A report on prime numbers

Orges Leka
Limburg, Hessen, Germany

June 27, 2025

Contents

1 Primes as Expected Values

Let $N \geq 3$ be a fixed natural number. Define

$$\Omega_N = \{y \in \mathbb{N} \mid 1 < y \leq N,\ \gcd(y, N) = 1\}.$$

Set

$$P(A) = \frac{|A|}{|\Omega_N|} = \frac{|A|}{\phi(N) - 1} \quad \text{for any } A \subseteq \Omega_N,$$

where ϕ is Euler's totient function. In particular, for a single value $y \in \Omega_N$,

$$P(y) = P(\{y\}) = \frac{1}{\phi(N) - 1}.$$

Define the counting-function

$$\chi(N, y) = \big|\{\, a \in \mathbb{N} \mid 1 < a \leq y,\ \gcd(a, N) = 1\}\big|,$$

so that

$$P\big(Y \leq y\big) = \frac{\chi(N, y)}{\phi(N) - 1}.$$

We draw with replacement m independent samples $y_1, \ldots, y_m$ from Ω_N, each with probability $1/(\phi(N) - 1)$. Let

$$Y_{\min} = \min\{y_1, \ldots, y_m\}.$$

Then the distribution of $Y_{\min}$ is given by the standard order–statistic formula:

$$P\big(Y_{\min} \leq y\big) = 1 - \big(1 - P(Y \leq y)\big)^m = 1 - \Big(1 - \tfrac{\chi(N,y)}{\phi(N)-1}\Big)^m.$$

Hence the probability mass function is

$$P\big(Y_{\min} = y\big) = \Big(1 - \tfrac{\chi(N,y-1)}{\phi(N)-1}\Big)^m - \Big(1 - \tfrac{\chi(N,y)}{\phi(N)-1}\Big)^m.$$

Let

$$r = \min \Omega_N.$$

By definition r is the smallest prime not dividing N. Clearly

$$\chi(N, r - 1) = 0, \quad \chi(N, r) = 1.$$

Thus

$$P\big(Y_{\min} = r\big) = (1 - 0)^m - \Big(1 - \tfrac{1}{\phi(N)-1}\Big)^m = 1 - \Big(1 - \tfrac{1}{\phi(N)-1}\Big)^m.$$

Passing to the limit,

$$\lim_{m \to \infty} P\big(Y_{\min} = r\big) = 1,$$

and for every other $y \neq r$,

$$\lim_{m \to \infty} P\big(Y_{\min} = y\big) = 0.$$

Proposition 1.1. *As $m \to \infty$, the expected value of $Y_{\min}$ converges to r:*

$$E(Y_{\min}) = \sum_{y \in \Omega_N} y\, P(Y_{\min} = y) \quad \Longrightarrow \quad \lim_{m \to \infty} E(Y_{\min}) = r.$$

Proof. Interchange limit and sum (all terms nonnegative):

$$\lim_{m \to \infty} E(Y_{\min}) = \sum_{y \in \Omega_N} y \lim_{m \to \infty} P(Y_{\min} = y) = r \cdot 1 + \sum_{y \neq r} y \cdot 0 = r.$$

□

2 The Primorial Case

Let $P_k = p_1 p_2 \cdots p_k$ be the k-th *primorial*, $k \geq 2$. Then

$$\Omega_{P_k} = \{ y \leq P_k : \gcd(y, P_k) = 1\} \quad \Longrightarrow \quad r = \min \Omega_{P_k} = p_{k+1}.$$

Hence in this special case,

$$\lim_{m\to\infty} E(Y_{\min}) = p_{k+1},$$

showing that by repeatedly sampling coprime residues modulo a primorial, in the large-sample limit one "discovers" the next prime.

3 A recursive formula for the n-th prime

Let $p_1 < p_2 < \cdots < p_n$ be the first n primes and set

$$P_n \;=\; \prod_{i=1}^{n} p_i.$$

A well-known characterization of the $(n+1)$-th prime is

$$p_{n+1} \;=\; \min\bigl\{ x > 1 : \gcd(x, P_n) = 1 \bigr\}.$$

We now show that one may "soften" the minimum via a log-sum-exp approximation and still recover p_{n+1} in the limit.

Theorem 3.1. *Define, for each real $\rho > 0$,*

$$S(\rho) \;=\; \sum_{x=2}^{\infty} \exp\bigl(-\rho\, x\bigr)\, \exp\Bigl(-\bigl(\gcd(P_n, x)\bigr)^{\rho}\Bigr).$$

Then

$$p_{n+1} \;=\; \lim_{\rho\to\infty} \Bigl[-\tfrac{1}{\rho}\, \log S(\rho)\Bigr].$$

Proof. We will sandwich $S(\rho)$ between two expressions whose $-\frac{1}{\rho}\log$ both tend to p_{n+1}.

1. Lower bound.
Since the summand is strictly positive for every $x \geq 2$, in particular the term $x = p_{n+1}$ appears:

$$S(\rho) \;\geq\; \exp\bigl(-\rho\, p_{n+1}\bigr)\, \exp\bigl(-\gcd(P_n, p_{n+1})^{\rho}\bigr) \;=\; \exp\bigl(-\rho\, p_{n+1}\bigr)\, \exp(-1),$$

because $\gcd(P_n, p_{n+1}) = 1$. Hence

$$-\tfrac{1}{\rho} \log S(\rho) \;\leq\; -\tfrac{1}{\rho}\bigl[-\rho\, p_{n+1} - 1\bigr] \;=\; p_{n+1} + \tfrac{1}{\rho},$$

and so

$$\limsup_{\rho\to\infty} \Bigl[-\tfrac{1}{\rho} \log S(\rho)\Bigr] \;\leq\; p_{n+1}.$$

2. Upper bound.
Split the sum at $x = p_{n+1}$:

$$S(\rho) = \sum_{x=2}^{p_{n+1}-1} e^{-\rho x}\, e^{-(\gcd(P_n,x))^{\rho}} \;+\; \sum_{x=p_{n+1}}^{\infty} e^{-\rho x}\, e^{-(\gcd(P_n,x))^{\rho}}.$$

- For $2 \le x < p_{n+1}$, every prime dividing x is among $p_1, \dots, p_n$, so $\gcd(P_n, x) \ge 2$, hence

$$e^{-(\gcd(P_n,x))^\rho} \;\le\; e^{-2^\rho}.$$

There are at most $p_{n+1} - 2 \le 2p_n$ such x, so

$$\sum_{x=2}^{p_{n+1}-1} e^{-\rho x}\, e^{-(\gcd(P_n,x))^\rho} \;\le\; 2p_n\, e^{-2^\rho}.$$

- For $x \ge p_{n+1}$, we have $\gcd(P_n, x) = 1$, so

$$e^{-(\gcd(P_n,x))^\rho} = e^{-1}.$$

Thus

$$\sum_{x=p_{n+1}}^{\infty} e^{-\rho x}\, e^{-1} = e^{-1} \sum_{k=0}^{\infty} e^{-\rho\,(p_{n+1}+k)} = e^{-1}\, e^{-\rho p_{n+1}} \sum_{k=0}^{\infty} e^{-\rho k} = e^{-1}\, e^{-\rho p_{n+1}} \frac{1}{1-e^{-\rho}}.$$

For sufficiently large ρ, $e^{-\rho} < \frac{1}{2}$, so $\frac{1}{1-e^{-\rho}} \le 1 + 2e^{-\rho}$. Hence

$$\sum_{x=p_{n+1}}^{\infty} e^{-\rho x}\, e^{-(\gcd(P_n,x))^\rho} \;\le\; e^{-\rho p_{n+1} - 1}\left(1 + 2e^{-\rho}\right).$$

Combining both parts,

$$S(\rho) \;\le\; 2p_n e^{-2^\rho} \;+\; e^{-\rho p_{n+1} - 1}\left(1 + 2e^{-\rho}\right).$$

Since $2p_n e^{-2^\rho}$ decays super-exponentially as $\rho \to \infty$, we get

$$-\tfrac{1}{\rho} \log S(\rho) \;\ge\; -\tfrac{1}{\rho} \log\!\left[e^{-\rho p_{n+1} - 1}(1 + 2e^{-\rho}) + o(1)\right] \;=\; p_{n+1} + \tfrac{1}{\rho} - \tfrac{1}{\rho}\log(1 + 2e^{-\rho}) + o(1/\rho),$$

and thus

$$\liminf_{\rho\to\infty} \left[-\tfrac{1}{\rho}\log S(\rho)\right] \;\ge\; p_{n+1}.$$

By the squeeze (sandwich) theorem,

$$\lim_{\rho\to\infty} \left[-\tfrac{1}{\rho}\log S(\rho)\right] \;=\; p_{n+1},$$

as claimed. □

Remark. Numerically one can verify this convergence in SageMath for moderate n and ρ. While elegant, this formula does not (so far) yield new effective bounds on primes, since the limit occurs at extremely large ρ.

4 Random primes, 13.01.2022

By Euclid and the sieve of Eratosthenes,

$$p_k = \min\{x > 1 : \gcd(x, p_1 \cdots p_{k-1}) = 1\}.$$

This process is deterministic. Now introduce randomness:

$$y = \min_{i=1,\dots,k} \{x_i\}, \quad x_i \overset{\text{iid}}{\sim} \text{Uniform}\{2, \dots, N\} \text{ subject to } \gcd(x_i, N) = 1.$$

Extreme-value theory studies $\min, \max$ of iid variables as $k \to \infty$. The Gumbel distribution often appears (see Wikipedia), and there is empirical work linking primes and EVT, e.g. arXiv:1301.2242.

In our model, the CDF is

$$F(y) = \frac{\chi(N, y)}{\phi(N) - 1}, \quad \chi(N, y) = |\{2 \le a \le y : \gcd(a, N) = 1\}|,$$

and

$$P(Y_{\min} = y) = (1 - F(y-1))^m - (1 - F(y))^m.$$

Hence

$$E\left(Y_{\min}^{(N,m)}\right) = \sum_{\substack{2 \le k \le N \\ \gcd(k,N)=1}} k\Big[(1 - F(k-1))^m - (1 - F(k))^m\Big].$$

As $m \to \infty$, for $N = P_k$ the expectation tends to p_{k+1}.

Empirical SageMath code:

```
def PK(k):
    return prod(primes(nth_prime(k)))

def FF(N,m,k):
    return 1/(euler_phi(N)-1)*len([a for a in range(2,k+1) if gcd(a,N)
        ==1])

def EE(N,m):
    return sum([k*((1-FF(N,m,k-1))**m - (1-FF(N,m,k))**m)
                for k in range(2,N+1) if gcd(k,N)==1])

for n in range(3,7):
    print(nth_prime(n), EE(PK(n), PK(n)))
```

Conclusion

We've modeled random draws of units mod N and studied the minimum. As the sample size grows, the minimum concentrates on the smallest prime not dividing N, recovering in particular the next prime after a primorial.

References

[1] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer, 1976.

[2] I. Niven and H. S. Zuckerman, *An Introduction to the Theory of Numbers*, Wiley, 5th edition, 1991.

[3] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 6th edition, 2008.