

Large Intervals Between Consecutive Primes

By J. H. Cadwell

Abstract. Some results in number theory, including the Prime Number Theorem, can be obtained by assuming a random distribution of prime numbers. In addition, conjectural formulae, such as Cherwell's for the density of prime pairs $(p, p + 2)$ obtained in this way, have been found to agree well with the available evidence. Recently, primes have been determined over ranges of 150,000 numbers with starting points up to 10^{16} . Statistical arguments are used to obtain a formula for the largest interval between consecutive primes in such a range, and it is found to agree well with recorded values. The same method is applied to predict the maximum interval between consecutive primes occurring below a given integer.

Introduction. Jones, Lal and Blundon [1] have tabled the largest interval between primes in each of the regions $(x, x + \Delta)$ with

$$\Delta = 150,000, \quad x = 10^n, \quad n = 8(1)15,$$

as well as the numbers of prime pairs $(p, p + 2)$, etc., lying in these ranges. They found that these counts of pairs, triples, etc. agree well with those given by the conjectured formulae of Hardy and Littlewood [2]. It is interesting to note that Cherwell [3] obtained the same formulae using statistical methods.

This note assumes that primes occur randomly in a specified region with a uniform distribution function. The largest interval between consecutive primes in the range $(x, x + \Delta)$ is investigated; for large x its mean value is

$$(1) \quad \log x \log(\Delta/\log x),$$

but for the ranges in [1] it pays to use the rather more elaborate result found below, and agreement with actual counts is good.

From Table 1 of [5], we can obtain values of the maximum interval between consecutive primes for the ranges

$$(10^m, 10^{m+1}), \quad m = 3(1)9,$$

and again the calculated maxima are in good agreement.

The method can be used to estimate $G(N)$, the maximum interval below N . The result is

$$G(N) \sim \log N(\log N - \log \log N).$$

Received December 7, 1970, revised March 25, 1971.

AMS 1969 subject classifications. Primary 1042.

Key words and phrases. Consecutive primes, interval between primes, maximum interval between primes.

Shanks in [4] defines $p(g)$ as the prime ending the maximum interval of length g occurring below $p(g)$. He suggests that $\sqrt{g} \sim \log p(g)$.

The gap associated with the selected endpoint $p(g)$ will be larger than that predicted by a formula applicable to an arbitrary endpoint, i.e. g should exceed $G[p(g)]$, so that the formulae are in general agreement about the largest interval up to a prescribed limit.

Random Division of an Interval. We consider the interval $(0, h)$ divided into $(m + 1)$ parts of lengths x_1, x_2, \dots, x_{m+1} by m internal points. The distribution of the first m x 's has the probability density function

$$(2) \quad \frac{m!}{h^m} dx_1 \cdot dx_2 \cdot dx_3 \cdot \dots \cdot dx_m,$$

with

$$x_i \geq 0 \text{ (all } i), \quad x_1 + x_2 + \dots + x_m \leq h.$$

If only those sets of x 's are counted for which

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_{m+1},$$

this p.d.f. must be multiplied by $(m + 1)!$.

The transformation defined by

$$\begin{aligned} x_1 &= y_1/(m + 1), \\ x_2 &= y_1/(m + 1) + y_2/m, \\ x_3 &= y_1/(m + 1) + y_2/m + y_3/(m - 1), \\ &\dots\dots\dots \\ x_{m+1} &= y_1/(m + 1) + y_2/m + \dots + y_{m+1}/1, \end{aligned}$$

leads to the results

$$(3) \quad \begin{aligned} y_i &\geq 0 \text{ (all } i), \\ y_1 + y_2 + \dots + y_{m+1} &= h, \\ x_{m+1} &= h - \left(\frac{m}{m + 1}\right)y_1 - \left(\frac{m - 1}{m}\right)y_2 - \dots - \frac{1}{2}y_m. \end{aligned}$$

The density function (2) is now transformed into the p.d.f. of y_1, y_2, \dots, y_m to give

$$\frac{m! (m + 1)!}{h^m} \cdot \frac{\partial(x_1 \cdot \dots \cdot x_m)}{\partial(y_1 \cdot \dots \cdot y_m)} dy_1 \cdot dy_2 \cdot \dots \cdot dy_m,$$

with the conditions

$$(4) \quad y_i \geq 0 \text{ (all } i), \quad y_1 + y_2 + \dots + y_m \leq h.$$

The Jacobian has the value $1/(m + 1)!$ and moments of the y 's can be found from the Dirichlet formula

$$\iint \dots \int F(\sum y_i) \prod (y_i^{\alpha_i - 1} dy_i) = \frac{\prod (\Gamma(\alpha_i))}{\Gamma(\sum \alpha_i)} \int_0^h \lambda^{\sum \alpha_i - 1} F(\lambda) d\lambda,$$

where the multiple integral is over the region defined by (4), and sums and products are, for i , from 1 to m .

The moments up to order 2 are found to be

$$E(y_i) = \frac{h}{m+1}, \quad \text{Var}(y_i) = \frac{mh^2}{(m+1)^2(m+2)},$$

$$\text{Cov}(y_i, y_j) = \frac{-h^2}{(m+1)^2(m+2)}.$$

It follows that the mean value of the largest interval x_{m+1} is

$$E(x_{m+1}) = h - \frac{h}{m+1} \left(\frac{m}{m+1} + \frac{m-1}{m} + \dots + \frac{1}{2} \right)$$

$$= \frac{h}{m+1} \left(\frac{1}{m+1} + \frac{1}{m} + \dots + 1 \right) \sim \frac{h}{m} (\log m + \gamma),$$

where m is large and γ is Euler's constant (0.577 ...). Similarly, for large m , we find that

$$\text{Var}(x_{m+1}) \sim \frac{\pi^2 h^2}{6m^2}.$$

The mean value can also be derived geometrically. It can be shown that, the intervals being in ascending order, the point $(x_1, x_2, \dots, x_{m+1})$ in $(m+1)$ -space is uniformly distributed throughout the contents of the m -dimensional simplex defined by the $(m+1)$ points

- $(0, 0, \dots, 0, 0, h),$
- $(0, 0, \dots, 0, h/2, h/2),$
- $(0, 0, \dots, h/3, h/3, h/3),$
-

the last point having all its coordinates equal to $h/(m+1)$. The mean centre of the contents of this simplex is also the mean centre of its vertices, hence the mean value of x_{m+1} follows at once.

Application to Prime Numbers. We assume that there are m primes between x and $(x + \Delta)$ so that from the Prime Number Theorem $m \sim \Delta/\log x$. Subtracting 2 from each of the $(m+1)$ intervals, we replace the m and h of our previous work by

$$m = \Delta/\log x, \quad h = \Delta - 2\Delta/\log x.$$

While the lengths of the intervals between primes are integers, we have considered continuously distributed quantities. However, as the expected value of the largest interval is of the order of 100, the truncation error will not be important.

Our previous formula, after allowing for the 2 subtracted from each interval, gives

$$(5) \quad 2 + (\log x - 2) \left\{ \log \left(\frac{\Delta}{\log x} \right) + \gamma \right\}$$

as the mean value of the largest gap. For large x , this reduces to (1). Application of the more accurate result gives the values of Table 1.

TABLE 1
Maximum Intervals Between Primes in the Range $(x, x + 150,000)$

x	10^8	10^9	10^{10}	10^{11}	10^{12}	10^{13}	10^{14}	10^{15}
<i>Observed</i>	168	176	182	148	222	234	300	276
<i>Calculated</i>	159	179	199	218	237	256	275	293
<i>Difference</i>	-9	+3	+17	+70	+15	+22	-25	+17

It will be seen that agreement is good. The variance of the maximum interval is 900 at about the middle of the table, so that differences are no greater than would be expected on the assumption of a random distribution.

Table 2 shows the results for relatively larger values of Δ .

TABLE 2
Maximum Intervals Between Primes in the Ranges $(10^m, 10^{m+1})$

m	3	4	5	6	7	8	9
<i>Observed</i>	36	72	114	154	220	282	354
<i>Calculated</i>	40	72	115	167	229	302	385

Agreement is reasonably good, but calculated values are almost all too big. As (5) assumes a fixed mean interval of $\log x$, while the mean interval increases somewhat over the range, we should expect under-estimates.

The Maximum Interval up to N . Formula (5) applies equally to the range $(x - \Delta, x)$ but should then lead to over-estimates. Applied to estimate $G(N)$, i.e. the maximum interval in $(0, N)$, and taking N to be large, it gives

$$G(N) \sim \log N \log (N/\log N).$$

While $\log N$ varies considerably over the range, for sufficiently large N , most of this variation is confined to a small part of $(0, N)$. Thus, for $N = 10^{100}$, $\log N$ changes from 225.7 to 230.3 over the interval $(10^{98}, 10^{100})$ covering 99 per cent of the whole range. The use of the value 230.3 throughout should thus produce only a moderate degree of over-estimation.

We have to be content with evaluating (5) for $N = 1.1096 \times 10^{11}$, the present limit of our knowledge of the maximum interval. The result is 462 as compared with the known value of 382.

1. M. F. JONES, M. LAL & W. J. BLUNDON, "Statistics on certain large primes," *Math. Comp.*, v. 21, 1967, pp. 103–107. MR 36 #3707.
2. 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.*, v. 44, 1923, pp. 1–70.
3. LORD CHERWELL, "Note on the distribution of the intervals between prime numbers," *Quart. J. Math. Oxford Ser.*, v. 17, 1946, pp. 46–62. MR 8, 136; 708.
4. D. SHANKS, "On maximal gaps between successive primes," *Math. Comp.*, v. 18, 1964, pp. 646–651. MR 29 #4745.
5. L. J. LANDER & T. R. PARKIN, "On first appearance of prime differences," *Math. Comp.*, v. 21, 1967, pp. 483–488. MR 37 #6237.