The Annals of Probability
1978, Vol. 6, No. 5, 870–875

A LOG LOG IMPROVEMENT TO THE RIEMANN HYPOTHESIS FOR THE HAWKINS RANDOM SIEVE

BY C. C. HEYDE

CSIRO Division of Mathematics and Statistics, Canberra

This paper is concerned with the Hawkins random sieve which is a probabilistic analogue of the sieve of Eratosthenes. Analogues of the prime number theorem, Mertens' theorem and the Riemann hypothesis have previously been established for the Hawkins sieve. In the present paper we give a more delicate analysis using iterated logarithm results for both martingales and tail sums of martingale differences to deduce a considerably improved log log replacement for the Riemann hypothesis result.

1. Introduction. The random sieve introduced by Hawkins [2] is analogous to the sieve of Eratosthenes and produces a random sequence with asymptotic properties similar in many ways to the primes. It is constructed as follows: Let $A_1 = \{2, 3, 4, \cdots\}$.

STAGE 1. Put $X_1 = \min A_1$. From the set $A_1 \backslash \{X_1\}$ each number in turn is (independently of the others) deleted with probability X_1^{-1} or not deleted with probability $1 - X_1^{-1}$. The set of elements of $A_1 \backslash \{X_1\}$ which remain is denoted by A_2.

STAGE n. Put $X_n = \min A_n$. From the set $A_n \backslash \{X_n\}$ each number in turn is (independently of the others) deleted with probability X_n^{-1} or not deleted with probability $1 - X_n^{-1}$. The set of elements of $A_n \backslash \{X_n\}$ which remain is denoted by A_{n+1}.

Define

$$Y_n = \prod_{1 \leq k \leq n} (1 - X_k^{-1})^{-1} .$$

Wunderlich [8], [9] has obtained the results

$$\lim_{n\to\infty} (n \log n)^{-1} X_n = 1 \quad \text{a.s.} \qquad \text{and} \qquad \lim_{n\to\infty} (\log n)^{-1} Y_n = 1 \quad \text{a.s.} ,$$

which are analogues of the prime number theorem and Mertens' theorem respectively, while Neudecker and Williams [6] have obtained an analogue of the Riemann hypothesis. Write

$$\operatorname{li}(x) = \lim_{\delta \downarrow 0} \left(\int_0^{1-\delta} + \int_{1+\delta}^{x}\right) \frac{dz}{\log z} \sim \frac{x}{\log x} \qquad \text{as} \quad x \to \infty ,$$

and recall that the "real" Riemann hypothesis about the zeros of the Riemann

Received April 14, 1977.

AMS 1970 *subject classifications*. Primary 60F15, 10H30; Secondary 60G45, 60J05.

Key words and phrases. Random sieve, prime numbers, Riemann hypothesis, martingale iterated logarithm laws.

zeta function is equivalent to

$$\mathrm{li}\,(p_n) = n + O(n^{\frac{1}{2}+\varepsilon})$$

for any $\varepsilon > 0$ where p_n denotes the nth prime. Neudecker and Williams showed that

$$L\,\mathrm{li}\,(X_n L^{-1}) = n + O(n^{\frac{1}{2}+\varepsilon}) \quad \text{a.s.}\,,$$

for any $\varepsilon > 0$ where L denotes the (random) limit

$$L = \lim_{n\to\infty} X_n \exp(-Y_n)\,,$$

which they show to exist and be nonzero a.s.

It is our object in this paper to sharpen this result to a log log form as given in the following theorem.

THEOREM. *For the Hawkins random sieve,*

$$\limsup_{n\to\infty} (2n \log\log n)^{-\frac{1}{2}} |L\,\mathrm{li}\,(X_n L^{-1}) - n| \leqq 3 \quad \text{a.s.}$$

The proof of this theorem rests heavily on the use of some recent results ([4], [5]) on the law of the iterated logarithm for martingales and for tail sums of martingale differences. The result, of course, suggests a candidate for the bound in the "real" Riemann hypothesis. It should be remarked that a similar type of log log error term also emerges for a structurally simpler diffusion analogue of the Hawkins sieve (Williams [7]).

2. Proof of the theorem. The proof requires a sharpening of the analysis of [6] whose notation we retain. The starting point is the fact that the process $\{(X_n, Y_n), n \geqq 1\}$ is Markovian with

$$P(X_{n+1} - X_n = j \mid F_n) = Y_n^{-1}(1 - Y_n^{-1})^{j-1}\,, \qquad j \geqq 1\,,$$

F_n being the σ-field generated by (X_j, Y_j), $j \leqq n$.

Set

$$Z_n = X_n - 1\,, \qquad U_{n+1} = (Z_{n+1} - Z_n)Y_n^{-1}\,, \qquad n \geqq 1\,.$$

In [6], Proposition 2, it is shown that if $H_n = \log Z_n - Y_n$, $n \geqq 1$, then

$$H_n = \log L + O(n^{-\frac{1}{2}+\varepsilon}) \quad \text{a.s.}$$

for any $\varepsilon > 0$, and we shall strengthen this to show that

$$H_n = \log L + \delta(n)(2n^{-1} \log\log n)^{\frac{1}{2}} \quad \text{a.s.}\,, \tag{1}$$

where $\limsup_{n\to\infty} \delta(n) = +1$ a.s., $\liminf_{n\to\infty} \delta(n) = -1$ a.s.

Writing $\alpha_n = Y_n Z_n^{-1}$, $\beta_n = Y_n Z_{n+1}^{-1}$, we have from [6] that $\alpha_n = O(n^{-1})$ a.s., $\beta_n = O(n^{-1})$ a.s. as $n \to \infty$ and then

$$\begin{aligned} H_{n+1} - H_n &= \log(1 + \alpha_n U_{n+1}) - \beta_n \\ &= \alpha_n U_{n+1} - \beta_n + R_n \end{aligned}$$

where

$$|R_n| \leqq \alpha_n^2 U_{n+1}^2 \quad \text{a.s.}$$

Further,

$$|\alpha_n - \beta_n| = |\alpha_n \beta_n U_{n+1}| \leqq \alpha_n^2 |U_{n+1}|$$

so that

$$H_{n+1} - H_n = \alpha_n(U_{n+1} - 1) + S_n ,$$

where

$$|S_n| \leqq \alpha_n^2(|U_{n+1}| + U_{n+1}^2) = O(n^{-2} \log n) \quad \text{a.s.} , \tag{2}$$

since, from [6], $|U_{n+1}| = O(\log n)$ a.s. Then

$$\begin{aligned} H_{n+1} &= \sum_{k=1}^n (H_{k+1} - H_k) - 2 \\ &= \sum_{k=1}^n \alpha_k(U_{k+1} - 1) + \sum_{k=1}^n S_k - 2 \end{aligned}$$

and, since $H_n \to_{\text{a.s.}} \log L$ (finite a.s.) as $n \to \infty$ and $\sum_1^\infty |S_k| < \infty$ a.s. in view of (2),

$$H_n - \log L = \sum_{k=n}^\infty \alpha_k(U_{k+1} - 1) + \sum_{k=n}^\infty S_k . \tag{3}$$

Now, from (2),

$$\begin{aligned} |(2n^{-1} \log\log n)^{-\frac{1}{2}} \textstyle\sum_{k=n}^\infty S_k| &= O(n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}} \textstyle\sum_{k=n}^\infty k^{-2} \log k) \\ &= o(1) \quad \text{a.s.} \end{aligned} \tag{4}$$

Furthermore,

$$\sum_{k=n}^\infty \alpha_k(U_{k+1} - 1) = \sum_{k=n}^\infty k^{-1}(U_{k+1} - 1) + \sum_{k=n}^\infty (\alpha_k - k^{-1})(U_{k+1} - 1) \tag{5}$$

and it is easily seen that

$$E(U_{n+1} - 1 \mid F_n) = 0 \quad \text{a.s.} ,$$

so that the $\{U_{n+1} - 1, n \geqq 1\}$ are martingale differences. To prove that

$$\lim_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}} \sum_{k=n}^\infty (\alpha_k - k^{-1})(U_{k+1} - 1) = 0 \quad \text{a.s.} , \tag{6}$$

it suffices, upon noting that the α_n are F_n-measurable so that the $\{(\alpha_k - k^{-1})(U_{k+1} - 1), k \geqq 1\}$ are still martingale differences, to show that (e.g. Lemma 1 of Heyde [4])

$$\sum_{k=1}^\infty k^{\frac{1}{2}}(\alpha_k - k^{-1})(U_{k+1} - 1) \quad \text{converges a.s.}$$

This holds if (e.g. Doob [1], page 320)

$$\sum_{k=1}^\infty k(\alpha_k - k^{-1})^2 E[(U_{k+1} - 1)^2 \mid F_k] < \infty \quad \text{a.s.} ,$$

and hence if

$$\sum_{k=1}^\infty k(\alpha_k - k^{-1})^2 < \infty \quad \text{a.s.} , \tag{7}$$

since $E[(U_{k+1} - 1)^2 \mid F_k] = 1 - Y_k^{-1}$ a.s. But,

$$\alpha_k - k^{-1} = Z_k^{-1}(Y_k - k^{-1}X_k + k^{-1}) = o((\log\log k)(k \log k)^{-1}) \quad \text{a.s.} ,$$

using the theorem of Heyde [3] and hence (7) and consequently (6) holds.

We now apply the iterated logarithm result of Theorem 1(b) of Heyde [4] to the tail sum $\sum_{k=n}^{\infty} k^{-1}(U_{k+1}-1)$ of martingale differences. We have

$$s_n^2 = \sum_{k=n}^{\infty} k^{-2}E[E\{(U_{k+1}-1)^2 \mid F_k\}] = \sum_{k=n}^{\infty} k^{-2}E(1-Y_k^{-1}) \sim n^{-1} \quad \text{a.s.}$$

as $n \to \infty$ and

$$(8) \qquad \sum_{n=1}^{\infty} n^{-1}[(U_{n+1}-1)^2 - E\{(U_{n+1}-1)^2 \mid F_n\}] \quad \text{converges a.s.},$$

again using [1], page 320, since $E(U_{n+1}-1)^4 \leqq 25$ for all $n \geqq 1$. This last result follows from

$$\begin{aligned} E(U_{n+1}-1)^4 &= E[E\{(U_{n+1}-1)^4 \mid F_n\}] \\ &= E\left[E\left\{\left(\frac{X_{n+1}-X_n}{Y_n}-1\right)^4 \middle| F_n\right\}\right] \\ &= E[\textstyle\sum_{j=1}^{\infty}(jY_n^{-1}-1)^4 Y_n^{-1}(1-Y_n^{-1})^{j-1}] \\ &\leqq E[\textstyle\sum_{j=1}^{Y_n} Y_n^{-1}(1-Y_n^{-1})^{j-1} + \sum_{j=Y_n+1}^{\infty}(jY_n^{-1})^4 Y_n^{-1}(1-Y_n^{-1})^{j-1}] \\ &\leqq 1 + EY_n^{-5}\textstyle\sum_{j=1}^{\infty} j^4(1-Y_n^{-1})^{j-1} \\ &\leqq 1 + 24EY_n^{-5}[1-(1-Y_n^{-1})]^{-5} \leqq 25 . \end{aligned}$$

The conditions required for the use of Theorem 1(b), Corollary 1 and Corollary 2 of [4] are then simply verified. In particular,

$$s_n^{-2}\sum_{k=n}^{\infty} k^{-2}E\{(U_{k+1}-1)^2 \mid F_k\} \rightarrow_{\text{a.s.}} 1 ,$$

$$E[|U_j-1| I(j^{-1}|U_j-1| > \varepsilon s_j)] \leqq 25(\varepsilon j s_j)^{-3}, \text{ and}$$

$$E(U_j-1)^2 I(j^{-1}|U_j-1| > \varepsilon s_n) \leqq 25(\varepsilon j s_n)^{-2}$$

so that

$$\sum_{j=1}^{\infty} s_j^{-1}E[j^{-1}|U_j-1| I(j^{-1}|U_j-1| > \varepsilon s_j)] < \infty , \qquad \text{and}$$

$$s_n^{-2}\sum_{j=n}^{\infty} E[j^{-2}(U_j-1)^2 I(j^{-1}|U_j-1| > \varepsilon s_n)] \to 0$$

as $n \to \infty$. We then obtain that the lim sup as $n \to \infty$ of $(2n^{-1}\log\log n)^{-\frac{1}{2}}\sum_{k=n}^{\infty} k^{-1}(U_{k+1}-1)$ is $+1$ a.s. while the lim inf is -1 a.s. The result (1) then follows via (3), (4), (5) and (6).

We now exponentiate (1) and rewrite it in the form

$$(9) \qquad Z_n = Le^{Y_n}\{1+\eta(n)(2n^{-1}\log\log n)^{\frac{1}{2}}\}$$

where $\limsup_{n\to\infty}\eta(n) = +1$ a.s., $\liminf_{n\to\infty}\eta(n) = -1$ a.s.

Next we have

$$Z_{n+1}Y_{n+1}^{-1} - Z_nY_n^{-1} = U_{n+1} - Y_{n+1}^{-1}, \qquad n \geqq 1 ,$$

so that summation gives

$$(10) \qquad Z_nY_n^{-1} = \sum_{k=2}^{n}(U_k-1) + n - \sum_{k=1}^{n} Y_k^{-1} .$$

The martingale law of the iterated logarithm (e.g. Heyde and Scott [5], Corollary 1) applies to $\sum_{k=2}^{n}(U_k-1)$ to give

$$\limsup_{n\to\infty}(2n\log\log n)^{-\frac{1}{2}}\sum_{k=2}^{n}(U_k-1) = +1 \quad \text{a.s.},$$

$$\liminf_{n\to\infty}(2n\log\log n)^{-\frac{1}{2}}\sum_{k=2}^{n}(U_k-1) = -1 \quad \text{a.s.},$$

since

$$\sum_{k=1}^{n} E(U_k - 1)^2 = \sum_{k=1}^{n} (1 - EY_k^{-1}) \sim n \quad \text{as} \quad n \to \infty$$

and $E(U_k - 1)^4 \leqq 25$ while (8) and $E[(U_{k+1} - 1)^2 \mid F_k] = 1 - Y_k^{-1}$ a.s. ensure that $n^{-1} \sum_{k=1}^{n} (U_k - 1)^2 \to_{\text{a.s.}} 1$ as $n \to \infty$. Then, from (10),

$$\begin{aligned} (11) \qquad & \limsup_{n\to\infty} (2n \log\log n)^{-\frac{1}{2}}(Z_n Y_n^{-1} - n + \textstyle\sum_{k=1}^{n} Y_k^{-1}) = +1 \quad \text{a.s.}, \\ & \liminf_{n\to\infty} (2n \log\log n)^{-\frac{1}{2}}(Z_n Y_n^{-1} - n + \textstyle\sum_{k=1}^{n} Y_k^{-1}) = -1 \quad \text{a.s.}, \end{aligned}$$

and using (9) and (11) we obtain

$$\begin{aligned} & \limsup_{n\to\infty} (2n \log\log n)^{-\frac{1}{2}} |Le^{Y_n} Y_n^{-1} - n + \textstyle\sum_{k=1}^{n} Y_k^{-1}| \\ & \quad \leqq \limsup_{n\to\infty} (2n \log\log n)^{-\frac{1}{2}} |Le^{Y_n} Y_n^{-1} - Z_n Y_n^{-1}| \\ & \qquad + \limsup_{n\to\infty} (2n \log\log n)^{-\frac{1}{2}} |Z_n Y_n^{-1} - n + \textstyle\sum_{k=1}^{n} Y_k^{-1}| \\ & \quad \leqq 2 \quad \text{a.s.} \end{aligned}$$

The remainder of the proof parallels that of [6]. We extend the random function Y from $\{1, 2, 3, \cdots\}$ to $(1, \infty)$ by linear interpolation,

$$Y_t = Y_n + (t - n)(Y_{n+1} - Y_n), \qquad n \leqq t \leqq n + 1 .$$

Then, it is easily checked that

$$LY_t^{-1}e^{Y_t} = \int_1^t (1 - Y_s^{-1})\, ds + f(t)$$

where $f(t) = (2t \log\log (t \vee 3))^{\frac{1}{2}}\alpha(t)$ with $\limsup_{t\to\infty} |\alpha(t)| \leqq 2$ a.s. Furthermore,

$$\begin{aligned} [L \operatorname{li} (\exp Y_s)]_1^t &= \int_1^t (1 - Y_s^{-1})\, d(LY_s^{-1}e^{Y_s}) \\ &= t - 1 + \int_1^t f'(s)(1 - Y_s^{-1})^{-1}\, ds \\ &= t - 1 + f(t)(1 - Y_t^{-1})^{-1} - f(1)(1 - Y_1^{-1})^{-1} \\ &\qquad - \int_1^t f(s) Y_s'(Y_s - 1)^{-2}\, ds , \end{aligned}$$

while

$$|f(s)Y_s'(Y_s - 1)^{-2}| = O((\log\log s)^{\frac{1}{2}} s^{-\frac{1}{2}} (\log s)^{-2}) \quad \text{as} \quad s \to \infty$$

so that

$$(t^{\frac{1}{2}} \log\log t)^{-1} \int_1^t |f(s)Y_s'(Y_s - 1)^{-2}|\, ds = o(1) \quad \text{as} \quad t \to \infty$$

and hence

$$(12) \qquad \limsup_{t\to\infty} (2t \log\log t)^{-\frac{1}{2}} |L \operatorname{li} (e^{Y_t}) - t| \leqq 2 \quad \text{a.s.}$$

The result of the theorem then follows from (9) and (12) since (9) gives

$$\limsup_{n\to\infty} (2n \log\log n)^{-\frac{1}{2}} |L \operatorname{li} (Z_n L^{-1}) - L \operatorname{li} (e^{Y_n})| \leqq 1 \quad \text{a.s.}$$

REFERENCES

[1] Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[2] Hawkins, D. (1958). The random sieve. *Math. Mag.* **31** 1–3.

[3] Heyde, C. C. (1976). On asymptotic behaviour for the Hawkins random sieve. *Proc. Amer. Math. Soc.* **56** 277–280.

[4] HEYDE, C. C. (1977). On central limit and iterated logarithm supplements to the martingale convergence theorem. *J. Appl. Probability* **14** 758–775.

[5] HEYDE, C. C. and SCOTT, D. J. (1973). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Probability* **1** 428–436.

[6] NEUDECKER, W. and WILLIAMS, D. (1974). The "Riemann hypothesis" for the Hawkins random sieve. *Composito Math.* **29** 197–200.

[7] WILLIAMS, D. (1974). A diffusion process motivated by the sieve of Eratosthenes. *Bull. London Math. Soc.* **6** 155–164.

[8] WUNDERLICH, M. C. (1974). A probabilistic setting for prime number theory. *Acta Arith.* **26** 59–81.

[9] WUNDERLICH, M. C. (1976). The prime number theorem for random sequences. *J. Number Theory* **8** 369–371.

DIVISION OF MATHEMATICS AND STATISTICS
CSIRO
P.O. BOX 1965, CANBERRA CITY
A.C.T. 2601
AUSTRALIA