ON STOCHASTIC APPROXIMATIONS TO THE DISTRIBUTION OF PRIMES AND PRIME METRICS

Roberto N. Padua and Merliza F. Libao, Philippines Volume 1, Issue No.1 June 2016 ISSN: 2546-0579

Abstract

The paper surveys some stochastic models used to describe the behaviour of primes and other prime metrics. The Prime Number Theorem is proved by modelling the primes as a Poisson process. Likewise, it is demonstrated that the Riemann hypothesis is equivalent to the hypothesis that the distribution of the prime gaps is exponential with rate parameter $\frac{1}{\log(n)}$. *Assuming that the prime gaps are exponentially distributed, the paper also proves the following assertions:*

$$\lim_{n\to\infty} sup \frac{d_n}{log^2(n)} = 1$$

$$\lim_{n\to\infty} inf \frac{d_n}{log(n)} = 0$$

Keywords: *primes, prime gaps, Erlang distribution, Poisson process, stochastic approximations*

1.0 Introduction

The gaps between consecutive primes are observed to obey an exponential distribution (Yamasaki, 1995) and since the primes can be expressed as the sum of the previous prime and these consecutive prime gaps, they too are often thought to behave in a pseudo-random fashion. In stochastic number theory, primes are considered pseudo-random quantities that obey certain probability models. Erdos (1932), Cramer (1919), Selberg (1943) and others have all viewed primes and prime gaps as essentially random and formulated useful results using this assumption. This paper examines some consequences and results that follow logically from stochastic assumptions about prime numbers. Later, the relationship between the Riemann hypothesis (1859) and the assumption of randomness in the distribution of prime gaps is also explored.

2.0 Some probability Models for Primes and Prime Metrics

2.1 The Beta (α,β) Distribution. Let π(x) be the number of primes less or equal to x and let $1<x_n<x$, $n=1, 2, 3,\dots,N$. The quantity $Y_n = \frac{\pi(x_n)}{\pi(x)}$ is observed to lie on the interval [0,1], and, hence, may be modelled by the Beta (α,β) Distribution:

$$f(y_n) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, y_n^{\alpha-1}(1-y_n)^{\beta-1}, \alpha,\beta > 0, 0 < y_n < 1 \quad (1)$$

Various choices of α and β yields different shapes of the beta distribution. If *α<β*, the density is more concentrated on the left (skewed to the right); if *α>β*, the reverse is true. If α=β, we have a symmetric distribution. The uniform distribution U(0,1)is obtained if *α* = *β* = 1.

The ratio of successive primes $z_n = \frac{p_n}{p_{n+1}}$ is also a quantity that lies on the interval [0,1] and may, thus, be modelled using a beta distribution Beta (α,β). The mean and variance of the beta distribution are:

$$\mu = \frac{\alpha}{\alpha+\beta}$$
$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2\,(\alpha+\beta+1)} \quad (2)$$

Moreover, it is known that the mean of the logarithm of a beta random variable is equal to the logarithm of the geometric mean:

$$E(\ln y) = \ln(G_y) = \frac{\partial \ln \Gamma(\alpha)}{\partial \alpha} - \frac{\partial \ln \Gamma(\alpha+\beta)}{\partial \alpha} \quad (3)$$

2.2 The Exponential Distribution and Erlang Distribution

Let $d_n = P_{n+1} - P_n$ be the gaps between successive primes $P_1, P_2, \ldots, P_n, P_{n+1}$. Empirical observations tend to confirm that these obey an exponential distribution. The random variable X has an exponential distribution if:

$$f(x) = \beta e^{-\beta x} \quad , x > 0, \beta > 0 \quad (4)$$

The mean and variance of the exponential distribution are:

$$\mu = \frac{1}{\beta}, \qquad \sigma^2 = \frac{1}{\beta^2} \quad (5)$$

The maximum-likelihood estimates of μ and σ² are:

$$\hat{\mu} = \bar{x}, \qquad \sigma^2 = \bar{x}^2 \quad (6)$$

where $\bar{x} = \sum_{i=1}^{n} x_i$ for a random sample $x_1, x_2, \ldots, x_n$ from f(x).

One can write $P_1 = d_1$, $P_1 = d_1 + P_2, \ldots, P_n = d_1 + d_2 + \ldots + d_n$ so that the nth prime is expressed as the sum of consecutive prime gaps. The probability distribution of P_n is, thus, given by:

$$f_{P_n}(t) = \frac{\beta^n t^{n-1} e^{-\beta t}}{(n-1)!}, \quad n = 1, 2, \ldots$$

which is the Erlang distribution. The mean and variance of the Erlang distribution are:

$$\mu = \alpha/\beta = n/\beta, \qquad \sigma^2 = \alpha/\beta^2 = n/\beta^2 \quad (7)$$

A related distribution that arises out of these considerations is the Poisson distribution. In the language of Stochastic Processes, the primes $\{P_n \mid n > 0\}$ constitute a renewal process, $0 < P_1 < P_2 < \ldots < P_n < \ldots$; the prime gaps $\{d_n \mid n > 0\}$ are the exponential inter-arrival times. The number of prime arrivals up to and including t, denoted by *N(t),* has the Poisson distribution:

$$P(N(t) = k) = \frac{e^{-\beta t}(\beta t)^k}{k!}, \quad k = 0, 1, 2, \ldots \quad (8)$$

The mean and variance of this Poisson distribution are:

$$\mu = \beta t, \qquad \sigma^2 = \beta t \quad (9)$$

2.3 Fractal or Power – Law Distribution. Power – law distributions are distributions of the Mandelbrot (1967) restriction:

$$\lambda = \frac{\log f(x)}{\log\left(\frac{\theta}{x}\right)}, x \geq \theta, \theta > 0 \quad (10)$$

We call the distribution f(x) below a fractal distribution with fractal dimension λ:

$$f(x) = \frac{\lambda - 1}{\theta}\left(\frac{x}{\theta}\right)^{-\lambda} \quad , \lambda > 1, x \geq \theta > 0 \quad (11)$$

An important relation exists between the fractal distribution (11) and the exponential distribution (4).

> **Result.** A random variable x has a fractal distribution if $y = \log\left(\frac{x}{\theta}\right)$ has an exponential distribution with rate parameter $\beta = \lambda - 1$.

Since the prime gaps $\{d_n\}$ has an exponential distribution (β), it follows that $x = \theta \exp(d_n)$ has a fractal distribution with fractal dimension $\lambda = 1 + \beta$.

The maximum likelihood estimator of λ given a random sample from *f(x)* is given by:

$$\hat{\lambda} = 1 + \frac{1}{\sum_{i=1}^{n} log\left(\frac{x_i}{\theta}\right)/n} \qquad (12)$$

It follows from (12) that

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n} log\left(\frac{x_i}{\theta}\right)/n} \qquad (13)$$

is a maximum-likelihood estimator of β based from the derived fractal distribution.

3.0 Distribution of Primes and Other Prime Metrics

For large n, we observe that:

$$\frac{1}{n}\sum_{i=1}^{n} log\left(\frac{x_i}{\theta}\right) = \frac{1}{n}\sum_{k=1}^{n} log\left(\frac{2k}{2}\right)$$

$$= \frac{1}{n}\sum_{k=1}^{n} log(k) < \frac{1}{n}.n\, logn = \log n$$

where x_i is the i^{th} prime gap. Hence,

$$\bar{x}_n \sim \log(n)$$

$$\hat{\beta} \sim \frac{1}{\log(n)} \qquad as\; n \to \infty. \qquad (14)$$

Relation (14) states that the average prime gap or spacing between primes is roughly log(n) for large n. The rate parameter β is asymptotically $\frac{1}{\log(n)}$. We now state a stochastic version of the Prime Number Theorem.

Stochastic Prime Number Theorem. SPNT Let the prime gaps $d_n=P_{n+1}-P_n$ be exponentially distributed with rate parameter β. Then, the number of primes less or equal to n is

$$\pi(n) \sim \frac{n}{\log(n)} \qquad (15)$$

Proof. Since the prime gaps $\{d_n\}$ are exponential, it follows that $\{N(t):t>0\}$ has a Poisson distribution with mean βt.
Hence,

$$\pi(n) = E(N(n)) = \beta n \sim \frac{n}{\log(n)} \qquad (16)$$

from (14).

Note that we have established the Prime Number Theorem (PNT) in (14) without appealing to the zeroes of the Riemann Zeta function. We can also derive a version of the Strong Law of Large Numbers (SLLN) for the primes. First, note that

$$Var(\overline{N(t)}) \sim \frac{1}{n}\, var(N(n))$$

$$\sim \frac{1}{n}\cdot\frac{n}{\log(n)} = \frac{1}{\log(n)} \qquad (17)$$

It follows that:

Strong Law of Large Numbers for Primes. The average number of primes less or equal to n approaches $\frac{n}{\log(n)}$ almost surely:

$$\overline{N(n)} \to \frac{n}{\log(n)} \qquad \text{a.s.} \qquad \text{as } n \to \infty. \quad (18)$$

Proof. From Chebychev's inequality:

$$P\left(\left|\overline{N(n)} - \frac{n}{\log(n)}\right| < \varepsilon\right) \geq 1 - \frac{var(\overline{N(n)})}{\varepsilon^2}$$

$$\geq 1 - \frac{1}{\varepsilon^2 \log(n)} \to 1 \qquad as\; n \to \infty$$

Since $\frac{n}{\log(n)}$ →∞ as n→∞, we cannot appeal to the Central Limit Theorem to establish asymptotic normality. In fact, what can be deduced from these probabilistic arguments is that there are indeed infinitely many primes (Euclid, 300 BC). The best statement that can be made along this line is the Erdos-kac Theorem:

Theorem (Erdos-kac). Let $\psi(n)$ denote the number of primes that divide n. Then:

$$\frac{\psi(n) - \sqrt{n \log\log n}}{\sqrt{n \log\log(n)}} \to N(0,1) \qquad \text{as } n \to \infty.$$

How large must n be so that $\left|\pi(n) - \frac{n}{\log(n)}\right| < \varepsilon$? we can use (18) to determine n. With probability (1-α), the error is less than ε when:

$$\log(n) \geq \frac{1}{\varepsilon^2\,(1-\alpha)} \qquad (19)$$

$$n \geq exp(\varepsilon^2\,(1-\alpha))^{-1}$$

For instance, if ε = *.01 (1%)* and α = *.01 (1-α = 99%),* then

$$n \geq exp((.01)^2(.99))^{-1} = \exp(10)^5 \qquad (20)$$

From the Erlang distribution of the primes $\{P_n\}$, we observe that:

$$E(P_n) \sim \frac{n}{\beta} = \frac{n}{\frac{1}{\log(n)}} = n \log n$$
$$E(P_n) \sim n \log n \qquad (21)$$

And also that:

$$Var\,(P_n) \sim n\,log^2(n) \qquad (22)$$

From (21), an estimate of the n[th] prime is provided by n log (n).

4.0 Bounds on Primes and Prime Gaps

The Stochastic Prime Number Theorem (SPNT) tells us that the estimate of the number of primes less or equal to n, $\pi(n) \sim \frac{n}{\log(n)}$, is a statistical average statement whereas the Prime Number Theorem (PNT) alludes to no such averaging process. The PNT is a deterministic statement, proved mathematically by using the zeta function. Thus, the requirement that the real part of the Riemann zeta function, have no zeroes on the line $Re(\zeta(s))$=1 needed by Hadamard's proof (1896) is replaced by the assumption that the prime gaps be exponentially distributed (β). In particular, the Riemann hypothesis: "The zeroes of the Riemann zeta function are located on the strip $\{\sigma | \sigma = \frac{1}{2} + it, t \in R\}$" is equivalent to the hypothesis that: "The prime gaps are exponentially distributed (β)". Evidence, so far, shows that the hypothesis of an exponential (β) prime gap distribution is consistent with available data on primes up to 1025 (Zhang, 2013) and so does the Riemann hypothesis. Both hypothesis are difficult (if not, impossible) to prove. The former requires knowledge of all primes while the latter requires mathematics yet to be invented. Much as progress in Biological Sciences continues without ever proving Darwin's Theory of Evolution, developments in Analytic Number Theory will continue while implicitly assuming the Riemann hypothesis.

Meanwhile, there are other conjectures that are equally deserving of serious consideration. These conjectures are of two types: (a) those that relate to small prime gaps e.g. the twin-prime conjecture, and (b) those that the relate to large prime gaps e.g. maximal prime gap conjectures. Equation (23) belongs to the first type while Equation (22) belongs to the second type:

$$\lim_{n\to\infty} sup \frac{d_n}{log^2(n)} = 1 \qquad (22)$$

$$\lim_{n\to\infty} inf \frac{d_n}{log(n)} = 0 \qquad (23)$$

We provide some stochastic answer to these conjectures:

Theorem 4.1 If the prime gaps obey an exponential distribution (β), then:

$$\lim_{n\to\infty} sup \frac{d_n}{log^2(n)} = 1$$

Proof. Let $\{d_n$:n≥1} obey an exponential distribution with $\beta \sim \frac{1}{\log(n)}$ for large n. Then:

$$P(d_n \geq x) = e^{-\beta x} = e^{-\frac{x}{\log(n)}}$$

By the Borel-Cantelli lemma, $P\left(\lim_{n\to\infty} sup \frac{d_n}{log^2(n)} \geq 1\right) = 1$ because

$$\sum_{n=2}^{\infty} exp\left(-\frac{log^2(n)}{\log(n)}\right) = \sum_{n=2}^{\infty} exp(-\log(n)) = \sum_{n=2}^{\infty} \frac{1}{n} \to \infty.$$

Moreover, $sup \frac{d_n}{log^2(n)} \leq 1$ for each n so that $\lim_{n\to\infty} sup \frac{d_n}{log^2(n)} \leq 1$. Combining the two results, yield (22).

Again, we point out that the hypothesis can be replaced by: "if the Riemann hypothesis is true, then...". The result remains valid in a deterministic sense.

Theorem 4.2 If the prime gaps obey an exponential distribution (β), then:

$$\lim_{n\to\infty} inf \frac{d_n}{log(n)} = 0.$$

Proof.

$$P\left(\inf_n d_n \geq \log(n)\right) = P(d_1 \geq \log(n)) \cdot P(d_2 \geq \log(n)) \cdots P(d_n \geq \log(n))$$

$$= \prod_{i=1}^{n} exp\left(-\frac{log(n)}{log(n)}\right)$$

$$= exp(\sum_{i=1}^{n}(-1))$$

$$= \exp(-n) \to 0 \qquad as\ n \to \infty$$

It follows that $\lim_{n\to\infty} inf \frac{d_n}{log(n)} = 0$.

Bounds on the prime gaps can be deduced based on the assumption of an exponential distribution. Bertrand (1845) gave a rough estimate of the prime gap d_n:

$$\text{Bertrand: } P_{n+1}-P_n \leq P_n \qquad (24)$$

Knowing that for large n, P_n~n log n, we deduce that:

$$P_{n+1} - P_n \leq n \log n, \text{ for large } n.$$

Likewise, we can show that:

$$P_{n+1} - P_n \leq log(\log(n))$$

$$P_{n+1} - P_n \leq \frac{\log\log(n)}{\log n} \qquad (25)$$

Proof.

$$P(d_n = P_{n+1} - P_n \leq \log(\log(n))) = \exp\left(-\frac{\log\log(n)}{\log n}\right) \to 1 \quad as \quad n \to \infty$$

Hence, P_{n+1} - $P_n \leq$ n log(log(n)) infinitely often. Similarly,

$$P\left(d_n \leq \frac{\log\log(n)}{\log n}\right) = exp\left(\frac{\log\log(n)}{\log^2(n)}\right) \to 1 \quad \text{as } n \to \infty,$$

Showing that:

$$P_{n+1} - P_n \leq \frac{\log(\log(n))}{\log(n)} \text{ infinitely often.}$$

In fact,

Theorem: P_{n+1} - $P_n \leq \varphi(n)$ infinitely often for all function $\varphi(n)$ such that:

$$\frac{\varphi(n)}{\mathrm{og}(n)} \to 0.$$

Proof.

$$P(d_n \leq \varphi(n)) = exp\left(\frac{\varphi(n)}{\log(n)}\right) \to 1 \quad \text{if} \left(\frac{\varphi(n)}{\log(n)}\right) \to 0.$$

References

______“Bounded gaps between primes”. Polymath. Retrieved 2013-07-21.

Baker, R. C.; Harman, G.; Pintz, J. (2001). “The difference between consecutive primes, II”. Proceedings of the London Mathematical Society 83 (3): 532–562.

Cheng, Yuan-You Fu-Rui (2010). “Explicit estimate on primes between consecutive cubes”. Rocky Mt. J. Math. 40: 117–153.

DHJ Polymath (2014). “Variants of the Selberg sieve, and bounded intervals containing many primes”. Research in the Mathematical Sciences 1 (12).

Ford, K., et al. (2014) “Large gaps between consecutive prime numbers”

Ford, K., et al. (2015). “Long gaps be tween primes”.

Huxley, M. N. (1972). “On the Difference between Consecutive Primes”. Inventiones Mathematicae 15 (2): 164–170.

Ingham, A. E. (1937). “On the difference between consecutive primes”. Quarterly Journal of Mathematics. Oxford Series 8 (1): 255–266. doi:10.1093/qmath/os-8.1.255.

Maynard, J. (2014) “Large gaps be tween primes”

Maynard, J. (2015). “Small gaps between primes”. Annals of Mathematics 181 (1): 383–413.

Pintz, J. (1997). “Very large gaps between consecutive primes”. J. Number Theory 63 (2): 286–301.

Shanks, D. (1964), “On Maximal Gaps

between Successive Primes", Mathematics of Computation (American Mathematical Society) 18 (88): 646–651.

Tchudakoff, N. G. (1936). "On the difference between two neighboring prime numbers". Math. Sb. 1: 799–814.

Zhang, Y. (2014). "Bounded gaps between primes". Annals of Mathematics 179 (3):1121-1174.