Welcome back to my random tidbits file! When I come up with interesting problems, I will put them here.

1 Probability Distributions and Weight Loss

I was keeping track of my own weight when I realized that my scale was sufficiently inconsistent that my weight loss was dominated by the statistical noise. So then I was curious what the best way of mitigating this is, mean or median of multiple measurements. One would suspect it's the mean, or one would know simply by having taken any real statistics class, but I'm curious.

1.1 Mean-based averaging

This one is easy. Assume we have n iid variables X_i with mean μ and variance σ^2, then the random variable corresponding to their average $\langle X_i \rangle$ has mean μ and variance $\frac{\sigma^2}{n}$, so standard deviation $\frac{\sigma}{\sqrt{n}}$. Thus, we have an unbiased estimator of the true mean and a variance that falls off like $\sim n^{-1/2}$.

1.2 Median-based averaging

This one is a bit more fun. Let's start with $n = 3$, then defining $f(x)$ the probability density function and $F_X(x) = f_X(X \leq x)$ the cumulative distribution function, the probability density of the median $f_\eta(y)$ is given

$$f_\eta(y) = 6 f_X(y) F_X(y) \left(1 - F_X(y)\right) \tag{1}$$

the probability we choose one value greater than y the median and one less, multiplied by 6 because there 3 ways to choose which element is the median and $\binom{2}{1}$ binomial coefficient for exactly one element on each side. This seems to be a bit difficult to verify to be normalized in the general case, or that

$$\int_{-\infty}^{\infty} f_\eta(y)\, \mathrm{d}y = \int_{-\infty}^{\infty} \left[6 f_X(y) \int_{-\infty}^{y} f_X(\xi)\, \mathrm{d}\xi \int_{y}^{\infty} f_X(\zeta)\, \mathrm{d}\zeta \right] \mathrm{d}y = 1 \tag{2}$$

Let's just verify this in the uniform distribution case, and leave the general oase as an exercise to brighter colleagues. We consider the normalized uniform distribution $f_X(x) = 1, x \in [0,1]$, or $F_X(x) = x, x \in [0,1]$. We confirm that the expression for f_η is normalized:

$$\int_0^1 6y(1-y)\, \mathrm{d}y = 1 \tag{3}$$

We then wish to examine whether $f_\eta(y)$ is an unbiased estimator of μ. Again, we begin with examining a sub-case, where $f_X(x)$ is symmetric about its mean μ. This yields that $F_X(\mu) = 0.5$ and is odd about μ[1] and so that $F_X(y)\left(1 - F_X(y)\right)$ is also even/symmetric about μ. Finally, this implies that $f_\eta(y)$ as defined in Equation 1 is also symmetric about μ and we are done.

[1] This is a slight abuse of terminology: we mean that $F_X(x - \mu) - 0.5 = -(F_X(-(x-\mu)) - 0.5)$.

However, this analysis breaks down in the asymmetric case. We see that $F_X(y)\left(1-F_X(y)\right)$ is *always* symmetric about the median η of f_X, since $F_X(\eta)=0.5$. In general, the mean and median of a probability distribution are not equal, so there is no guarantee that $\langle f_\eta(y)\rangle = \langle f(y)\rangle$, and indeed we can verify for some contrived probability distribution such as

$$f_X(x)=\begin{cases}2 & 0\le x\le 0.25\\ 1 & 0.5\le x\le 1\\ 0 & \text{else}\end{cases} \tag{4}$$

that $\langle f_X(x)\rangle = 0.4375$ while

$$\langle f_\eta(y)\rangle = \int_0^{0.25} 24y^2(1-2y)\,\mathrm{d}y + \int_{0.5}^{1} 6y^2(1-y)\,\mathrm{d}y \tag{5}$$

$$\approx 0.4218 \tag{6}$$

This should not have surprised us: we're trying to use a median to estimate the mean of a distribution, and the two are equal when the PD is symmetric and unequal otherwise.

The above analyses probably generalizes to median-of-n trials, where with a symmetric PD we have the being a unbiased estimator of the mean and with an asymmetric an biased estimator, for any parity of n, but I'm too lazy to check this out and will take it on faith. For reference, we assert the generalization of Equation 1 below for odd $N=2m+1$ trials below

$$f_{\eta,2m+1}(y)=N\binom{2m}{m}f_X(y)\left(F_X(y)\right)^m\left(1-F_X(y)\right)^m \tag{7}$$

which is simply generalizing to the concept of "m elements on either side of y."

It seems difficult to compare these median-based results (many of which could probably be strengthened) to the mean based results in the case of an arbitrary PDF, so let's specialize to a few tractable cases.

1.3 Uniform Distribution

I'm tired of not obtaining usable results, so let's simplify the discussion considerably and assume that we have a uniform probability distribution, or that $X\in[\mu-a,\mu+a]$. In this case the median-of-three also provides for an unbiased estimator as shown above. What is the variance of this estimator then?

1.3.1 Mean-based

Let's first examine the results of a mean-based estimation of μ. Call $\hat{\mu}_N$ the estimator generated by averaging N samplings, then we know that $\langle\hat{\mu}_N\rangle=\mu$ by linearity of expectation and $\sigma^2_{\hat{\mu}_N}=\dfrac{\sigma_X^2}{N}$ by linearity of

variance, so it remains to compute σ_X^2, which is given by

$$\sigma_X^2 = \left\langle X^2 \right\rangle - \langle X \rangle^2 \tag{8}$$

$$= \int_{\mu-a}^{\mu+a} \frac{1}{2a} x^2 \, \mathrm{d}x - \mu^2 \tag{9}$$

$$= \frac{6\mu^2 a + 2a^3}{6a} - \mu^2 \tag{10}$$

$$= \frac{a^2}{3} \tag{11}$$

Thus, $\sigma_{\hat{\mu}_N}^2 = \dfrac{a^2}{3N}$.

1.3.2 Median-based, $N = 3$

Now for the median-based approach. Denote $\tilde{\mu}_N$ the estimator generated by taking the median of N samplings, then we know that $\langle \tilde{\mu}_N \rangle = \mu$ nonetheless because the uniform PD is a symmetric probability distribution. It thus remains to compute $\left\langle \sigma_{\tilde{\mu}_N}^2 \right\rangle$. This seems nontrivial, so let's start with $N = 3$:

$$\left\langle \tilde{\mu}_3^2 \right\rangle = \int_{-\infty}^{\infty} 6 f_X(x) F_X(x) \left(1 - F_X(x)\right) \, x^2 \mathrm{d}x \tag{12}$$

$$= \int_{\mu-a}^{\mu+a} \frac{6}{2a} \frac{x - (\mu - a)}{2a} \frac{(\mu + a) - x}{2a} x^2 \, \mathrm{d}x \tag{13}$$

$$= \int_{-a}^{a} \frac{6}{2a} \frac{a + y}{2a} \frac{a - y}{2a} (y + \mu)^2 \, \mathrm{d}y \tag{14}$$

$$= \int_{-a}^{a} \left[\frac{6}{8a^3} \left(a^2 y^2 + a^2 2y\mu + a^2\mu^2 - y^4 - 2\mu y^3 - y^2\mu^2 \right) \right] \, \mathrm{d}y \tag{15}$$

$$= \frac{6}{8a^3} \left[\frac{(a^2 - \mu^2) y^3}{3} - \frac{y^5}{5} \right]_{-a}^{a} + \frac{3\mu^2}{2} \tag{16}$$

$$= \frac{6}{8a^3} \left[\frac{\left(a^2 - \mu^2\right) 2a^3}{3} - \frac{2a^5}{5} \right] + \frac{3\mu^2}{2} \tag{17}$$

$$= \mu^2 + \frac{a^2}{5} \tag{18}$$

and so $\sigma_{\tilde{\mu}_3}^2 = \dfrac{a^2}{5}$. Compare this to $\sigma_{\hat{\mu}_3}^2 = \dfrac{a^2}{9}$ and we see that the mean-based estimation has lower uncertainty.

1.3.3 Median-based, arbitrary N

Armed with this, let's also try to compute for arbitrary, odd $N = 2m + 1$, for which we have

$$\left\langle \tilde{\mu}_N^2 \right\rangle = N\binom{2m}{m} \int_{-a}^{a} \frac{1}{2a} \left(\frac{a^2 - y^2}{4a^2} \right)^m (y + \mu)^2 \, \mathrm{d}y \tag{19}$$

Now, there's probably a cool combinatorial way to evaluate this, but let's just care about asymptotic behavior. Then

$$\lim_{N\to\infty} \left\langle \tilde{\mu}_N^2 \right\rangle \approx N \frac{2^{2m}\sqrt{2m}}{m} \int_{-a}^{a} \frac{1}{2a} \frac{1}{4^m} \left(1 - \frac{y^2}{a^2} \right)^m (y + \mu)^2 \, \mathrm{d}y \tag{20}$$

$$\approx \frac{1}{2a}\sqrt{8m} \int_{-a}^{a} \left(1 - \frac{y^2}{a^2} \right)^m (y + \mu)^2 \, \mathrm{d}y \tag{21}$$

where we approximate $N \approx 2m$. Now, we know that $\left(1 - \frac{y^2}{a^2} \right)^m$ is going to fall off sharply to 0 as y increases, so we can approximate (for some normalization factor A)

$$\int_{-a}^{a} \left(1 - \frac{y^2}{a^2} \right)^m \mathrm{d}y \sim A \int_{-a/\sqrt{m}}^{a/\sqrt{m}} 1 - \frac{my^2}{a^2} \, \mathrm{d}y \tag{22}$$

$$\lim_{N\to\infty} \left\langle \tilde{\mu}_N^2 \right\rangle \approx \frac{A}{2a}\sqrt{8m} \int_{-a/\sqrt{m}}^{a/\sqrt{m}} \left(1 - \frac{my^2}{a^2} \right) (y + \mu)^2 \, \mathrm{d}y \tag{23}$$

To compute A, we require that the coefficient of μ^2 be 1 so that the difference $\left\langle \tilde{\mu}_N^2 \right\rangle - \left\langle \tilde{\mu}_N \right\rangle^2$ does not depend in first order on μ. It's clear that since the integral is symmetric, we need only consider even powers of y, and so our integral becomes

$$\lim_{N\to\infty} \left\langle \tilde{\mu}_N^2 \right\rangle = \frac{A\sqrt{8m}}{2a} \int_{-a/\sqrt{m}}^{a/\sqrt{m}} \left(1 - \frac{my^2}{a^2} \right) \left(y^2 + \mu^2 \right) \, \mathrm{d}y \tag{24}$$

$$= \frac{A\sqrt{8m}}{2a} \int_{-a/\sqrt{m}}^{a/\sqrt{m}} \mu^2 + \left(1 - \frac{m\mu^2}{a^2} \right) y^2 - \frac{my^4}{a^2} \, \mathrm{d}y \tag{25}$$

$$= \frac{A\sqrt{8m}}{2a} \left[\frac{2\mu^2 a}{\sqrt{m}} + \left(1 - \frac{m\mu^2}{a^2} \right) \left(\frac{2}{3} \frac{a^3}{m^{3/2}} \right) - \frac{2ma^5}{5a^3 m^{5/2}} \right] \tag{26}$$

$$= A\frac{\sqrt{32}}{3}\mu^2 + A\frac{4\sqrt{2}}{15}\frac{a^2}{m} \tag{27}$$

and so we find that $A = \dfrac{3}{\sqrt{32}}$ and finally

$$\sigma^2_{\tilde{\mu}^2_N} = \frac{a^2}{5m} \tag{28}$$

The agreement for $N = 3, m = 1$ is a bit uncanny, but let's try to verify this computationally before jumping for joy.

This is a polynomial relationship on m, so we can sample m logarithmically to computationally verify our result. The obtained results are as follows in Figure 1.

The histogram is plotted merely out of curiosity, but seems to suggest a normal distribution per the Law of Large Numbers. Nonetheless, Equation 28 seems to be slightly off. It perfectly agrees in the $N = 3$ case as can be verified by simulation, but eventually grows to be a factor of approximately 2 off.

So it turns out our uncanny success for $N = 3, m = 1$ was a pure stroke of luck, and our expression isn't precisely correct. Nonetheless, we can make a few plots to figure out numerically how well median vs. mean based averaging performs, and the degredation of our estimate over N. These plots are

1.3.4 Median-based, arbitrary N, reworked

Let's try to include the truncated terms in $\left(1 - \dfrac{y^2}{a^2}\right)^m$, since they really are rather non-small compared to the leading term that we kept. Where we had before put $1 - \dfrac{my^2}{a^2}$, we should instead put

$$\left(1 - \frac{y^2}{a^2}\right)^m = \sum_{k=0}^{m} \binom{m}{k} \left(-\frac{y^2}{a^2}\right)^k \tag{29}$$

$$\lim_{N\to\infty} \left\langle \tilde{\mu}^2_N \right\rangle \approx \frac{A}{2a}\sqrt{8m} \int_{-a/\sqrt{m}}^{a/\sqrt{m}} \sum_{k=0}^{m} \binom{m}{k} \left(-\frac{y^2}{a^2}\right)^k (y+\mu)^2 \, \mathrm{d}y \tag{30}$$

We approximate $\dbinom{m}{k} \approx \dfrac{m^k}{k!}$ since higher terms in k are attenuated anyways. Using the same parity argument to kill the term odd in y, we rewrite

$$\lim_{N\to\infty} \left\langle \tilde{\mu}^2_N \right\rangle \approx \frac{A}{2a}\sqrt{8m} \int_{-a/\sqrt{m}}^{a/\sqrt{m}} \sum_{k=0}^{m} \frac{m^k}{k!} \left(-\frac{y^2}{a^2}\right)^k (y^2+\mu^2) \, \mathrm{d}y \tag{31}$$

Examine first the μ^2 coefficient

$$1 = \frac{A}{2a}\sqrt{8m}\sum_{k=0}^{m}\frac{m^k}{k!}\int_{-a/\sqrt{m}}^{a/\sqrt{m}}\left(-\frac{y^2}{a^2}\right)^k \mathrm{d}y \tag{32}$$

$$= \frac{A}{2a}\sqrt{8m}\sum_{k=0}^{m}\frac{m^k}{k!(2k+1)}2\left(\frac{a}{m^{k+1/2}}\right)(-1)^k \tag{33}$$

$$= A\sqrt{8}\sum_{k=0}^{m}\frac{(-1)^k}{k!(2k+1)} \tag{34}$$

and the other term

$$\lim_{N\to\infty}\sigma^2_{\tilde{\mu}_N} \approx \frac{A}{2a}\sqrt{8m}\int_{-a/\sqrt{m}}^{a/\sqrt{m}}\sum_{k=0}^{m}\frac{m^k}{k!}\left(-\frac{y^2}{a^2}\right)^k y^2\,\mathrm{d}y \tag{35}$$

$$= \frac{A}{2a}\sqrt{8m}\sum_{k=0}^{m}\frac{m^k}{k!(2k+3)}2\left(\frac{a^3}{m^{k+3/2}}\right)(-1)^k \tag{36}$$

$$= \frac{A\sqrt{8}a^2}{m}\sum_{k=0}^{m}\frac{(-1)^k}{k!(2k+3)} \tag{37}$$

$$= \frac{a^2}{m}\frac{\sum_{k=0}^{m}\frac{(-1)^k}{k!(2k+3)}}{\sum_{k=0}^{m}\frac{(-1)^k}{k!(2k+1)}} \tag{38}$$

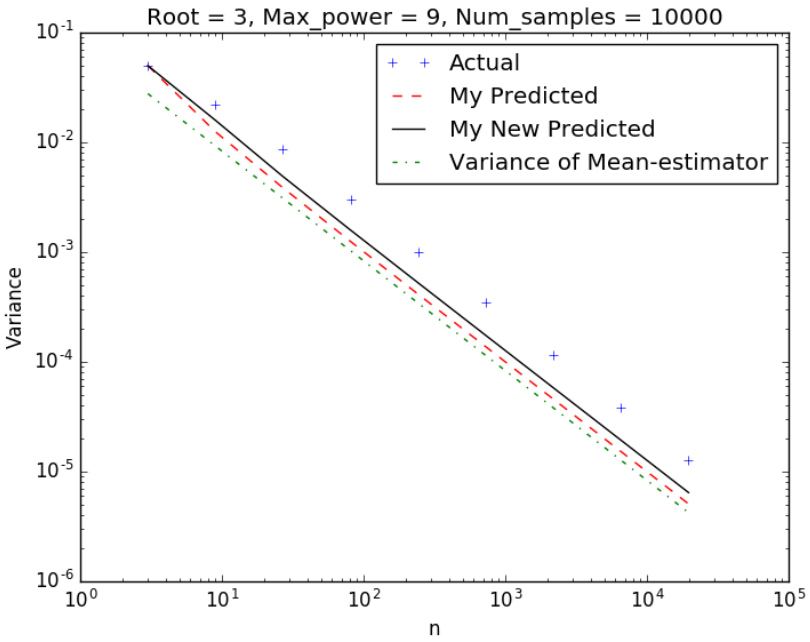
and we find that we reproduce our previous result for $N = 3$. Crunching the numbers, we get something slightly better, though since factorials fall off so quickly the change is very slight. The results are shown in Figure 2.

1.3.5 Further ruminations (TBC)

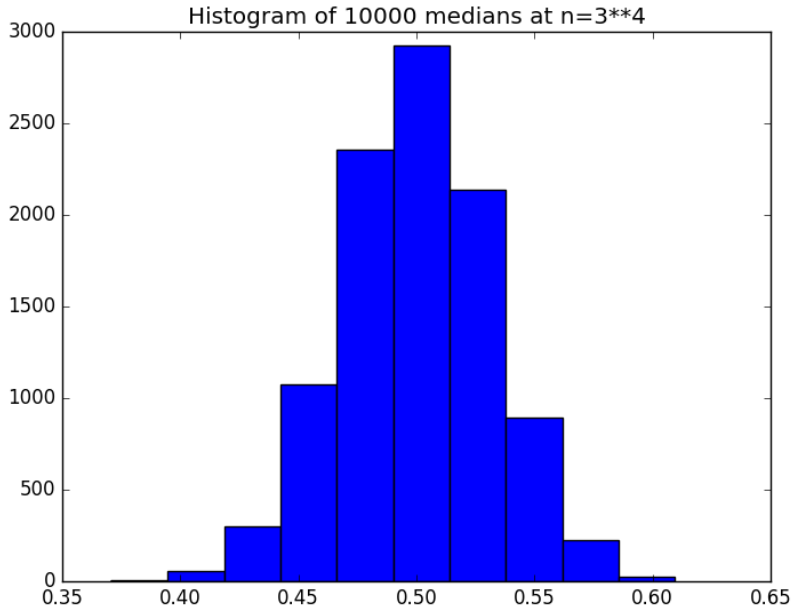
The approximation where we took the integral over interval $[-a/\sqrt{m}, a/\sqrt{m}]$ seems to be the last point of contention, as it bears noting that if we allow a degree of freedom in the choice of range $\left[-Ba/\sqrt{m}, Ba/\sqrt{m}\right]$ that our choice of B propagates as a factor of B^{2k+3} to the summation in the numerator of **??** and B^{2k+1} to the summation in the denominator. Thus, our choice of B has nontrivial implications on the exact prefactor we obtain.

1.4 Open Questions

- Is there any way to find the missing factor on median-based averaging for a uniform-distribution and arbitrary N?
- If we have discretized measurements, what are the statistics of mode-based averaging?
- Did I actually normalize the median-based averaging correctly, for a general probability distribution?

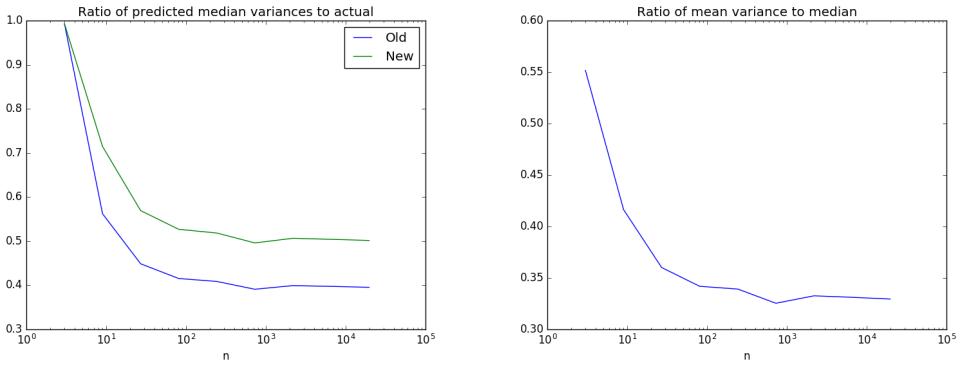


(a) Plot of medians as a function of N



(b) Histogram of 1000 medians at a single value of N

Figure 1: Computational results for our medians result. Used $\mu = 0.5$, $a = 0.5$, or a uniform sampling $[0, 1]$. Sampled over $n = 3^{[1,9]}$ with 10000 samples at each value of n.



(a) Seems asymptotically close to $2/5$, $1/2$ maybe?

(b) Seems asymptotically close to $1/3$?

Figure 2: A couple ratios of interest. Same sampling as in Figure 1.

2 Feynman-style number theory

In case you have not yet seen `http://www.lbatalha.com/blog/feynman-on-fermats-last-theorem` yet, it's quite a fun read! Would recommend. That sort of thinking inspired this section.

2.1 Asymptotic behavior of primes

Call $\Pi(N)$ the prime number counting function, how many primes are below N. The Prime Number Theorem is a well known result that postulates two approximations to $\Pi(N)$:

$$\Pi(N) \approx \frac{N}{\log N} \approx \int_2^N \frac{1}{\log x}\,\mathrm{d}x \tag{39}$$

We will attempt to derive the latter approximation. Consider $P(N)$ the probability density that N is a prime, roughly the statement "if I randomly choose a number near N, what is the probability it is a prime?" The relationship between $P(N)$ and $\Pi(N)$ is then

$$P(N) = \frac{\mathrm{d}\Pi}{\mathrm{d}N} \tag{40}$$

To attempt to derive $P(N)$, consider that a number N is prime iff it is not divisible by any primes less than it. Thus, we have that

$$P(N) \approx \prod_{p \in primes}^{N} \left(1 - \frac{1}{p}\right) \tag{41}$$

Taking a leap of faith, we recognize that two consecutive contributions to the product above differ roughly by $\frac{1}{P(p)}$, the local inverse probability density that p is prime. Thus, we can rewrite each contribution as $\frac{1}{P(p)}$ contributions of $\left(1 - \frac{1}{p}\right)^{P(p)}$, and then allow p to run over all integers. We thus propose the approximation

$$P(N) \approx \prod_{k=2}^{N} \left(1 - \frac{1}{k}\right)^{P(k)} \tag{42}$$

Taking the logarithm of both sides, we obtain

$$\log P(N) = \sum_{k=2}^{N} P(k) \log\left(1 - \frac{1}{k}\right) \tag{43}$$

Approximating the right hand side with an integral, we obtain

$$\log P(N) = \int_2^N P(k) \log\left(1 - \frac{1}{k}\right)\,\mathrm{d}k \tag{44}$$

Differentiating both sides now, we obtain

$$\frac{P'(N)}{P(N)} = P(N)\log\left(1-\frac{1}{N}\right) \tag{45}$$

$$\frac{\mathrm{d}P}{\mathrm{d}N} = P^2\log\left(1-\frac{1}{N}\right) \tag{46}$$

$$\frac{\mathrm{d}P}{P^2} = \mathrm{d}N\log\left(1-\frac{1}{N}\right) \tag{47}$$

$$-\frac{1}{P} = N\log\left(1-\frac{1}{N}\right) - \log(N-1) \tag{48}$$

$$P(N) = \frac{1}{\log(N-1)+O(1)} \tag{49}$$

$$\approx \frac{1}{\log N} \tag{50}$$

This recovers the expression $\Pi(N) = \int_2^N P(N)\,\mathrm{d}N = \int_2^N \frac{1}{\log N}\,\mathrm{d}N$.

2.2 Scratch work

What follows is me working out loud, which is a lot less interesting.

It's a well-known result (Prime Number Theorem) that the number of primes below N is approximated by $\Pi(N) = N/\log(N)$. Can we try to get a handle on this behavior via application of continuum analysis?

One way of thinking of the problem is to instead look at it from a probabilistic standpoint, that arbitrarily choosing a number n, it has some probability of being prime. Can we estimate this probability and recover the prime number theorem? We should be able to obtain

$$\frac{\mathrm{d}\Pi}{\mathrm{d}N} \approx \frac{\log N - 1}{\log^2(N)} \tag{51}$$

2.2.1 First attempt

Let's consider the probability that some large number N is divisible by some divisor d; this is just $\frac{1}{d}$. We might think that the probability that N is prime then just the product of probabilities it is not divisible by any number smaller than it

$$P(N) = \prod_{k=2}^{N}\left(1-\frac{1}{k}\right) \tag{52}$$

To try to evaluate this product, we take the logarithm of both sides

$$\log P(N) = \sum_{k=2}^{N} \log\left(1 - \frac{1}{k}\right) \tag{53}$$

$$\approx \int_{k=2}^{N} \log\left(1 - \frac{1}{k}\right)\, \mathrm{d}k \tag{54}$$

$$\tag{55}$$

To compute this antiderivative, it's easiest to separate the integrand

$$\int \log\left(\frac{k-1}{k}\right)\, \mathrm{d}k = \int \log(k-1)\, \mathrm{d}k - \int \log k\, \mathrm{d}k \tag{56}$$

$$= (k-1)\log(k-1) - k - k\log(k) + k + C \tag{57}$$

$$= k\log\left(1 - \frac{1}{k}\right) - \log(k-1) + C \tag{58}$$

with C some undetermined constant that becomes irrelevant when we consider the definite integral. Thus, we return to our primary expression

$$\log P(N) \sim N\log\left(1 - \frac{1}{N}\right) - \log(N-1) \tag{59}$$

where we drop the evaluation of the antiderivative at $k = 2$ since it's a constant in the scaling. Then, we find

$$P(N) \sim \frac{\left(1 - \frac{1}{N}\right)^N}{N-1} = \frac{1/e}{N-1} \tag{60}$$

In fact, a quick google search shows that Equation 52 evaluates to $\frac{1}{N}$, and so our result is pretty reasonable; we're off by a constant factor since our integral approximation Equation 54 misestimates by a constant factor, no surprise there. So where did we go wrong?

2.2.2 Second attempt

The iusse, as some people smarter than me may have noticed, is that our expression Equation 52 is faulty: we should only be multiplying *over primes*! While this is correct, primes are not divisible by any primes smaller than them, it's a bit difficult to handle under our present formalism, where we only attach a probability to a number's being prime or not.

Let's think carefully about how to integrate this into our formalism. If a number k is not prime, it should contribute 1 to our product, and if it is prime then it should contribute $\left(1 - \frac{1}{k}\right)$. Since we're doing products, the natural way to "average" is via geometric mean, so we modify expression Equation 52 to

$$P(N) = \prod_{k=2}^{N} \left(1 - \frac{1}{k}\right)^{P(k)} \tag{61}$$

where we average each k-th contribution as $\left(1-\frac{1}{k}\right)^{P(k)}(1)^{1-P(k)}$ geometric mean[2]. Doing the usual trick,

$$\log P(N) = \int_2^N P(k)\log\left(1-\frac{1}{k}\right)\,\mathrm{d}k \tag{62}$$

Differentiating both sides,

$$\frac{P'(N)}{P(N)} = P(N)\log\left(1-\frac{1}{N}\right) \tag{63}$$

$$\frac{\mathrm{d}P}{\mathrm{d}N} = P^2(N)\log\left(1-\frac{1}{N}\right) \tag{64}$$

$$\frac{\mathrm{d}P}{P^2} = \log\left(1-\frac{1}{N}\right)\mathrm{d}N \tag{65}$$

$$-\frac{1}{P} = N\log\left(1-\frac{1}{N}\right) - \log\left(N-1\right) \tag{66}$$

$$P(N) \approx \frac{1}{\log N} \tag{67}$$

Interestingly, this expression is a better approximation to $\Pi(N)$ than the aforementioned $\Pi(N) \approx \frac{N}{\log(N)}$, so it looks like this is a satisfactory conclusion, namely that

$$\Pi(N) \approx \int_2^N \frac{1}{\log(m)}\,\mathrm{d}m \tag{68}$$

However, we pursue one last direction of thought out of curiosity.

2.2.3 Third attempt

In Equation 61, maybe we only need to check up until $\sqrt{N}$ in the product. Continuing our thought above, we obtain

$$\frac{P'(N)}{P(N)} = P(\sqrt{N})\log\left(1-\frac{1}{\sqrt{N}}\right) \tag{69}$$

$$\approx -\frac{P(\sqrt{N})}{\sqrt{N}} \tag{70}$$

At this point, our expression doesn't seem particularly amenable to solution, but we can at least check

[2]Intuitively, this means that we need to multiply $\frac{1}{P(k)}$ of these factors before getting a single one that contributes, i.e. the distance between primes.

how well $P(N) \sim \dfrac{1}{\log N}$ works:

$$\frac{-\frac{1}{N\log^2 N}}{\frac{1}{\log N}} = -\frac{2}{\sqrt{N}\log N} \tag{71}$$

$$-\frac{1}{N\log N} = \frac{2}{\sqrt{N}\log N} \tag{72}$$

which doesn't seem to work too well. How about the original estimate $P(N) \sim \dfrac{\log N - 1}{\log^2 N}$?

$$\frac{\frac{2-\log N}{N\log^3 N}}{\frac{\log N-1}{\log^2 N}} = -\frac{\frac{\log\sqrt{N}-1}{\log^2\sqrt{N}}}{\sqrt{N}} \tag{73}$$

$$\frac{2-\log N}{N\log N\left(\log N-1\right)} = \frac{2\left(2-\log N\right)}{\sqrt{N}\log^2 N} \tag{74}$$

which is even worse. The obvious problem is that the $\sqrt{N}$ has nowhere tho go since the probability density P depends only on the logarithm of N. So interesting, considering the further optimization of only going up to $\sqrt{N}$ ruins the accuracy of our prediction!