COMPOSITIO MATHEMATICA

WERNER NEUDECKER
DAVID WILLIAMS

The "Riemann hypothesis" for the Hawkins random Sieve

Compositio Mathematica, tome 29, nº 2 (1974), p. 197-200
<http://www.numdam.org/item?id=CM_1974__29_2_197_0>

COMPOSITIO MATHEMATICA, Vol. 29, Fasc. 2, 1974, pag. 197–200
Noordhoff International Publishing
Printed in the Netherlands

THE 'RIEMANN HYPOTHESIS' FOR THE HAWKINS RANDOM SIEVE

Werner Neudecker and David Williams

1

Hawkins ([1, 2]) and Wunderlich ([6, 7]) have studied the sequence $\{X_n : n \in N\}$ of 'random primes' caught in the Hawkins Sieve which operates inductively as follows.

Let

$$A_1 = \{2, 3, 4, 5, 6, \cdots\}.$$

Stage 1. Declare $X_1 = \min A_1$. From the set $A_1 \backslash \{X_1\}$, each number in turn (and each independently of the others) is deleted with probability X_1^{-1}. The set of elements of $A_1 \backslash \{X_1\}$ which remain is denoted by A_2.

Stage n. Declare $X_n = \min A_n$. From the set $A_n \backslash \{X_n\}$, each number in turn (and each independently of the others) is deleted with probability X_n^{-1}. The set of elements of $A_n \backslash \{X_n\}$ which remain is denoted by A_{n+1}.

Define

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

Notation and conventions. For x > 1, we write li (x) for the value of the logarithmic integral at x:

$$\operatorname{li}(x) = \lim_{\delta \downarrow 0} \left(\int_0^{1-\delta} + \int_{1+\delta}^{x} \right) \frac{dz}{\log z} \sim \frac{x}{\log x}.$$

Qualifying 'with probability one' phrases will be suppressed. An equation involving the symbol 'ε' is to be understood as true for every positive ε.

2

Recall that the 'real' Riemann Hypothesis about the zeros of the Riemann zeta-function is equivalent to the statement:

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

(where p_n denotes the nth prime) and that equation (1) is the result which Riemann really wished to prove. See for example Ingham [3].

THEOREM: *The following 'Riemann Hypothesis for the Hawkins Sieve' holds*:

$$L \operatorname{li}(X_n L^{-1}) = n + O(n^{\frac{1}{2}+\varepsilon}) \tag{2}$$

where L denotes the random limit

$$L = \lim_{n\to\infty} X_n \exp(-Y_n). \tag{3}$$

This theorem was motivated by the exactly analogous 'diffusion' result in Williams [5]. The proof now given mirrors that in [5].

3

The process $\{(X_n, Y_n): n \in \boldsymbol{N}\}$ is *Markovian* with

$$P[X_{n+1} - X_n = j | \mathscr{F}_n] = Y_n^{-1}(1-Y_n^{-1})^{j-1} \qquad (j \in \boldsymbol{N}),$$
$$Y_{n+1} = Y_n(1-X_{n+1}^{-1})^{-1} = Y_n(1+Z_{n+1}^{-1}) \tag{4}$$

and $X_1 = Y_1 = 2$. We have written (for $n \in \boldsymbol{N}$):

$$Z_n = X_n - 1, \qquad \mathscr{F}_n = \sigma\{(X_k, Y_k) : k \leqq n\},$$

the latter equation signifying that $\mathscr{F}_n$ is the smallest σ-algebra with respect to which X_k and Y_k are measurable for every $k \leqq n$. Introduce

$$U_{n+1} = (Z_{n+1} - Z_n)Y_n^{-1} \qquad (n \in \boldsymbol{N}). \tag{5}$$

Elementary properties of geometric distributions now make things particularly neat. For $x > 0$ and $n \in \boldsymbol{N}$,

$$P[U_{n+1} > x | \mathscr{F}_n] \leqq (1-Y_n^{-1})^{-1}(1-Y_n^{-1})^{xY_n} \leqq 2e^{-x}.$$

(Recall that $Y_n \geqq 2$ for every n.) By the Borel-Cantelli Lemma,

$$U_{n+1} = O(\log n) = O(n^{\varepsilon}). \tag{6}$$

Because

$$E[(U_{n+1}-1)|\mathscr{F}_n] = 0,$$

$\{(U_{n+1}-1) : n \in \boldsymbol{N}\}$ is a family of *orthogonal* random variables. Since also

$$E[(U_{n+1}-1)^2|\mathscr{F}_n] = 1 - Y_n^{-1} \leqq 1,$$

Theorem 33B(ii) of Loève [4] provides the estimate:

$$\sum_{k \leqq n} (U_{k+1} - 1) = O(n^{\frac{1}{2}+\varepsilon}). \tag{7}$$

In particular,

$$n^{-1}\sum_{k\leqq n} U_{k+1} \to 1. \tag{8}$$

4

The remainder of the proof is divided into three stages.

PROPOSITION 1: *$Y_n \uparrow \infty$ and $Z_n \sim nY_n$.*

PROOF: If $Y_n \uparrow Y < \infty$, then we could conclude from (4) that $\sum Z_n^{-1} < \infty$ and from (5) and (8) that (in contradiction) $n^{-1}Z_n \to Y$.

Thus $Y_n \uparrow \infty$ and

$$Z_{n+1}Y_{n+1}^{-1} - Z_n Y_n^{-1} = U_{n+1} - Y_{n+1}^{-1} = U_{n+1} + o(1) \tag{9}$$

so that (from (8)) $Z_n \sim nY_n$.

PROPOSITION 2: *If $H_n = \log Z_n - Y_n$ ($n \in N$), then*

$$H_n = C + O(n^{-\frac{1}{2}+\varepsilon}) \tag{10}$$

for some (random) C.

PROOF: Since $x(1+x)^{-1} \leqq \log(1+x) \leqq x$ for $x \geqq 0$,

$$\begin{aligned} H_{n+1} - H_n &= \log(1+\alpha_n U_{n+1}) - \alpha_n(1+\alpha_n U_{n+1})^{-1} \\ &= \beta_n(U_{n+1}-1) + O(\alpha_n^2 U_{n+1}^2) \end{aligned}$$

where

$$\alpha_n = Y_n Z_n^{-1} = O(n^{-1}) \quad \text{and} \quad \beta_n = Y_n Z_{n+1}^{-1} = O(n^{-1}).$$

Proposition 2 now follows by partial summation using (6), (7) and the further estimate:

$$\beta_{n-1} - \beta_n = \beta_{n-1}\beta_n U_{n+1} - \beta_{n-1}Z_{n+1}^{-1} = O(n^{-2+\varepsilon}).$$

Note. From equation (3), in which of course $L = \exp(C)$, and Proposition 1, it follows that

$$Y_n \sim \log n, \qquad X_n \sim n\log n. \tag{11}$$

In other words, 'Mertens' Theorem' and the 'Prime Number Theorem' hold. (One could not expect the e^{γ} factor which is the rather tantalising feature of the *real* Mertens Theorem.) Wunderlich obtained both results at (11) by a more complicated method.

5

On summing equation (9) over n and utilising the exponentiated form:

$$Z_n = L \exp(Y_n)(1+O(n^{-\frac{1}{2}+\varepsilon})) \tag{12}$$

of equation (10), we obtain

$$LY_n^{-1}\exp(Y_n) = n - \sum_{k \leq n} Y_k^{-1} + O(n^{\frac{1}{2}+\varepsilon}).$$

Extend the random function Y from $\{1, 2, 3, \cdots\}$ to $(1, \infty)$ by linear interpolation. Then it is easily checked that

$$LY_t^{-1}\exp(Y_t) = \int_1^t (1 - Y_s^{-1})\mathrm{d}s + f(t)$$

where $f(t) = O(t^{\frac{1}{2}+\varepsilon})$. But now we may compute

$$[L\,\mathrm{li}(\exp(Y_s))]_1^t = \int_1^t (1-Y_s^{-1})^{-1}\mathrm{d}[LY_s^{-1}\exp(Y_s)]$$

$$= t - 1 + \int_1^t f'(s)[1-Y_s^{-1}]^{-1}\mathrm{d}s.$$

Integration by parts using $Y_s' = O(s^{-1})$ establishes the following strong form of 'Mertens' Theorem':

$$L\,\mathrm{li}(\exp(Y_t)) = t + O(t^{\frac{1}{2}+\varepsilon}). \tag{13}$$

Equation (2) now follows on combining equations (12) and (13).

REFERENCES

[1] D. Hawkins: The random sieve. *Math. Mag.*, *31* (1958) 1–3.
[2] D. Hawkins: Random Sieves II. *J. Number Theory 6 No. 3* (1974) 192–200.
[3] A. E. Ingham: *The distribution of prime numbers*. Cambridge University Press, 1932. Reprinted by Stechert-Hafner Service Agency, Inc., New York and London, 1964.
[4] M. Loève: *Probability Theory*, van Nostrand, Princeton, N.J., 1963.
[5] D. Williams: A study of a diffusion process motivated by the Sieve of Eratosthenes (to appear in *Bull. London Math. Soc.*).
[6] M. C. Wunderlich: The prime number theorem for random sequences (to appear in *J. Number Theory*).
[7] M. C. Wunderlich: A probabilistic setting for prime number theory (to appear in *Acta Arithmetica 27*, 1).

(Oblatum 30–V–1974)

University College of Swansea
Swansea, Wales, Great Britain