Hardy-Ramanujan Journal
Vol. 33 (2010) pp. 2–22

THE COMBINATORICS OF MOMENT CALCULATIONS

Hugh L. Montgomery*

1. Introduction

If X is a discrete random variable with mass function p_X, then the k^{th} moment $E\big[X^k\big]$ is given by the familiar formula

$$E\big[X^k\big] = \sum_x p_X(x) x^k \,. \tag{1}$$

For most purposes in probability theory, all information that one wants concerning this moment can be derived readily from the above formula. However, in analytic number theory one is often dealing with quantities that are nearly, but not exactly independent, and analyses of such situations leads to interesting formulæ for moments. In §2 we consider the distribution of primes in intervals of length $\asymp \log x$; this leads to insights concerning Poisson random variables. In §3 we consider reduced residues (mod q) in short intervals, which similarly motivates us to derive further information concerning moments of binomial random variables. This discussion is extended, and new ground is broken, in §4, where we consider the distribution of primes in intervals of length $\asymp x^\theta$ where $0 < \theta < 1$. Our analysis gives rise to a family of polynomials, whose properties we explore in §5. In §6 we apply the information gained to complete the investigation begun in §4.

Concerning primes in short intervals, we recall that Cramér [2] proposed a probabilistic model in which an integer $n > 1$ would be taken to be 'prime' with probability $1/\log n$, independently over n. If almost all such sequences have a certain property, then one may conjecture that the actual primes also have the same property. By means of this mechanism, we obtain the following conjectures:

1. For any fixed real number $\lambda > 0$, the distribution of $\pi(x + \lambda \log x) - \pi(x)$ tends to Poisson λ as $x \to \infty$.

*Author supported in part by NSF FRG grant DMS–0244660.

2. For any fixed $c > 0$, the number of $n \le N$ such that $p_{n+1} - p_n > c \log p_n$ is asymptotic to $e^{-c}N$ as $N \to \infty$.

3. ('Cramér's conjecture')

$$\limsup_{p_n \to \infty} \frac{p_{n+1} - p_n}{(\log p_n)^2} \;=\; 1\,.$$

4. If c is a fixed real number, $c > 2$, then $\pi(x+h) - \pi(x) \sim h/\log x$ uniformly for $(\log x)^c \le h \le x$.

5. As $x \to \infty$,

$$\lim \left\{ \begin{matrix} \sup \\ \inf \end{matrix} \right\} \frac{\pi(x) - \mathrm{li}(x)}{\sqrt{\dfrac{x \log\log x}{\log x}}} \;=\; \pm\sqrt{2}\,.$$

6. If $X^\varepsilon \le h \le X^{1-\varepsilon}$, then

$$\int_1^X \Bigl(\pi(x+h) - \pi(x) - \int_x^{x+h} \frac{dt}{\log t}\Bigr)^2 dx \;\sim\; \frac{hX}{\log X}.$$

It is to be expected that Cramér's model is subject to some limitations, partly because it fails to recognize the effect of sieving by small primes (e.g., if $n > 2$ is prime, then n is odd and therefore $n+1$ cannot be prime), and also because it does not recognize that the error term in the prime number theorem can be expressed as a sum over zeros of the zeta function. Briefly, we expect that assertion 1. is true. Since 2. is an elementary consequence of 1., we also believe that 2. is true. Assertion 3. was believed for many years, but is now believed to be false, although Granville [7] and later Tenenbaum & Mendès France [22, pp. 61–66] constructed more elaborate probabilistic models that takes account of sieving by small primes, which has led to the suggestion that 3. would be correct if the right hand side were replaced by $2e^{-C_0} = 1.1229\ldots$. Here $C_0 = 0.5772\ldots$ is Euler's constant. From the work of Maier [10], we know that 4. is false when $h \le (\log x)^A$, no matter how large A is. Possibly 4. is true under the slightly stronger hypothesis that $(\log h)/\log\log x \to \infty$ with $h \le x$. Assertions 5. and 6. are consistent with everything known today, but are believed to be false. Assertion 5. implies the Riemann Hypothesis (RH), which we do believe, but 5. reflects the law of the iterated logarithm for random walks, while if RH is true, then the error term in the Prime Number Theorem has a very regular almost periodic behavior that is quite different from a random walk. Indeed, instead of 5., Montgomery [12] conjectured that

$$\lim \left\{ \begin{matrix} \sup \\ \inf \end{matrix} \right\} \frac{\pi(x) - \mathrm{li}(x)}{\dfrac{x^{1/2}(\log\log\log x)^2}{\log x}} \;=\; \pm\,\frac{1}{2\pi}\,.$$

With regard to 6., we recall that Goldston & Montgomery [6] showed that if RH is true, then the pair correlation conjecture of Montgomery [11] is equivalent to the assertion that

$$\int_1^X \Big(\pi(x+h)-\pi(x)-\int_x^{x+h}\frac{dt}{\log t}\Big)^2\,dx \;\sim\; \frac{hX\log X/h}{(\log X)^2} \tag{2}$$

for $X^\varepsilon \le h \le X^{1-\varepsilon}$. This is the same order of magnitude as in 6., but with a smaller constant: If $h = X^\theta$, then the right hand side above is smaller than that in 6. by a factor $1-\theta$. The rationale behind 6. is that the distribution of $\pi(x+h)-\pi(x)$ should be approximately normal, with the indicated variance. In discarding 6. in favor of (2) we have a perhaps more likely conjecture concerning the variance, but have lost a conjecture concerning the distribution of $\pi(x+h)-\pi(x)$. To address this, Montgomery & Soundararajan [13] studied the higher moments, and came to the conjecture that the distribution of $\pi(x+h)-\pi(x)$ is approximately normal, but with the smaller variance given in (2). This analysis is the subject of §§4–6 below.

Concerning the disparity between 6. and (2), we have not only considerable heuristic evidence in favor of (2), but also numerical evidence. Odlyzko [16] computed zeros of the zeta function, and found that they are distributed as predicted by the pair correlation conjecture. Also, Montgomery & Soundararajan [13] tabulated the distribution of $\pi(x+h)-\pi(x)$ for $0 \le x \le 10^{10}$ and $h = 10^5$, and found an extraordinarily good fit to normal.

We note in passing that the pair correlation analysis can be extended to consideration of n-level correlations, where a combinatorial lemma of Spitzer [21] is useful; see Rudnick & Sarnak [19, 21].

2. Primes in short intervals — I

We recall that the Prime Number Theorem asserts that $\pi(x) \sim x/\log x$ as $x \to \infty$. Hence if p_n denotes the n^{th} prime number, then

$$\frac{1}{\pi(x)}\sum_{p_n\le x}(p_{n+1}-p_n) \;\sim\; \log x,$$

which is to say that the average of $p_{n+1}-p_n$ is $\log p_n$. Gallagher [5] explored the distribution of primes in intervals $(n, n+h]$ of length $h \asymp \log x$ by considering moments:

$$\sum_{n=1}^N \big(\pi(n+h)-\pi(n)\big)^k \;=\; \sum_{n=1}^N \sum_{\substack{p_1,\dots,p_k\\ n<p_i\le n+h}} 1\,. \tag{3}$$

Let $\mathcal{D}$ denote the set of differences $p_i - n$. These differences may not all be distinct, so we put $r = \operatorname{card}\mathcal{D}$. Thus $1 \le r \le k$, and we let $d_1 < d_2 < \cdots < d_r$ denote the elements of $\mathcal{D}$. Put

$$\pi_{\mathcal{D}}(N) \;=\; \sum_{\substack{n=1\\ n+d_i \text{ prime}\\ (1\le i\le r)}}^N 1\,.$$

By grouping k-tuples of primes according to the set $\mathcal{D}$ of differences generated, we see that the right hand side of (3) is

$$= \sum_{r=1}^{k} S(k,r) r! \sum_{\substack{\mathcal{D} \subseteq \{1,\dots,h\} \\ \operatorname{card} \mathcal{D} = r}} \pi_{\mathcal{D}}(N) \tag{4}$$

where $S(k,r)$ denotes the Stirling numbers of the second kind. The prime r-tuple conjecture asserts that $\pi_{\mathcal{D}}(N)$ tends to infinity with N, provided that $\mathcal{D}$ is *admissible*, which is to say if there is no prime p such that the members of $\mathcal{D}$ cover every residue class (mod p). This condition is obviously necessary, for if it fails then there is a fixed prime p such that for any integer n there is at least one i such that p divides $n + d_i$. In 1922, Hardy & Littlewood [8] developed heuristics that suggested a quantitative form of the prime r-tuple conjecture, namely that

$$\pi_{\mathcal{D}}(N) \;=\; \big(\mathfrak{S}(\mathcal{D}) + o(1)\big) \frac{N}{(\log N)^r} \tag{5}$$

where $\mathfrak{S}(\mathcal{D})$ is the 'singular series'

$$\mathfrak{S}(\mathcal{D}) \;=\; \prod_{p} \Big(1 - \frac{\nu_p(\mathcal{D})}{p}\Big)\Big(1 - \frac{1}{p}\Big)^{-r}. \tag{6}$$

Here $\nu_p(\mathcal{D})$ denotes the number of distinct residue classes (mod p) represented by the members of $\mathcal{D}$. Since $\nu_p(\mathcal{D}) = r$ for all sufficiently large primes, the product above is absolutely convergent. Hence its value is positive, unless there is a prime p for which $\nu_p(\mathcal{D}) = p$, which is to say the set $\mathcal{D}$ is not admissible. One may wonder why $\mathfrak{S}(\mathcal{D})$ is called the singular series, since in (6) we see that $\mathfrak{S}(\mathcal{D})$ is patently a product, not a series. The simple explanation is that $\mathfrak{S}(\mathcal{D})$ arises initially as

$$\mathfrak{S}(\mathcal{D}) \;=\; \sum_{\substack{q_1,\dots,q_r \\ 1 \le q_i < \infty}} \Big(\prod_{i=1}^{r} \frac{\mu(q_i)}{\varphi(q_i)} \Big) \sum_{\substack{a_1,\dots,a_r \\ 1 \le a_i \le q_i \\ (a_i,q_i)=1 \\ \sum a_i/q_i \in \mathbb{Z}}} e\Big(\sum_{i=1}^{r} \frac{a_i d_i}{q_i} \Big). \tag{7}$$

Here $e(\theta) = e^{2\pi i \theta}$ is the complex exponential with period 1. Hardy & Littlewood showed that this complicated sum is in fact equal to the rather simpler (and more suggestive) product in (6).

Although (5) is unproved, it is strongly supported not only by heuristics but also by numerical evidence (see Brent [1]), so we use it as guide. In (4) we replace $\pi_{\mathcal{D}}(N)$ by the conjectured approximation $\mathfrak{S}(\mathcal{D})N/(\log N)^r$, which leads us to expect that the left hand side of (3) is approximately

$$N \sum_{r=1}^{k} \frac{S(k,r) r!}{(\log N)^r} \sum_{\substack{\mathcal{D} \subseteq \{1,\dots,h\} \\ \operatorname{card} \mathcal{D} = r}} \mathfrak{S}(\mathcal{D}). \tag{8}$$

Gallagher showed that

$$\sum_{\substack{\mathcal{D}\subseteq\{1,\dots,h\}\\ \text{card }\mathcal{D}=r}} \mathfrak{S}(\mathcal{D}) \ \sim \ \frac{h^r}{r!} \tag{9}$$

as $h \to \infty$. Since the number of summands here is $\binom{h}{r} \sim h^r/r!$, the above may be interpreted as saying that the mean value of $\mathfrak{S}(\mathcal{D})$ is asymptotically 1. That this should be the case is not so obvious when one starts from (6), but it is not surprising when one considers (7). Among the many terms, there is one, with $a_i = q_i = 1$ for all i, that contributes the constant 1. All other terms involve at least one factor of the form $e(a_i d_i/q_i)$ with $q_i > 1$. This is a root of unity, and has mean value tending to 0 as d_i runs over a long interval.

On inserting (9) into (8), we come to the conclusion that we should expect that

$$\frac{1}{N}\sum_{n=1}^{N}\big(\pi(n+h)-\pi(n)\big)^k \ \sim \ \sum_{r=1}^{k} S(k,r)\Big(\frac{h}{\log N}\Big)^r . \tag{10}$$

¿From the Cramér prediction 1. we would expect to see the k^{th} moment of a Poisson variable here. Let X denote a Poisson variable with parameter λ. The X has the mass function $p_{_X}(n) = e^{-\lambda}\lambda^n/n!$ for non-negative integers n. Let $m_k(\lambda) = E\big[X^k\big]$ denote the k^{th} moment of X. It is familiar that

$$m_k(\lambda) \ = \ e^{-\lambda}\sum_{n=0}^{\infty}\frac{n^k}{n!}\lambda^n . \tag{11}$$

What is not so familiar is that this moment can also be written as

$$= \ \sum_{r=1}^{k} S(k,r)\lambda^r . \tag{12}$$

Thus the right hand side of (10) is $m_k(h/\log N)$. Since the moment generating function

$$\sum_{k=0}^{\infty}\frac{m_k(\lambda)}{k!}z^k \ = \ e^{-\lambda}e^{\lambda e^z}$$

is entire, it follows from (10) that the distribution of $\pi(n+h)-\pi(n)$ is approximately Poisson with parameter $\lambda = h/\log N$, when $h \asymp \log N$.

Although this derivation is only heuristic, it could be made rigorous if the prime r-tuple conjecture (5) would hold uniformly when the d_i are no bigger than $C\log N$. Indeed, a good deal less would suffice. Let $E_{\mathcal{D}}(N)$ denote the error term in (5), which is to say that

$$\pi_{\mathcal{D}}(N) \ = \ \mathfrak{S}(\mathcal{D})\frac{N}{(\log N)^r} \ + \ E_{\mathcal{D}}(N) . \tag{13}$$

If for every fixed positive r we have

$$\sum_{\substack{\mathcal{D}\subseteq\{1,\dots,h\}\\ \text{card }\mathcal{D}=r}} |E_{\mathcal{D}}(N)| \;=\; o\Big(\frac{h^r N}{(\log N)^r}\Big) \tag{14}$$

when $h \asymp \log N$, then (10) and the Cramér predictions 1. and 2. both hold.

3. Reduced residues in short intervals

We say that a residue class $a \pmod q$ is a 'reduced residue class' if $(a,q)=1$. Since there are $\varphi(q)$ reduced residue classes modulo q, if a residue class is chosen at random, then the probability that it is a reduced residue class is $P=\varphi(q)/q$. Among h consecutive residue classes, on average the number of reduced residue classes is hP. To examine the distribution of the number of reduced residue classes in an interval, we form the moment

$$M_K(q;h) \;=\; \frac{1}{q}\sum_{n=1}^{q}\Big(\sum_{\substack{m=1\\(m+n,q)=1}}^{h} 1 \;-\; hP\Big)^K. \tag{15}$$

Hausman & Shapiro [9] showed that

$$M_2(q;h) \;=\; P^2\sum_{\substack{r\mid q\\ r>1}}\mu(r)^2\Big(\prod_{\substack{p\mid q\\ p\nmid r}}\frac{p(p-2)}{(p-1)^2}\Big)\frac{r^2}{\varphi(r)^2}\Big\{\frac{h}{r}\Big\}\Big(1-\Big\{\frac{h}{r}\Big\}\Big)$$

where $\{x\}$ denotes the fractional part of x, $\{x\}=x-[x]$. Since $\{x\}(1-\{x\})\le x$ for all $x\ge 0$, we see from the above that $M_2(q;h)\le hP$. This is encouraging, for if the summands in the inner sum in (15) behave like independent independent Bernoulli variables, then the moment $M_K(q;h)$ should be approximately the moment of a binomial variable (centered about its mean). In particular, for $K=2$ we have the variance, which for a binomial variable with parameters h and P is hP. Thus in our arithmetic setting, the variance is no bigger than it would be in the corresponding probabilistic situation. The problem of bounding the higher moments proved to be much more problematic, but Montgomery & Vaughan [15] showed that

$$M_K(q;h) \;=\; O_K\big((hP)^{K/2}\big) \qquad\qquad (1/P\le h\le q). \tag{16}$$

This has an interesting application to the gaps between consecutive reduced residues. Let $1=a_1<a_2<\cdots<a_{\varphi(q)}<q$ denote the reduced residues in increasing order. The average of $a_{i+1}-a_i$ is

$$\frac{1}{\varphi(q)}\sum_{i=1}^{\varphi(q)}(a_{i+1}-a_i) \;=\; \frac{q}{\varphi(q)} \;=\; \frac{1}{P}.$$

¿From (16) it follows that

$$\sum_{i=1}^{\varphi(q)} (a_{i+1} - a_i)^\gamma \;=\; O_\gamma\big(P^{-\gamma}\big)$$

for any real $\gamma > 0$. Since the left hand side is trivially $\geq P^{-\gamma}$, the estimate above is best possible, apart from determining the dependence of the implicit constant on γ. The estimate above was conjectured by Erdős, who offered (and subsequently paid) \$500 for its solution.

In the Montgomery–Vaughan proof of (16) one finds two methods. One method is effective in treating q composed entirely of small primes, while the other method is useful in dealing with q composed only of large primes. Fortunately, the two methods are sufficiently flexible to allow one to write $q = q_1 q_2$, where q_1 is composed only of small primes, q_2 is composed only of large primes, and the methods combine to treat any q. It is possible that a complete proof might be constructed using only one of the methods, but as far as is known at this point, both methods are needed. Although the method for q composed of small primes has many interesting features, we ignore it and dwell on the method having to do with large primes, as it relates to sums of variables that are nearly independent.

In considering the distribution of reduced residues, it is a simple matter to reduce to a square-free modulus, so we may assume, without loss of generality, that q is square-free. By definition, the number of n, $0 < n \leq q$, such that $(n, q) = 1$ is $\varphi(q) = Pq$. Similarly, the number of n, $0 < n \leq q$, such that $(n + 1, q) = 1$ is $\varphi(q) = Pq$. By the Chinese remainder theorem we see that the number of n, $0 < n \leq q$, such that $(n, q) = (n + 1, q) = 1$ is

$$\prod_{p|q} (p - 2) \;=\; q \prod_{p|q} \Big(1 - \frac{2}{p}\Big)\,.$$

Now $1 - 2/p$ is approximately $(1 - 1/p)^2$ if p is large, and thus we see that the above is approximately $P^2 q$ if all the prime factors of q are large. It is with these kinds of considerations in mind that we pursue the estimation of $M_K(q; h)$ for those square-free q with the property that $p > y$ for all prime factors p of q, where y is a parameter such that $y > h$.

By the binomial theorem,

$$M_K(q; h) \;=\; \frac{1}{q} \sum_{k=0}^{K} \binom{K}{k} (-hP)^{K-k} \sum_{n=1}^{q} \Bigg(\sum_{\substack{m=1 \\ (m+n,q)=1}}^{h} 1 \Bigg)^k .$$

By writing the final k^{th} power as a product of k sums, we see that the sum over n is

$$= \sum_{d_1=1}^{h} \cdots \sum_{d_k=1}^{h} \sum_{\substack{n=1 \\ (n+d_i,q)=1 \\ (1 \leq i \leq k)}}^{q} 1\,.$$

Let $\mathcal{D} = \{d_1, \dots, d_k\}$, and put $r = \operatorname{card} \mathcal{D}$. Thus $1 \le r \le k$. Suppose that $p|q$. Since $p > y > h$, we see that the r members of $\mathcal{D}$ are not only distinct as integers, but also represent distinct residue classes (mod p). Thus by the Chinese remainder theorem it follows that

$$\sum_{\substack{n=1 \\ (n+d_i,q)=1}}^{q} 1 \;=\; \prod_{p|q}(p-r) \;=\; q\prod_{p|q}\Big(1-\frac{r}{p}\Big)\,.$$

Each factor here is approximately $(1-1/p)^r$, so the above is approximately qP^r, with an error term that can be made explicit in terms of the parameter y. Thus we see that $M_K(q;h)$ is approximately

$$\sum_{k=0}^{K}\binom{K}{k}(-hP)^{K-k}\sum_{r=1}^{k}\binom{h}{r}S(k,r)r!P^r\,. \tag{17}$$

For purposes of comparison, suppose that X is a binomial random variable with parameters h and P, which is to say that

$$X = X_1 + \cdots + X_h \tag{18}$$

where the X_i are independent Bernoulli variables with parameter P. Then X has expectation $E[X] = hP$, and we let $\mu_{_K}(h,P) = E\big[(X-hP)^K\big]$ denote the K^{th} moment of X about its mean. Of course X has mass function $p_{_X}(n) = \binom{h}{n}P^n(1-P)^{h-n}$ for $n = 0, 1, \dots, h$, and from this it is immediate that

$$\mu_{_K}(h,P) \;=\; \sum_{n=0}^{h}\binom{h}{n}P^n(1-P)^{h-n}(n-hP)^K\,. \tag{19}$$

We can derive a second formula for this moment by mimicking the calculation just completed. First, by the binomial theorem we see that

$$\mu_{_K}(h,P) \;=\; E\big[(X-hP)^K\big] \;=\; \sum_{k=0}^{K}\binom{K}{k}(-hP)^{K-k}E\big[X^k\big]\,.$$

In view of (18), we can write X^k as a k-fold sum,

$$X^k = \sum_{i_1=1}^{h}\cdots\sum_{i_h=1}^{h}X_{i_1}\cdots X_{i_h}\,.$$

Since X_i takes only the values 0 and 1, we know that $X_i^m = X_i$ for $m = 1, 2, \dots$. Thus the multiplicity to which an index occurs is unimportant, but the number of distinct indices

is significant, for if $X_1, \dots, X_r$ are r independent Bernoulli variables with parameter P, then $E\big[X_1 \cdots X_r\big] = P^r$. Let $\mathcal{D} = \{i_1, \dots, i_k\}$, and once again set $r = \operatorname{card} \mathcal{D}$, so that

$$E\big[X^k\big] \;=\; \sum_{r=1}^{k} S(k,r) r! \sum_{\substack{\mathcal{D} \subseteq \{1,\dots,h\} \\ \operatorname{card} \mathcal{D} = r}} E\Big[\prod_{d \in \mathcal{D}} X_d\Big] \;=\; \sum_{r=1}^{k} \binom{h}{r} S(k,r) r! P^r \,.$$

Hence the expression (17) is precisely $\mu_{K}(h,P)$. The expression (17) is also interesting, since it is a polynomial in the two variables h and P, unlike the right hand side of (19). In (17) one may collect monomial terms according to the power of h, and thus write

$$\mu_{K}(h,P) \;=\; \sum_{k=0}^{K} F_{K,k}(P) h^k$$

where $F_{K,k}$ is a polynomial in P. In order to obtain the needed estimates, Montgomery & Vaughan showed that

$$\deg F_{K,k} \;\le\; K, \tag{20}$$

$$P^k (1-P)^k \mid F_{K,k}(P), \tag{21}$$

$$F_{K,k}(P) \equiv 0 \text{ if } k > [K/2], \tag{22}$$

$$F_{2k,k}(P) = 1 \cdot 3 \cdots (2k-1) P^k (1-P)^k \,. \tag{23}$$

These properties of the $F_{K,k}$ are easily derived by induction, by means of a recurrence of Romanovsky [18], which asserts that

$$\mu_{K+1}(h,P) \;=\; KhP(1-P)\mu_{K-1}(h,P) + P(1-P)\frac{\partial}{\partial P}\mu_{K}(h,P) \,. \tag{24}$$

Romanovsky established this by using (19) and familiar properties of binomial coefficients. Montgomery & Vaughan [15, pp. 328–329] gave a second proof, based on (17) and elementary properties of the $S(k,r)$.

Let X be a binomial variable with parameters h and P, and set

$$Y \;=\; \frac{X - hP}{\sqrt{hP(1-P)}} \,.$$

Then $E[Y] = 0$, $\operatorname{Var}(Y) = 1$, and

$$E\big[Y^K\big] \;=\; \frac{\mu_{K}(h,P)}{(hP(1-P))^{K/2}} \,.$$

¿From (22) and (23) we deduce that

$$\lim_{h\to\infty} E\big[Y^K\big] \;=\; \mu_{_K} \tag{25}$$

where

$$\mu_{_K} \;=\; \begin{cases} \frac{K!}{2^{K/2}(K/2)!} & K \text{ even}, \\ 0 & K \text{ odd}. \end{cases} \tag{26}$$

Let Z be a normal random variable with $\mu = 0$ and $\sigma = 1$. Since the right hand side above is precisely the K^{th} moment of Z, and since the moment generating function of Z is entire, it follows that the distribution function of Y tends to normal as $h \to \infty$. This, of course, is nothing more than the most classical case of the Central Limit Theorem.

In a similar vein we note that in (17), the sum over r is $E\big[X^k\big]$. If we set $P = \lambda/h$, and let h tend to infinity, then we obtain the expression (12), which is $m_k(\lambda)$, the k^{th} moment of a Poisson variable with parameter λ. It is of course a familiar elementary fact that the distribution of such a sequence of binomial variables tends in the limit to the distribution of a Poisson variable with parameter λ.

4. Primes in short intervals — II

We return now to the subject of primes in short intervals, but instead of considering intervals of length $\asymp \log x$, we discuss now intervals whose length is a fractional power of x. Rather than count primes with weight 1, as we have done thus far, it is more convenient now to count primes by means of the von Mangoldt function,

$$\Lambda(n) \;=\; \begin{cases} \log p & \text{if } n = p^k, \\ 0 & \text{otherwise.} \end{cases}$$

It is traditional to set $\psi(x) = \sum_{n\le x} \Lambda(n)$. In this notation, the Prime Number Theorem is expressed by the relation $\psi(x) \sim x$, the Cramér prediction would be that

$$\int_1^X (\psi(x+h) - \psi(x) - h)^K \, dx = (\mu_{_K} + o(1))X(h\log X)^{K/2} \tag{27}$$

for $X^\varepsilon \le h \le X^{1-\varepsilon}$ with the $\mu_{_K}$ give by (26), while the result of Goldston & Montgomery [6] previously mentioned is that (assuming RH) the strong form of the Pair Correlation Conjecture is equivalent to the relation

$$\int_1^X (\psi(x+h) - \psi(x) - h)^2 \, dx \;\sim\; Xh\log\frac{X}{h} \tag{28}$$

for $X^\varepsilon \le h \le X^{1-\varepsilon}$. As already noted, these estimates are inconsistent, and we believe that (28) is more plausible. However, this leaves us without a conjecture concerning the higher moments. By the binomial theorem as in the preceding section, we see that

$$\sum_{n=1}^{N} \Big(\sum_{m=1}^{h} \Lambda(m+n) \ - h \Big)^K \ = \ \sum_{k=0}^{K} \binom{K}{k} (-h)^{K-k} \sum_{n=1}^{N} \Big(\sum_{m=1}^{h} \Lambda(m+n) \Big)^k . \tag{29}$$

Suppose that $k > 0$. Then the sum over n is

$$\sum_{m_1=1}^{h} \cdots \sum_{m_k=1}^{h} \sum_{n=1}^{N} \prod_{i=1}^{k} \Lambda(n+m_i) .$$

Let $\mathcal{D} = \{m_1, \dots, m_k\}$, let $r = \operatorname{card} \mathcal{D}$, let $1 \le d_1 < \cdots < d_r \le h$ denote the members of $\mathcal{D}$, and let M_i denote the number of j, $1 \le j \le k$, for which $m_j = d_i$. Thus the above is

$$= \sum_{r=1}^{k} \sum_{\substack{\mathcal{D} \subseteq \{1,\dots,h\} \\ \operatorname{card} \mathcal{D} = r}} \sum_{\substack{M_1,\dots,M_r \\ M_i \ge 1 \\ \sum M_i = r}} \binom{k}{M_1 \ \cdots \ M_r} \sum_{n=1}^{N} \prod_{i=1}^{r} \Lambda(n+d_i)^{M_i} .$$

Since higher powers of primes are comparatively rare, and since the d_i are small compared with N, we replace $\Lambda(n+d_i)^{M_i}$ by $\Lambda(n+d_i)(\log n)^{M_i - 1}$, and in doing so we introduce a small and easily estimated error term. Thus the above is approximately

$$\sum_{r=1}^{k} S(k,r) r! \sum_{\substack{\mathcal{D} \subseteq \{1,\dots,h\} \\ \operatorname{card} \mathcal{D} = r}} \sum_{n=1}^{N} (\log n)^{k-r} \prod_{i=1}^{r} \Lambda(n+d_i) . \tag{30}$$

We assume that the prime r-tuple conjecture holds in the quantitative form

$$\sum_{n=1}^{N} \prod_{i=1}^{r} \Lambda(n+d_i) \ = \ (\mathfrak{S}(\mathcal{D}) + o(1))N \tag{31}$$

where $\mathfrak{S}(\mathcal{D})$ is defined as in (6) or (7). The error term here may be quite small (possibly $O\big(N^{1/2+\varepsilon}\big)$), and even to the extent that it isn't small, there may be considerable cancellation among the error terms when the above is applied over a wide range of $\mathcal{D}$. Assuming the above, it follows by integration by parts that

$$\sum_{n=1}^{N} (\log n)^j \prod_{i=1}^{r} \Lambda(n+d_i) \ = \ (\mathfrak{S}(\mathcal{D}) + o(1)) I_j(N)$$

where

$$I_j(N) \;=\; \int_1^N (\log u)^j \, du\,. \tag{32}$$

We insert this approximation in (30), ignore any error terms introduced by doing so, and continue by studying the main term

$$\sum_{r=1}^{k} S(k,r) r! I_{k-r} \sum_{\substack{\mathcal{D}\subseteq\{1,\dots,h\}\\ \text{card }\mathcal{D}=r}} \mathfrak{S}(\mathcal{D})\,.$$

The above would still hold if we were to allow also $r = 0$ in the sum, since $S(k,0) = 0$ when $k > 0$. This is advantageous, because with this change, the above is also valid when $k = 0$, provided that we adopt the convention that $S(0,0) = 1$. Thus our main term, for all k, is

$$\sum_{r=0}^{k} S(k,r) r! I_{k-r} \sum_{\substack{\mathcal{D}\subseteq\{1,\dots,h\}\\ \text{card }\mathcal{D}=r}} \mathfrak{S}(\mathcal{D})\,. \tag{33}$$

In discussing (7), we noted that a constant term 1 arises when $q_i = 1$ for all i. It is now essential to separate those i for which $q_i = 1$ from those for which $q_i > 1$. To this end, we put

$$\mathfrak{S}_0(\mathcal{D}) \;=\; \sum_{\substack{q_1,\dots,q_r\\ 1<q_i<\infty}} \Big(\prod_{i=1}^{r} \frac{\mu(q_i)}{\varphi(q_i)}\Big) \sum_{\substack{a_1,\dots,a_r\\ 1\le a_i\le q_i\\ (a_i,q_i)=1\\ \sum a_i/q_i\in\mathbb{Z}}} e\Big(\sum_{i=1}^{r}\frac{a_i d_i}{q_i}\Big)\,. \tag{34}$$

Note that here $q_i > 1$ for all i. Hence

$$\mathfrak{S}(\mathcal{D}) \;=\; \sum_{\mathcal{J}\subseteq\mathcal{D}} \mathfrak{S}_0(\mathcal{J}), \tag{35}$$

$$\mathfrak{S}_0(\mathcal{D}) \;=\; \sum_{\mathcal{J}\subseteq\mathcal{D}} (-1)^{\text{card }\mathcal{J}} \mathfrak{S}(\mathcal{J}), \tag{36}$$

where it is understood that $\mathfrak{S}_0(\emptyset) = \mathfrak{S}(\emptyset) = 1$. Alternatively, one could take (36) to be the definition of $\mathfrak{S}_0$, but this might seem to be somewhat unmotivated. An alternative approach (cf Montgomery & Soundararajan [14]) would be to note that part of the awkwardness of the situation is due to the fact that $\Lambda(n)$ has a non-zero mean value,

which would be eliminated if we would work instead with $\Lambda_0(n) = \Lambda(n) - 1$; then (31) is equivalent to

$$\sum_{n=1}^{N} \prod_{i=1}^{r} \Lambda_0(n + d_i) \;=\; (\mathfrak{S}_0(\mathcal{D}) + o(1))N\,.$$

We substitute (35) into (33), group subsets $\mathcal{I}$ by cardinality, with $s = \operatorname{card} \mathcal{I}$, and observe that for any given such $\mathcal{I}$, there exist exactly $\binom{h-s}{r-s}$ sets $\mathcal{D} \subseteq \{1, \dots, h\}$ with $\operatorname{card} \mathcal{D} = r$ and $\mathcal{I} \subseteq \mathcal{D}$. Thus we see that the expression (33) is

$$\sum_{r=0}^{k} S(k,r) r! I_{k-r} \sum_{s=0}^{r} \frac{1}{s!} \binom{h-s}{r-s} R_s(h)$$

where

$$R_s(h) \;=\; \sum_{\substack{1 \le d_1, \dots, d_s \le h \\ d_i \text{ distinct}}} \mathfrak{S}_0(\mathcal{D}) \tag{37}$$

for $s > 0$, and $R_0(h) = 1$. Hence our best guess is that the moment (29) should be approximately

$$\mathcal{M}_K(h, P) \;=\; \sum_{s=0}^{K} R_s(h) \sum_{j=0}^{K-s} I_j(N) P_{K,s,j}(h) \tag{38}$$

where

$$P_{K,s,j}(h) \;=\; \sum_{i=s}^{K-j} \binom{K}{i+j} S(i+j, i) \frac{i!}{s!} (-h)^{K-i-j} \binom{h-s}{i-s}\,. \tag{39}$$

This can also be written as

$$P_{K,s,j}(h) \;=\; \sum_{i=s}^{K-j} \binom{K}{i+j} \binom{i}{s} S(i+j, i) (-h)^{K-i-j} (h-s)(h-s-1) \cdots (h-i+1)$$

provided that one interprets the product $(h-s) \cdots (h-i+1)$ to be 1 when $i = s$.

Montgomery & Soundararajan [14] refined the work of Montgomery & Vaughan [15] to show that

$$R_s(h) \;=\; \mu_s (-h \log h + Ah)^{s/2} + O_s\big(h^{s/2 - 1/(7s) + \varepsilon}\big) \tag{40}$$

where $A = 2 - C_0 - \log 2\pi$. In order to continue our investigation we require further knowledge of the polynomials $P_{K,s,j}(h)$.

5. The polynomials $P_{K,s,j}(h)$

The main properties of the polynomials $P_{K,s,j}$ can be derived by using the following basic recurrence.

Theorem 1. *For $K \geq 0$, $s \geq 0$, $j \geq 0$, and $s + j \leq K$, let $P_{K,s,j}(h)$ be defined as in (39), and put $P_{K,s,j}(h) = 0$ otherwise. Then $P_{K,s,K-s}(h) = S(K,s)$ where by convention, $S(0,0) = 1$, and if $K \geq s + j$, then*

(41)
$$\begin{aligned} P_{K+1,s,j}(h) \;=\; & (j - s - K)P_{K,s,j}(h) \;+\; P_{K,s-1,j}(h) \\ & +\; (K - j + 1)P_{K,s,j-1}(h) \;-\; hKP_{K-1,s,j}(h) \;+ hKP_{K-1,s,j-1}(h)\,. \end{aligned}$$

Proof. The first assertion is immediate from (39). As for the second, by the identity $\binom{K+1}{i+j} = \binom{K}{i+j} + \binom{K}{i+j-1}$ we see that

$$\begin{aligned} P_{K+1,s,j}(h) \;=\; & \sum_{i=s}^{K+1-j} \binom{K}{i+j} S(i+j,i) \frac{i!}{s!} (-h)^{K+1-i-j} \binom{h-s}{i-s} \\ & +\; \sum_{i=s}^{K+1-j} \binom{K}{i+j-1} S(i+j,i) \frac{i!}{s!} (-h)^{K+1-i-j} \binom{h-s}{i-s}\,. \end{aligned}$$

In the first sum we can restrict i to $i \leq K - j$, since $\binom{K}{i+j} = 0$ when $i = K + 1 - j$. In the second sum we reindex by writing i for $i - 1$. Thus the above is

$$\begin{aligned} =\; & \sum_{i=s}^{K-j} \binom{K}{i+j} S(i+j,i) \frac{i!}{s!} (-h)^{K+1-i-j} \binom{h-s}{i-s} \\ & +\; \sum_{i=s-1}^{K-j} \binom{K}{i+j} S(i+j+1,i+1) \frac{(i+1)!}{s!} (-h)^{K-i-j} \binom{h-s}{i+1-s} \\ =\; & \Sigma_1 + \Sigma_2, \end{aligned} \tag{42}$$

say. Here

$$\Sigma_1 = -hP_{K,s,j}(h)\,. \tag{43}$$

By the recurrence $S(i+j+1,i+1) = (i+1)S(i+j,i+1) + S(i+j,i+j)$ we see that

$$\begin{aligned} \Sigma_2 \;=\; & \sum_{i=s-1}^{K-j} \binom{K}{i+j} (i+1) S(i+j,i+1) \frac{(i+1)!}{s!} (-h)^{K-i-j} \binom{h-s}{i+1-s} \\ & +\; \sum_{i=s-1}^{K-j} \binom{K}{i+j} S(i+j,i)(i+1)!s!(-h)^{K-i-j} \binom{h-s}{i+1-s} \\ =\; & \Sigma_{21} + \Sigma_{22}, \end{aligned} \tag{44}$$

say. In Σ_{21} we replace $i+1$ by i to see that

$$\Sigma_{21} \;=\; \sum_{i=s}^{K+1-j} \binom{K}{i+j-1} i S(i+j-1,i) \frac{i!}{s!} (-h)^{K+1-i-j} \binom{h-s}{i-s}.$$

We write the factor i in the above as $i = -(K-i-j+1)+(K-j+1)$. Since $\binom{K}{i+j-1}(K-i-j+1) = K\binom{K-1}{i+j-1}$, we deduce that

$$\begin{aligned}\Sigma_{21} \;&=\; -K \sum_{i=s}^{K+1-j} \binom{K-1}{i+j-1} S(i+j-1,i) \frac{i!}{s!} (-h)^{K+1-i-j} \binom{h-s}{i-s} \\ &\quad +\; (K-j+1) \sum_{i=s}^{K+1-j} \binom{K}{i+j-1} S(i+j-1,i) \frac{i!}{s!} (-h)^{K+1-i-j} \binom{h-s}{i-s} \\ &=\; hK P_{K-1,s,j-1}(h) \;+\; (K-j+1) P_{K,s,j-1}(h). \qquad (45)\end{aligned}$$

On writing

$$\frac{(i+1)!}{s!} \;=\; \frac{i!}{s!}(i+1-s) \;+\; \frac{i!}{(s-1)!},$$

we see that

$$\begin{aligned}\Sigma_{22} \;&=\; \sum_{i=s}^{K-j} \binom{K}{i+j} S(i+j,i) \frac{i!}{s!} (-h)^{K-i-j} \binom{h-s}{i+1-s} (i+1-s) \\ &\quad +\; \sum_{i=s-1}^{K-j} \binom{K}{i+j} S(i+j,i) \frac{i!}{(s-1)!} (-h)^{K-i-j} \binom{h-s}{i+1-s} \\ &=\; \Sigma_{221} \;+\; \Sigma_{222}, \qquad (46)\end{aligned}$$

say. Since

$$\binom{h-s}{i+1-s}(i+1-s) \;=\; \binom{h-s}{i-s}(h-i) \;=\; \binom{h-s}{i-s}(h-K+j) \;+\; \binom{h-s}{i-s}(K-i-j),$$

it follows that

$$\Sigma_{221} \;=\; (h-K+j) P_{K,s,j}(h) + \sum_{i=s}^{K-j} \binom{K}{i+j} (K-i-j) S(i+j,i) \frac{i!}{s!} (-h)^{K-i-j} \binom{h-s}{i-s}.$$

But $\binom{K}{i+j}(K-i-j) = K\binom{K-1}{i+j}$, so this last sum is $-hK P_{K-1,s,j}$. Thus

$$\Sigma_{221} \;=\; (h-K+j) P_{K,s,j}(h) \;-\; hK P_{K-1,s,j}(h). \qquad (47)$$

Finally, we note that

$$\binom{h-s}{i+1-s} \;=\; \binom{h+1-s}{i+1-s} \;-\; \binom{h-s}{i-s},$$

from which we deduce that

$$\Sigma_{222} \;=\; P_{K,s-1,j}(h) \;-\; sP_{K,s,j}(h)\,. \tag{48}$$

The stated identity for $P_{K+1,s,j}(h)$ now follows by combining (42)–(48).

With this recurrence as a useful tool, we are able to show that the polynomials $P_{K,s,j}$ have the following important properties.

Theorem 2. *For $K \ge 0$, $s \ge 0$, $j \ge 0$, and $s+j \le K$, let $P_{K,s,j}(h)$ be defined as in (39), and put $P_{K,s,j}(h) = 0$ otherwise. Then* $\deg P_{K,s,j} \le K-s-j$, *and also* $\deg P_{K,s,j} \le [(K-s)/2]$. *Moreover, for $0 \le j \le k-s$, the leading term of $P_{2k,2s,j}(h)$ is*

$$(-1)^{k-s-j}\binom{k-s}{j}\binom{k}{s}\frac{1\cdot 3\cdots(2k-1)}{1\cdot 2\cdots(2s-1)}h^{k-s}, \tag{49}$$

and the leading term of $P_{2k+1,2s+1,j}(h)$ is

$$(-1)^{k-s+j}\binom{k-s}{j}\binom{k}{s}\frac{1\cdot 3\cdots(2k+1)}{1\cdot 3\cdots(2s+1)}h^{k-s}\,. \tag{50}$$

Proof. The first assertion is immediate from (39), since each term in the sum has degree $\le K-s-j$. The second assertion follows easily from (41) by induction on K.

We prove the last two assertions in a single induction on K, with the understanding that K and s have the same parity. Since $P_{0,0,0}(h) = 1$ and $P_{1,1,0}(h) = 1$, we have the basis for the induction. Suppose that (49) and (50) hold for all $K \le 2k$, and we wish to prove it for $K = 2k+1$. By (41) we see that

$$\begin{aligned} P_{2k+1,2s+1,j}(h) = (j-2s-2k-2)P_{2k,2s+1,j}(h) + P_{2k,2s,j}(h) + (2k-j+1)P_{2k,2s+1,j}(h) \\ - 2hkP_{2k-1,2s+1,j}(h) + 2hkP_{2k-1,2s+1,j-1}(h)\,. \end{aligned}$$

Here the first and third terms on the right each have degree $\le k-s-1$. By the inductive hypothesis, the other three terms on the right hand side have degree $k-s$, and the combined leading coefficient is

$$\begin{aligned} &(-1)^{k-s-j}\binom{k-s}{j}\binom{k}{s}\frac{1\cdots(2k-1)}{1\cdots(2s-1)} \\ +\;&(-1)^{k-s-j}\binom{k-1-s}{j}\binom{k-1}{s}\frac{1\cdots(2k-1)}{1\cdots(2s+1)}2k \\ +\;&(-1)^{k-s-j}\binom{k-1-s}{j-1}\binom{k-1}{s}\frac{1\cdots(2k-1)}{1\cdots(2s+1)}2k\,. \end{aligned}$$

Since $\binom{k-1-s}{j} + \binom{k-1-s}{j-1} = \binom{k-s}{j}$, the last two terms combine to form a single term, and we see that the above is

$$(-1)^{k-s-j}\binom{k-s}{j}\frac{1\cdots(2k-1)}{1\cdots(2s+1)}\left(\binom{k}{s}(2s+1)+\binom{k-1}{s}2k\right).$$

Since $\binom{k-1}{s}k = \binom{k}{s}(k-s)$, the quantity inside the large parentheses is $\binom{k}{s}(2k+1)$, which gives the stated formula.

Now suppose that (49) and (50) hold for all $K \le 2k+1$, and we wish to prove it for $K = 2k+2$. By (41) we see that

$$\begin{aligned} P_{2k+2,2s,j}(h) = (j-2s-2k-1)P_{2k+1,2s,j}(h) + P_{2k+1,2s-1,j}(h) + (2k+2-j)P_{2k+1,2s,j-1}(h) \\ - (2k+1)hP_{2k,2s,j}(h) + (2k+1)hP_{2k,2s,j-1}(h)\,. \end{aligned}$$

Here the first and third terms on the right hand side each has degree $\le k-s$, while by the inductive hypothesis the other three terms each has degree $k-s+1$, with combined leading coefficient

$$\begin{aligned} &(-1)^{k-s-j+1}\binom{k-s+1}{j}\binom{k}{s-1}\frac{1\cdots 2k+1}{1\cdots(2s-1)} \\ &+ (-1)^{k-s-j+1}\binom{k-s}{j}\binom{k}{s}\frac{1\cdots(2k+1)}{1\cdots(2s-1)} \\ &+ (-1)^{k-s-j+1}\binom{k-s}{j-1}\binom{k}{s}\frac{1\cdots(2k+1)}{1\cdots(2s-1)}\,. \end{aligned}$$

Since $\binom{k-s}{j} + \binom{k-s}{j-1} = \binom{k-s+1}{j}$, the second and third terms combine to form a single term. Also, since $\binom{k}{s-1} + \binom{k}{s} = \binom{k+1}{s}$, we conclude that the above is

$$(-1)^{k+1-s-j}\binom{k+1-s}{j}\binom{k+1}{s}\frac{1\cdots(2k+1)}{1\cdots(2s-1)},$$

as desired.

On comparing (17) with (39), we find that

$$\mu_K(h,P) \;=\; \sum_{j=0}^{K} P^{K-j}P_{K,0,j}(h)\,. \tag{51}$$

In particular, since $\mu_K(h,1) = 0$ for $K \ge 1$, it follows that

$$\sum_{j=0}^{K} P_{K,0,j}(h) \;=\; 0\,. \tag{52}$$

The case $s = 0$ of (41) can be recovered from Romanovsky's recurrence (24). It would be interesting to know if there is a proof of the general case of (41) via generating functions.

6. Completion of the argument

We return to the expression $\mathcal{M}_K(h,P)$ in (38) that forms our best guess as to the size of the moment

$$\int_1^X (\psi(x+h)-\psi(x)-h)^K\,dx\,.$$

¿From Theorem 2 we see that $\deg P_{K,s,j} < (K-s)/2$ if $j > (K-s)/2$. Thus by (40) we see that the summands in (38) are

$$O\Big((h\log h)^{s/2}h^{[(K-s)/2]}N(\log N)^{(K-s)/2}\Big) \;=\; O\Big(N(h\log N)^{K/2}\Big)\,.$$

Moreover, this estimate can be improved by $h^{1/2}$ when K and s have opposite parity, and by (40) can be improved by $h^{1/(7s)-\varepsilon}$ when s is odd. Thus if K is odd, then

$$\mathcal{M}_K(h,P) \;=\; O\Big(Nh^{K/2-\delta}(\log N)^{K/2}\Big) \tag{53}$$

when $h \ge N^(\varepsilon)$. Suppose now that K is even. Terms of the magnitude $\ge Nh^{K/2}$ occur only when s is even and $j \le (K-s)/2$. In addition any term of $P_{K,s,j}$ other than the leading term will contribute an amount that is smaller by a factor of at least h. Thus by (49) we see that $\mathcal{M}_{2k}(h,P)$ is approximately

$$\sum_{s=0}^{k}\mu_{2s}(-h\log h+Ah)^s\sum_{j=0}^{k-s}I_j(N)(-1)^{k-s-j}\binom{k-s}{j}\binom{k}{s}\frac{\mu_{2k}}{\mu_{2s}}h^{k-s}$$

$$=\mu_{2k}h^k\int_1^N\sum_{s=0}^{k}\binom{k}{s}\sum_{j=0}^{k-s}\binom{k-s}{j}(-1)^{k-s-j}(-\log h+A)^s(\log u)^j\,du$$

where $\mu_{2k}=1\cdot 3\cdots(2k-1)$ is given by (26). By the trinomial theorem, the above is

$$=\mu_{2k}h^k\int_1^N(\log u-\log h+A-1)^k\,du\,.$$

That is,

$$\mathcal{M}_{2k}(h,P) \;=\mu_{2k}h^k\int_1^N(\log u-\log h+B)^k\,du+O\big(Nh^{k-\delta}\big) \tag{54}$$

when $N^\varepsilon \le h \le N$, where $B=A-1=1-C_0-\log 2\pi$.

It is instructive to note how the above can be interpreted in terms of zeros of the Riemann zeta function. It is classical that

$$\psi(x) \;=\; x-\sum_\rho\frac{x^\rho}{\rho} \;-\; \log 2\pi \;-\; \frac{1}{2}\log(1-1/x^2)$$

for $x > 1$. Here $\rho = \beta + i\gamma$ denotes the generic non-trivial zero of the zeta function. We assume RH, which is to say that $\beta = 1/2$ for all ρ, difference the above, and divide by $x^{1/2}$ to see that

$$\frac{\psi(x+h) - \psi(x) - h}{x^{1/2}} \;=\; -\sum_{\rho} \frac{(x+h)^{i\gamma} - x^{i\gamma}}{\rho} \;+\; O\big(x^{-5/2}\big) \;+\; O(h(\log x)^2/x)\,.$$

Here the sum over ρ is approximately

$$\sum_{\gamma>0} w(\gamma) \cos \frac{\gamma}{2} \log(x(x+h)) \tag{55}$$

where

$$w(u) \;=\; \frac{-4 \sin(\frac{u}{2} \log(1 + h/x))}{u}\,.$$

We suppose that this is distributed like

$$\sum_{\gamma>0} w(\gamma) \cos \theta_\gamma$$

where the θ_γ are independent random variables, each one uniformly distributed on $[0, 2\pi]$. Thus by the Central Limit Theorem, the distribution is approximately normal with variance

$$\frac{1}{2} \sum_{\gamma>0} w(\gamma)^2\,.$$

Let $N(T)$ denote the number of zeros of the zeta function with $0 < \gamma \le T$. Then

$$N(T) \;=\; \frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi} + \frac{7}{8} + O(\log T),$$

so

$$\begin{aligned} \frac{1}{2} \sum_{\gamma>0} w(\gamma)^2 &= \frac{1}{2} \int_0^\infty w(t)^2 \, dN(t) \\ &= \frac{1}{2} \int_0^\infty w(t)^2 \log \frac{t}{2\pi} \, dt + O\big(h^2 x^{-2} \log x/h\big)\,. \end{aligned}$$

Put $\delta = \log(1 + h/x)$. Then

$$\int_0^\infty w(t)^2 \log \frac{t}{2\pi} \, dt = \frac{2\delta}{\pi} \int_0^\infty \Big(\frac{\sin u}{u}\Big)^2 \, du + \frac{2\delta \log \pi\delta}{\pi} \int_0^\infty \Big(\frac{\sin u}{u}\Big)^2 \log u \, du\,.$$

Here the first integral is $\pi/2$, and the second one is $\frac{1}{2}\pi(1 - C_0 - \log 2)$, so we are led to expect that the distribution of $\psi(x+h) - \psi(x) - h$ is approximately normal with variance

approximately $h(\log x/h + B)$, which is precisely what is suggested by (54). In other words, we expect that the sum (55) is distributed in the same way that it would be if it were a sum of independent random variables.

Montgomery & Soundararajan [14] observed that the suggestion that (55) should be distributed like a sum of independent reandom variables is also predicted by random matrix theory. Let $U(N)$ denote the classical group of $N \times N$ unitary matrices. Rains [17] has shown that if A is Haar-distributed in $U(N)$ and M is an integer, $|M| > N$, then the eigenvalues of A^M are exactly uniformly distributed. As a consequence, the distribution of $\Re \operatorname{trace} A^M$ is exactly the same as the distribution of

$$\sum_{n=1}^{N} \cos(2\pi X_n)$$

where the X_n are independent random variables, each one uniformly distributed on $[0, 1]$.

References

1. R. P. Brent, *Irregularities in the distribution of primes and twin primes*, Math. Comp. **29** (1975), 43–56; Correction **30** (1976), 198.
2. H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. **2** (1936), 23–46.
3. P. J. Forrester & A. M. Odlyzko, *Gaussian unitarz ensemble eigenvalues and Riemann ζ function zeros: a nonlinear equation for a new statistic*, Phys. Rev. E (3) **54** (1996), R4493–R4495.
4. J. B. Friedlander & D. A. Goldston, *Some singular series averages and the distribution of Goldbach numbers in short intervals*, Illinois J. Math. **39** (1995), 158–180.
5. P. X. Gallagher, *On the distribution of primes in short intervals*, Mathematika **23** (1976), 4–9; Corrigendum **28** (1981), 86.
6. D. A. Goldston & H. L. Montgomery, *On pair correlations of zeros and primes in short intervals*, Analytic Number Theory and Diophantine Problems, Stillwater, OK, July 1984 (A. C. Adolphson, J. B. Conrey, A. Ghosh, R. I. Yager, eds.), Prog. Math. 70, Birkhäuser, Boston, 1987, pp. 183–203.
7. A. Granville, *Harald Cramér and the distribution of prime numbers*, Scand. Actuarial J. **1** (1995), 12–28.
8. G. H. Hardy & J. E. Littlewood, *Some problems of Partitio Numerorum (III): On the expression of a number as a sum of primes*, Acta Math. **44** (1922), 1–70.
9. M. Hausman & H. N. Shapiro, *On the mean square distribution of primitive roots of unity*, Comm. Pure App. Math. **26** (1973), 539–547.
10. H. Maier, *primes in short intervals*, Michigan Math. J. **32** (1985), 221–225.
11. H. L. Montgomery, *The pair correlation of zeros of the zeta function*, Analytic Number Theory, St. Louis, 1972, Proc. Sympos. Pure Math. 24, Amer. Math. Soc., Providence, 1973, pp. 181–193.
12. ———, *The zeta function and prime numbers*, Proc. Queen's Number Theory Conference (Kingston, 1979), Queen's Papers in Pure and Appl. Math. 54, Queen's Univ., Kingston, 1980, pp. 1–31.
13. H. L. Montgomery & K. Soundararajan, *Beyond pair correlation*, Paul Erdős and his Mathematics I, Mathematical Studies 11, Bolyai Society, Budapest, 2002, pp. 507–514.
14. ———, *Primes in short intervals*, Comm. Math. Phys. **252** (2004), 589–617.
15. H. L. Montgomery & R. C. Vaughan, *On the distribution of reduced residues*, Annals of Math. **123** (1986), 311–333.
16. A. M. Odlyzko, *On the distribution of spacings between zeros of the zeta function*, Math. Comp. **48** (1987), 273–308.

17. E. M. Rains, *High powers of random elements of compact Lie groups*, Probab. Theory Related Fields **107** (1997), 219–241.
18. V. Romanovsky, *Note on the moments of a binomial* $(p+q)^n$ *about its mean*, Biometrika **15** (1923), 410.
19. Z. Rudnick & P. Sarnak, *The n-level correlations of zeros of the zeta function*, C. R. Acad. Sci. Sér. I Math. **319** (1994), 1027–1032.
20. ———, *Zeros of principal L-functions and random matrix theory. A celebration of John F. Nath, Jr.*, Duke Math. J. **81** (1996), 269–322.
21. F. Spitzer, *A combinatorial lemma and its applications to probability theory*, Trans. Amer. Math. Soc. **82** (1956), 323–339.
22. G. Tenenbaum & M. Mendès France, *The Prime Numbers and their Distribution*, Student Math. Library 6, Amer. Math. Soc., Providence, 2000.

Department of Mathematics, University of Michigan, Ann Arbor MI 48109–1043, USA
E-mail address: hlm@umich.edu

Received on April 10, 2009
Accepted on May 5, 2010