UCRL 16825
cy 5

University of California

Ernest O. Lawrence Radiation Laboratory

TWO-WEEK LOAN COPY

This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 5545

Berkeley, California

16825

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Submitted to Mathematical Monthly

UCRL-16825

UNIVERSITY OF CALIFORNIA

Lawrence Radiation Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

FREQUENCY DISTRIBUTION OF PRIME NUMBERS IN INTERVALS

Edwin M. McMillan

6

April 20, 1966

FREQUENCY DISTRIBUTION OF PRIME NUMBERS IN INTERVALS

Edwin M. McMillan

Lawrence Radiation Laboratory
University of California
Berkeley, California

April 20, 1966

Stein and Ulam[1] have reported some interesting empirical results on the distribution of prime numbers. They consider intervals of length $E[\ln^2 k/(\ln k - 1)]$, chosen to contain on the average about one prime, and find the relative frequencies of occurrence $\nu_m(n)$ of m primes in all such intervals contained in number sequences extending from $k = 1$ to $k = n$. The results are tabulated for several values of n from $n = 10{,}000$ to $n = 100{,}000{,}000$. From their table one can find the corresponding frequencies $\nu_m(n_1, n_2)$ for all intervals in a sequence $n_1 < k < n_2$. In the table below, in the columns labeled "S & U", values of $\nu_m(n_1, n_2)$ are given for two ranges of n, chosen from the lowest and highest parts of the table of Stein and Ulam.

	$n = 10{,}000$ to $n = 50{,}000$		$n = 80{,}000{,}000$ to $n = 100{,}000{,}000$	
	S & U	Calc.	S & U	Calc.
ν_0	.2742	.2746	.3068	.3093
ν_1	.4480	.4436	.4166	.4121
ν_2	.2333	.2389	.2158	.2160
ν_3	.0430	.0429	.0541	.0554
ν_4	.0013	0	.0064	.0069
ν_5	.0000	0	.0003	.0003
mean m	1.049		1.038	

The purpose of this note is to show that the empirical results are consistent with a simple semirandom model for prime number distribution. In a purely random model, an interval of length s would have s possible places for primes, each with an equal probability c for occupancy. The mean value of m would be equal to sc, and ν_m would have a binomial distribution, i.e., $\nu_0 = (1-c)^s$, $\nu_1 = sc(1-c)^{s-1}$, $\nu_2 = [s(s-1)/2]\ s^2(1-c)^{s-2}$, etc. In the two ranges of k considered, $s = 11$ (over most of the range) and $s = 19$, respectively. It is easily seen that this model, with these values, gives a poor fit to the empirical data. It is also obvious that the model is wrong; a sequence of contiguous integers cannot all be prime! The elimination of the even numbers reduces the possible number of primes in a given interval to an average of half the length of the interval. Further elimination of factors of 3 reduces the fraction to 1/3, of factors of 5 to 4/15, and so on, each additional prime p eliminated multiplying the fraction by $(p-1)/p$. At each stage of elimination the resulting sequence is periodic, with a repeat length equal to the product of the primes eliminated. Thus, while the fraction of possible primes decreases slowly in successive stages of elimination, the repeat length increases rapidly.

The idea of the model discussed in this note is to construct the "partial sieve of Eratosthenes" to some stage, and to treat the distribution of primes in the remaining spaces as random, justifying the randomness on the basis that regularity on a large scale can appear random on a small scale. Having constructed the "partial sieve" to a suitable stage (to be discussed later), we find the numbers of possible primes for all positions of the given interval along a repeat length; the resulting

numbers can then be given weighting factors equal to their relative frequencies of occurrence. For each number of possible primes there can then be constructed a binomial distribution, with a common value for the probability of occurrence of a prime, satisfying the condition that its product with the mean "possible number" should be equal to the mean m. These distributions are then added, with the appropriate weighting factors. One would expect that, if this procedure is carried out for successively increasing stages, the result should become a successively better approximation, but I have not carried out such a computation, since it appears that a simplified procedure gives a good account of the main features of the empirical data.

The simplification is illustrated by the distribution for $n = 10,000$ to $n = 50,000$. First, we fit $\nu_0 = (1-c')^{s'}$ by a single s' (not necessarily an integer) and c', satisfying $c's'$ = mean m. It turns out that s' is very close to 3. Then we use the single value 3 as the basis for a binomial distribution, with the results given in the column marked "calc." They fit the empirical results very well, except for ν_4. The more detailed calculation would involve three binomial distributions based on 2, 3, and 4, with 4 having a small weighting factor; a rough estimate, carrying the elimination through 7, gives $\nu_4 = .001$; ν_5 is equal to zero.

For the interval $n = 80,000,000$ to $n = 100,000,000$, the same procedure leads to an s' close to 4.5; this is approximated by using two binomial distributions based on 4 and 5, with a common value of c' and with equal weights. The results are again given under "calc.", and the fit is good for all tabulated values. The largest "possible number" that can occur with $s = 19$ is 6, and an estimate of its frequency indicates

that the first significant figure of ν_6 will occur in the sixth decimal place.

In these two cases, the ratios s'/s are 0.27 and 0.24 respectively, corresponding approximately to the fractions of places left after eliminating through 5 and 7 respectively, leaving sequences with repeat lengths of 30 and 210. Further elimination leads to very long repeat lengths while changing the ratio only slightly. I would expect that, for any range likely to be accessible to empirical test, the ratio will not go below 0.2; this can be used to estimate the rapidity of approach to a Poisson distribution, i.e., the expected distribution will be a binomial distribution based on $s' \sim s/5$. Estimation of the highest ν_m will require more detailed consideration. For example, ν_7 cannot exceed zero for $s < 21$, so that 7 primes in an interval cannot occur below $n = 470{,}000{,}000$.

REFERENCE

1. M. Stein and S. M. Ulam, An Observation on the Distribution of Primes, Am. Math. Monthly,

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.