A SURPRISINGLY ACCURATE NEW FORMULA FOR ESTIMATING THE PRIMES AND ACCURACY TESTS

by

Anirban DasGupta
Purdue University

Technical Report #06-05

Department of Statistics
Purdue University

August 2005

A Surprisingly Accurate New Formula for Estimating the Primes and Accuracy Tests

August 14,2005

Anirban DasGupta, Purdue University

ABSTRACT

Based on a theoretical result on the asymptotic order of the primes, a new estimate of the nth prime is given. The estimate is obtained after a numerical search for the best set of nice constants in a more general class of estimates. The formula estimates all of the first 25000 primes with an absolute error of at most 250 and an average percentage error of 1.47 %. The estimate of the prime counting function $\pi(x)$ obtained from inversion of this formula is also seen to be accurate for x upto five million in the tests that we conducted. These accuracy tests and the explicit prime estimate formula are provided in the note.

1 Introduction

The purpose of this note is to present an estimating formula for the nth prime based on some theory and then essentially an extensive search for a formula that gives the best fit to the actual primes in the region we tested. What is worthwhile about this formula is that it is more elaborate than a very simple approximation like $n \log n$, and that in the region we tested the formula is really quite accurate, sometimes intriguingly accurate. As an example, the exact value of the 15000th prime is 163841, and the formula estimates it to be 163849, an error of .005%. The exact value of the 100000th prime is 1299709, and our formula estimates it to be 1299589, anerror of .009%. As another example, if we invert the formula to produce an estimate of the prime π function, the exact value of $\pi(50000000)$ is 3001134, and inversion of our formula estimates it to be 3000718, an error of only 416. In contrast, the simple estimate $\frac{n}{\log n}$ estimates $\pi(50000000)$ to be 2820470, which is an error of 180664. The formula seems to have something to offer in the range of primes up to about 5 millions. If we stretch the range , say, to 10^8, the formula fails to have as good accuracy. However, the search can be refined in order to produce another and better formula applicable to such a range. Needless to say, we should not expect accuracy of equal quality in predicting an erratic sequence like the primes over a very large range. We were in fact slightly surprised that upto 5 millions, our formula worked quite well.

2 Background and the Formula

The textbook estimate of $p(n)$, the nth prime is of course $n \log n$(perhaps rounded to the nearest odd integer). It is always an underestimate, i.e., $p(n) > n \log n$ for any $n > 1$; see Rosser(1938). It is actually not a very accurate estimate, numerically. A better estimate is $n(\log n + \log\log n - 1)$. A neat result in Dusart(1999) is that this too is always an underestimate. An available result, perhaps less well known, is that $p(n) = n(\log n + \log\log n - 1 + o(\frac{\log\log n}{\log n}))$; see Ribenboim(1991). Based on this result, we looked for an estimate $\hat{p}(n)$ of the form $\hat{p}(n) = n(\log n + a\log\log n - b + c\frac{\log\log n}{\log n}) + k$, for $a, b \approx 1, c \approx 0$, and k an integer that is intended to improve on the accuracy in some average sense rather than just using $k = 0$. Subjectively, we imposed on ourselves the accuracy requirement that each of the first $25,000$ primes should be estimated within an absolute error of at most 250. We were able to find one such combination of a, b, c and k on a search. Here is our estimating formula :

$$\hat{p}(n) = n(\log n + 1.02\log\log n - .99 - .1\frac{\log\log n}{\log n}) + 185.$$

If the constants $1.02, .99, .1$ and 185 are perturbed slightly, we still get comparable accuracy, but in our search these constants with at most two decimals in them gave us the 250 accuracy we wanted. So that is the one we report.

3 Numerical Report of Performance

We simply state a subset of the performance results we obtained.

Fact

a) $\frac{1}{24999}\sum_{n=2}^{25000}(\hat{p}(n) - p(n)) = 33.95;$

b) $\frac{1}{24999}\sum_{n=2}^{25000}(\hat{p}(n) - p(n))/p(n) = .0147;$

c) $\max\{\hat{p}(n) - p(n)\} = 250 = -\min\{\hat{p}(n) - p(n)\}$, where the max and the min are for $2 \leq n \leq 25000$;

d) Number of n such that $p(n) = \hat{p}(n)$ in the above range is 100;

e) Number of n such that $\hat{p}(n)$ is a prime is 2217;

f) Number of n such that $|\hat{p}(n) - p(n)| \leq 10$ is 2099;

g) Number of n such that $|\hat{p}(n) - p(n)| \leq 20$ is 4146;

h) Number of n such that $|\hat{p}(n) - p(n)| \leq 30$ is 6130;

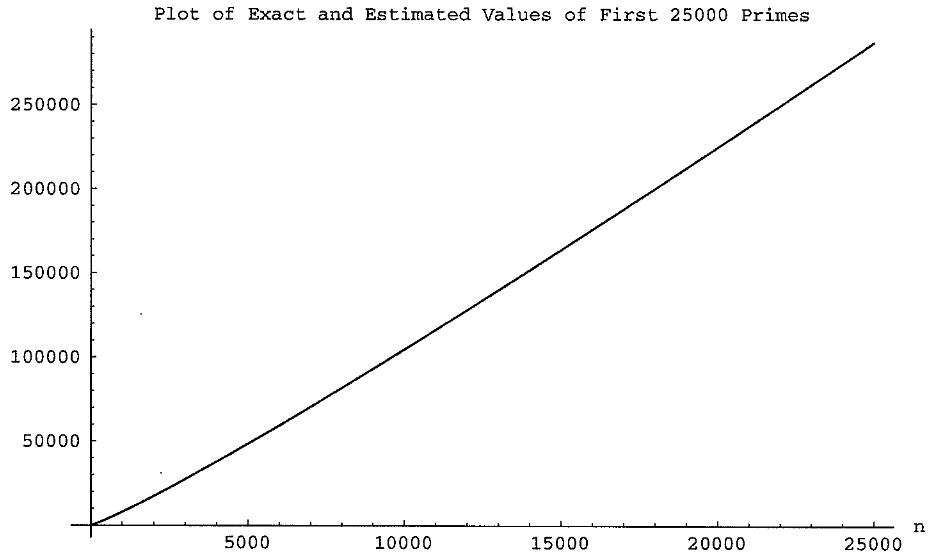
i) For the following selected values of $n, p(n)$ and $\hat{p}(n)$, with the associated % errors are as follows :

n	$p(n)$	$\hat{p}(n)$	% error
1000	7919	8046	1.6
2000	17389	17491	.59
5000	48611	48620	.02
10000	104729	104795	.06
15000	163841	163849	.005
20000	224737	224767	.01
25000	287117	287066	.02
50000	611953	612023	.01
100000	1299709	1299589	.009

j) If $\hat{\pi}(x)$ denotes the estimate of $\pi(x)$ obtained from an inversion of $\hat{p}(n)$, then for the following selected values of $n, \pi(n), \hat{\pi}(n)$ are as follows :

n	$\pi(n)$	$\hat{\pi}(n)$
10000	1229	1214
50000	5133	5127
100000	9592	9585
1000000	78498	78536
2500000	183072	183112
5000000	348513	348452

A plot of the exact values of the first 25,000 primes and the corresponding estimated values from our formula is given below for a visual examination. The two functions, when superimposed, look virtually identical in this plot.



Plot of Exact and Estimated Values of First 25000 Primes

4 Summary

Based on a theoretical result on the asymptotic behavior of the primes, we conducted a numerical search for an estimating formula for the primes of a specific form suggested by the theorem. The formula seems to have good numerical accuracy in estimating the primes as well as $\pi(x)$ for about the first 50,000 or even the first 100,000 primes and for x about 5 million. A refinement of this approach may produce another accurate formula over a bigger range.

Acknowledgement I am grateful to my teacher Gutti Jogesh Babu for his input, and his inspiration.

Bibliography

Dusart,P.(1999). The kth prime is greater than $k(\log k+\log\log k)$, Math. Comp., 68, 255, 411-15.

Ribenboim,P.(1991). The Little Book of Big Primes, Springer, New York.

Rosser,J.B.(1938). The nth prime is greater than $n\log n$, Proc. London Math. Soc., 45, 21-44.