8.044: Statistical Physics I

Delon Shen

Notes and assignments for Greytak's Statistical Physics I course on MIT OCW. If you have any comments let me know at `hi@delon-shen.com`.

Last updated on February 6, 2021

Blundell and Blundell Chapter 3: Probability

January 26, 2021

Let x be a discrete random variable (e.g. can only take a finite number of values). We'll say that x can become x_i with probability P_i[1]. We can define the expectation value of x_i as

$$\langle x\rangle = \sum_i x_i P_i$$

A weighted sum. This is naturally extended to expectation values of functions of x_i

$$\langle f(x)\rangle = \sum_i f(x_i) P_i$$

We can extend these notions to continuous random variables where we have a probability $P(x)dx$ of find x as a value between x and $x+dx$. In this case we have

$$\langle x\rangle = \int xP(x)dx \qquad \langle f(x)\rangle = \int f(x)P(x)dx$$

If we relate two expectation values with a linear transformation then the expectation values transform as we'd expect as well. Let y also be a random variable and a, b be constants

$$y = ax+b \Rightarrow \langle y\rangle = a\langle x\rangle + b$$

Now what if we want to quantify the spread of values. Our first guess may be just to find the average $x-\langle x\rangle$. However note that

$$\langle x-\langle x\rangle\rangle = \langle x\rangle - \langle x\rangle = 0$$

Our next guess then could be the $(x-\langle x\rangle)^2$. We define $\left\langle (x-\langle x\rangle)^2\right\rangle$ as the *variance*. The variance of x, which we'll denote as σ_x^2, is the *mean squared deviation* and $\sqrt{\sigma_x^2}$ to be the *standard deviation*. A useful identity to keep in mind is

$$\begin{aligned}
\sigma_x^2 &= \left\langle ((x-\langle x\rangle)^2\right\rangle \\
&= \left\langle x^2 + \langle x\rangle^2 - 2x\langle x\rangle\right\rangle \\
&= \left\langle x^2\right\rangle + \langle x\rangle^2 - 2\langle x\rangle^2 \\
&= \left\langle x^2\right\rangle - \langle x\rangle^2
\end{aligned}$$

The effect of linear transformations on the variance can be found as follows. Again let $y = ax+b$. After some algebra grinding we get

$$\sigma_y^2 = a^2\sigma_x^2$$

Namely the b parameter doesn't affect anything. This should make sense since b is an arbitrary shift of the random variable x which should have no impact on the "spread" of the values that

[1] *edit 2021-02-03: it's probably worth noting that the lectures use a different notation. Here we denote the probability density as $P(x)$ wheras in the lectures we use $p(x)$. $P(x)$ in the lectures is reserved for proability distribution functions*

the variance measures.

Independent random variables are variables that are independent! Knowing the value of one variable gives us no information about the other variables. For these kind of variables we can just multiply them together and things work out the way we'd expect them to

$$\langle uv \rangle = \langle u \rangle \langle v \rangle$$

Lets consider a random variable Y that is the sum of n independent random variable X_i such that $\langle X_i \rangle = X$. What is the variance of Y? First the easy one

$$\langle Y \rangle^2 = n \langle X \rangle^2$$

Now for $\left\langle Y^2 \right\rangle$ we just have to be a bit clever. In Y^2 there are n X_i^2 terms and then $n(n-1)$ $X_i X_j$ terms where $i \neq j$. Thus in $\left\langle Y^2 \right\rangle$ we'll have $n \left\langle X^2 \right\rangle$ terms and $n(n-1)$ $\langle X \rangle^2$ terms. So

$$\left\langle Y^2 \right\rangle = n \left\langle X^2 \right\rangle + n(n-1) \langle X \rangle^2$$

From this we can calculate the variance

$$\sigma_y^2 = \left\langle Y^2 \right\rangle - \langle Y \rangle^2 = n \left\langle X^2 \right\rangle + n(n-1) \langle X \rangle^2 - n \langle X \rangle^2 = n \left\langle X^2 \right\rangle - n \langle X \rangle^2 = n \sigma_x^2$$

This has applications to experiments. Lets say you independently measure the same quantity X multiple times. Let X_i be the measured value of the i^{th} run. This means that if you use the average of the measured X values $Y = \frac{1}{n} \sum_i X_i$ then the standard deviation of Y would be $\sigma_x / \sqrt{n}$. This also has applications to random walks. Lets say we have some discrete random variable X which can either be $-a$ or a with equal probability. This means that $\langle X \rangle = 0$. Thus the variance is $\left\langle X^2 \right\rangle = a^2$. Now lets consider a drunk man desperately trying to get to his home on the other side of a tight rope(over the grand canyon of course.) Each step the drunk man takes is either forward or backwards a units but because he's drunk the probability that he steps forwards or backwards is equal. The standard deviation of the sum of multiple X_i in this case would correspond to the root mean squared distance that degenerate has walked. Lets now calculate the variance

$$Y = \sum_{i=1}^{n} X_i \Rightarrow \sigma_Y^2 = n \sigma_X^2 \Rightarrow \sigma_Y = \sqrt{n} \sigma_X = \sqrt{n} \left\langle X^2 \right\rangle = a \sqrt{n}$$

After n steps the rms length this guy has travelled is $a\sqrt{n}$.

Lecture 1: One Random Variable

Started: January 27, 2021. Finished: January 28, 2021

A *random variable* is a quantity whose state is determined by a processes which can be analyzed and thus allow us to assign some probability to the variable taking a certain state. For our purposes randomness can be introduced from

(a) Insufficient information: e.g. where my cat is. I knew he was sleeping on the couch at 1:00PM but where could he be now?

(b) Quantum Mechanics: uncertainty principles.

We can classify random variables into three categories

(i) Discrete: the variable can only take a finite set of values. For example how many children does Ms. Smith have? I hope she doesn't have a half of a child....

(ii) Continuous: the variable can take a value from a continuum of possibilities. For example how hot will it be the next time an election takes place.

(iii) Mixed: the variable can take discrete values for some conditions and continuous variables for other conditions. For example how many grapes will Chuck eat tomorrow? Well it depends on if he blends the grapes or not. If he eats grapes noramlly then the values are discrete, but if he blends them up and and puts them in a smoothie then we have a continuous random variable. (I don't know if this example is correct.)

Lets consider a system of n similarly prepared systems. We'll let x_i denote the output of the i^{th} system. Now imagine creating a histogram that counts how many x_i lie in some bin. This is the notion of a probability density function $\rho_x(\xi)$.

$$\rho_x(\xi) = \lim_{M\to\infty\ d\xi\to 0} \frac{\text{Number of times } x \text{ landed in } [\xi, \xi + d\xi)}{M d\xi}$$

We'll drop the $p_x(\xi)$ for the much nicer $p(x)$ notation when it should be clear. From this we can also define a probability distribution function $P_x(\xi)$ where

$$P_x(\xi) = \int_{-\infty}^{\xi} \rho_x(\xi) d\xi \Rightarrow \frac{dP_x(\xi)}{d\xi} = \rho_x(\xi)$$

Where we used Leibniz integral rule implicitly assume that $\rho_x(\xi)$ vanishes at $-\infty$. The probability density completley specified the random variable. There are two questions we could ask about random variables. How do we find the probability distribution of a random variable and what can we gleam about this random variable from this probability distribution.

Example 1: (Radioactive Decay) Lets say we have some radioactive source. The probability density function to detect the first decay t seconds after beginning the experiment is

$$\rho(t) = \begin{cases} \tau^{-1} e^{-\frac{t}{\tau}} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Where τ is inverse the mean rate of events per second. We note that the probability is normalized

$$\int_{-\infty}^{\infty} \rho(t) dt = 1$$

And the probability distribution is what we would expect to see (as time increases the probability that we see an event occuring also increases)

$$P(t) = \int_0^t \rho(t) dt = 1 - e^{-\frac{t}{\tau}}$$

To model discrete or mixed random variables we can use dirac delta functions.

$$\rho(x) = \sum_i P_i \delta(x - x_i)$$

This gives us all the properties we'd expect when dealing with probability distributions and probability densities.

When measuring physical quantities usually we want to extract *ensemble averages* and the deviation of the ensemble average from the mean via *variance* and *standard deviation*. (All of these derived in B&B Ch 3).

Example 2: (Poisson Density) We're in a situation where in a interval ΔX the probability that an event happens in an interval $X + \Delta X$ is proportional to the size of the interval $P = r\Delta X$. Also the probability that an event occurs in one interval $X + \Delta X$ is independent of all other intervals. It will be derived soon that the pobability that n events will occur in an interval of length L is given be

$$p(n) = \frac{1}{n!}(rL)^n e^{-rL} \qquad n \in \mathbb{N}$$

We can use this probability to create a probability density function for a random variable y (where y is how many events occur) with delta functions

$$P(y) = \sum_i p(i)\delta(y - i)$$

Lets check normalization, mean, and variance

$$\int_{-\infty}^{\infty} P(y)dy = \sum_i p(i) \int \delta(y - i) = \sum_i p(i) = e^{-rL} \sum_i \frac{1}{i!}(rL)^i = e^{-rL} e^{rL} = 1$$

We see that the normalization is what we expect

$$\langle n \rangle = \int_{-\infty}^{\infty} yP(y)dy = \sum_{i=1}^{\infty} p(i) \int_{-\infty}^{\infty} y\delta(y - i)dy = e^{-rL} \sum_{i=1}^{\infty} \frac{1}{(i-1)!}(rL)^i$$

We can pull out a rL term and let $j = i - 1$. The bounds go from 0 to ∞ and we again have our definition of the exponential

$$\langle n \rangle = e^{-rL} \sum_{i=1}^{\infty} \frac{1}{(i-1)!}(rL)^i = e^{-rL}(rL) \times \sum_{j=0}^{\infty} \frac{1}{j!}(rL)^j = rL$$

Now the variance

$$\left\langle n^2 \right\rangle = \int_{-\infty}^{\infty} y^2 P(y)dy = \sum_{i=1}^{\infty} p(i) \int_{-\infty}^{\infty} y^2 \delta(y - i)dy = e^{-rL} \sum_{i=1}^{\infty} \frac{i}{(i-1)!}(rL)^i$$

To evaluate this lets use $\langle n \rangle$. First note that our above result suggests that

$$\sum_{i=1}^{\infty} \frac{1}{(i-1)!}(rL)^i = rLe^{rL}$$

Lets take the derivative of this with respect to r

$$\frac{d}{dr}\left\{\sum_{i=1}^{\infty}\frac{1}{(i-1)!}(rL)^i\right\}=\sum_{i=1}^{\infty}\frac{i}{(i-1)!}(rL)^{i-1}L=\frac{d}{dr}\left\{rLe^{rL}\right\}=Le^{rL}+rL^2e^{rL}$$

Now lets multiply both sides by r giving us

$$\sum_{i=1}^{\infty}\frac{i}{(i-1)!}(rL)^i=Lre^{rL}+r^2L^2e^{rL}$$

Plugging this result in gives us

$$\left\langle n^2\right\rangle=e^{-rL}\left(Lre^{rL}+r^2L^2e^{rL}\right)=rL+r^2L^2$$

Which finally allows us to compute the variance

$$\left\langle n^2\right\rangle-\langle n\rangle^2=rL$$

It should be noted that we can write $p(n)$ in terms of its variance (or equivalently is mean) meaning that the Poisson distribution can characterized by its variance alone

$$p(n)=\frac{1}{n!}\langle n\rangle^n e^{-\langle n\rangle}$$

Lecture 2: Two Random Variables

Started: January 28, 2021. Finished: January 31, 2021

Lets introduce the concept of *joint probability distributions*. There's a lot they have in common a single random variable. For example

$$\begin{aligned}
p_{x,y}(\xi,\eta)d\xi d\eta &= \text{Probability } \xi\leq x\leq\xi+d\xi \text{ and } \eta\leq y\leq\eta+d\eta\\
P_{x,y}(\xi,\eta) &= \int_{-\infty}^{\xi}\int_{-\infty}^{\eta}p_{x,y}(\xi,\eta)d\eta d\xi\\
p_{x,y}(\xi,\eta) &= \frac{\partial}{\partial\xi}\frac{\partial}{\partial\eta}P_{x,y}(\xi,\eta)\\
\langle f(x,y)\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p_{x,y}(\xi,\eta)d\xi d\eta
\end{aligned}$$

We can also reduce the probability distribution to one varibale

$$p(x)=\int_{-\infty}^{\infty}p(x,y)dy$$

Now lets talk about conditional probability. The lecture notes approach this from a different viewpoint from what I'm going to do. I'll try to approach it from a discrete viewpoint and then extend it to the continuous case. Lets consider an example.

Example 3: (Oh Fuck I Tested Positive for Corona) Lets say that you're an unfortunate soul who has to take Senior Lab during the pandemic to graduate on time and your lab partner tested positive for corona! So you go to CVS for some rapid COVID-19 testing and, what's this? You've tested positive! Now don't fret yet. Rapid testing isn't the most accurate thing in the world. Given that you tested positive, what is the probability that you actually have corona? Lets say (these aren't real numbers) that if you have corona you have an 80% chance of testing positive from CVS. Furthermore 10% of people who don't have corona test positive for corona. We also know that 40% of the population of the world has corona. So what are the chances you have corona given that you tested positive? So first we need to figure out what's the probability that you test positive P_{pos}.

$$\begin{aligned} P_{\text{pos}} = &\ (\text{\% of People who have Virus}) \times (\text{Probability of Testing Positive with Virus}) \\ &+ (\text{\% of People who don't have the Virus}) \times (\text{Probability of Testing Positive w/o Virus}) \\ &= \frac{4}{10} \times \frac{8}{10} + \frac{6}{10} \times \frac{1}{10} = \frac{19}{50} \end{aligned}$$

Now this basically becomes or whole probability space. So to figure out the probability that you have the virus given that you tested positive can be found by dividing the blue term by P_{pos}

$$P(\text{have virus}|\text{tested positive}) = \frac{P\,(\text{Testing Positive with Virus}) \times P\,(\text{Have Virus})}{P\,(\text{Testing Positive})} = \frac{16}{19}$$

Nice!

Lets notice a few things aboout the above result. We can restate the final statement as

$$\boxed{(\text{Probability that } \xi = x \text{ given that } \eta = y) = p(\xi = x|\eta = y) = \frac{p(\xi = x, \eta = y)}{p(\eta = y)}}$$

This is *Bayes' Theorem* and is true in the continuous case as well as the discrete case. We can look at this result in two ways.

(a) Using the conditional probability to construct the joint probability distribution

(b) Using the joint probability distribution to guess the conditional probability

Bayes's Rule also gives as a good notion of statistical independence. We know that statistically independent means that having information on one variable gives no information on the other varaible. What this means in the context of Bayes' rule is that knowing $\eta = y$ gives us no information on whether $\xi = x$ which means

$$\text{If } p(\xi = x|\eta = y) = p(\xi = x) \text{ then } x, y \text{ are statistically independent}$$

Probability is god awful to think about without examples so lets take a look at some

Example 4: (People keep getting blown up on my friend's farm!) You're a young man living in the fictional war torn country of Bratlastinavia. You know that there's a perfectly circular area of radius 1km in the corn field by your friend's house that's just full of land mines. In fact the probability of stepping a land mine (assume they're infinitesimally small land mines) is uniform in the circle.

$$p(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \le 1 \\ 0 & x^2+y^2 > 1 \end{cases}$$

One night you have some particularly incurable insomnia so you try to figure out some things. If you just walk in a straight line parallel to one of the axes through the danger zone, at any given point what's the probability that you'll step on a land mine(find $p(x)$ and $p(y)$). Also you want to know whether x and y are statistically independent[a].

$$p(x) = \int p(x,y)dy = \frac{1}{\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi}\sqrt{1-x^2} \qquad x^2+y^2 \le 1$$

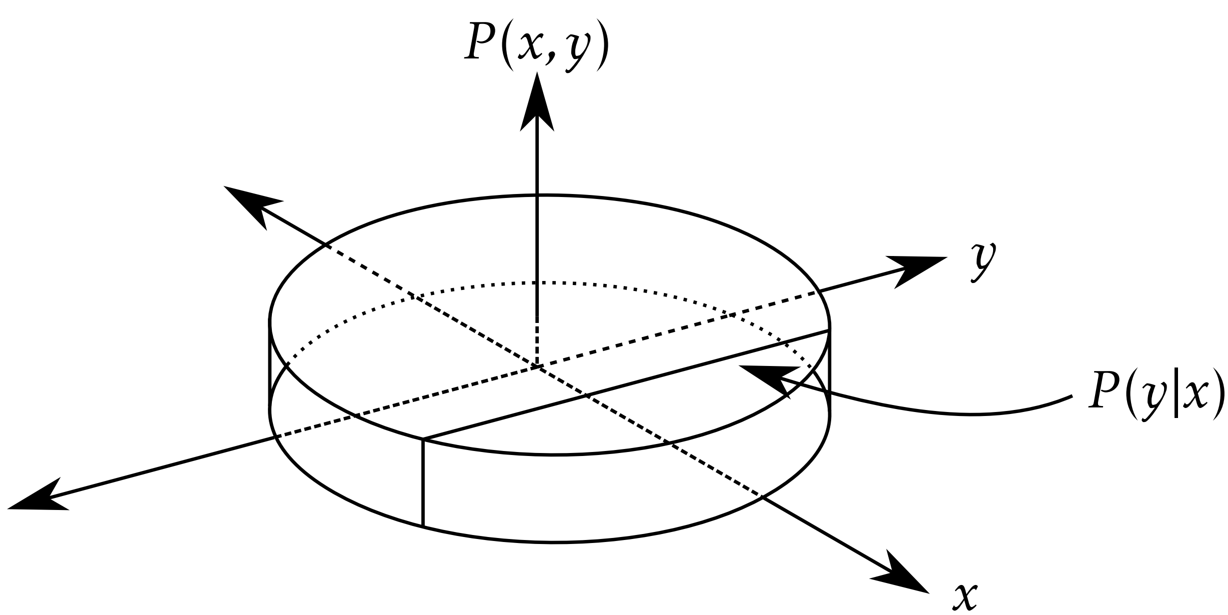
By symmetry we can also say that

$$p(y) = \frac{2}{\pi}\sqrt{1-y^2} \qquad x^2+y^2 \le 1$$

Now applying our favorite new rule for S.I. we have

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{1/\pi}{\frac{2}{\pi}\sqrt{1-y^2}} = \frac{1}{2\sqrt{1-y^2}} \ne p(x) \qquad x^2+y^2 \le 1$$

The fact that the conditional probability is constant might be suprising. However if we consider a visualization of the conditional probability below the reason should become clearer.



[a] I couldn't think of what this physically means :(

Example 5: (Poisson Density Again...) Recall the defining properties that given the Poisson distribution

- In the limit $\Delta x \to 0$ the probability that one and only one event occurs between X and $X + \Delta X$ is given by $r\Delta X$ where r is a given constnat independent of X
- The probability of an event occuring in some interval ΔX is statistically independent of events in all other portion of the line

We now want to show a few thing

(a) Find $p(n = 0, L)$, the probability that no events occur in a region of length L. Then divide L into infinite number of S.I. intervals and calculate the joint probability that none of the intervals contain an event

(b) Obtain $\frac{d}{dL}p(n;L) + rp(n;L) = rp(n-1;L)$

(c) Show that $p(n;L) = \frac{1}{n!}(rL)^n e^{-rL}$. Is the solution unique?

Lets get started then

(a) Our first guess is to do something like

$$p(n = 0, L) = 1 - \int_0^L r\,dx = 1 - rL$$

However this might not be the entire solution. We were already given $p(n;L)$ and expanding that for small L we have

$$p(0;L) \approx 1 - rL + \ldots$$

So really what we found above is only true for small L like an infinitesmial dL. I think what went wrong is the fact that rdx is the probability that one event happens in the region dx, not the probability that *at least* one event happens in the region dx. Lets try again. We'll only use $p(n = 0; dL) = 1 - rdL$ and then multiply these together for all dL.

$$p(n = 0, L) = (1 - rdL)^{\frac{L}{dL}}$$

This almost looks like the definition of e. Actually let $-rdL = \frac{1}{n}$ giving us

$$p(n = 0; L) = \left(1 + \frac{1}{n}\right)^{-nrL}$$

Now in the limit $dL \to 0$ we have $n \to \infty$ so this is exactly e

$$p(n = 0; L) = e^{-rL}$$

(b) I think we can do this one by induction. But to do that we need to prove the base case. And to do that we need to find $p(n = 1; L)$. What the probability that only one event occurs in the region dL? I think it might be something like

$$p(n = 1; L) = rdL(1 - rdL)^{\frac{L}{dL}-1} \times \frac{L}{dL}$$

In one of the dL an event takes place and the rest of the intervals have no events. We multiply by an $\frac{L}{dL}$ factor since we could pick any interval. Let $-rdL = \frac{1}{n}$ again.

$$p(n=1;L) = rL \times \left(1+\frac{1}{n}\right)^{-rnL} \times (1-rdL)^{-1}$$

Now taking the limit $dL \to 0 \Leftrightarrow n \to \infty$ we get

$$p(n=1;L) = rL \times e^{-rL}$$

Actually do we have to use the recursive formula. We could just derive a combinatioráil expression right? We can imagine for $p(n;L)$ we're trying to divide n balls into $\frac{L}{dL}$ bins. Lets say we put all the balls into one bin. Then the contribution to the probability would be

$$(rdL)^n(1-rdL)^{\frac{L}{dL}-1} \times \frac{L}{dL}$$

Now lets say we put all the balls into two bins. Then the contribution would be

$$(rdL)^n(1-rdL)^{\frac{L}{dL}-2} \times \binom{\frac{L}{dL}}{2}$$

and so on. Since we'll eventually take $dL \to 0$ we'll ignore the case when $n > \frac{L}{dL}$. So then we have

$$p(n;L) = (rdL)^n \sum_{i=1}(1-rdL)^{\frac{L}{dL}-i} \times \binom{\frac{L}{dL}}{i} = (rdL)^n \sum_{i=1}(1-rdL)^{\frac{L}{dL}-i} \times \frac{\left(\frac{L}{dL}\right)!}{i!\left(\frac{L}{dL}-i\right)!}$$

Now lets consider

$$(rdL)^n \times \left(\frac{L}{dL}\right)! = (r)^n dL^n \left(\frac{L}{dL}\right)\left(\frac{L}{dL}-1\right)\dots$$

Hmmm, I don't think this is the way to go. Now that I re-read the defining properties I realize there's a "one and only one event" statement in there meaning that there should only be one term that contributes to the probability.

$$(rdL)^n(1-rdL)^{\frac{L}{dL}-n} \times \binom{\frac{L}{dL}}{n}$$

The above should be $p(n;L)$ instead of that whole summing business going on before. A single bin can't have multiple events. Actually I think it's because the chance of multiple events in one bin times no events happening in other bins (due to the $(1-rdL)^{\cdots}$ term) is so small compared to $p(1,dL)$ that we can neglect it in our summation since they'll be higher order terms in dL. Anyway I've hand-waved my way to

$$p(n;L) = (rdL)^n(1-rdL)^{\frac{L}{dL}-n} \times \frac{\left(\frac{L}{dL}\right)!}{n!\left(\frac{L}{dL}-n\right)!}$$

From here I think the argument I was doing before works. Consider

$$(rdL)^n \frac{\left(\frac{L}{dL}\right)!}{\left(\frac{L}{dL}-n\right)!} = r^n(L)(L-dL)(L-2dL)(\dots)(L-(n-1)dL)$$

Where the equality comes from the fact that the division creates n terms each of which we shove a dL into. Now taking the limit $dL \to 0$ we get

$$\boxed{p(n;L) = \frac{1}{n!}(rL)^n e^{-rL}}$$

No differential equations needed! I'm sure there's some glaring mistake I'm making and things just happened to work out.

LECTURE 3: FUNCTIONS OF A RANDOM VARIABLE

STARTED: February 01, 2021. FINISHED: February 02, 2021

After some experimenting we found the probability density function for the velocity of atoms, $p_v(\xi)$. Given this how would we find the probability density function for a function of the velocities like the kinetic energy $T = \frac{1}{2}mv^2$. Well I don't know but the lecture notes gave a formula without proof. Lets say we have $x, p_x(\xi)$ and we want the probability distribution for $f(x)$ denoted by $p_f(\eta)$.

(a) Sketch $f(x)$ and find region of x where $f(x) < \eta$

(b) Integrate p_x over the indicated region in order to find the cumulative distribution function for f, $P_f(\eta)$

(c) Differentiate the distribution function to get the density function

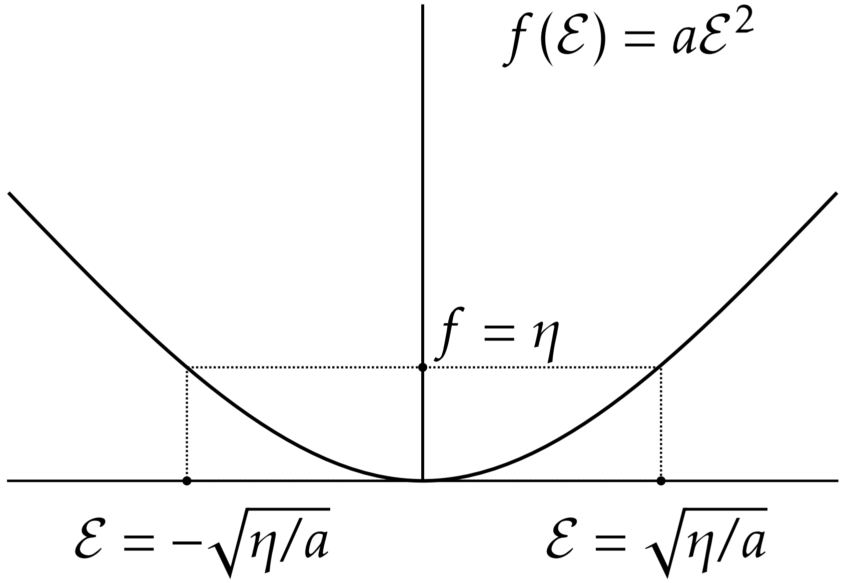
EXAMPLE 6: (CLASSICAL INTENSITY OF POLARIZED THERMAL LIGHT) The intensity of a linearly polarized electromagnetic wave is

$$I = a\mathcal{E}^2$$

Where $\mathcal{E}$ is the amplitude of the wave. The probability density function of the amplitude is Gaussian.

$$p(\mathcal{E}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mathcal{E}^2}{2\sigma^2}}$$

Lets try to find $p(I)$. We'll start by drawing a graph.



Next we'll find the probability distribution of $\mathcal{E}$

$$P_I(\eta) = \int_{-\sqrt{\frac{\eta}{a}}}^{\sqrt{\frac{\eta}{a}}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mathcal{E}^2}{2\sigma^2}} d\mathcal{E}$$

Now we need to take the derivative wrt. η to get $p_I(\eta)$. By Leibnitz's integral rule we know that we can ignore the integral and we're left with

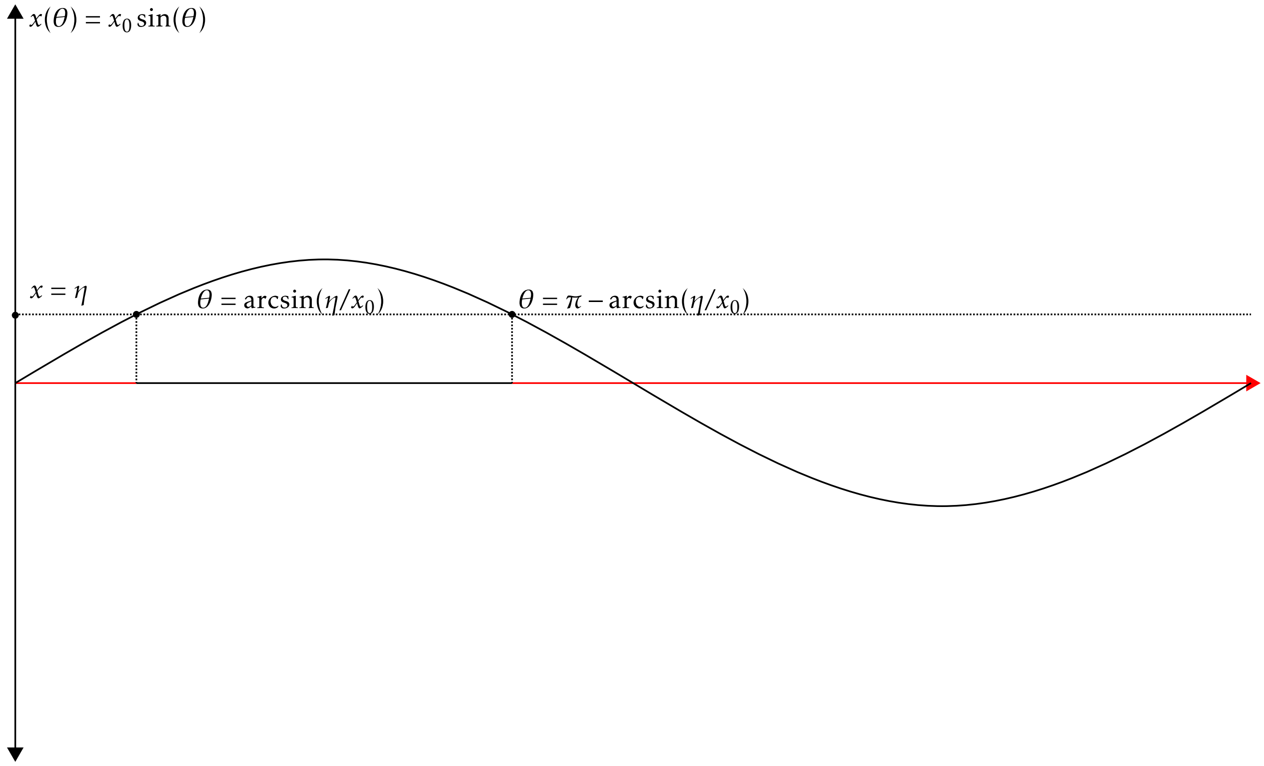
$$p_I(\eta) = \frac{dP_I(\eta)}{d\eta} = \frac{e^{-\frac{\eta}{2a\sigma^2}}}{\sqrt{2\pi\sigma^2}}\sqrt{\frac{1}{a\eta}} = \frac{1}{\sqrt{2\pi\sigma^2 a\eta}} \times \exp\left\{-\frac{\eta}{2a\sigma^2}\right\}$$

Not so bad I think. Lets try another one.

Example 7: (Harmonic Motion) Today, you decide, is the day you'll finally get into magic. The first trick you come up with is guessing the position of a harmonic oscillator with fixed total energy without ever looking at the harmonic oscillator. Your nephew is in the audience. He's also a big nerd so he knows that

$$p(\theta) = \frac{1}{2\pi} \qquad 0 \le \theta < 2\pi$$

Now he want to find the probability that you guess correctly given that you're truly randomly guessing (I guess he's setting up a null hypothesis?) Lets try to help him along.



Instead of integrating over the two red regions what we can do is integrate over the complement

$$P_x(\eta) = 1 - \int_{\arcsin(\eta/x_0)}^{\pi - \arcsin(\eta/x_0)} p(\theta) d\theta$$

From here we take the derivative wrt. η to get

$$p_x(\eta) = -\frac{d(\pi - \arcsin(\eta/x_0))}{d\eta} p(\pi - \arcsin(\eta/x_0)) + \frac{d(\arcsin(\eta/x_0))}{d\eta} p(\arcsin(\eta/x_0))$$

Now what in the world is the derivative of arcsin? First lets let $y = \arcsin(x)$ meaning that $\sin(y) = x$. Then we know that

$$1 = \frac{d(\sin(y))}{dx} = \frac{d\sin(y)}{dy}\frac{dy}{dx} = \cos(y)\frac{dy}{dx} = \cos(y)\frac{d(\arcsin(x))}{dx}$$

We can find $\cos(y)$ from basic trig

$$\cos(y) = \cos(\arcsin(x)) = \sqrt{1 - x^2}$$

And plugging this all in gives us

$$\frac{d(\arcsin(x))}{dx} = \frac{1}{\sqrt{1 - x^2}} \Rightarrow \frac{d(\arcsin(\eta/x_0))}{d\eta} = \frac{1}{\sqrt{1 - (\eta/x_0)^2}}\frac{1}{x_0}$$

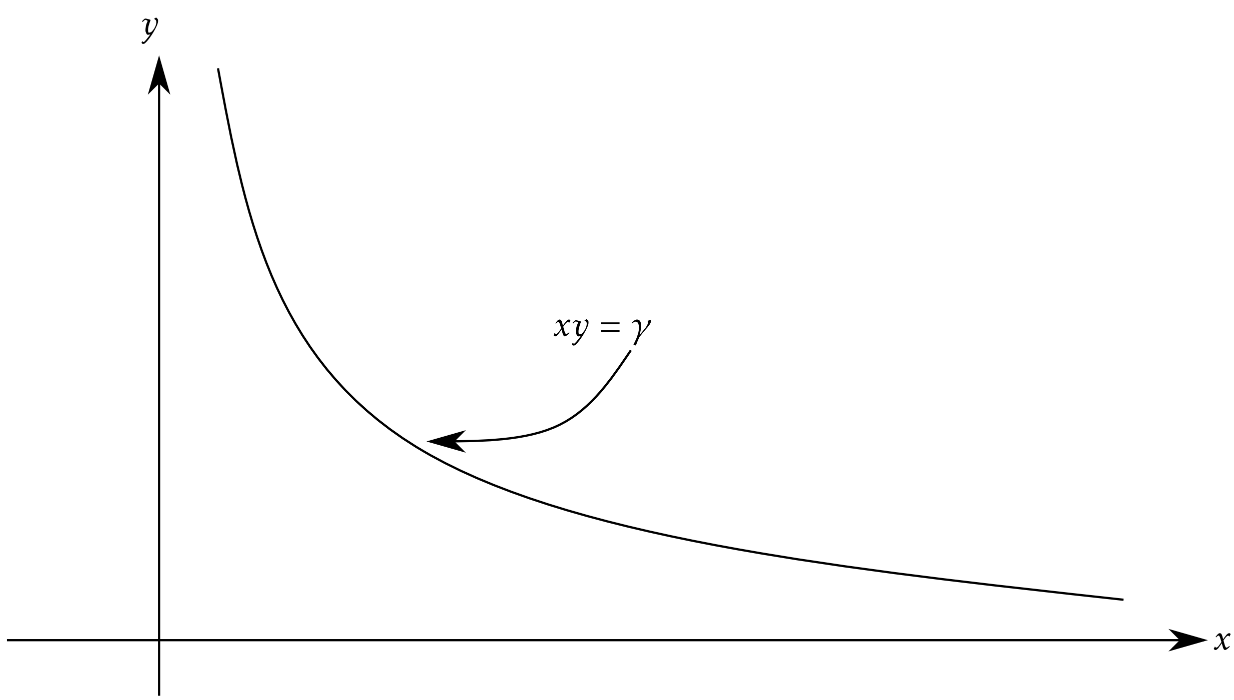
Plugging this all in gives us

$$p_x(\eta) = \frac{1}{\pi\sqrt{x_0^2 - \eta^2}} \qquad x > 0$$

The case when $x < 0$ is pretty much the same.

That was fun and all but in the previous few examples everything was done with a single random variable. What happens when we functions of multiple random variables

Example 8: (A Gambling Game like Pachinko except it's noting like Pachinko) You're in a casino playing with some slot machines. Howevere the machine you're playing with is kinda strange. The game goes as follows: the machine generates two random numbers x and y and multiplies them together. You know the probability $p(x,y)$. It is your job to bet some money. If you bet more than $x \times y$ then you get $x \times y$ and your money back. If not you lose the money you bet. Lets assume that the casino has unlimited money so that they can go as high as they like and you have the regular amount of money a person in a casino would have. To bet most efficiently you decide to figure out $p(z)$ where $z = x \times y$. How do you do this? We'll restrict ourselves to $x, y, z > 0$.



The region we want to integrate over is

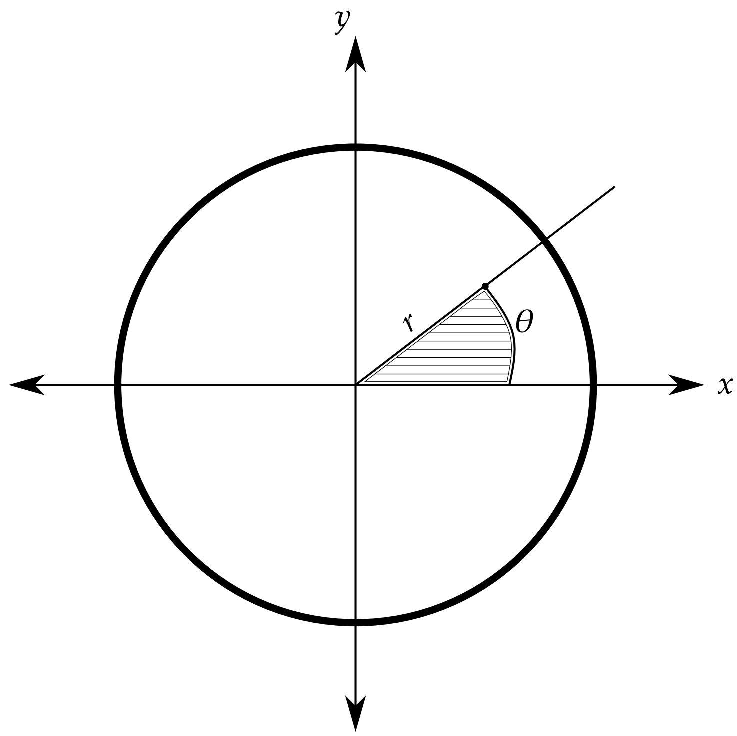
$$P_z(\gamma) = \int_0^\infty \int_0^{\gamma/y} p(x,y)dxdy \Rightarrow p_z(\gamma) = \frac{dP_z}{d\gamma} = \int_0^\infty dy \left(\frac{1}{y} p(\gamma/y, y)\right)$$

One useful application of functions of random variables is that they can be used to chnage of variables.

Example 9: (People keep getting blown up on my friend's farm!(But this time it's polar)) Same idea as last time

$$p(x,y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1 \\ 0 & x^2 + y^2 > 1 \end{cases}$$

But this time we want to use this to find the joint probability distribtuion $p(r, \theta)$.



From the picture we can compute the probabilty

$$P(r,\theta) = \underbrace{\left(\pi r^2\right)}_{\text{total area}} \underbrace{\left(\frac{\theta}{2\pi}\right)}_{\text{percent filled}} \underbrace{\left(\frac{1}{\pi}\right)}_{p(x,y)}$$

From this we can compute some probability densities

$$p(r,\theta) = \frac{\partial}{\partial r}\frac{\partial}{\partial \theta}P(r,\theta) = \frac{r}{\pi}$$

$$p(r) = 2r \qquad p(\theta) = \frac{1}{2\pi}$$

Using our rule for statistical independence we also note that $p(r)$ and $p(\theta)$ are statistically independent.

Assignment 1

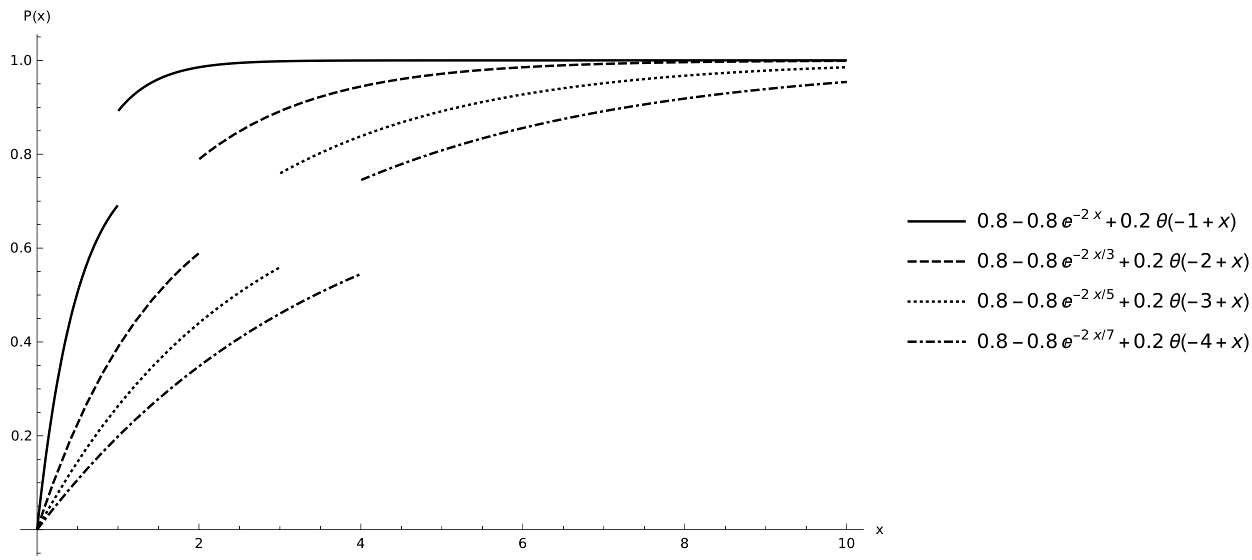
Started: February 03, 2021. Finished: February 04, 2021

Doping a Semiconductor

(a) They say we don't need to give an analytic expression for $P(x)$ but we'll do it anyways since we have mathematica. We have

$$P(x) = 0.2 \times \theta(x-d) - 0.8e^{-\frac{x}{l}} + 0.8$$

Where θ is the Heaviside function. Plotting this in mathematica as well



(b) To find $\langle x\rangle$ we just take an integral

$$\boxed{\langle x\rangle=\int_0^\infty p(x)x\,dx=\int_0^\infty x\left(\left(\frac{0.8}{l}\right)\exp\left\{-\frac{x}{l}\right\}+0.2\delta(x-d)\right)dx=0.2d+0.8l}$$

(c) The variance we found in lecture is

$$\sigma_x^2=\left\langle x^2\right\rangle-\langle x\rangle^2$$

We find that

$$\left\langle x^2\right\rangle=\int_0^\infty p(x)x^2dx=0.2d^2+1.6l^2$$

Putting this together gives us

$$\boxed{\sigma_x^2=.16d^2-.32dl+.96l^2}$$

(d) Taking another integral

$$\boxed{\left\langle\exp\left\{-\frac{x}{s}\right\}\right\rangle=0.2\times\exp\left\{-\frac{d}{s}\right\}+\frac{0.8s}{l+s}}$$

Peculiar Probability Distribution

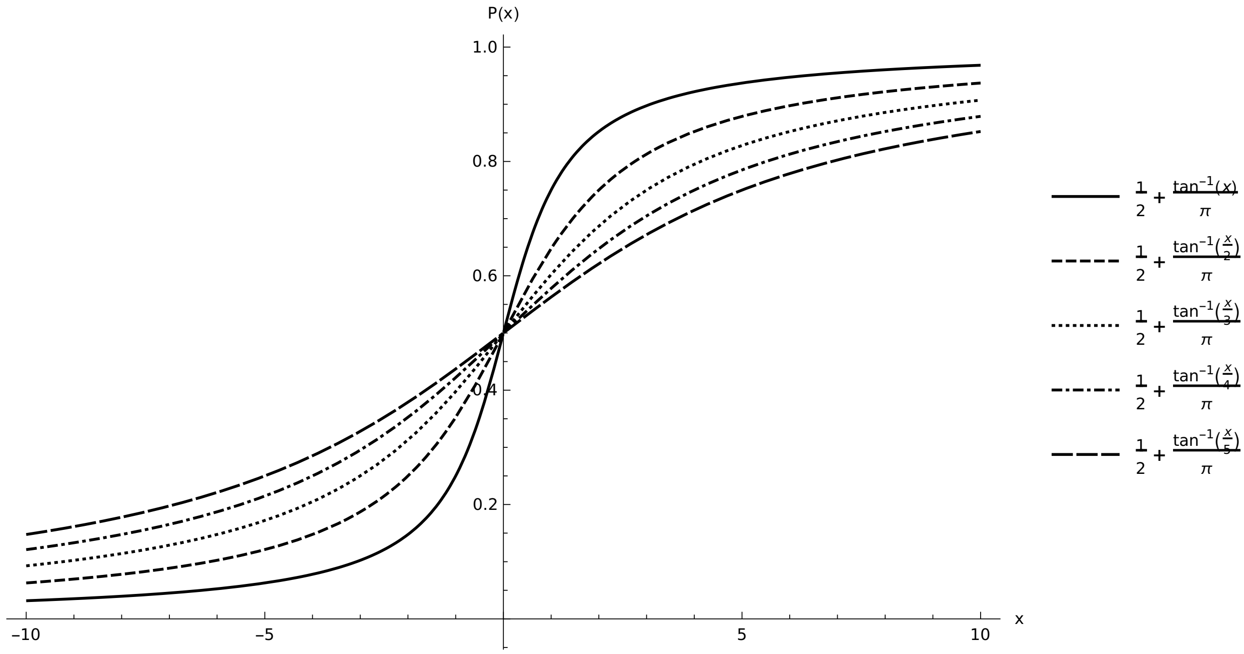
(a) Lets integrate

$$1=\int_{-\infty}^\infty p(x)dx=a\sqrt{\frac{1}{b^2}}\pi\Rightarrow\boxed{a=\frac{\sqrt{b^2}}{\pi}}$$

(b) Taking another integral

$$P(x)=\int_{-\infty}^x p(x)dx=\frac{1}{2}+\frac{\arctan\left(\frac{x}{b}\right)}{\pi}$$

We can plot this



(c) I don't think we need to take an integral here. $p(x)$ is an even function so

$$\boxed{\langle x \rangle = 0}$$

(d) We want to solve

$$p(x) = \frac{a}{2b^2} = \frac{1}{2\pi b} \Rightarrow x = \pm b$$

(e) Well lets just look at the integral real quick

$$\left\langle x^2 \right\rangle = \int_{-\infty}^{\infty} \frac{bx^2}{(b^2 + x^2)\pi}$$

As $x \to \pm\infty$ we have the integrand converging to a constant. So it can't be convergent. This means that

$$\text{Var}(x) \text{ diverges}$$

Probability Density for a Clasical Harmonic Oscillator

(a) We can take the derivative of $x(t)$ wrt t to find the velocity

$$\frac{dx}{dt} = \omega x_0 \cos(\omega t + \phi)$$

To find the speed we can square things and then take the square root.

$$v^2 = (\omega x_0)^2 \underbrace{(1 - \sin^2(\omega t + \phi))}_{\cos^2(\omega t + \phi)} = \omega^2(x_0^2 - x^2) \Rightarrow \boxed{\text{Speed} = \omega\sqrt{x_0^2 - x^2}}$$

(b) We know the probability of finding a particle in $[x, x+dx]$ is inverseley proportional to the speed at x. Thus we have

$$p(x) = \frac{A}{\omega\sqrt{x_0^2 - x^2}}$$

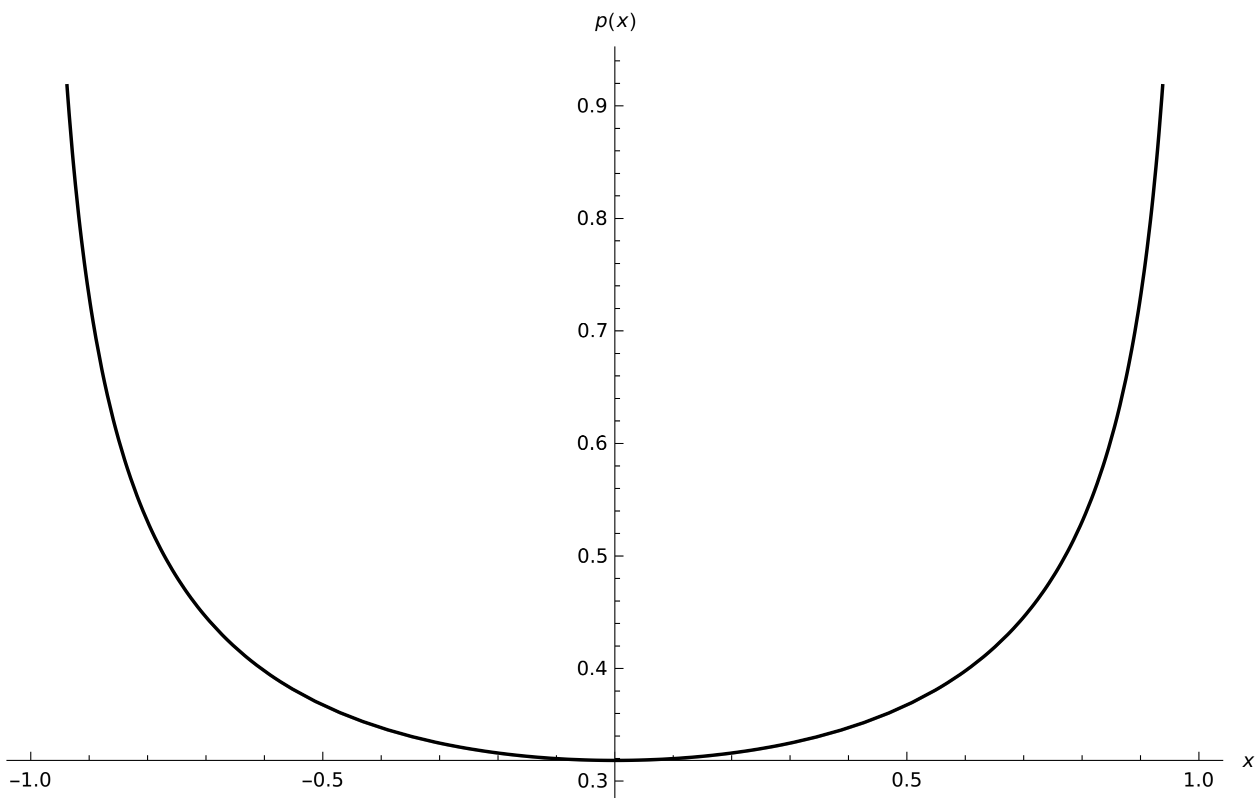
Where A is a normalization constant. We can find A by integrating between the bounds of x

$$1 = \int_{-x_0}^{x_0} p(x)dx = \frac{A\pi}{\omega} \Rightarrow A = \frac{\omega}{\pi}$$

This gives us

$$\boxed{p(x) = \frac{1}{\pi\sqrt{x_0^2 - x^2}}}$$

(c) Using mathematica we can plot



We can see that the most probable values of x are when the particle is moving the slowest (at the max displacements) and the least probable is when x is moving the fastest ($x = 0$). The mean is $\langle x \rangle = 0$ by symmetry.

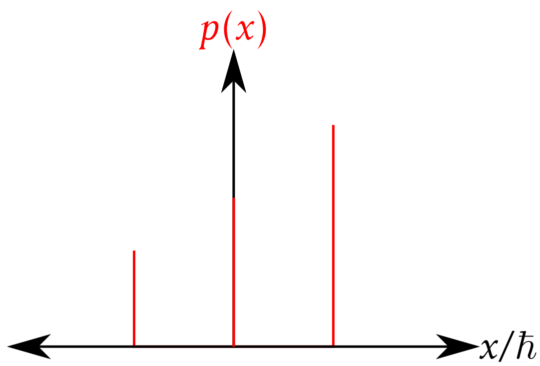
Quantized Angular Momentum

(a) All we have is $\langle L_x \rangle = \frac{1}{3}\hbar$ and $\left\langle L_x^2 \right\rangle = \frac{2}{3}\hbar$. Since we have only two unknowns(the third is determined by $1 - p_1 - p_0$) this should be enough. Lets setup our system of equations

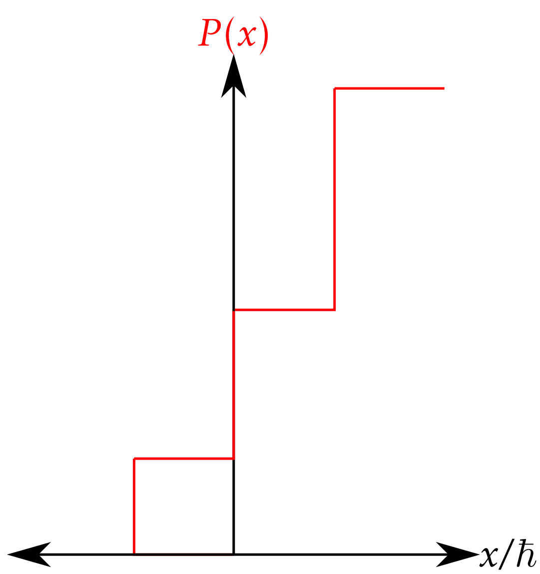
$$-p_{-1} + p_1 = \frac{1}{3}$$
$$p_{-1} + p_1 = \frac{2}{3}$$
$$\Rightarrow p_1 = \frac{1}{2} \text{ and } p_{-1} = \frac{1}{6} \Rightarrow p_0 = \frac{1}{3}$$

We can then write $p(L_x)$ as

$$\boxed{p(L_x) = \frac{\delta(x+\hbar)}{6} + \frac{\delta(x)}{3} + \frac{\delta(x-\hbar)}{2}}$$



(b) We can sketch the cumulative function as



COHERENT STATE OF A QUANTUM HARMONIC OSCILLATOR

(a) The terms with i in the exponential are just gonna vanish so we're left with

$$\Psi^*\Psi = \frac{1}{\sqrt{2\pi x_0^2}}\exp\left\{-\frac{(x-2\alpha x_0\cos\omega t)^2}{2x_0^2}\right\}$$

(b) Looks kinda like a Gaussian wave packet huh?

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} \Rightarrow \sigma = x_0 \text{ and } a = 2\alpha x_0\cos\omega t$$

From the properties of Gaussians' we have

$$\langle x\rangle = 2\alpha x_0\cos\omega t \qquad \mathrm{Var}(x) = \sigma^2 = x_0^2$$

(c) It's just a oscillating Gaussian wave packet.

BOSE-EINSTEIN STATISTICS

(a) Lets first normalize

$$1 = (1-a)\sum_n a^n$$

Remember those slick derivative tricks we did in lecture 1 to calculate properties of the poisson distribution? Lets try that here

$$\frac{d}{da}\frac{1}{1-a} = \frac{1}{(1-a)^2} = \frac{d}{da}\sum_{n=0}^{\infty}a^n = \sum_{n=0}^{\infty}a^{n-1}\times\underbrace{(n-1+1)}_{=n} = \sum_{n=0}^{\infty}a^{n-1}\times(n-1)+\sum_{n=0}^{\infty}a^{n-1}$$

In the LHS lets let $m = n-1$ giving us

$$\frac{d}{da}\sum_{n=0}^{\infty}a^n = \underbrace{\cancel{-a^{-1}}+\sum_{m=0}^{\infty}a^m\times(m)}_{\sum_{m=-1}^{\infty}a^m\times m}+\underbrace{\cancel{a^{-1}}+\sum_{m=0}^{\infty}a^m}_{\sum_{m=-1}^{\infty}a^m}$$

Now that the red term is $\langle m\rangle/(1-a)$. Furthermore by normalization we know that the blue term is $1/(1-a)$. We'll set what we have above equal to the green term.

$$\frac{1}{(1-a)^2} = (1-a)^{-1}(\langle m\rangle+1) \Rightarrow \frac{1}{1-a}-1 = \langle m\rangle \Rightarrow \langle m\rangle = \boxed{\langle n\rangle = \frac{a}{1-a}}$$

(b) For the variance we need $\langle n^2\rangle$. We're given the hint to take the derivative of $\langle n\rangle$ wrt a so lets try that

$$\frac{d\langle n\rangle}{da} = \underbrace{\frac{1}{1-a}}_{\langle n\rangle+1}+\underbrace{\frac{a}{(1-a)^2}}_{\langle n\rangle^2} = \frac{d}{da}\left((1-a)\sum_{n=0}^{\infty}a^n\times n\right) = (1-a)\times\underbrace{\langle n^2\rangle/a}_{\sum_{n=0}^{\infty}a^{n-1}\times n^2} - \underbrace{\frac{\langle n\rangle}{(1-a)}}_{\sum_{n=0}^{\infty}a^n\times n}$$

Setting the LHS equal to the green term ant noting $(1-a)/a = \frac{1}{\langle n \rangle}$ we get

$$\frac{1}{1-a} + \frac{a}{(1-a)^2} = \frac{\left\langle n^2 \right\rangle}{a} - \frac{a}{(1-a)^2} \Rightarrow \left\langle n^2 \right\rangle = \frac{a}{1-a} + \frac{2a^2}{(1-a)^2} = \langle n \rangle + 2\langle n \rangle^2$$

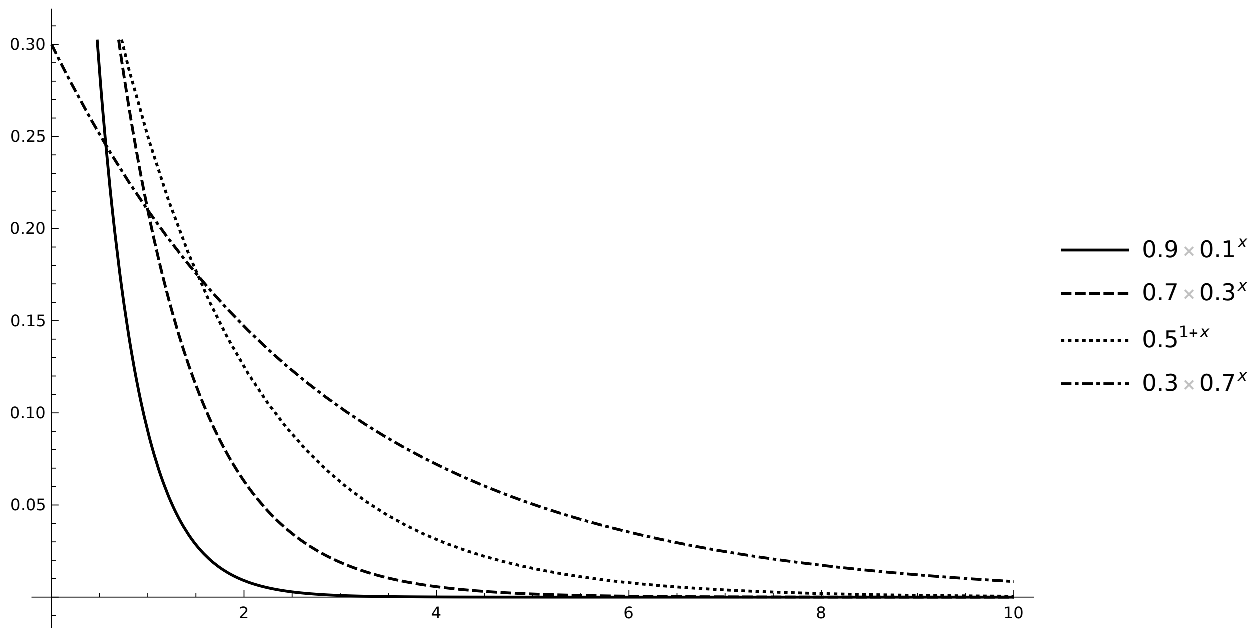
Putting everything together gives us

$$\boxed{\mathrm{Var}(n) = \left\langle n^2 \right\rangle - \langle n \rangle^2 = \langle n \rangle (1 + \langle n \rangle)}$$

(c) We are told to write the probability density as the product of an envelope function $F(x)$ times a train of δ functions.

$$p(x) = F(x) \times \left(\sum_{n=0}^{\infty} \delta(x-n) \right) = \underbrace{\sum_{n=0}^{\infty} \delta(x-n)(1-a)a^n}_{\substack{\text{original} \\ \text{probabilty}}}$$

Lets for a second imagine that the probabilty distribution is continuous. Would would it look like then? Using mathematica we plot the continous limit of the original probability for a few values of a



That does look pretty exponentially decreasing. Since the sum of δ takes care of discretizing everything we're trying to find

$$F(x) = \lambda e^{-x/\phi} = (1-a)a^x \Rightarrow \lambda = (1-a) \text{ and } \underbrace{e^{\ln(a)x}}_{a^x} = e^{-x/\phi} \Rightarrow \phi = -\frac{1}{\ln(a)}$$

The red terms are the parameters of our exponential. Lets try to get $\langle n \rangle$ in here somehow. First lets try to solve a in terms of $\langle n \rangle$. From our result in part (a) we have

$$\langle n \rangle = \frac{a}{1-a} \Rightarrow a = \frac{\langle n \rangle}{1 + \langle n \rangle}$$

Plugging this into our formula for ϕ we get

$$\phi = -\left(\ln\left\{\frac{\langle n\rangle}{1+\langle n\rangle}\right\}\right)^{-1}$$

Lets rewrite the argument of ln as

$$\frac{\langle n\rangle}{1+\langle n\rangle} = \left(1 - \frac{1}{\langle n\rangle}\right)^{-1} = (1+\omega)^{-1}$$

Note as $\langle n\rangle \to \infty$ we have $\omega \to 0$ so we can do a binomial expansion

$$\frac{\langle n\rangle}{1+\langle n\rangle} = 1 + \omega + \ldots$$

This for large $\langle n\rangle$ we have

$$\ln\left\{\frac{\langle n\rangle}{1+\langle n\rangle}\right\} \approx \ln\{1+\omega\}$$

Now lets expand ln for $\omega \to 0$ again. This gives us

$$\ln\left\{\frac{\langle n\rangle}{1+\langle n\rangle}\right\} \approx \omega = -\frac{1}{\langle n\rangle}$$

Thus as $\langle n\rangle \to \infty$ we get

$$\boxed{\phi \to \langle n\rangle}$$

Lecture 4: Sums of Random Variables

Started: February 06, 2021. Finished: TODO

Lets start by proving some things about means! Consider a bunch of statistically independent random varaibles $\{x_i\}$ where the probability density of each x_i can be different. The sum of these random variables is $S_n = \sum_{i=1}^n x_i$ which is a new random variable. Lets consider the properties of S_n

$$\begin{aligned}
\langle S_n\rangle &= \int dx_1 \ldots \int dx_i \ldots \int dx_n \left(\sum_{i=1}^n x_i\right) \underbrace{p(x_1,\ldots,x_i,\ldots,x_n)}_{\substack{p(x_1)\times\cdots\times p(x_i)\times\cdots\times p(x_n) \\ \text{due to S.I.}}} \\
&= \sum_{i=1}^n \left(\underbrace{\int x_i p(x_i) dx_i}_{\langle x_i\rangle} \times \prod_{i\neq j} \underbrace{\int p(x_j) dx_j}_{1}\right) \\
&= \sum_{i=1}^n \langle x_i\rangle
\end{aligned}$$

$$\langle S_n^2\rangle = \int dx_1 \ldots \int dx_i \ldots \int dx_n \underbrace{\left(\sum_{i=1}^{n} x_i^2 + \sum_{i\neq j} 2x_i x_j\right)}_{\left(\sum_{i=1}^{n} x_i\right)^2} p(x_1)\times\cdots\times p(x_i)\times\ldots p(x_n)$$

$$= \sum_{i=1}^{n} \underbrace{\left(\underbrace{\int x_i^2 p(x_i) dx_i}_{\langle x_i^2\rangle} \times \prod_{i\neq j} \underbrace{\int p(x_j) dx_j}_{1} \right)}_{\langle x_i^2\rangle} + 2\sum_{i\neq j} \underbrace{\int x_i dx_i}_{\langle x_i\rangle} \times \underbrace{\int x_j dx_j}_{\langle x_j\rangle} \times \prod_{i\neq j\neq k} \underbrace{\int p(x_k) dx_k}_{1}$$

$$= \sum_{i=1}^{n} \langle x_i^2\rangle + \sum_{i\neq j} 2\langle x_i\rangle\langle x_j\rangle$$

From this we can find the variance

$$\mathrm{Var}(S_n) = \langle S_n^2\rangle - \langle S_n\rangle^2 = \sum_{i=1}^{n}\langle x_i^2\rangle + \cancel{\sum_{i\neq j} 2\langle x_i\rangle\langle x_j\rangle} - \left(\sum_{i=1}^{n}\langle x_i\rangle^2 + \cancel{\sum_{i\neq j} 2\langle x_i\rangle\langle x_j\rangle}\right) = \sum_{i=1}^{n} \underbrace{\langle x_i^2\rangle - \langle x_i\rangle^2}_{\sigma_i^2}$$

So from all this mess we find that

$$S_n = \sum_{i=1}^{n} x_n \Rightarrow \langle S_n\rangle = \sum_{i=1}^{n}\langle x_i\rangle \text{ and } \mathrm{Var}(S_n) = \sum_{i=1}^{n}\mathrm{Var}(x_i)$$

In the special case where we're summing the same variable we have

$$\langle S_n\rangle = n\langle x\rangle \text{ and } (S_n) = n\times\mathrm{Var}(x)$$

The probability density of S_n trends to spiking at $\langle S_n\rangle$ for large n since the mean grows as n while the width of the peak grows as $\sqrt{n}$.