Why does thinking about the primes lead naturally to inventing the Riemann zeta function and discovering the product formula?

A standard answer to this question is to "observe" Euler's product formula, and that it gives a new proof that there are infinitely many primes and more besides. Of course, once one has seen this simple argument, one cannot doubt that the zeta function is intimately related to the sequence of primes, but one can still ask: is the definition of the zeta function a brilliant trick, or is it bound to emerge sooner or later if people start thinking about the distribution of the primes?

I know nothing about how Euler actually thought of the zeta function. Perhaps he was so good at algebraic manipulations that he discovered it just by messing around. What I shall argue here is that it *was* inevitable that the function would be defined. This is hardly surprising, and my argument is certainly not original - any expert on zeta functions would be able to give a good explanation of why they are the 'obvious' thing to consider. However, my experience is that such explanations, while they exist in the minds of those who truly understand a subject, rarely make it into textbooks. So, if you know all about analytic number theory, what follows will be old hat. If you are an interested beginner, then I hope it will offer something.

Suppose, then, that we wish to understand about the distribution of the primes. It is natural to try to find some simple formula or algorithm for generating the sequence of primes, but if one plays around for a while, one soon becomes convinced that this is very difficult, and probably not even possible. There is a standard mathematical response to this situation:

> *If there seems to be no hope of working something out exactly, then try to estimate it instead.*

In this case, one quantity we might wish to evaluate, or at least estimate, is the size of the n^{th} largest prime, or equivalently the number of primes up to m.

Even this looks difficult. However, before we start, it is a good idea to make a few guesses. Here is another very standard mathematical reaction:

> *Before tackling a problem, guess the answer by looking at data, or using vague, non-rigorous, half-formed arguments.*

For example, working out and then looking at tables of the primes led Gauss to conjecture the prime number theorem - quite an achievement in the days before computers. However, even by looking at a list of all primes up to a thousand, one notices that, although they are distributed in a somewhat random way, they are reasonably *uniform.* That is, there do not tend to be enormous gaps between consecutive primes, and on average the density seems to go down steadily.

There comes a moment with any problem where one has to ask two questions.

> *What do I know for certain, and what techniques are there for showing what I wish to prove?*

Roughly speaking, these questions focus on the beginning and end respectively of some imagined, hoped-for proof. When it comes to the primes, we find that we do not have a good feeling for which numbers are primes, but we do know of a very interesting property that they have - the fundamental theorem of arithmetic. As for what we are trying to prove, some sort of uniform distribution, it suggests the use of Fourier analysis. (If it does not suggest to you the use of Fourier analysis, then you should read another of my essays, which unfortunately doesn't exist yet, "What is natural about Fourier analysis?")

Have we gained anything from these thoughts? At first sight it would appear that we haven't, because the fundamental theorem of arithmetic is a very algebraic statement. It is hard to see how it could have much to do with Fourier analysis. Since it is rather remarkable that there is a close connection between the two, let me say at this point that I am not claiming that the idea of connecting the two should be one of the first ideas to occur to any half-intelligent person. Indeed, I have ignored another handle one appears to have on the primes - that they can be produced by the sieve of Erastosthenes. This suggests a multitude of approaches to thinking about the primes. For example, events of the form 'x is divisible by p' appear to be approximately independent for randomly chosen x in some interval and small primes p. This raises the hope that there might be a direct, probabilistic approach to the distribution of primes. However, getting such an approach to work is very hard indeed, so it is not unreasonable to suppose that somebody who goes down that route will eventually be assaulted by the following thought:

> *I am having great difficulty with this idea. Is there anything else I could try?*

Returning to the connection between the fundamental theorem of arithmetic and Fourier analysis, let us remember two things about the latter:

1. *Information about the distribution of a set is often encapsulated by the Fourier*

coefficients of its characteristic function.

2. *One of the great advantages of the Fourier transform is that it converts convolution, a complicated operation, into pointwise multiplication, a simple one.*

The first of these principles suggests that we should be interested in the characteristic function of the primes. Unfortunately, this function is too large to have a Fourier transform, but for now let us be guided by another useful habit:

Work out what's going on first and worry about convergence later.

The second principle tells us that Fourier analysis may be helpful if we can express the fundamental theorem of arithmetic in terms of convolutions. Can we?

Let us mess around for a bit. Write P for the characteristic function of the primes (so $P(x) = 1$ if x is a prime and 0 otherwise.) What is $P * P$, the convolution of P with itself? By the definition of convolution, $P * P(x) = \sum_{y=1}^{x} P(y)P(x-y) = \sum_{p+q=x} 1$, where the second sum is over primes p and q. In other words, $P * P(x)$ is the number of ways of writing x as a sum of two primes.

This is not much use, since the fundamental theorem of arithmetic is about products and not sums. Should we therefore give up? Of course not. We are in the following situation.

A. We wish to express the fundamental theorem of arithmetic as a statement about convolutions.

B. $P * P(x)$ is the number of ways of writing x as a sum of two primes.

C. If the word "sum" had been "product" we would have been much happier, because the fundamental theorem of arithmetic tells us exactly how many ways there are of writing any given number x as a product of two primes: it is 0 if x is not a product of two primes, 1 if $x = p^2$ for some prime p and 2 if $x = pq$ for distinct primes p and q.

It is now difficult not to think

D1. Therefore, I would like somehow to convert sums into products.

Alternatively, one could think

D2. Therefore, I would like to rephrase the fundamental theorem of arithmetic as a statement about sums.

As it happens, one can pursue either thought D1 or thought D2, and the results will be more or less equivalent. I shall go for D2 here. How does one turn products into sums?

One uses logarithms. (Even if one didn't know this, one might well invent it since it is basically a unique process.)

Here then is a different way of writing the fundamental theorem of arithmetic. If x is the logarithm of a positive integer, then there is exactly one way of writing $x = \sum_{i=1}^{r} a_i \log p_i$, where $a_1, \ldots, a_r$ are positive integers and $p_1, \ldots, p_r$ are primes with $p_1 < \ldots < p_r$.

There now seems to be some chance of writing this as a statement about convolutions. Let Q be the characteristic function of $\log P$, so $Q(x) = 1$ if x is the logarithm of a prime and 0 otherwise, and let us calculate $Q * Q$. (Strictly speaking, Q is perhaps better thought of as an atomic measure, where each $\log p$ has measure 1. At any rate, by $Q * Q(x)$ I mean the number of ways of writing x as $a + b$ with a and b in Q.) We find that $Q * Q(\log x)$ is the number of ways of writing $\log x$ as $\log p + \log q$, where p and q are primes. This is the number of ways of factorizing x as a product of two primes - but where the order makes a difference.

This is a little unfortunate - $Q * Q(\log 6) = 2$, whereas $Q * Q(\log 9) = 1$, for example. The fundamental theorem of arithmetic is concerned with the number of ways of writing an integer as a product of powers of primes, where the primes themselves are in increasing order. This, even after taking logarithms, is not a convolution after all.

However, another golden rule of mathematical research is

> *Never be put off at an early stage by technical problems. If necessary just ignore them.*

Let us therefore ignore this problem, and pretend that $Q * Q(\log x) = 2$ whenever x is a product of two primes. If we want to look at products of three primes, we should obviously consider $Q * Q * Q(\log x)$. This works out as the number of ways of writing $\log x$ as $\log p + \log q + \log r$, with p, q and r all prime, which is 6 if $x = pqr$ with p, q and r all distinct, 3 if $x = p^2 q$ and 1 if $x = p^3$.

A pattern is emerging. If x is a product of r distinct primes, then the s-fold convolution of Q evaluated at $\log x$ will be zero if $s \neq r$ and $r!$ if $s = r$. It follows that for any x without repeated factors, we have

$$Q(\log x) + \frac{1}{2!} Q * Q(\log x) + \frac{1}{3!} Q * Q * Q(\log x) + \ldots = 1 \ .$$

This is such a suggestive formula that we must surely be on to something. Indeed, the

Fourier transform of the function

$$Q + \frac{Q*Q}{2!} + \frac{Q*Q*Q}{3!} + \dots$$

is

$$\hat{Q} + \frac{\hat{Q}^2}{2!} + \frac{\hat{Q}^3}{3!} + \dots = \exp \hat{Q} - 1 \ .$$

Let us try to "add 1" to both sides. For this we want to know what it means to convolve Q with itself zero times. Since the function that is 1 at 0 and 0 otherwise acts as the identity under convolution, it is the obvious candidate. Let us call this function Δ_1. Then the formula

$$\Delta_1(\log x) + Q(\log x) + \frac{1}{2!}Q*Q(\log x) + \frac{1}{3!}Q*Q*Q(\log x) + \dots = 1$$

is still valid for integers x with no repeated factors, but now it is valid when $x = 1$ as well. What is more, the Fourier transform of Δ_1 is identically 1, so the Fourier transform of

$$\Delta_1(\log x) + Q(\log x) + \frac{1}{2!}Q*Q(\log x) + \frac{1}{3!}Q*Q*Q(\log x) + \dots$$

is $\exp(\hat{Q})$. What could we deduce if we did not have the irritating problem of integers with repeated factors? We would have

$$\Delta_1(\log x) + Q(\log x) + \frac{1}{2!}Q*Q(\log x) + \frac{1}{3!}Q*Q*Q(\log x) + \dots = 1$$

for every positive integer x, and therefore that

$$\Delta_1 + Q + \frac{Q*Q}{2!} + \frac{Q*Q*Q}{3!} + \dots$$

was the characteristic function of the set $\{\log n : n \in \mathbb{N}\}$. Therefore, after taking the Fourier transform, we would find that $\exp \hat{Q}$ was the Fourier transform ϕ of the characteristic function of this set. It would be given by the formula

$$\phi(\sigma) = \sum_{n \in \mathbb{N}} \exp(i\sigma \log n) = \sum_{n \in N} n^{i\sigma} \ .$$

As the reader will no doubt have noticed, this is $\zeta(i\sigma)$, at least in a formal sense. However, perhaps now is the time to think a little about convergence. It is a definite drawback of our idea of using the Fourier transform is that we cannot apply it to the characteristic function of an infinite set. On the other hand, the questions we are asking about primes do not in fact concern all primes at once, so perhaps this is not an insuperable problem. Indeed, we could solve it at a stroke by simply considering all primes up to some large N and ignoring the rest.

Another mathematical instinct should make itself felt now:

> *If I just chop off all primes after N, then I am introducing an ugly boundary at N, which is likely to make all my calculations messy.*

Fortunately, there is another standard idea for dealing with a problem like this.

> *If I want to smooth off a boundary, then I must introduce weights. Chopping off all primes after N is the same as weighting them by zero. Instead, I ought to choose weights that decay more gradually.*

We are now forced to ask ourselves what properties we should be looking for in our weights. Once we have introduced them, we will have replaced the function Q, which is 01-valued, by a more general function, which will take non-zero values only when Q does. We now apply another old idea.

> *Make an arbitrary choice, do some calculations, and see if desirable properties emerge.*

Let us define a function R by letting $R(\log p) = w(p)$ when p is a prime, and $R(x) = 0$ otherwise. What could we calculate? An obvious choice is $R * R$. We find that

$$R * R(\log x) = \sum_{\log p + \log q = \log x} w(p)w(q)$$

which is $2w(p)w(q)$ if $x = pq$, $w(p)^2$ if $x = p^2$ and 0 otherwise. More generally, if $x = p_1 \dots p_r$ with all the p_i distinct primes, then the r-fold convolution of R, evaluated at $\log x$, will be $r!w(p_1)\dots w(p_r)$. Therefore, if we replace Q by R, we will be replacing the characteristic function of $\{\log n : n \in \mathbb{N}\}$ by the function f that takes n to the product of the weights associated with the prime factors of n. We want this function to decay gradually to zero, which means that all the weights $w(p)$ should be less than 1. It would also be very natural for f to decrease monotonically. If we ask for this, then there is no choice but to take $w(p) = p^{-s}$ for some $s > 0$, which means that $f(n) = n^{-s}$. If in addition we want to be sure that the Fourier transform of f exists, then we should take $s > 1$ in order to make it absolutely summable.

After introducing this sequence of weights, we can write another incorrect formula: $\exp \hat{R} = \phi_s$, where

$$\phi_s(\sigma) = \sum_{n\in\mathbb{N}} n^{-s} \exp(-i\sigma \log n) = \sum_{n\in\mathbb{N}} n^{-s-i\sigma} = \zeta(s+i\sigma) \ .$$

This formula is so beautiful and concise that most mathematicians will be guided by the following instinct.

> *If a beautiful formula is reasonably close to being true, then it ought to be possible to correct it so that it becomes true.*

What do we have to correct? Well, we could simply work out what $1+R+\frac{1}{2!}R*R+\frac{1}{3!}R*R*R+\ldots$ actually is and try to transform that, but then we would lose the very nice right hand side of our formula. A better approach is surely to try to alter the function R in order to make the formula true.

It will be easier to go back to Q. What can we do to 'correct' the definition of Q so that it will give the right answer for numbers with repeated factors? Let us write

$$E(x) = Q(\log x) + \frac{1}{2!}Q*Q(\log x) + \frac{1}{3!}Q*Q*Q(\log x) + \ldots$$

and focus on the problem that $E(4) = 1/2$, when we want it to be 1. Recall that this is because $Q*Q(\log 4)$ is the number of ways of writing $\log 4$ as $\log p + \log q$ where p and q are primes, and there is only one such way rather than 2. One thing we could do to sort this out is replace Q by a modified function Q' with the single difference that $Q'(\log 4) = 1/2$, and replace E by E' accordingly. It is easy to see that $E'(4) = 1$, so we have the right answer at 4, but are we messed up elsewhere? What, for example, is the value of $E'(12)$? Previously, we would have noted that 12 has the three prime factors 2, 2 and 3, so the only contribution to E comes from $Q*Q*Q$. Since there are 3 distinct ways of writing $12 = pqr$ with p, q and r prime, we would have calculated that $E(12) = 3/3! = 1/2$. Now, however, with $Q'(\log 4)$ set to $1/2$, we also have a contribution from $Q'*Q'$. Indeed, $Q'*Q'(12) = 1$, since we can write $\log 12$ as $\log 3 + \log 4$ or as $\log 4 + \log 3$ and each such expression contributes $1/2$. So, rather pleasantly, we discover that $E'(12) = 1$. Similarly, it can be checked that if n is an odd number with no repeated factors, then $E'(4n) = 1$.

It is now tempting to carry out further adjustments. For example, why not define Q'' by letting $Q''(\log x) = 1$ if x is prime and $1/2$ if x is the square of a prime? This is not a bad idea, and one can check that if E'' is the corresponding 'exponential convolution', then $E''(n) = 1$ if no prime occurs more than twice in the prime factorization of n.

One can then consider $E''(8)$. It turns out to equal $2/3$, so a further adjustment suggests itself, which is to set $Q'''(8) = 1/3$. Inspired by our experiences so far, we might also set $Q'''(3 \log p)$ to be $1/3$ for every prime p. Doing so produces an E''' which equals 1 on every n in which no prime appears more than three times.

It is now easy to guess the pattern. For each prime p and each integer r let $Q(r \log p) = 1/r$. (Note that I have now abandoned my original Q, and used the letter for the new improved function.) Let us calculate $E(n)$, where $n = p_1^{a_1} \dots p_k^{a_k}$.

At first, this seems a slightly tricky combinatorial problem, but the following observation simplifies it considerably. We are counting the number of ways of writing n as a product of prime powers, with each product weighted in a certain way. However, doing this is equivalent to writing $p_1^{a_1}$ as a product of prime powers, then $p_2^{a_2}$ and so on, and multiplying the corresponding weights. That is, $E(n) = E(p_1^{a_1})E(p_2^{a_2}) \dots E(p_k^{a_k})$. (Equivalently, the function E is *multiplicative* in the usual sense that $E(ab) = E(a)E(b)$ whenever a and b are coprime.) It is not hard to give a better proof of this statement than my vague sketch.

It therefore remains to calculate $E(p^a)$ for a prime p and positive integer a. Of course, we hope that the answer will be 1. However, it presents itself to us as the sum over all k-tuples $(a_1, \dots, a_k)$ such that $a_1 + \dots + a_k = a$ of $a_1^{-1} \dots a_k^{-1}/k!$. Now the sum over all those k-tuples should remind us of multiplication of power series. Indeed, if we look at the series $f(x) = x + x^2/2 + x^3/3 + \dots$ then what we have calculated is $1/k!$ times the coefficient of x^a in $f(x)^k$. Hence, summing over k gives us the coefficient of x^r in $\exp(f(x))$. But $f(x)$ is just $-\log(1-x)$, so $\exp(f(x)) = (1-x)^{-1} = 1 + x + x^2 + \dots$ and the relevant coefficient is 1, as desired. We have therefore proved that, with the modified definition of Q, it is indeed true that $E(n) = 1$ for all n.

Just to be clear about what exactly we have achieved with the modified function Q, let us repeat an earlier argument. Our new Q has the property that

$$\Delta_1(x) + Q(x) + \frac{1}{2!} Q * Q(\log x) + \frac{1}{3!} Q * Q * Q(\log x) + \dots = 1$$

whenever x is the logarithm of a positive integer. (For the previous version of Q we could say that only if the positive integer had no repeated factor.) On taking Fourier transforms, we find that

$$\exp(\hat{Q}) = \sum_{n\in\mathbb{N}} \exp(-i\sigma\log n) = \sum_{n\in\mathbb{N}} n^{-i\sigma} = \zeta(i\sigma)$$

at least formally. In order to deal with convergence problems, we can modify Q by introducing weights, just as we did with the previous, incorrect version of Q. Similar considerations to the ones before tell us that we should choose $s > 1$ and then, if $y = \log m$ for some prime power m, then we should set $R_s(y) = m^{-s}Q(y)$. Doing that, we find, for any positive integer n with $x = \log n$, that

$$\Delta_1(x) + R_s(x) + \frac{1}{2!}R_s * R_s(\log x) + \frac{1}{3!}R_s * R_s * R_s(\log x) + \ldots = n^{-s} .$$

This time, taking Fourier transforms gives us the (entirely correct at last) formula

$$\exp(\hat{R}_s) = \sum_{n\in\mathbb{N}} n^{-s}\exp(-i\sigma\log n) = \sum_{n\in\mathbb{N}} n^{-s-i\sigma} = \zeta(s+i\sigma) .$$

Having gone through a somewhat messy and disorganized argument (which is supposed to reflect the messy and disorganized way that mathematicians think) let us try to tidy it up. We shall find that the seemingly clever identity we have discovered can be demonstrated far more easily, but with the cost that it just appears out of the blue.

Recall that when we were calculating $E(\log n)$ for the modified function Q, we observed that E was multiplicative, and then calculated it for prime powers. Let us consider this part of the argument more carefully. What we did was to split Q into pieces and consider the pieces separately. Given a prime p, let Q_p be the function that takes $r\log p$ to $1/r$ and everything else to zero. Then $Q = Q_2 + Q_3 + Q_5 + \ldots$. Now when we calculate $E(p^a)$ for some prime power p^a, it is clear (by the fundamental theorem of arithmetic) that the only part of Q that makes any contribution is Q_p. Therefore, when we showed that $E(p^a) = 1$, we were showing that the convolution-exp of Q_p was the function that took the value 1 at every $r\log p$. We proved this by noticing that convolutions of Q_p were basically the same as multiplications of the power series of $-\log(1-x)$. Can we make this connection with logarithms more precise?

Well, one way to do it is to take Fourier transforms again. This means that we shall replace convolution-logs and convolution-exps with real logs and real exps. So what is the function $\hat{Q}_p$? It works out as

$$\hat{Q}_p(\sigma) = \sum_{a=1}^{\infty} a^{-1}\exp(-i\sigma a\log p) = \sum_{a=1}^{\infty} a^{-1}p^{-ai\sigma} = -\log(1-p^{-i\sigma}) .$$

(I have gone back to my habit of forgetting about convergence.) And how about the Fourier transform of

$$\Delta_1 + Q_p + \frac{1}{2!}Q_p * Q_p + \frac{1}{3!}Q_p * Q_p * Q_p + \ldots ?$$

This will be the exp of what we have just calculated, or in other words

$$(1 - p^{-i\sigma})^{-1} = 1 + p^{-i\sigma} + p^{-2i\sigma} + \ldots .$$

Now what happens when we consider not Q_p but Q? Well,

$$\hat{Q}(\sigma) = \sum_p \hat{Q}_p(\sigma) = -\sum_p \log(1 - p^{-i\sigma})$$

and

$$\exp(\hat{Q}(\sigma)) = \prod_p \exp(\hat{Q}_p(i\sigma)) = \prod_p (1 - p^{-i\sigma})^{-1} .$$

But we *also* know that $\exp(\hat{Q}(\sigma)) = \zeta(i\sigma)$, since that was what we constructed it to be. Hence, we have 'discovered' the product formula:

$$\zeta(i\sigma) = \sum_n n^{-i\sigma} = \prod_p (1 - p^{-i\sigma})^{-1}$$

at least for purely imaginary numbers. But now, when we put in the weights to make everything converge properly, we obtain the product formula for all numbers with real part greater than one. (I omit the routine details of this.)

Having discovered the product formula, it now becomes an easy matter to reverse the whole process and construct a function whose convolution-exponential is the characteristic function of the set $\{\log n : n \in \mathbb{N}\}$. Since the Fourier transform of this function is simply ζ, all we need to do is take the log of ζ and transform back. We are greatly helped in this task by the fact that ζ splits as a product of easily understood functions, since taking logarithms of a product is easy. So, just to spell it out, the steps are:

$$\zeta(s + i\sigma) = \prod_p (1 - p^{-s-i\sigma})^{-1} .$$

$$\log \zeta(s + i\sigma) = -\sum_p \log(1 - p^{-s-i\sigma}) .$$

But

$$\log(1-p^{-s+i\sigma}) = p^{-s-i\sigma} + \frac{1}{2}p^{-2s-2i\sigma} + \frac{1}{3}p^{-3s-3i\sigma} + \dots$$
$$= p^{-s}\exp(-i\sigma\log p) + \frac{1}{2}p^{-2s}\exp(-i\sigma.2\log p) + \dots$$

which one can recognise as the Fourier transform of the function $R_{s,p}$ which takes the value $\frac{1}{a}p^{-as}$ at $a\log p$. Hence, summing over all p we obtain a function R_s which takes the value $\frac{1}{a}p^{-as}$ at *any* number of the form $a\log p$ (rather than for one specific value of p). This function exp-convolves to the function that transforms to the function $\sigma \mapsto \zeta(s+i\sigma)$. But

$$\zeta(s+i\sigma) = \sum_{n\in\mathbb{N}} n^{-s}\exp(-i\sigma\log n)$$

which once again is directly written as a Fourier transform, this time of the function that takes the value n^{-s} at $\log n$. Wickedly setting $s=0$, we obtain the characterteristic function of $\{\log n : n \in \mathbb{N}\}$, while R_0 is the function Q that we so laboriously built earlier.

Postscript.

It might seem unfortunate that we have passed from our first attempt at Q, which was the characteristic function of the set of all logarithms of primes, to the modified function Q, which included prime powers, with appropriate weights. However, this is not a problem at all, since the density of prime powers is very small compared with that of primes. Therefore, if we could understand the distribution of the modified function Q, we would understand the distribution of the primes as well.

But does this not put us in a great position? We have a very simple set, namely $\mathbb{N}$, to which we have done a very natural process (take 'Fourier transforms' take the logarithm and transform back) to obtain a function whose distribution is more or less the same as that of the primes. Unfortunately, a great deal more work is needed to turn this promising situation into a proof of the prime number theorem. At some stage, I may write more on where the difficulties arise.