

# Feedback, Control, and the Distribution of Prime Numbers

SUSAN H. MARSHALL

Monmouth University  
West Long Branch, NJ 07764-1898  
smarshal@monmouth.edu

DONALD R. SMITH

Monmouth University  
West Long Branch, NJ 07764-1898

We explore the system of prime numbers from an unusual viewpoint, that of an applied mathematician employing mathematical modeling to study a natural phenomenon. We liken the prime numbers to a physical system that can be empirically observed and, based on our observations, we derive a mathematical model in the form of a differential equation to describe the system:

$$f'(x) = \frac{-f(x)f(\sqrt{x})}{2x}, \quad (1)$$

where  $f(x)$  represents the “density of primes at  $x$ .” The model predicts well-known results concerning the distribution of prime numbers and has been discovered at least twice previously. The model seems to have been forgotten by the number theory community, but the distribution of primes is mentioned as an application in the differential equation literature. (See, for example, [7, p. 237] or [16].) We endeavor to contribute a deeper understanding of the connection between equation (1) and number theory using the framework of feedback and control, a field within engineering.

Lord Cherwell, a prominent British physicist and scientific advisor to Churchill during World War II, discovered equation (1) in the context of prime numbers in about 1942. Cherwell did not publish his finding, but shared it with British mathematician and number theorist E. M. Wright. This led Wright to study a class of differential equations related to (1) in a series of papers spanning a decade [19, 20, 21, 22, 23, 24]. In the final paper of this series, Wright credits Cherwell with introducing him to the equations but does not explain how they are derived from number-theoretic concerns [24, p. 66].

Another discovery was made by G. Hoffman de Visme, a British electrical engineer, who published his work in a 1961 note in the *Mathematical Gazette* [12]. Somewhat fortuitously, this note was sent to Wright to referee, and in the next volume of the *Gazette* appears a note by Wright [25] discussing Cherwell’s contribution as well as some of Wright’s own work on the differential equation. Wright also recounts the story in his 1988 retrospective of his friendship and collaboration with Cherwell [26].

The repeated discoveries—all by applied mathematicians with a side interest in number theory—reveal that there is something quite natural about this approach, given the right perspective. The system of prime numbers exhibits randomness combined with deep structure. We propose that this can be understood by viewing the primes as a feedback and control system. We do not know whether Cherwell or Hoffman de Visme viewed the primes in this way; however, their derivation of equation (1) is essentially the same as ours.

Both Cherwell and Hoffman de Visme sought to use the model to explain why the prime number theorem holds. Wright's technical work on the differential equation is instrumental in seeing how the model predicts this famous theorem. Using Wright's work, we are also able to show that the model predicts a surprising result due to Littlewood. Both of these results require sophisticated techniques of proof. The model provides insight into why they follow from simple observations of the system of primes.

On the other hand, the model has limitations. As in any modeling process, we make simplifying assumptions about our system. For example, although the prime numbers are clearly a discrete phenomenon, we assume a smooth version of their density function so that we can use the tools of calculus. Moreover, the element of randomness within the system of prime numbers is not directly incorporated into our model. (Instead we consider "perturbed" solutions, a technique borrowed from physics.) The model does not capture all aspects of the system; for example, some implications of the model are in conflict with the Riemann Hypothesis. We explore these limitations further in the final section of this paper. Despite these limitations, the model appears to give at least a first-order understanding of the system of primes.

## Feedback and control systems

A system that uses measurements of its own behavior to adjust its future behavior is a *feedback and control system*.

For example, consider the system of a car and driver. As the driver notices that she is too far to the right, she steers left, in an attempt to keep the car on track. The driver is constantly encountering small "perturbations" (such as a pebble in the road) and we can imagine different reactions as the driver attempts to steer back to the ideal path:

- **Cool, calm, and collected:** This driver gets closer and closer to the line without actually crossing it.
- **Whimsical:** This driver crosses the line before she is able to steer back, and then repeatedly crosses the line in such a way that the distance from the line in successive crossings becomes smaller and smaller.
- **Panicky and Overreacting:** This driver also crosses the line before she is able to steer back, but then overreacts and yanks the wheel back and forth successively more violently. The successive crossings go farther and farther from the line.

Other reactions of the driver are possible, but the reactions listed above represent some typical behaviors of feedback and control systems. For example, a well-designed and maintained car suspension hitting a bump in the road behaves like our cool, calm, and collected driver, while a car with worn shocks resembles our whimsical driver. The panicky and overreacting driver is represented in examples such as the Tacoma Narrows Bridge that collapsed in 1940, or the audio feedback we hear when a microphone gets too close to a speaker.

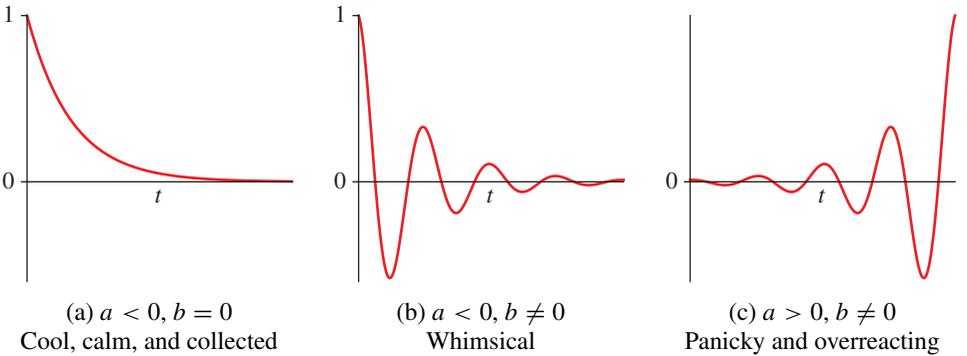
In our examples, we focused on the reaction of a driver to a single perturbation, even though an actual car is constantly bombarded by random, small changes. Despite the unpredictability of these perturbations, we claim that knowledge of the reaction to just one perturbation is enough to predict at least large-scale behavior of the system. For some systems, like the cool, calm, and collected driver or the whimsical driver, perturbations tend to die out, so that only the most recent perturbations affect the path and their cumulative effect is bounded. For other systems, such as the panicky and overreacting driver, perturbations result in explosively growing behavior that eventually

overwhelms other factors. Wild oscillations are inevitable, even if we cannot predict their exact magnitude or when they will occur.

Mathematically, feedback and control systems are often modeled as differential equations, sometimes involving delay (when time is needed to incorporate the feedback into the system). The solutions of the differential equations tell us about the nature of the system. For example, a solution might be of the form  $f(t) = e^{\lambda t}$ , where  $\lambda = a + bi$  is a complex number and  $t$  is a real variable. This function can be written as

$$f(t) = e^{\lambda t} = e^{at} (\cos bt + i \sin bt).$$

When  $\lambda$  is real (that is,  $b = 0$ ), the complex exponential degenerates to a real exponential. If  $b$  is nonzero, then the trigonometric functions lead to oscillatory behavior. Various choices for  $a$  and  $b$  lead to the typical behaviors for our car and driver, where we view  $f(t) = 0$  as the ideal path. FIGURE 1 shows these behaviors.



**Figure 1** Typical solutions to feedback and control systems

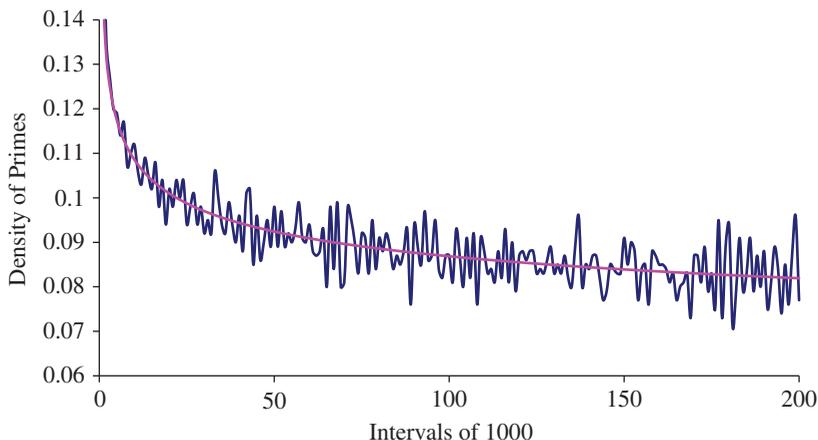
## Primes as a feedback and control system

Why would we suspect the primes of being a feedback and control system? Systems that are subject to randomness do not stay on track without some kind of feedback and control mechanism. When a car stays on the road despite many random perturbations, it is reasonable to infer that there is a driver making corrections.

The distribution of the prime numbers appears to have an element of randomness. Yet, the primes do “stay on track” in a very specific way. Gauss observed around 1792 that the density of primes appears to be inversely proportional to the natural logarithm (see, for example, [8, pp. 2–3] or [18]). The evidence is quite striking. Gauss tabulated the number of primes in intervals of length 1000, called *chiliads*, and if we plot the average density of the primes in successive chiliads, a beautiful logarithmic shape emerges. In FIGURE 2, the data points are connected and graphed against the function  $1/\ln x$ . We can see the feedback and control, as the actual density of primes in these intervals jumps above and below the ideal path line of  $1/\ln x$ .

Gauss’s estimate for the density of primes can be used to obtain an approximation for the total number of primes less than or equal to  $x$ , traditionally denoted  $\pi(x)$ . A first approximation is found by simply multiplying by  $x$  and comparing  $\pi(x)$  to  $x/\ln x$ . A more sophisticated approximation is found by integrating to obtain the *logarithmic integral*,

$$\text{Li}(x) := \int_2^x \frac{1}{\ln t} dt.$$



**Figure 2** Average density of primes in successive chiliads versus  $1/\ln x$

As seen in TABLE 1, these estimates compare favorably to the actual number of primes, with  $\text{Li}(x)$  faring a bit better.

TABLE 1: The total number of primes compared to Gauss's estimates

$x$	$\pi(x)$	$x/\ln x$	$\text{Li}(x)$
10	4	4.3	6.2
$10^2$	25	21.7	30.1
$10^3$	168	144.8	177.6
$10^4$	1 229	1 085.7	1 246.1
$10^5$	9 592	8 685.9	9 629.8
$10^6$	78 498	72 382.4	78 627.6
$10^7$	664 579	620 420.7	664 918.4
$10^8$	5 761 455	5 428 681.0	5 762 209.4
$10^9$	50 847 534	48 254 942.4	50 849 235.0
$10^{10}$	455 052 511	434 294 481.9	455 055 614.6

A century after Gauss's observations, it was shown that both of these estimates are "good" in the sense that their relative errors,

$$\frac{\pi(x) - x/\ln x}{\pi(x)} \quad \text{and} \quad \frac{\pi(x) - \text{Li}(x)}{\pi(x)},$$

both tend to 0 as  $x \rightarrow \infty$ . This relationship is described by saying that these functions are *asymptotic* to one another, written symbolically as  $\pi(x) \sim x/\ln x \sim \text{Li}(x)$ . This connection between the primes and the natural logarithm is known as the *prime number theorem*. We prefer to state the theorem in terms of the density of primes. The

relationship  $\pi(x) \sim x/\ln x$  implies that the average density  $\pi(x)/x$  is asymptotic to  $1/\ln x$ .

**THE PRIME NUMBER THEOREM.** *Let  $x > 0$  be a real number. The average density of primes less than  $x$  is asymptotic to  $1/\ln x$ .*

The absolute error,  $\pi(x) - \text{Li}(x)$ , behaves rather differently than the relative error. Despite numerical evidence showing  $\text{Li}(x)$  as an overestimate for  $\pi(x)$ , Littlewood [15] proved in 1914 that the difference between  $\pi(x)$  and  $\text{Li}(x)$  changes sign infinitely often. (In fact, all current numerical evidence shows  $\text{Li}(x)$  as an overestimate [14].) More precisely, he showed the existence of a positive constant  $k$  such that we can find arbitrarily large values of  $x$  that satisfy

$$\pi(x) - \text{Li}(x) > \frac{k\sqrt{x} \ln \ln x}{\ln x} \quad (2)$$

and arbitrarily large values of  $x$  that satisfy

$$\pi(x) - \text{Li}(x) < \frac{-k\sqrt{x} \ln \ln x}{\ln x}. \quad (3)$$

The function on the right side of (2) is increasing (slowly) for all sufficiently large positive values of  $x$ . Thus, Littlewood's theorem tells us not only that the difference changes sign infinitely often, but it also indicates that there are arbitrarily large oscillations between the values of the two functions.

**LITTLEWOOD'S THEOREM.** *The absolute value of  $\pi(x) - \text{Li}(x)$  is unbounded and the sign of  $\pi(x) - \text{Li}(x)$  changes infinitely often.*

The apparent randomness of the primes, coupled with the strong structure implied by both the prime number theorem and Littlewood's theorem, suggest a feedback and control mechanism at work. Furthermore, the behavior of the primes fits the pattern of typical feedback and control systems, reinforcing the idea that the primes are governed by feedback and control. The density of primes behaves like the driver who gets successively closer to the ideal path line  $1/\ln x$ . However, when we consider the difference between the total number of primes and *its* ideal path line  $\text{Li}(x)$ , Littlewood's theorem is reminiscent of our panicky and overreacting driver.

## Modeling the feedback

We would like to understand how the feedback and control mechanism of the primes works, in order to model it. What is it about the system of prime numbers that seems to notice that the density of primes is "off-track" and steers it back? The essential idea is this: *"Too many" primes in one interval tend to decrease the likelihood of primes in higher intervals while "too few" increase the likelihood of future primes.*

The *sieve of Eratosthenes* illustrates this intuitive idea of feedback. The sieve is a tool for finding prime numbers less than a given number. It begins with a list of candidate primes,  $\{2, 3, 4, \dots\}$ . At each iteration, the smallest number  $p$  on the list that is not yet marked prime must be prime, since it cannot be written as a multiple of numbers smaller than it. We mark  $p$  as prime. Then, we remove all larger multiples of  $p$  from the list, and move on to the next iteration.

When we begin the sieving process, we consider all numbers as potentially prime. When we remove the multiples of a prime  $p$ , we decrease the density of potential primes. Since this process removes  $1/p$  of the remaining potential primes, it changes

the density by a factor of  $1 - 1/p$ . For example, sieving by  $p = 2$  removes the even composite numbers and decreases the density of potential primes by  $1/2$ , and then sieving by  $p = 3$  reduces the density by an additional  $1/3$ . Furthermore, a prime  $p$  only begins to affect the density of primes after its square, since smaller multiples of  $p$  have been removed at earlier iterations of the sieve.

Armed with these observations, we are now ready to derive a differential equation to model the density of primes at  $x$ . Let  $f(x)$  be a differentiable function that approximates the density of primes, in the sense that it satisfies the properties we discovered from the sieve of Eratosthenes. The function  $f(x)$  should be interpreted as a point density rather than an average density—that is, it approximates the density of primes in the immediate neighborhood of  $x$ , not the cumulative density in the interval from 2 to  $x$ .

To derive the differential equation, we will estimate the change in our density function over a certain interval in two different ways. One of the key facts we learned from the sieve of Eratosthenes is that a prime begins to affect the density only after its square. Thus, we consider the following two intervals.

- **Interval A** is the interval on the real line given by endpoints  $x$  and  $x + dx$ .
- **Interval B** is the square of Interval A, and thus has endpoints  $x^2$  and  $(x + dx)^2$ .

As usual, the quantity  $dx$  is meant to be an infinitesimal quantity, or at least very small compared to  $x$ . We will estimate the change in density on Interval B; that is, we will approximate the quantity  $f((x + dx)^2) - f(x^2)$ .

We consider only the effect of primes from Interval A on the density along Interval B. Each prime in Interval A changes the density on Interval B by a factor of approximately  $1 - 1/x$ , since all primes in the small Interval A are roughly of size  $x$ . Thus, each prime subtracts  $f(x^2)/x$  from the density. There are about  $f(x)dx$  primes in Interval A, and so our first approximation of the change in density on Interval B is

$$f((x + dx)^2) - f(x^2) \approx \frac{-f(x^2)f(x)dx}{x}. \quad (4)$$

We can also estimate the change in density on Interval B using derivatives. The change in density is approximately the derivative of  $f$  at  $x^2$  multiplied by the length of Interval B. The length of the interval is  $(x + dx)^2 - x^2 = 2x dx + (dx)^2$ . We may safely ignore the second-order term, and our second approximation of the change in density on Interval B is

$$f((x + dx)^2) - f(x^2) \approx f'(x^2) \cdot 2x dx. \quad (5)$$

We equate the right-hand sides of (4) and (5), and solve for the derivative to find

$$f'(x^2) = \frac{-f(x^2)f(x)}{2x^2}.$$

Changing variables so that our equation is in terms of  $x$ , we recover equation (1) from the introduction:

$$f'(x) = \frac{-f(x)f(\sqrt{x})}{2x}.$$

This equation is our model of the system of primes. The function  $f(x)$  represents the density of primes at  $x$  and we clearly see the dependence of the change in density on the density at  $x$ , the density at  $\sqrt{x}$ , and on  $x$  itself.

As mentioned in the introduction, a previous derivation of the above differential equation was given by G. Hoffman de Visme in 1961 [12]. Mathematically, his derivation is essentially the same as ours, yet his is often described as “probabilistic” in the literature. The term probabilistic arises because we can think of the density correction factor  $1 - 1/p$  as a probability, namely the probability that a number is not divisible by the prime  $p$ .

## The prime number theorem

*Proofs of the prime number theorem are extremely hard to follow, and leave the impression, at least among amateurs, that the essential property of prime numbers, namely their primeness, plays very little part in the argument.*

—G. Hoffman de Visme (1961)

An intuitive justification of the prime number theorem was the motivation for Hoffman de Visme’s work. His thoughts are summarized in the first sentence of his *Gazette* article (quoted above) and Wright attributes this motivation to Chermak’s work as well [25]. The original proofs of the prime number theorem, independently given by Hadamard and de la Vallée-Poussin in 1896, were inspired by Riemann’s famous memoir of 1859 and involved complex analysis and the Riemann zeta-function. Even among nonamateurs, the complex-analytic proofs of this theorem were unsatisfying and an “elementary” (not to be confused with easy) proof was sought. Success came with the work of Erdős [9] and Selberg [17], who gave independent elementary proofs in 1949. (For an historical survey of these achievements, see [1, 10].) Our modeling approach illuminates the prime number theorem from yet another angle.

The prime number theorem tells us that the average density of primes less than  $x$  is asymptotic to the function  $1/\ln x$ . The relationship  $\pi(x) \sim \text{Li}(x)$  suggests that the density at  $x$  is also asymptotic to  $1/\ln x$ . (The computations with Gauss’s chiliads illustrated in FIGURE 2 further support this assertion.) Thus, we expect  $f(x) = 1/\ln x$  to be a solution to our differential equation (1). A quick computation shows, in fact, that it is.

To fully validate our model against the prime number theorem, however, we need to go a bit further. We do not expect that the true density of primes is *equal* to  $1/\ln x$ —rather, we expect  $1/\ln x$  to be the *ideal path* of the true density. If this density gets off-track, we want to know that the feedback and control mechanism brings it back on course. From a differential equation point of view, we must show that the solution  $1/\ln x$  is stable under perturbations.

In the system of prime numbers, a perturbation is the random appearance (or non-appearance) of a prime that may cause the density to deviate from  $1/\ln x$ . We want to understand what happens after this perturbation—does our model predict that the density will return to  $1/\ln x$ ? To analyze this mathematically, we consider “perturbed solutions” of our model differential equation (1). That is, we consider a function  $f(x)$  that satisfies  $f(x) = 1/\ln x$  initially, up to some value  $x = x_0$ , and then jumps arbitrarily to some other value  $f(x_0) \neq 1/\ln(x_0)$  as the result of a single perturbation. Afterward, for  $x > x_0$ , the behavior of  $f(x)$  is governed by (1). If  $f(x)$  returns to its path as  $x$  increases—that is, if the relative error between  $f(x)$  and  $1/\ln x$  is eventually negligible—then the solution  $1/\ln x$  is said to be *stable under perturbations*.

Since the true density of primes is represented by a perturbed solution  $f(x)$ , the stability of  $1/\ln x$  will also predict the prime number theorem and validate our model against empirical evidence. Relying on Wright, we will compute the limit of the

relative error,

$$\lim_{x \rightarrow \infty} \frac{f(x) - 1/\ln x}{1/\ln x}, \quad (6)$$

and show that the limit is equal to zero.

To compute the limit in (6), we first make an advantageous change of variables by letting  $x = e^{2^u}$  and defining  $g(u) = f(e^{2^u})$ . This change of variables will have the convenience of relating  $g'(u)$  to the value of  $g$  at  $u$  and  $u - 1$ , rather than the relation of  $f'(x)$  to the value of  $f$  at  $x$  and  $\sqrt{x}$ . We provide a roadmap for the computations involved in the change of variables. The only tools required are calculus and algebra (and a pencil and paper), but the details can be safely skipped on a first reading.

**Change of variables** Let  $x = e^{2^u}$  and  $g(u) = f(e^{2^u})$ . Applying the chain rule to  $g(u)$  (twice), we find that

$$g'(u) = (\ln 2)2^u e^{2^u} f'(e^{2^u}). \quad (7)$$

Substituting for the derivative  $f'(e^{2^u})$  in (7) using equation (1) and converting back to  $g$ -notation, we find a differential equation for  $g$ :

$$g'(u) = -(\ln 2)2^{u-1}g(u-1)g(u). \quad (8)$$

Under the change of variables, our ideal solution  $1/\ln x$  becomes  $2^{-u}$ , so that the limit (6) we seek to compute transforms as

$$\lim_{u \rightarrow \infty} \frac{g(u) - 2^{-u}}{2^{-u}} = 2^u g(u) - 1, \quad (9)$$

where  $g(u)$  represents a perturbed solution of (8).

**The relative error function** We let  $h(u)$  represent the relative error function appearing in (9), so that

$$h(u) = 2^u g(u) - 1. \quad (10)$$

Our main tool in computing the limit of  $h(u)$  as  $u \rightarrow \infty$  is a differential-delay equation for  $h(u)$ . To find it, we first differentiate (10) using the product rule,

$$h'(u) = 2^u g'(u) + (\ln 2)2^u g(u). \quad (11)$$

We then substitute for  $g'(u)$  in (11) using equation (8) to find

$$\begin{aligned} h'(u) &= -2^u (\ln 2)2^{u-1}g(u-1)g(u) + (\ln 2)2^u g(u) \\ &= -(\ln 2)2^u g(u)(2^{u-1}g(u-1) - 1). \end{aligned}$$

Converting back to  $h$ -notation, we find our differential-delay equation for the relative error:

$$h'(u) = -(\ln 2)(h(u) + 1)h(u - 1). \quad (12)$$

Equation (12) is a particular member in a family of differential-difference equations studied by Wright in his 1955 paper [24]. This family of equations bears his name in deference to his contribution to their study.

**Computing the limit** We now return to our goal of computing the limit (6) of the relative error between a perturbed solution  $f(x)$  and our ideal solution  $1/\ln x$ . In terms of our new variables, we must compute the limit of the relative error function,  $h(u)$ , given in (10). We assume  $h(u) = 0$  initially, up to some value  $u = u_0$ , and then jumps arbitrarily to a nonzero value  $h(u_0)$  as the result of a single perturbation. Afterward, for  $u > u_0$ , the behavior of  $h(u)$  is governed by (12). To show that  $1/\ln x$  is stable under perturbations, we must show that the relative error eventually returns to 0; that is, we must show that the limit of  $h(u)$  as  $u \rightarrow \infty$  is 0.

Equation (12) severely restricts the possible values of the limit of  $h(u)$  as  $u \rightarrow \infty$ . If  $h(u) \rightarrow \ell$  for some  $\ell \in \mathbb{R}$ , then the derivative must be 0 for large  $u$  and thus, equation (12) tells us that

$$-(\ln 2)\ell(\ell + 1) = 0.$$

Hence,  $\ell$  must equal either 0 or  $-1$ . Given reasonable initial conditions on the interval  $-1 \leq u \leq 0$ , Wright shows that equation (12) has a unique, continuous solution [24, Thm 1 (i), p. 66] and that the limiting value of  $h$  is determined by the value of  $h(0)$ . If  $h(0)$  exceeds  $-1$ , then the limit is 0 [24, Thm 3, p. 67]. In our change of variables,  $x = e^{2^u}$  and so  $u = 0$  corresponds to  $x = e$ . The value  $h(0) = -1$  corresponds to  $f(e) = 0$ . Since  $f(x)$  represents a density function, we are in the situation where  $h(0) > -1$  and Wright tells us that  $h(u) \rightarrow 0$  as  $u \rightarrow \infty$ .

Thus, the solution  $1/\ln x$  is stable under perturbations. This tells us that the density of prime numbers reacts to a single perturbation like the cool, calm, and collected driver, or like the whimsical driver—smoothly adjusting to the bump in the road and getting back on track to  $1/\ln x$ . The model predicts the prime number theorem!

Solutions of  $h'(u) = -(\ln 2)(h(u) + 1)h(u - 1)$

Before we turn our attention to Littlewood’s theorem, it will be useful to have more information about the unique solution of (12) guaranteed by Wright. Equation (12) is a nonlinear differential-delay equation, and such equations are difficult to solve explicitly. However, we can get the information we need using numerical and asymptotic methods. We have already seen that any solution to (12) eventually tends to 0, but here we will see exactly *how* the solution goes to 0.

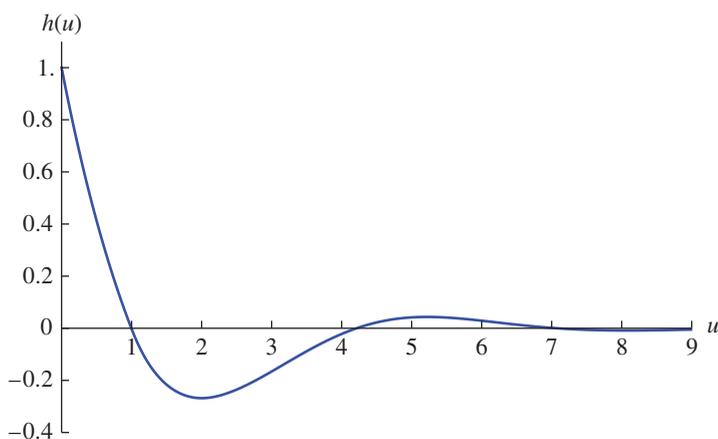
**Numerical solutions** We use a tangent line approximation to obtain a numerical solution to (12). The tangent line approximation at  $u = u_0$  is given by

$$\begin{aligned} h(u) &\approx h(u_0) + h'(u_0)(u - u_0) \\ &= h(u_0) - (\ln 2)(h(u_0) + 1)h(u_0 - 1)(u - u_0), \end{aligned} \tag{13}$$

where we substitute for the derivative  $h'(u_0)$  using (12). Knowledge of the values of  $h(u)$  on an interval of length one combined with (13) allow us to compute the values of  $h(u)$  on the next such interval. In particular, observe that if  $h(u) = 0$  on an interval of length one, then (13) tells us that the value will remain 0 on the next such interval. Since  $h(u)$  represents the relative error between  $f(x)$  and  $1/\ln x$ , this tells us that once the system is on track, it stays on track!

When the system is perturbed, however, we expect nonzero values for  $h(u)$ , and thus use a nonzero initial condition to represent a perturbation. We assume that  $h(u) = 0$  for  $u < -1$  and set  $h(-1) = 1$ . (Equation (12) is invariant under translations, which means that a perturbation at any value  $u_0$  has the same effect as at any other value of  $u_0$ ,

just displaced along the  $u$ -axis. Therefore, we lose no generality in taking  $u_0 = -1$  in what follows.) Tracing back through our change of variables, this represents perturbing  $1/\ln x$  by a factor of 2. We then use (13) to determine values of  $h$  when  $u > -1$ . Note that  $h(u) = 1$  on the entire interval  $-1 \leq u \leq 0$ , since (12) tells us that  $h'(u) = 0$  on that interval. FIGURE 3 shows the approximation on the interval  $0 \leq u \leq 9$ . (Recall that a  $u$ -value of 9 corresponds to an  $x$ -value of  $e^{2^9} \approx 2.28 \times 10^{222}$ .) From the graph in FIGURE 3, we not only see that  $h(u) \rightarrow 0$  as  $u \rightarrow \infty$ , but we also see *how*  $h(u)$  is tending to zero. It looks like our old friend, the whimsical driver! Different initial conditions with  $h(0) > -1$  lead to similar approximations.



**Figure 3** Numerical solution to  $h'(u) = -(\ln 2)(h(u) + 1)h(u - 1)$

**Asymptotic solutions** Because  $h(u) \rightarrow 0$  as  $u \rightarrow \infty$ , we expect solutions of (12) will be well approximated, when  $u$  is large, by solutions of the associated *linear* differential-delay equation, found by removing the quadratic term of equation (12):

$$h'(u) = -(\ln 2) h(u - 1). \quad (14)$$

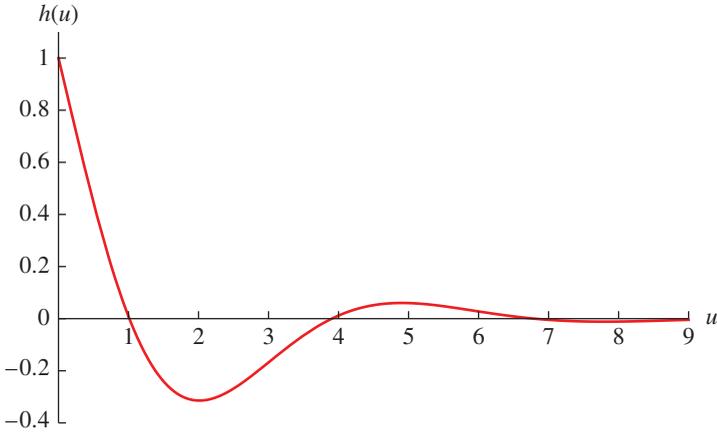
We have much more hope of finding an explicit solution in the linear case. Indeed, we can find solutions to (14) using a common technique in differential equations. We set  $h(u) = Ae^{\omega u}$ , where  $\omega = a + bi$  is a complex number and substitute this form into (14). Simplifying, we find that solutions of

$$\omega e^{\omega} = -\ln 2 \quad (15)$$

give rise to solutions of (14). Equations such as (15) arise in many applications and have been studied extensively. In particular, Wright studied them in conjunction with equation (12). There are countably many solutions to (15) occurring in conjugate pairs, all of which have negative real part. (See, for example, [24, Theorem 5, p. 72].) This is encouraging; the negative real part of  $\omega$  leads to an exponentially-damped oscillatory solution, matching our numerical data. The solution with the least negative real part is  $\omega = -0.5716236091 \pm 1.086461157i$ . (This numerical solution was found using the multi-valued complex-valued function, Lambert- $W$ , implemented in Maple. For a nice overview of this function and its many uses, see [6].) The corresponding solution to (14) is the most dominant asymptotically, and so we expect that the relative error will resemble this solution in the long run. Indeed, Wright *proved* that any solution to (12) is asymptotic to

$$Ae^{-0.5716236091u} \sin(1.086461157u + B) \quad (16)$$

for appropriate constants  $A$  and  $B$  (provided  $h(0) > -1$ ) [24, Thm 7 and Eqn (6.7) on p. 77 combined with Thm 8 on p. 78]. Equation (16) is graphed in FIGURE 4, with  $A$  and  $B$  chosen so that  $h(0) = 1$ , matching the initial condition of the numerical approximation in FIGURE 3.



**Figure 4** Asymptotic solution to  $h'(u) = -(\ln 2)(h(u) + 1)h(u - 1)$

### Littlewood’s theorem

Armed with more specific information about the relative error  $h(u)$ , we turn our attention to Littlewood’s theorem. In particular, Littlewood’s theorem tells us that the difference between  $\pi(x)$  and its approximation  $\text{Li}(x)$  changes sign infinitely often. The function  $\text{Li}(x)$  is defined by integrating  $1/\ln x$ . The function  $\pi(x)$  can similarly be represented as an integral using our model. A solution  $f(x)$  to (1) represents the density of primes at  $x$  and so  $\pi(x)$  is approximated by integrating  $f(x)$ . Thus,

$$\pi(x) - \text{Li}(x) \approx \int_2^x (f(t) - 1/\ln t) dt. \tag{17}$$

To study this integral, we apply the previously-used change of variables ( $t = e^{2^s}$ ) and rewrite the integrand in terms of  $h$  using equation (10):

$$\begin{aligned} (f(t) - 1/\ln t)dt &= (\ln 2)2^s e^{2^s} (g(s) - 2^{-s}) ds \\ &= (\ln 2)e^{2^s} h(s) ds. \end{aligned} \tag{18}$$

Since  $h$  is asymptotic to an exponentially-damped trigonometric function, our approximation to  $\pi(x) - \text{Li}(x)$  is (eventually) found by integrating an exponentially-growing trigonometric function. (The double-exponential appearing in (18) will eventually dominate the exponential appearing in (16)). The peaks and troughs of such a function become successively more pronounced, like our overreacting and panicky driver in FIGURE 1c. Now imagine integrating such a function. When we integrate, the negative area of a trough is canceled by a portion of the positive area of the next peak, and vice-versa. Thus, each zero of the integrand will give rise to a sign change of the integral function. The infinitely many zeros of (16) predict infinitely many sign changes of  $\pi(x) - \text{Li}(x)$ !

## Limitations of the model

Thus far, we have seen that our model predicts the prime number theorem and the infinite crossings of Littlewood's theorem. The model agrees with empirical data, at least regarding gross behavior of the system. However, this section will illustrate the need for caution in using the model to predict the finer details of the distribution of primes.

In the last section, we saw that each zero of  $h(u)$  gives rise to a zero of  $\pi(x) - \text{Li}(x)$ . The asymptotic form of  $h(u)$  given in (16) tells us that there are infinitely many such zeros. The specific frequency and period of the trigonometric function in (16) allow us to deduce more precise information. For example, the specific frequency, along with a change back to the variable  $x$ , predicts that the number of sign changes of  $\pi(x) - \text{Li}(x)$  on the interval  $(2, x)$  is at least

$$\frac{1.086461157u}{\pi} = \frac{1.086461157 \ln \ln x}{\pi \ln 2} \approx \frac{1}{2} \ln \ln x. \quad (19)$$

We expect (19) to be a lower bound for the number of sign changes, as the random nature of the system may lead to several sign changes in the area of a single sign change predicted by the continuous model. Indeed, Kaczorowski [13] proved in 1985 that, for  $x$  sufficiently large, the number of sign changes of  $\pi(x) - \text{Li}(x)$  on  $(2, x)$  is at least  $c \ln x$ , for some positive constant  $c$ . Thus, our model is predicting fewer sign changes than what is currently known. Despite knowing that there are infinitely many sign changes between  $\pi(x)$  and  $\text{Li}(x)$ , no one yet knows where the first one occurs. (Currently, the best result is that the first sign change occurs for some  $x < 1.39 \times 10^{316}$  [2, 4].) Without numerical data regarding actual sign changes, we cannot confirm that the discrepancy between our bound and that of Kaczorowski can be explained by random sign changes clustered about a single sign change predicted by the model.

Littlewood's theorem includes more information than just the infinitely many sign changes of  $\pi(x) - \text{Li}(x)$ . Littlewood gave precise information concerning how much  $\pi(x) - \text{Li}(x)$  differs from zero, in the form of functional bounds found in inequalities (2) and (3). Our model gives a similar result. Using the exact form of the solution to  $h(u)$  in (16), we can estimate the integral in (17) at specific  $x$ -values, which in turn allows us to estimate  $\pi(x) - \text{Li}(x)$ . For arbitrarily small  $\epsilon > 0$ , the model predicts the existence of a positive constant  $k_\epsilon$  such that we can find arbitrarily large values of  $x$  that satisfy

$$\pi(x) - \text{Li}(x) > k_\epsilon x^{2-\epsilon} (\ln x)^{-0.825}, \quad (20)$$

and arbitrarily large values of  $x$  that satisfy

$$\pi(x) - \text{Li}(x) < -k_\epsilon x^{2-\epsilon} (\ln x)^{-0.825}. \quad (21)$$

Both Littlewood's theorem and our model suggest that the total number of primes  $\pi(x)$  behaves like our panicky and overreacting driver with respect to the ideal path of  $\text{Li}(x)$ . The functions in (20) and (21), however, predict more volatile oscillations than Littlewood's theorem. This result warrants skepticism, though, because it is in conflict with the Riemann hypothesis, perhaps the most famous open conjecture in mathematics today.

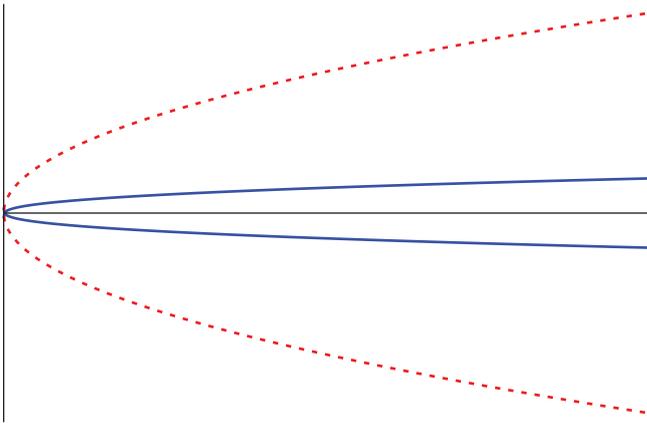
The Riemann hypothesis is a conjecture about the placement of the zeros of the Riemann zeta function, a function of a complex variable with deep connections to the distribution of primes. Indeed, the Riemann hypothesis implies that the prime number theorem and the proofs of Hadamard and de la Vallée-Poussin from 1896 rely on partial

information about the location of the zeros, rather than the precise placement predicted by Riemann in 1859. An equivalent version of the Riemann hypothesis gives bounds on the absolute error between  $\pi(x)$  and its approximation  $\text{Li}(x)$ , the quantity we are interested in. This is the version we state.

**RIEMANN HYPOTHESIS.** *For all  $\epsilon > 0$ , there exists a positive constant  $C_\epsilon$  such that*

$$|\pi(x) - \text{Li}(x)| \leq C_\epsilon x^{\frac{1}{2} + \epsilon}.$$

Littlewood’s theorem tells us that the value of  $\pi(x) - \text{Li}(x)$  jumps above and below the solid curve in FIGURE 5 while the Riemann hypothesis predicts the value stays *within* the dotted curve. Our model predicts the value jumps above and below a curve that vastly exceeds the dotted line in FIGURE 5, giving rise to the conflict. The Riemann hypothesis is an open conjecture, not a proved theorem, yet a large amount of supporting evidence exists. Thus, we should view our model’s prediction with caution, at the very least. (For more information about the Riemann hypothesis, a good starting point is Bombieri’s article [3], which can be found on the Clay Mathematics Institute’s website [5], along with other interesting resources.)



**Figure 5** Comparison of bounds for  $\pi(x) - \text{Li}(x)$ ; Littlewood (solid) and Riemann Hypothesis (dotted).

A natural next step in any modeling process is to fine-tune the model based on what we’ve learned about it. We based our derivation of equation (1) on the Sieve of Eratosthenes, considering the effect of a prime  $x$  on the density at  $x^2$ . A finer analysis—for example, incorporating the effect of a prime  $x$  on the density at powers of  $x$  beyond 2—might improve the situation. However, there may be a limit to how well we can describe the system using this approach. By multiplying the density correction factors together, the Sieve of Eratosthenes provides an approximation for the density of primes less than  $x$ :

$$\prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right). \tag{22}$$

The product in (22) is known to be asymptotic to

$$\frac{2e^{-\gamma}}{\ln x},$$

where  $\gamma$  is the Euler-Mascheroni constant given by

$$\gamma := \lim_{n \rightarrow \infty} \left( 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n \right).$$

Unfortunately, this does not quite give the correct answer, as  $2e^{-\gamma}$  is about  $1.12292\dots$  rather than 1 as expected from the prime number theorem. (For more about this approximation, see [11, p. 13].) The use of more sophisticated tools from number theory may lead to a model with better predictive power. However, this may be at the expense of the intuitive insight that this model brings to the table.

**Acknowledgment** The authors would like to thank Alexander Perlis and the anonymous referees for their helpful suggestions on both the style and substance of earlier versions of this article.

## REFERENCES

1. Paul T. Bateman and Harold G. Diamond, A hundred years of prime numbers, *Amer. Math. Monthly* **103** (1996) 729–741.
2. Carter Bays and Richard H. Hudson, A new bound for the smallest  $x$  with  $\pi(x) > \text{li}(x)$ , *Math. Comp.* **69** (2000) no. 231, 1285–1296 (electronic).
3. Enrico Bombieri, The Riemann hypothesis, in *The Millennium Prize Problems*, Clay Mathematics Institute, Cambridge, MA, 2006, 107–124.
4. Kuok Fai Chao and Roger Plymen, A new bound for the smallest  $x$  with  $\pi(x) > \text{li}(x)$ , *Int. J. Number Theory* **6** (2010) 681–690.
5. Clay Mathematics Institute, “Millennium problems—Riemann hypothesis,” [http://www.claymath.org/millennium/Riemann\\_Hypothesis/](http://www.claymath.org/millennium/Riemann_Hypothesis/), 2011.
6. R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, On the Lambert  $W$  function, *Adv. Comput. Math.* **5** (1996) 329–359.
7. R. D. Driver, *Ordinary and Delay Differential Equations*, Springer-Verlag, New York, 1977. <http://dx.doi.org/10.1007/978-1-4684-9467-9>
8. H. M. Edwards, *Riemann’s Zeta Function*, Academic Press, New York, 1974.
9. P. Erdős, On a new method in elementary number theory which leads to an elementary proof of the prime number theorem, *Proc. Nat. Acad. Sci. USA* **35** (1949) 374–384.
10. L. J. Goldstein, A history of the prime number theorem, *Amer. Math. Monthly* **80** (1973) 599–615.
11. Andrew Granville, Harald Cramér and the distribution of prime numbers, *Scand. Actuar. J.* (Harald Cramér Symposium, Stockholm, 1993) (1995) 12–28.
12. G. Hoffman de Visme, The density of prime numbers, *Math. Gaz.* **45** (1961) 13–14.
13. J. Kaczorowski, On sign-changes in the remainder-term of the prime-number formula, II, *Acta Arith.* **45** (1985) 65–74.
14. Tadej Kotnik, The prime-counting function and its analytic approximations:  $\pi(x)$  and its approximations, *Adv. Comput. Math.* **29** (2008) 55–70.
15. J. E. Littlewood, Sur la distribution des nombres premiers, *C. R. Acad. Sci. Paris* **158** (1914) 1869–1875.
16. Eduardo Liz, Gonzalo Robledo, and Sergei Trofimchuk, Wright’s conjecture: between the distribution of primes and population growth, *Cubo Mat. Educ.* **3** (2001) 89–107.
17. Atle Selberg, An elementary proof of the prime-number theorem, *Ann. of Math.* (2) **50** (1949) 305–313.
18. Yuri Tschinkel, About the cover: on the distribution of primes—Gauss’ tables, *Bull. Amer. Math. Soc.* (N.S.) **43** (2006) 89–91.
19. E. M. Wright, The non-linear difference-differential equation, *Quart. J. Math.*, Oxford Ser. **17** (1946) 245–252.
20. ———, The linear difference-differential equation with asymptotically constant coefficients, *Amer. J. Math.* **70** (1948) 221–238.
21. ———, Linear difference-differential equations, *Proc. Cambridge Philos. Soc.* **44** (1948) 179–185.
22. ———, The linear difference-differential equation with constant coefficients, *Proc. Roy. Soc. Edinburgh. Sect. A.* **62** (1949) 387–393.
23. ———, The stability of solutions of non-linear difference-differential equations, *Proc. Roy. Soc. Edinburgh. Sect. A.* **63** (1950) 18–26.
24. ———, A non-linear difference-differential equation, *J. Reine Angew. Math.* **194** (1955) 66–87.
25. ———, A functional equation in the heuristic theory of primes, *Math. Gaz.* **44** (1960) 15–16.
26. ———, Number theory and other reminiscences of Viscount Chermwell, *Notes and Records Roy. Soc. London* **42**(2) (1988) 197–204.

**Summary** A differential-difference equation modeling the density of primes has been independently discovered several times by applied mathematicians with an interest in number theory, including Lord Cherwell. In this article, we use the framework of feedback and control to explain the natural connection between this equation and the distribution of prime numbers. Building on work of E. M. Wright, who learned of the equation from Cherwell and studied it for over a decade, we explain how the model predicts both the prime number theorem and a theorem of Littlewood. We close with known limitations of the model.

**SUSAN H. MARSHALL** teaches mathematics at Monmouth University, located on the Jersey Shore. Her primary research area is number theory, and in particular she finds the interplay between number theory and other mathematical subjects (such as geometry) fascinating. As an undergraduate at Wake Forest University, she had a particular fondness for differential equations, but was lured to arithmetic geometry as a graduate student at the University of Arizona. This project exposed her to somewhat unexpected connections between number theory and the fields of differential equations and mathematical modeling.

**DONALD R. SMITH** is Associate Professor of Management and Decision Sciences at the Leon Hess Business School of Monmouth University. He originally intended to major in mathematics, but transitioned to physics while an undergraduate at Cornell University, and then to operations research for his graduate work, earning a Masters degree from Columbia University and a Ph.D. from the University of California at Berkeley. His primary research interests have been in modeling stochastic systems. He has always been fascinated by the distribution of primes, and found this work fun because it put an operations research modeling perspective on this captivating topic.

### Maybe We're All the Same at Heart

I can tell that 41 and 107 are different integers, because they have different digits.

But if I multiply both of them by 3,

$$41 \times 3 = 123 \quad 107 \times 3 = 321$$

then the results have the same digits! Only their order distinguishes them.

Let's say that two positive integers  $m$  and  $n$  are *really the same at heart* if there is a positive integer  $c$  such that the products  $mc$  and  $nc$  have the same digits. Let's be more precise: We insist that the decimal representations of  $mc$  and  $nc$  have the same number of 1's, the same number of 2's, and so on up to the same number of 9's. We ignore 0's, because  $mc$  and  $nc$  can each have as many leading zeros as we like.

Then 41 and 107 are really the same at heart, because we can choose  $c = 3$ . (And if we don't like  $c = 3$ , we can try  $c = 9006$ .)

Which pairs  $m$  and  $n$  are really the same at heart?

**Fair warning:** This might not be easy. The answer will be in the solutions to this year's USA Mathematical Olympiad, in our October issue. The problem was proposed by Richard Stong.