COMPUTATION & NUMBER THEORY

Prime Analysis: Analytic and Probabilistic Approaches

Ethan Ashbrook, Collin Jung, Akash Pardeshi
James Schmidt (Graduate Student Mentor)

University of Illinois at Urbana-Champaign

Illinois Geometry Lab

Final Presentation
June 27, 2019

We would like to thank the David Frankel ('76) Fund for making this experience possible for Uni High Students.

Primes

- Motivation
- Q: Explore where primes are in a nontraditional way

Prime Number Theorem (PNT)

- Asymptotically, how many primes
- $\pi(x) =$ #primes less than or equal to x
- PNT: $\pi(x) \sim \frac{x}{\log(x)}$
- Weaker/Approximate notion of a function behaving like another one near infinity

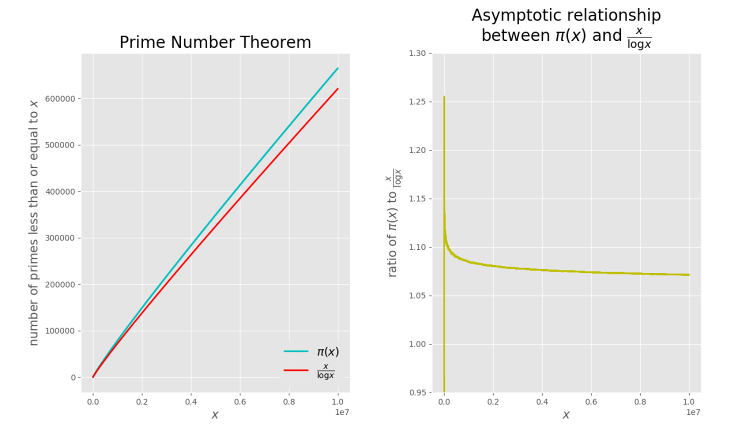

Figure: $\pi(x)$ and $x \log x$ are plotted to $x = 10^7$. Their ratio is shown as well and converges to 1 in the limit.

- Pretend that the primes are random.
- Then we can use ideas from probability to analyze them.
- Define a function $\gamma(n)$ that outputs the nth prime gap: $p_{n+1} - p_n$

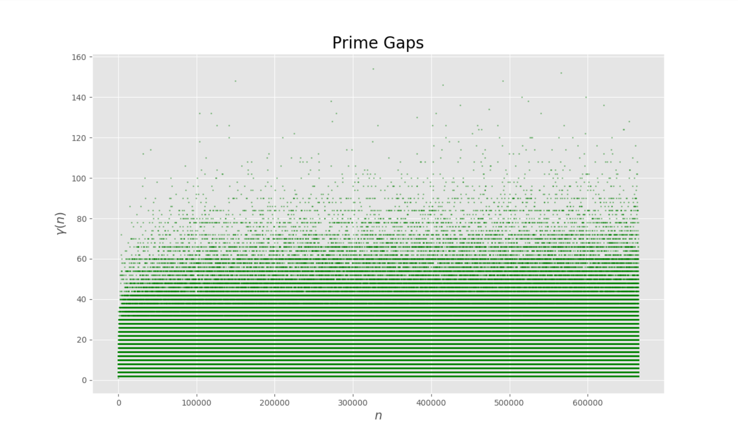

Figure: The prime gaps to $n = 10^7$

- Let $\mathcal{D}_n = \{\gamma(1), \gamma(2), \ldots, \gamma(n)\}$ be a data set.
- Define a function $\alpha(n)$ to be the mean of $\mathcal{D}_n$: $\alpha(n) = \frac{1}{n}\sum_{i=1}^{n} \gamma(i)$
- Define a function $\beta(n)$ to be the variance of $\mathcal{D}_n$: $\beta(n) = \frac{1}{n}\sum_{i=1}^{n} (\alpha(n) - \gamma(i))^2$

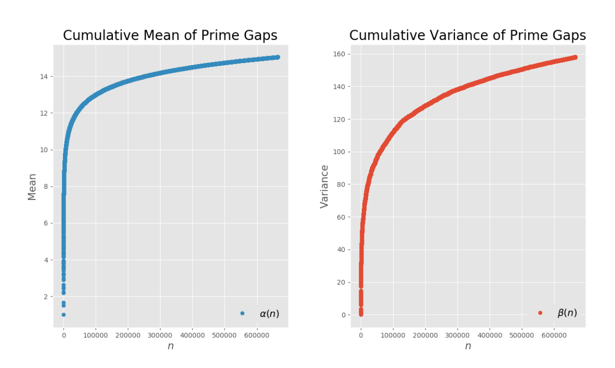

Figure: $\alpha(n)$ and $\beta(n)$ are shown up to $n = 10^7$

- Model $\alpha(n)$ with $a \log(n)^c + b$
- Perform regression on $\alpha(n)$ with this model.

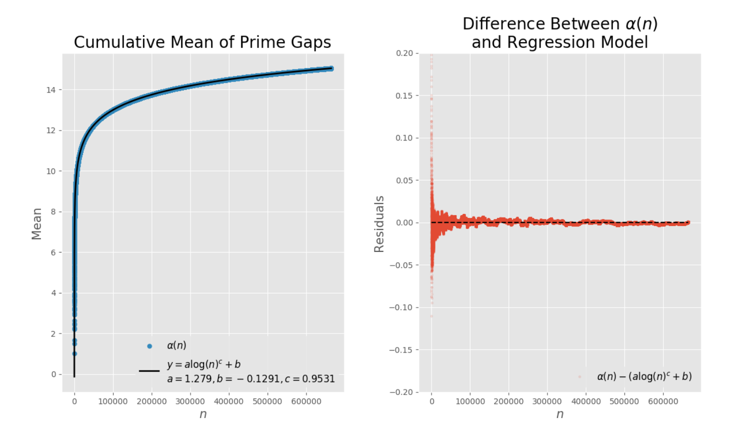

Figure: $\alpha(n)$ is shown to $n = 10^7$, along with the regression model. Their difference is also shown and seems to converge to 0.

- We compare how well each method predicts where the primes are
- Each of the two methods provide approximations for where the primes are
- The PNT implies the following: $p_n \sim n \log n$. Note that $p_n \nrightarrow n \log(n)$; the best we can conclude from the PNT is the nth prime is approximately $n \log n$
- The probabilistic approach suggests the following approximation of the nth prime: $p_{n-1} + \alpha(n-1)$
- Denote $\hat{p}_n$ as an approximation for the nth prime. Let $Er(n) = |p_n - \hat{p}_n|$ define a notion of an error on a prediction

$$\mathrm{Er}_{\mathrm{PNT}}(n) = |p_n - n\log(n)| \qquad \mathrm{Er}_{\mathrm{PRB}}(n) = |p_n - (p_{n-1} + \alpha(n-1))|$$

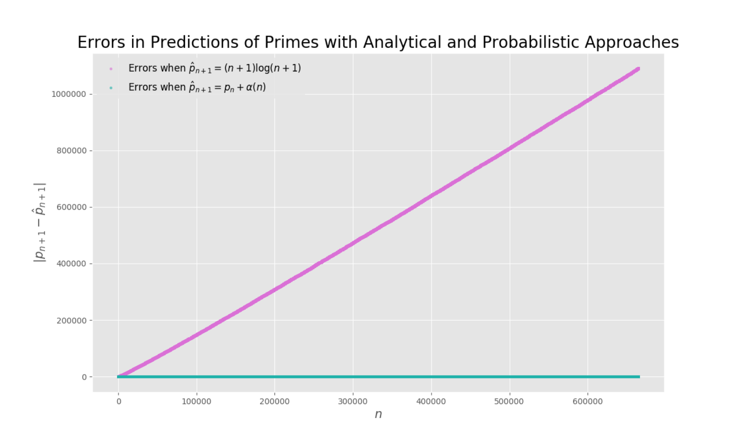

Figure: The errors for the two types of predictions are shown to $n = 10^7$

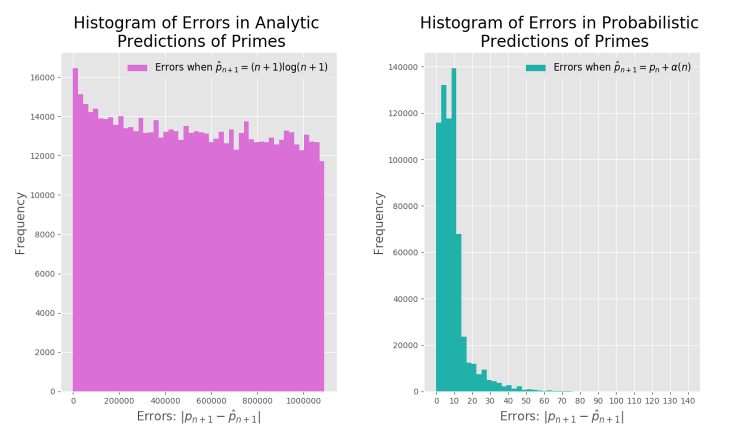

Figure: Above are histograms for the first ten million errors generated by the analytic and probabilistic approaches.

- Recall the analytic approximation: $\hat{p}_n = n\log(n)$
  and the probabilistic approximation: $\hat{p}_n = p_{n-1} + \alpha(n-1)$.
- Note that probabilistic approach required a stronger hypothesis, namely information about the previous prime. However, the probabilistic approach gave a much better prediction for the primes.
- Primes can be understand through different means and each approach yields unique insights into the nature of primes.

- Analyze higher moments of the gap data ($\mathcal{D}_n$)
- Formalize the notion of weighted least-squares regression for asymptotics

- $\varepsilon - \delta$ visualizations
- **Theorem**: if functions $f(n)$ and $an + b$ are asymptotic, then the coefficients of regression on the set $\{(1, f(1)), (2, f(2)), \ldots, (n, f(n))\}$ converge to the true parameters a and b.

$$f(n) \sim an + b \Rightarrow a, b = \lim_{n \to \infty} \arg\min_{a_n, b_n} \sum_{i=1}^{n} (f(i) - (a_n i + b_n))^2$$

- Lots of code