t\to+\infty} \limsup_{n\to+\infty} \left(\prod_{T_{k+t}<p\leq T_n} E_{T_n}\left[|g(p^{\beta_p(m)})|^{2t_n}\right]\right)^{\frac{1}{t_n}},$$

and hence to show that α is bounded it suffices to show that

$$\lim_{t\to+\infty} \limsup_{n\to+\infty} \prod_{T_{k+t}<p\leq T_n} E_{T_n}\left[|g(p^{\beta_p(m)})|^{2t_n}\right]$$

is bounded.

$$E_{T_n}\left[|g(p^{\beta_p(m)})|^{2t_n}\right] = 1 - (1 - |g(p)|^{2t_n})\frac{\lfloor T_n/p\rfloor}{T_n} \leq$$

$$\leq \begin{cases} 1 - \frac{1-|g(p)|^{2t_n}}{p} & \text{if} \quad |g(p)| \geq 1, \\ 1 - \frac{1-|g(p)|^{2t_n}}{p} + \frac{1-|g(p)|^{2t_n}}{T_n} & \text{if} \quad |g(p)| < 1. \end{cases}$$

Let $\prod^{**}$ denote product over all p's with $|g(p)| < 1$, and $b_p(\omega) = 1$ if $e_p(\omega) \geq 1$, 0 otherwise. Then

(10)

$$\prod_{T_{k+t}<p\leq T_n} E_{T_n}\left[|g(p^{\beta_p(m)})|^{2t_n}\right] \leq$$

$$\leq \prod_{T_{k+t}<p\leq T_n} E\left[|g(p^{b_p(\omega)})|^{2t_n}\right] \prod_{T_{k+t}<p\leq T_n}^{**} \left[\frac{1 - \frac{1-|g(p)|^{2t_n}}{p} + \frac{1-|g(p)|^{2t_n}}{T_n}}{1 - \frac{1-|g(p)|^{2t_n}}{p}}\right] \leq$$

$$\leq \prod_{T_{k+t}<p\leq T_n} E\left[|g(p^{b_p(\omega)})|^{2t_n}\right] \prod_{T_{k+t}<p\leq T_n}^{**} \left[1 + \frac{p}{p-1}\frac{1}{T_n}\right] \leq$$

$$\leq \prod_{T_{k+t}<p\leq T_n} E\left[|g(p^{b_p(\omega)})|^{2t_n}\right] e^{\sum^{**}_{T_{k+t}<p\leq T_n} \frac{2}{T_n}} =$$

$$= E\left[\prod_{T_{k+t}<p\leq T_n} |g(p^{b_p(\omega)})|^{2t_n}\right] e^{\sum^{**}_{T_{k+t}<p\leq T_n} \frac{2}{T_n}}.$$

Since $|G_{T_k}(\omega)|$ converges to zero almost surely, the first term in (10) goes to zero as $n \to +\infty$, whereas the second term goes to one since the exponent of e is at most t_n/T_n that also goes to zero as $n \to +\infty$. Hence

$$\lim_{t \to +\infty} \limsup_{n \to +\infty} \prod_{T_{k+t} < p \leq T_n} E_{T_n} \left[|g(p^{\beta_p(m)})|^{2t_n} \right] = 0,$$

that implies α is bounded and consequently (8) holds, completing the proof of our claim.

Next we assume that $H(x)$, the limiting distribution of $|G_{T_k}|$, is continuous at 0 but not proper. In this case we consider $|G_{T_k}|^{-1}$. Since $g(m) \neq 0$, this new random variable is well defined. Since $H(x)$ is not proper, by Theorem 3 $|G_{T_k}|^{-1}$ converges to 0 almost surely, and as in the previous case this implies that the limiting distribution of the multiplicative function $|g(m)|^{-1}$ is also degenerate at 0 which contradicts our initial assumption on $g(m)$. Hence weak convergence of $|g(m)|$ implies weak convergence of $|G_N(\omega)|$, the convergence of the three series in (5), and by the first part of our proof weak convergence of $g(m)$.

Clearly, if $g(m)$ has a limiting distribution so does $|g(m)|$, reducing this case to the one we have just proved. The proof of Theorem 1 is now complete.

References

[1] **Бакштис А.,** О предельных законах распределения мультипликативных арифметических функций, *Литовский мат. сб.,* **8** (1) (1968), 5-20. (Bakstys A., On limiting distribution laws of multiplicative arithmetical functions, *Lietuv. Mat. Rinkinys,* 8 (1) (1968), 5-20.)

[2] **Elliott P.D.T.A.,** *Probabilistic number theory I*, Springer, 1979.

[3] **Erdős P. and Wintner A.,** Additive arithmetical functions and statistical independence, *Amer. J. Math.,* **61** (1939), 713-721.

[4] **Galambos J.,** Limiting distribution of sums of (dependent) random variables with applications to arithmetical functions, *Z. Wahrscheinlichkeits. verw. Gebiete,* **18** (1971), 261-270.

[5] **Galambos J.,** *Advanced probability theory*, Marcel Dekker, New York, 1995.

[6] **Kubilius J.,** *Probabilistic methods in the theory of numbers,* Amer. Math. Soc. Translations of Math. Monographs **11**, Providence, R.I., 1964.

[7] **Simonelli I.,** Convergence and symmetry of infinite products of independent random variables, *Statistics & Probability Letters,* **55** (2001), 45-52.

(Received March 28, 2002)

J. Galambos
Department of Mathematics
Temple University TU 038-16
Philadelphia, PA 19122
USA

I. Simonelli
Department of Mathematics
Texas A&M University-Commerce
Commerce, TX 75429-3011
USA