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 $\dfrac{\sigma^2}{n}$, so standard deviation $\dfrac{\sigma}{\sqrt{n}}$. 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 $F_X(x) = f_X(X \le x)$ 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 $\dbinom{2}{1}$ 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

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

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:

$$\int\limits_{0}^{1} 6y(1-y)\,\mathrm{d}y = 1 \tag{3}$$

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

$$\begin{aligned} \langle f_\eta(y)\rangle &= \int\limits_0^{0.25} 24y^2(1-2y)\,\mathrm{d}y + \int\limits_{0.5}^{1} 6y^2(1-y)\,\mathrm{d}y \qquad (5) \\ &\approx 0.4218 \qquad (6) \end{aligned}$$

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

$$\sigma_X^2 = \left\langle X^2 \right\rangle - \left\langle X \right\rangle^2 \tag{8}$$
$$= \int\limits_{\mu-a}^{\mu+a} \frac{1}{2a} x^2 \,\mathrm{d}x - \mu^2 \tag{9}$$
$$= \frac{6\mu^2 a + 2a^3}{6a} - \mu^2 \tag{10}$$
$$= \frac{a^2}{3} \tag{11}$$

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

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 $\left\langle \sigma^2_{\tilde{\mu}_N} \right\rangle$. This seems nontrivial, so let's start with $N = 3$:

$$\left\langle \tilde{\mu}_3^2 \right\rangle = \int\limits_{-\infty}^{\infty} 6 f_X(x) F_X(x) \left(1 - F_X(x)\right) \, x^2 \mathrm{d}x \tag{12}$$
$$= \int\limits_{\mu-a}^{\mu+a} \frac{6}{2a} \frac{x - (\mu - a)}{2a} \frac{(\mu + a) - x}{2a} x^2 \,\mathrm{d}x \tag{13}$$
$$= \int\limits_{-a}^{a} \frac{6}{2a} \frac{a + y}{2a} \frac{a - y}{2a} (y + \mu)^2 \,\mathrm{d}y \tag{14}$$
$$= \int\limits_{-a}^{a} \left[\frac{6}{8a^3} \left(a^2 y^2 + a^2 2y\mu + a^2\mu^2 - y^4 - 2\mu y^3 - y^2\mu^2 \right) \right] \,\mathrm{d}y \tag{15}$$
$$= \frac{6}{8a^3} \left[\frac{(a^2 - \mu^2) y^3}{3} - \frac{y^5}{5} \right]_{-a}^{a} + \frac{3\mu^2}{2} \tag{16}$$
$$= \frac{6}{8a^3} \left[\frac{\left(a^2 - \mu^2\right) 2a^3}{3} - \frac{2a^5}{5} \right] + \frac{3\mu^2}{2} \tag{17}$$
$$= \mu^2 + \frac{a^2}{5} \tag{18}$$

and so $\sigma^2_{\tilde{\mu}_3} = \dfrac{a^2}{5}$. Compare this to $\sigma^2_{\hat{\mu}_3} = \dfrac{a^2}{9}$ 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

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

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

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

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

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)

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

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

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

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

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

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

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

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

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

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

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

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

Examine first the μ^2 coefficient

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

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

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

and the other term

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

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

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

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

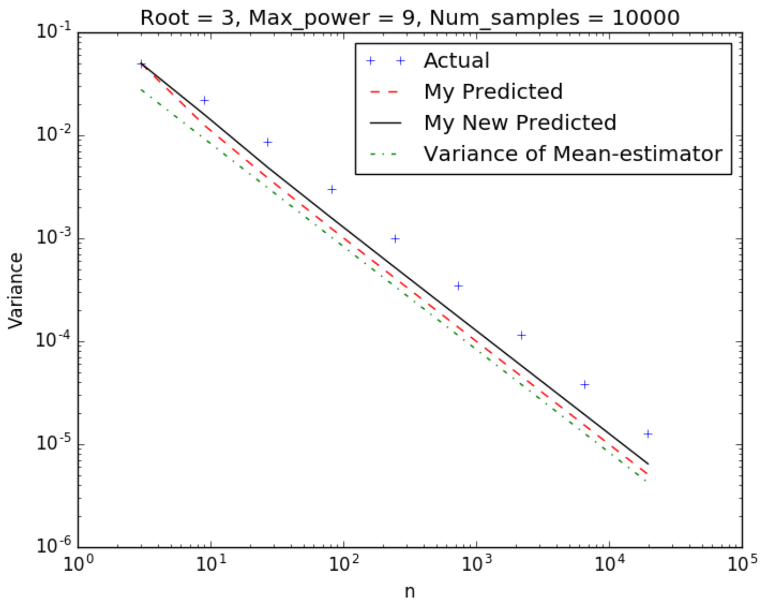
and we find that we reproduce our previous result for $N = 3$. Crunching the numbers, we get something slightly better, though since factorials fall off so quickly the change is very slight. The results are shown in Figure 2.

1.3.5 Further ruminations (TBC)

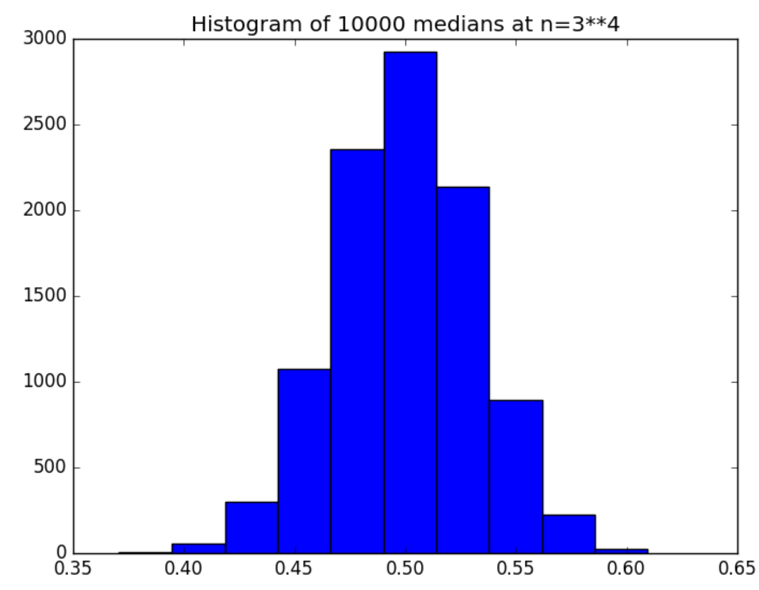
The approximation where we took the integral over interval $[-a/\sqrt{m}, a/\sqrt{m}]$ seems to be the last point of contention, as it bears noting that if we allow a degree of freedom in the choice of range $\left[-Ba/\sqrt{m}, Ba/\sqrt{m}\right]$ that our choice of B propagates as a factor of B^{2k+3} to the summation in the numerator of Equation 38 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?

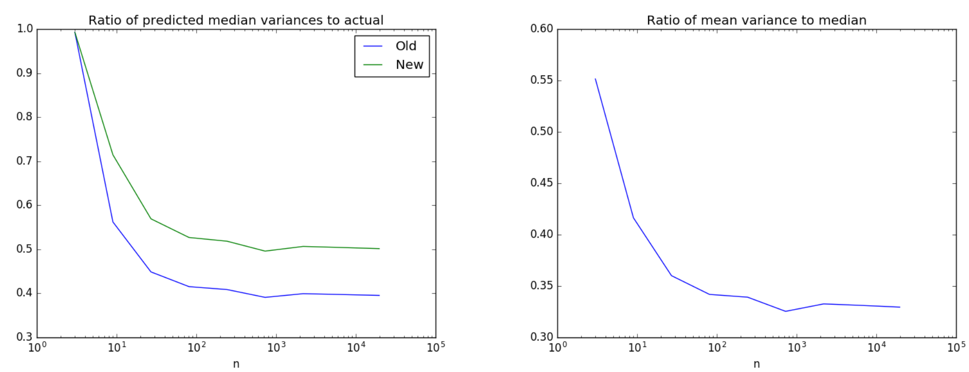


(a) Plot of medians as a function of N



(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 $\Pi(N)$:

$$\Pi(N) \approx \frac{N}{\log N} \approx \int\limits_2^N \frac{1}{\log x}\,\mathrm{d}x \tag{39}$$

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

$$\log P(N) = \int\limits_2^N P(k) \log\left(1 - \frac{1}{k}\right)\,\mathrm{d}k \tag{44}$$

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 $\Pi(N) = \int\limits_2^N P(N)\,\mathrm{d}N = \int\limits_2^N \frac{1}{\log N}\,\mathrm{d}N$.

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!

3 Ellipsoidal surface areas

We all know that ellipses do not have a closed form for their arclength, but their enclosed area is well defined, namely $A = \pi ab$. This can be seen by defining an ellipse as a projection of a circle by unevenly scaling the axes, and noting that an area element $\mathrm{d}x\mathrm{d}y$ scales linearly with the projection factors.

One series approximation to the arclength can be computed by noting the following: if S is the arclength of an ellipse, then $S\mathrm{d}n$ for some small $\mathrm{d}n$ estimates the change in area by enlarging the ellipse.

Systematically, exhibit an ellipse with axis lengths a, b, such that its area is πab. Then, say that we extend both axes by some $\mathrm{d}\epsilon$, then its area becomes $\pi ab + \pi(a+b)\mathrm{d}\epsilon + \mathcal{O}(\mathrm{d}\epsilon^2)$, and the change in area is $\pi(a+b)\mathrm{d}\epsilon + \mathcal{O}(\mathrm{d}\epsilon^2)$. This implies that the arclength of an ellipse to first order is $\pi(a+b)$, which seems to make sense for $a = b$.

This isn't particularly radical, and neither is this entire section, but we can verify it to be reasonable for three dimensions as well:

$$V = \frac{4}{3}\pi abc + \frac{4}{3}\pi(ac + bc + ab)\mathrm{d}\epsilon + \mathcal{O}(\mathrm{d}\epsilon^2) \tag{75}$$

$$S = \frac{4}{3}\pi\left(ac + bc + ab\right) + \mathcal{O}(\mathrm{d}\epsilon) \tag{76}$$

which again agrees with intuition for $a = b = c$

4 12/04/16—Musings on Hamiltonian Chaos

We learned in our chaos readings that given an integrable Hamiltonian (can be written in terms of action-angle variables, has N constants of motion for $2N$ dimensional phase space), a small perturbation generally breaks the toroidal phase space trajectory into chaotic motion. Let's see how much of this we can actually understand.

4.1 Action-Angle variables

I don't have my 106 notes handy, so let's rederive some action-angle stuff. The archetypal Hamiltonian to use is the SHO $H = (p^2 + q^2)/2$. While we may have suspicions for the choice of action-angle, we look up that

$$I = \frac{1}{2\pi}\oint \mathrm{d}(pq) \tag{77}$$

the integral over one period. For us, let's note that $E = H = \dfrac{p^2+q^2}{2}$ is a constant of motion, thus we can write

$$\begin{aligned}
p &= \sqrt{2E - q^2} && (78)\\
I &= \frac{1}{2\pi}\left[2\int\limits_{-\sqrt{2E}}^{\sqrt{2E}} \sqrt{2E-q^2}\,\mathrm{d}q\right] && (79)\\
&= E && (80)
\end{aligned}$$

where we recognize the integral to just be the integral of the circle. This makes sense, as we recognize that the action integral I is just the area of phase space enclosed within a full period, which for us is just $2\pi E$ since we enclose a circle in phase space with radius $r^2 = 2E$.

The angle θ must be such that the above expression also holds, i.e.

$$\oint \mathrm{d}(pq) = \oint \mathrm{d}(I\theta) \tag{81}$$

so that the phase space volume enclosed in one rotation is the same for both variables. We can then differentiate both sides by I to obtain

$$\theta = \frac{\partial}{\partial I}\oint \mathrm{d}pq. \tag{82}$$

Since the $\oint$ depends only on the bounds of integration, we see that θ is simply the limit on the integral, which further algebra shows to be $\arctan\dfrac{q}{p}$. We can verify that this is canonical by computing the PB

$$\begin{aligned}
\frac{\partial I}{\partial p}\frac{\partial \theta}{\partial q} - \frac{\partial I}{\partial q}\frac{\partial \theta}{\partial p} &= p\frac{1}{p}\frac{1}{1+\left(\frac{q}{p}\right)^2} - q\left(-\frac{q}{p^2}\right)\frac{1}{1+\left(\frac{q}{p}\right)^2} && (83)\\
&= 1 && (84)
\end{aligned}$$

so we're in good shape.

4.2 Multi-dimensional SHOs

In more generality, if we have a multidimensional SHO, we see that the Hamiltonian is just their sum, and so H in terms of action angle variables is still the sum of the actions, while their angles evolve separately.

What is the rate at which the angle evolves? For our above single-dimensional oscillator, it's easy to simply solve the EOM and find that $\frac{q}{p} = \tan t$, and so that the angle evolves with unit frequency. More generally, if the Hamiltonian is of form $H = p^2 + Cq^2$, it is easy to associate $C = \omega^2$ thanks to Hamilton's canonical equations $\dot{p} = H_q, \dot{q} = -H_p$ or something like that up to a sign. Thus, in general our Hamiltonian takes on form

$$H = \frac{\sum_j p_j^2 + \omega_j^2 q_j^2}{2} = \sum_j \omega_j I_j \tag{85}$$

with each of the I_j having a corresponding angle θ_j that evolves at ω_j.

4.3 With perturbation (incorrect result and faulty method)

How can we handle the perturbation? I have no idea, but I'll give it a shot. Let's adopt a phase space where each (q_j, p_j) are components of a complex number. Then the I_j are the magnitudes of each component, the θ_j the phases, and the Hamiltonian acts simply to rotate each component. It's easy to write down a system of equations that reproduces this behavior, but can we express it in terms of the Hamiltonian? Put another way, is there a way we can cast the $2N$-dimensional real Hamiltonian system above into an N-dimensional complex system with similar rules?

The defining property of a Hamiltonian system is Hamilton's canonical equations $\dot{p} = \frac{\partial H}{\partial q}, \dot{q} = -\frac{\partial H}{\partial p}$. If each dynamical variable is instead $z_j = p_j + iq_j$, we instead want a property that looks something like $\dot{z}_j = i\frac{\partial H}{\partial z_j}$[3].

Specializing to the SHO, given an ω in the SHO, we should make the correspondence $z_j = p_j + i\omega_j q_j$. Then we can write $H = \sum_j \omega_j z_j^2/2$ which gives us results something like $\dot{p}_j = -\omega_j q_j, \omega_j \dot{q}_j = p_j$ which is in accordance with what we expect. Thus, for an SHO we have

$$\dot{z}_j = i\frac{\partial H}{\partial z_j} = i\omega_j z_j \tag{86}$$

In other words, the evolution of the system is fully diagonal with eigenvalues ω_j. This is awfully reminescent of quantum mechanics! However, in QM, first order perturbation theory always gives us a result for a new orthogonal basis; why would any tori in classical mechanics break down? What does it even mean for a torus to break down? Why am I so stupid? These are not rhetorical questions but rather live musings as I type this up.

Consider if we then perturb the above EOM to something like

$$\dot{z}_j = i\omega_j z_j + i\epsilon\frac{\partial \delta H\left(\vec{z}\right)}{\partial z_j} \tag{87}$$

[3] We have made a choice of convention in putting the i with the partial derivative rather than in the Hamiltonian, motivated by keeping the Hamiltonian clean and most in analog with the real-variable Hamiltonian

Let's do the sensible thing and linearize δH so that $\delta H(\vec{z}) = \delta\mathbf{H}\vec{z}$. We seek a new set of z_j such that the EOM is again diagonal. If we treat the z_j coordinates as vectors in a vector space, we can do this via perturbation theory analogous to QM. Let's write down ansatz for new eigenbasis

$$\vec{z}_j = \vec{z}_j + \sum_k A_{jk}\vec{z}_k \tag{88}$$

We then seek that $\dfrac{\partial(\mathbf{H}+\delta\mathbf{H})(\vec{z}_j')}{\partial z_j'} = \omega_j'\vec{z}_j'$, and so (goodness the algebra below is so so wrong, but suck it up)

$$\frac{\partial \mathbf{H}\vec{z}_j}{\partial z_j} + \frac{\partial \delta\mathbf{H}\vec{z}_j}{\partial z_j} + \frac{\partial}{\partial z_j}\mathbf{H}\sum_k A_{jk}\vec{z}_k = \omega_j z_j\hat{\jmath} + \delta\omega_j\vec{z}_j + \omega_j\sum_k A_{jk}\vec{z}_k \tag{89}$$

$$\sum_k(\delta H)_{kj}\hat{k} + \sum_k A_{jk}\omega_k\vec{z}_k = \delta\omega_j z_j\hat{\jmath} + \omega_j\sum_k A_{jk}\vec{z}_k \tag{90}$$

$$A_{jk} = \frac{(\delta H)_{jk}}{\omega_j - \omega_k} \tag{91}$$

which is in line with the result from QM. I don't trust the algebra but I think the result is reasonable.

Not really sure where to go from here, but I found old lecture notes on the topic so I'll just consult them. Oops. It's been fun!

The correct approach (which we may or may not work through) attacks this from the generating function perspective, computing the Hamilton-Jacobi generating function for the canonical transformation in terms of the small perturbation, then showing that for purely rational ω_j, the perturbation theory fails to converge. It would appear then that the transformation of coordinates we propose above in general is not canonical for a rational winding number or something like that, if H_1 has a sufficiently high frequency component.

The moral of the story is that in classical mechanics, to change coordinates we must approach from a generating function to show that the transformation is canonical. Duly noted.

I continue this discussion in a separate section in my chaos notes, and conclude this discussion to pursue more interesting musings about the classical-quantum correspondence we began to uncover above.

5 12/07/16—Musings on Quantum-Classical correspondence

I'll repeat a small amount of the earlier discussion for my own benefit as I write this up.

The definiting characteristic of the Hamiltonian formalism are Hamilton's canonical equations

$$\frac{\partial H}{\partial p_i} = \dot{q}_i \qquad\qquad \frac{\partial H}{\partial q_i} = -\dot{p}_i \tag{92}$$

Above, we tried to map $z_i = q_i + ip_i$. However, part of what we did incorrectly in the above section was to assume that $\frac{\partial H}{\partial z_i}$ exists, i.e. H is analytic in the z_i. This is generally not true, obviously, so we need to do some work here.

I still want to try and make some sort of z_i transformation, so let's consider the following: let's try to construct some $\mathcal{H} = H + iH'$ such that $\mathcal{H}$ is an analytic function of the $z_i = q_i + ip_i$. Recall the Cauchy-Riemann equations for differentiability of a function $f(z) = u(z) + iv(z), z = x + iy$:

$$\frac{\partial v}{\partial y} = \frac{\partial u}{\partial x} \qquad\qquad i\frac{\partial v}{\partial x} = -i\frac{\partial u}{\partial y} \tag{93}$$

and so for us

$$\frac{\partial H'}{\partial p_i} = \frac{\partial H}{\partial x_i} \qquad\qquad \frac{\partial H'}{\partial q_i} = -\frac{\partial H}{\partial p_i} \tag{94}$$

We then wish to compute $\frac{\partial \mathcal{H}}{\partial z_i}$. Since $\mathcal{H}$ is analytic, it doesn't matter along which direction in the complex plane we approach z_i, so we choose to fix p_i, and we obtain

$$\left.\frac{\partial \mathcal{H}}{\partial z_i}\right|_{p_i} = \frac{\partial H(q_i, p_i)}{\partial q_i} + i\frac{\partial H'(q_i, p_i)}{\partial q_i} \tag{95}$$

$$= \frac{\partial H}{\partial q_i} - i\frac{\mathrm{d}H}{\mathrm{d}p_i} \tag{96}$$

Hey, that's super nice, since now we can write $\dot{z}_i = -i\frac{\partial \mathcal{H}}{\partial z_i}$, or in slightly more familiar form

$$i\frac{\partial z_i}{\partial t} = \frac{\partial \mathcal{H}}{\partial z_i} \tag{97}$$

But wait! We can do even better. Since for any N-dimensional system, we have N conserved quantities, we can always make canonical transformation for $H(p_i, q_i) \to H(I_i)$ the action-angle transformation. I won't bother checking canonical-ness here, but the new Hamiltonian $H(I_i) = \omega_i I_i$.

Furthermore, from here we can always make a canonical transformation $H(I_i) \to H(a_i, b_i)$ where $a_i = \sqrt{2I_i}\cos\theta, b_i = \sqrt{2I_i}\sin\theta$ with θ the angle variable. Let's verify that this is canonical

$$\frac{\partial a_i}{\partial I_i}\frac{\partial b_i}{\partial \theta_i} - \frac{\partial b_i}{\partial I_i}\frac{\partial a_i}{\partial \theta_i} = \left(\frac{\cos\theta}{\sqrt{2I_i}}\right)\left(\sqrt{2I_i}(-\cos\theta)\right) - \left(\frac{\sin\theta}{\sqrt{2I_i}}\right)\left(\sqrt{2I_i}\sin\theta\right) \tag{98}$$

$$= -1 \tag{99}$$

which means the Hamiltonian becomes

$$H = \sum_i \frac{\omega_i}{2}(a_i^2 + b_i^2) \tag{100}$$

which is basically an SHO in a_i, b_i. But what's nice about this formulation in particular is that

$$\frac{\partial \mathcal{H}}{\partial z_i} = \frac{\partial H}{\partial a_i} - i\frac{\partial H}{\partial b_i} \tag{101}$$

$$= \omega_i(a_i + ib_i) = \omega_i z_i \tag{102}$$

and ergo we obtain

$$i\frac{\partial z_i}{\partial t} = \omega_i z_i \tag{103}$$

which is exactly the Schödinger equation in the energy eigenbasis! In other words, write the matrix $H_{QM} = \mathrm{diag}(\hbar\omega_i)$, then we have

$$i\hbar\frac{\partial \vec{z}}{\partial t} = H_{QM}\vec{z} \tag{104}$$

I'm off by a sign, which is simply a convention of the definition of z_i, or alternatively a definiton on i. But we can start to see the emergence of the QM interpretation: there are certain dynamical orbits that are invariant under evolution by the Hamiltonian, and these correspond to the energy levels in QM that have energy $\hbar\omega$ associated with them.

It should be noted in hindsight that our jubilance is slightly overeager: we have showed that the Hamiltonian in classical mechanics can be interpreted to dictate each component of $\vec{z}$ in a way that corresponds to ω_i the rate of change of the angle variable, but we have in no way derived any energy levels or the sort. This should not surprise us, as the quantization of energy levels is a purely quantum mechanical phenomenon, but the resemblance in Equation 104 is still slightly superficial.

5.1 Significance of $\frac{\partial \mathcal{H}}{\partial z_i} \Rightarrow H_{QM}\vec{z}$

Note that to arrive at (104), we made what amounted to the correspondence $\frac{\partial \mathcal{H}}{\partial z_i} \Rightarrow (H_{QM}\vec{z})_i$ up to $\hbar$. This result makes perfect sense when H is a quadratic sum of the q_i, p_i, but how the hell do the other operators enter the picture?

It would seem that the key identification we must make here is how the complex $\vec{z}$ vector relates to dynamical measurements in classical mechanics, since the correspondence between $\vec{z}$ and state vector $|\psi\rangle$ is clear. In order to measure, say, the q_j position of the system given $\vec{z}$, we need simply to compute the change-of-basis operation from the $z_i = a_i + ib_i$ to the q_i, p_i basis, then project out the jth index.

Another way to extract q_j that perhaps seems more promising is to construct the generating function S (which exists for all canonical transformations). Let's choose the generating function $S(a, b)$ such that $q = S_a, p = S_b$. It seems clear that we can use the same complexification trick above to obtain a $\mathcal{S}$ such that

$$q + ip = \frac{\partial \mathcal{S}(a + ib)}{\partial(a + ib)} \tag{105}$$

Let's call $w = q + ip$, then we find that the q_j component can be computed by computing

$$q_j = \mathrm{Re}\left(-i\hat{w}_j \cdot \frac{\partial \mathcal{S}}{\partial \vec{z}}\right) \tag{106}$$

where I use the shorthand $\left[\frac{\partial \mathcal{S}}{\partial \vec{z}}\right]_j = \frac{\partial \mathcal{S}}{\partial z_j}$. Compare this to the QM result, which would be $\langle q_j \mid \psi \rangle$, where in order to evaluate the dot product we must compute the components of ψ in the $\{q_j\}$ basis. We now interpret $\frac{\partial \mathcal{S}}{\partial \vec{z}}$ to be the $\vec{z}$ under the basis tranformation dictated by $\mathcal{S}$. The quantum mechanical equivalent of this operator would be $\sum\limits_k |q_k\rangle \langle q_k|$ the resolution of identity but also the change of basis operator.

There seems to be at first glance a trend here: quantum mechanical operators tend to correspond to classical functions that act on a variable by *differentiating* about that variable. In other words, the classical correspondence of some operator Ω that acts on a wavefunction $|\psi\rangle$ is $\mathbf{\Omega} = \frac{\partial \Omega}{\partial [\,]}$ to be such that $\mathbf{\Omega}\vec{v} = \frac{\partial \Omega}{\partial \vec{v}}$ and where Ω is some complexified analytic function of its real/Hamiltonian counterpart. We have seen this to be true for the resolution of identity and for the Hamiltonian itself.

Let's note that, heuristically, this explains the mapping of Poisson Brackets to commutators. Recall that a set of coordinates is classically canonical if its PB with q, p is unity, and a set of variables is conjugate if their commutator is $i\hbar$. These are both phase space-preserving constraints, so it's clear that the latter is really the stipulation that the commutator of conjugate variables must equal $[X, P] = i\hbar$, just as the PB is just that the change of coordinates has Jacobian with unit determinant. Let's try to flesh this out a bit more.

5.2 Classical Observables and Quantum Mechanical Operators

Let's formalize the above a little bit. When we speak of a classical dynamical variable a, it means that given a full specification of a state $\{x_i, p_i\}$, we can measure a for the given state. Similarly, when we speak of a quantum mechanical Hermitian operator (an observable) A, we mean that given a $|\psi\rangle$ we can determine the probability of measuring A given $|\psi\rangle$ and observing any eigenvalue λ of A, the probability $|\psi\rangle$ is eigenstate $|\lambda\rangle$.

Under our above formalism, we mean the following: when we speak of a classical dynamical variable a, we mean that the value of a in a classical state described by the complexified $\vec{z}_i$ can be computed by

$$a = \mathrm{Re}\left(-i\hat{a} \cdot \frac{\partial \mathcal{S}}{\partial \vec{z}}\right) \tag{107}$$

where $\mathcal{S} = S + iS'$, S being the generating function describing a canonical transformation from $\{x_i, p_i\}$ to some set of canonical variables involving a and S' such that $\mathcal{S}$ is analytic.

For instance, if we want to compute x_j, we note that the generating function has form $S = x_i p_i$, and so

$$\left[\frac{\partial \mathcal{S}}{\partial \vec{z}}\right]_j = \frac{\partial \mathcal{S}}{\partial z_j} \tag{108}$$

$$= p_j + ix_j \tag{109}$$

$$x_j = \mathrm{Re}(-i * (p_i + ix_j)) = x_j \tag{110}$$

The insertion of the $-i$ is a bit inconvenient, and clearly would have been different were we trying to measure p_j. This seems a bit inconvenient, until we remember that we can probably absorb these factors into $\mathcal{S}$ by defining different operators; indeed, X, P are different operators in QM, and there seems to be little reason for us to use the same $\mathcal{S}$ for them!

Thus, under our formalism, classical dynamical variables a map to second-order tensors (matricies), which is again right for our correspondence.

As one last parting shot of the merit of our formalism, let's observe that if we define a canonical transformation $\{x_i, p_i\} \to \{z_i, z_i^*\}$ a classic trick, where $z_i = \dfrac{x_i + p_i}{\sqrt{2}}$, we obtain that the PB is equal to i. The significance of this eludes me, but certainly the agreement with QM up to an omnipresent $\hbar$ ought to be encouraging?

5.3 Normalizing the state vector

An astute reader (all zero of you) would have objected a while ago that our Hamiltonian only superficially resembled a quantum mechancical one, since as I mentioned, the frequencies above correspond to the rate of change of the action variable, which is a constant of the Hamiltonian and the *system*, not the state of the system. Contrast this with quantum mechanics, where the $\hbar\omega$ eigenvalues of the Hamiltonian depend on the *state* of the system equally well as the Hamiltonian/configuration of the system itself.

We can see why we err: the ω above does not change when the amplitude of motion changes, but the energy of the system clearly increases. To better understand where this comes from, let's focus on a single dimension for now, so two dimensions of phase space. If we recall how we wrote down the formula to measure x_j above, it relied on the fact that the x_j component of some vector was equal to the x_j position of the taste. But $x_j \in (-\infty, \infty)$, so this won't do if we try to normalize the state vector. The only way we can get around this is by turning the single x_j component into an infinity of components, one for each value of x_j. Then we can ask whether the system's x_j coordinate is some x_0 value by examining whether the x_0 component is nonzero! (Now we see why we prefer to discuss only a single dimension for now...)

Consider our system to be in a state $z_0 = x_0 + ip_0$, and map it to a continuous function $\psi(z = x + ip) = \delta(x - x_0)\delta(p - p_0)$. It is easy to verify that ψ is normalized $\int \left|\psi^2\right| = 1$, and we may measure the position of ψ by integrating it against an operator

$$x_0 = \int x \left|\psi\right|^2 \, \mathrm{d}x\mathrm{d}p \tag{111}$$

Lest I be accused of cherry-picking my normalization criterion for ψ, I remind you this is the standard normalization procedure for vectors. My argument above holds equally well whether using $\int \psi$ or $\int \left|\psi\right|^2$, since we're only dealing with δ functions. So I guess it's a tossup.

We can find similar expressions to measure the energy etc. of the system, and we see that we have moved from a $\mathbb{C}^n$-component vector representation of the state of the system to a representation that has unit norm albeit is a continuous function on $\mathbb{C}^n$. One can easily contest that this is an obfuscation of the physics, but is a natural starting ground for quantum mechanical states, where the phase space is discretized and we can have linear superpositions of states.

5.4 Afterword

We are officially at an impasse: without foresight of the quantization that quantum mechanics brings, without supposing that systems have finite precision in conjugate dynamical variables, it is impossible to proceed. These assumptions have no place in a classical worldview, and it seems that we have largely exhausted the range of speculations we can accomplish. But what a rich theory it is already! We have shown that it is possible to

- Map a $2N$ real phase space to a normalized state vector without losing any expressiveness.
- Proposed a second-order tensor form for all classical dynamical variables that act similarly to quantum observables.
- Accidentally come across the canonical commutation relation and an eerily remeniscent version of the Schrödinger Equation.

A symplectic manifold is a system endowed with an antisymmetric two-form, called the symplectic form. I think that this association of the antisymmetric two-form with a complexified function is generalizable and perhaps broadly applicable as a useful way of thinking about these two-forms. Of course, n-forms are probably just generalizations of the complex numbers, on second thought, but alas that is far beyond the scope of my intellect.

With this I close my analysis. Thanks for listening!

6 12/14/16—Matricies

We saw earlier in quantum chaos that the way that the continuous energy spectrum breaks down into discrete eigenvalues (which governs the power spectrum of the trajectory) is described by random matrix theory, specifically the distribution of gaps between eigenvalues of random matricies follows the same distribution. Let's see how much ground we can make on these results.

6.1 Prelim: Distributions of combinations of random variables

Suppose we want the difference of two random variables X, Y. For simplicity, let's say that they're normally distributed. Of course, if we're just talking expectation and variance then this is a trivial problem; how about the shape of the distribution? Differences of normal variables are also normally distributed, but let's see one way we can describe this. Call $Z = X - Y$, with both X, Y chosen with zero mean and standard deviation σ, then we should obtain

$$P_Z(z) = \iint P_X(x)P_y(y)\delta(x - y - z)\;\mathrm{d}x\mathrm{d}y \tag{112}$$

$$= \int\limits_{-\infty}^{\infty} P_X(x)P_y(x - z)\;\mathrm{d}x \tag{113}$$

and if we specialize $P_X = P_Y = N(0, \sigma)$ then (dropping normalizations)

$$P_Z(z) \propto \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}} \,\mathrm{d}x \tag{114}$$

$$\propto \int_{-\infty}^{\infty} \exp\left[\frac{-(2x^2 - 2xz + z^2)}{2\sigma^2}\right] \,\mathrm{d}x \tag{115}$$

$$\propto \int_{-\infty}^{\infty} e^{-\frac{2(x-z/2)^2}{2\sigma^2}} e^{-\frac{z^2/2}{2\sigma^2}} \,\mathrm{d}x \tag{116}$$

and since the former exponent integrates to a constant and the latter a Gaussian of width $\sigma\sqrt{2}$, we haeve our answer. It is worth noting the former exponent integrates to a constant with an extra factor of $\sqrt{2}$ that provides exactly the correct normalization for the latter.

What about the product of two normally distributed variables?