Chebyshev's Bias

Steven Finch

April 26, 2006

How do we quantify irregularities in the distribution of prime numbers? Define

$$\pi_{q,a}(n) = \#\{p \leq n : p \equiv a \bmod q\}$$

where $\gcd(a, q) = 1$. A well-known result:

$$\lim_{n\to\infty} \frac{\ln(n)}{n}\pi_{q,a}(n) = \frac{1}{\varphi(q)}$$

informs us that primes are asymptotically equidistributed modulo q, where $\varphi(q)$ is the Euler totient. There is, however, unrest beneath the surface of such symmetry. For fixed a_1, a_1, …, a_r and q, define

$$S_N = \#\{n \leq N : \pi_{q,a_1}(n) > \pi_{q,a_2}(n) > \ldots > \pi_{q,a_r}(n)\}$$

and

$$P(a_1 > a_2 > \ldots > a_r \bmod q) = \lim_{N\to\infty} \frac{1}{\ln(N)} \sum_{n\in S_N} \frac{1}{n}.$$

As the notation suggests, P is to be interpreted as a probability (via logarithmic measure). Rubinstein & Sarnak [1], assuming both the Generalized Riemann Hypothesis and the Grand Simplicity Hypothesis [2], succeeded in proving that

$$P(3 > 1 \bmod 4) = 0.9959280...,$$

$$P(2 > 1 \bmod 3) = 0.9990633....$$

Feuerverger & Martin [3] further proved that

$$P(3 > 5 > 7 \bmod 8) = P(7 > 5 > 3 \bmod 8) = 0.1928013...,$$

$$P(3 > 7 > 5 \bmod 8) = P(5 > 7 > 3 \bmod 8) = 0.1664263...,$$

$$P(5 > 3 > 7 \bmod 8) = P(7 > 3 > 5 \bmod 8) = 0.1407724...$$

and

$$P(5>7>11 \bmod 12)=P(11>7>5 \bmod 12)=0.1984521...,$$
$$P(7>5>11 \bmod 12)=P(11>5>7 \bmod 12)=0.1799849...,$$
$$P(5>11>7 \bmod 12)=P(7>11>5 \bmod 12)=0.1215630...;$$

thus it is more probable that 5 will occupy the middle position for mod 8, and 7 will occupy the middle position for mod 12!

New constants do not always emerge: we have, for example,

$$P(1>4 \bmod 5)=P(2>3 \bmod 5)=\frac{1}{2}$$

which is due to 1, 4 being squares mod 5 and 2, 3 being nonsquares mod 5. Also

$$P(1>2>4 \bmod 7)=P(3>5>6 \bmod 7)=\frac{1}{6}$$

which is due to 1, 2, 4 being squares mod 7 and 3, 5, 6 being nonsquares mod 7. Examples with exact probabilities $1/r!$, where $r>3$, have not been found.

Define the logarithmic integral

$$\operatorname{li}(x)=\int_2^x \frac{1}{\ln(t)}dt$$

for $x \geq 2$ and

$$T_N=\#\left\{n \leq N : \pi_{1,0}(n)>\operatorname{li}(n)\right\}.$$

In another demonstration of their methods, Rubinstein & Sarnak [1] showed that

$$\lim_{N\to\infty}\frac{1}{\ln(N)}\sum_{n\in T_N}\frac{1}{n}=0.00000026...=1-0.99999973....$$

Further results have been obtained by Ng [4], as reported in [5]; we shall discuss these at a later time.

0.1. Addendum. Let us return to the usual sense of probability (via uniform measure). Brent [6] conjectured that, for random $0<N<n$, we have

$$\lim_{n\to\infty}\mathrm{P}\left(\frac{\operatorname{li}(N)-\pi_{1,0}(N)}{\sqrt{N}/\ln(N)}<x\right)=F(x)$$

where the probability distribution F has mean $\mu = 1$ and variance $\sigma^2 \approx (0.21)^2$. If the Riemann hypothesis is true, then it can be shown that [7]

$$\begin{aligned} \sigma^2 &= 2 - \ln(4\pi) + \gamma = (0.2149218879...)^2 \\ &= 0.0461914179... = 2(0.0230957089...) \end{aligned}$$

which we have seen elsewhere [8, 9]. An open question is whether F is the normal distribution; a density plot [1] and a time series graph [5] suggest that the answer might be yes. We also wonder about extensions of this probabilistic result to $\pi_{q,a}(n)$ for arbitrary a and q.

If, in the definition of $\pi_{q,a}(n)$, the symbol p is understood to encompass *semiprimes* (products of two primes) rather than primes, then with formulas for S_N and P exactly as before [10, 11],

$$P(3 > 1 \bmod 4) = 0.10572....$$

Hence the bias for semiprimes is reversed from that of primes, although it is less pronounced. Information on the asymptotics of $\pi_{q,a}(n)$ here would be gratefully received. The terms *2-almost prime* or *biprime* are often encountered; a less common term *quasi-prime* appears in [10]

References

[1] M. Rubinstein and P. Sarnak, Chebyshev's bias, *Experiment. Math.* 3 (1994) 173–197; MR1329368 (96d:11099).

[2] S. R. Finch, Quadratic Dirichlet L-series, unpublished note (2005).

[3] A. Feuerverger and G. Martin, Biases in the Shanks-Rényi prime number race, *Experiment. Math.* 9 (2000) 535–570; MR1806291 (2002d:11111).

[4] N. C. Ng, Limiting Distributions and Zeros of Artin L-functions, Ph.D. thesis, University of British Columbia, 2000.

[5] A. Granville and G. Martin, Prime number races, *Amer. Math. Monthly* 113 (2006) 1–33; Spanish version in *La Gaceta de la Real Sociedad Matematica Espanola* 8 (2005) 197–240; MR2158415 (2006f:11114).

[6] R. P. Brent, Irregularities in the distribution of primes and twin primes, *Math. Comp.* 29 (1975) 43–56; correction 30 (1976) 198; MR0369287 (51 #5522) and MR0396437 (53 #302).

[7] R. P. Brent, Comments on "Irregularities...", http://maths-people.anu.edu.au/~brent/pub/pub024.html.

[8] S. R. Finch, Apéry's constant, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 40–53.

[9] S. R. Finch, Stieltjes constants, *Mathematical Constants*, Cambridge Univ. Press, 2003, pp. 166–171.

[10] K. Ford and J. Sneed, Chebyshev's bias for products of two primes, *Experim. Math.* 19 (2010) 385–398; arXiv:0908.0093; MR2778652 (2012a:11141).

[11] G. Martin and J. Scarfy, Comparative prime number theory: A survey, arXiv:1202.3408.