Week 7

1 July 2021

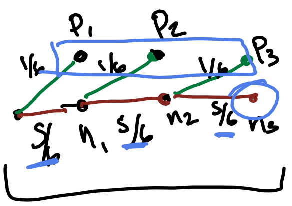
34 Geometric Distribution

We start off this week with a question which was posed in chapter 1 of the book, but I waited until now to talk about. Say you're rolling a die multiple times and you want to see how many times you have to roll it until you get a 6. The first roll, you have a $p = \frac{1}{6}$ chance of getting a six. That means you have a 16.666...% chance to roll a six in 1 roll and a $1 - 16.66\ldots = 83.33\ldots\%$ chance to roll a six in 2 or more rolls.

1st not a 6 — 2nd a 6

The chances of a six showing up on the second roll is: $\frac{5}{6} \cdot \frac{1}{6}$ where the $\frac{5}{6}$ is coming from you *not* rolling a six in the first roll. So to roll a six in the first two rolls gives us $\frac{1}{6} + \frac{5}{6}\frac{1}{6} = \frac{11}{36} \approx 0.3056$. Therefore, there's a 30.56% chance of rolling a six in your first two rolls.

P_i = Getting a 6 for the 1st time on ith roll



$1 - n_3 = 1 - q^3$

What about three? Well, we do the same. We want to add a third roll to our list. In this case, we need to fail twice and succeed the third time so we get: $\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} = \frac{25}{216}$. This gives us the probability of rolling a six only on the third roll. Adding this to our second sum gives:

Diff from Birthday Problem

$$\boxed{\frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6}} = \frac{91}{216} \approx 0.4213 \quad 42.13\%$$

Birth: sampling w/o replacement

which is the probability we roll a six in our first three rolls.

Geo: sampling w/ replacement.

Notice how we're getting a little pattern here. For k rolls, we have

$$\sum_{i=1}^{k} \left(\frac{5}{6}\right)^{i-1} \frac{1}{6} = \underbrace{\frac{1}{6}}_{p} \sum_{i=1}^{k} \underbrace{\left(\frac{5}{6}\right)}_{(1-p)}^{i-1}$$

If we replace $\frac{1}{6}$ with p and $\frac{5}{6}$ with $(1-p)$, we end up with $p \sum_{i=1}^{k} (1-p)^{i-1}$

But, with a little trick, we can actually simplify this. Let $q = 1 - p$ and then we get that $p = 1 - q$. Plugging this in we have:

$$\underbrace{(1-q)}_{p}\sum_{i=1}^{k} \underbrace{q}_{1-p}^{i-1} = \sum_{i=1}^{k} q^{i-1} - \sum_{i=1}^{k} q^{i} = 1 - q^{k}$$

In other words, the chance of rolling a six in the first 3 rolls is:

$$1 - q^{k} = 1 - \left(\frac{5}{6}\right)^{3} = 0.4213$$

With our notation we also have that the chance of rolling a six *on* the fifth roll is $\left(\frac{5}{6}\right)^{4}\frac{1}{6}$. Generalizing this gives us a probability distribution which is known as the *geometric distribution*. Let X be the number of times that something happens with probability p and let $q = 1 - p$.

Wikipedia: Geometric distribution

$$P(X=k) = q^{k-1}p \qquad P(X \leq k) = 1 - q^{k}$$

35 Discrete Distribution

Notice how for the geometric distribution, we can let k be any number! In other words, our sample space is no longer finite. This is the next step we take in our study of distributions from a discrete angle. A *discrete distribution* is a probability distribution where the sample space is the set of non-negative integers $\{0, 1, 2, 3, \ldots\}$ with a sequence of probabilities $p_0, p_1, p_2, p_3, \ldots$ such that each $p_i \geq 0$ and $\sum_i p_i = 1$.

Wikipedia: Discrete distribution

Examples of discrete distributions are:

- Discrete uniform distribution — rolling a die $\quad p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$, $p_i = 0$ for all other i.
- Bernoulli distribution — flipping an (unfair) coin
- Binomial distribution — n-indepent trials of a Bernoulli dist.
- Geometric distribution — multiple trials until certain outcome (w/ replacement)
- Poisson distribution — $e^{-\mu}\frac{\mu^{k}}{k!} \leftarrow P(X=k)$

Note that we do need to introduce one rule in order to make discrete distributions a distribution.

Theorem 35.1 (Infinite sum rule) *If an event A is partitioned into $A_1, A_2, A_3, \ldots$:*

$$A_1 \sqcup A_2 \sqcup A_3 \sqcup \cdots$$

$A_i \sqcup A_j \qquad A_i \cup A_j$ s.t. $A_i \cap A_j = \emptyset$

then

$$P(A) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

With that, all of our rules are satisfied for discrete distributions to be distributions. non-negative ✓ addition ✓ total sum ✓

Note that finite distributions are discrete distributions! For example, if we want to just roll a dice the distribution involves only six numbers: 1, 2, 3, 4, 5, 6. Where

$$P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = P(X = 6) = \frac{1}{6}$$

In order to view this as a discrete distribution, you just need to let $P(X = i) = 0$ whenever $i \notin \{1, 2, 3, 4, 5, 6\}$.

Example 35.2 Let's say that your friend decides to tell you that they're luckier than you. They pick up a dice and say, "I bet you I can get a six before you can". Before you agree, you decide to calculate the chances that they're right. What's the probability that your friend will win?

friend wins if $1^{st}, 3^{rd}, 5^{th}, 7^{th}, \ldots$ or odd roll is a 6

you win if $2^{nd}, 4^{th}, 6^{th}, \ldots$ or even roll is a 6

Let X be getting a 6 in x^{th} roll (for 1^{st} time)

↳ Geometric

$$P(X=k) = (1-p)^{k-1}p = q^{k-1}p \quad \begin{bmatrix} q = 5/6 \\ p = 1/6 \end{bmatrix}$$

$$P(\underbrace{\tfrac{X}{2} \notin \mathbb{Z}}_{\text{not an even number}}) = P(X=1) + P(X=3) + P(X=5) + \cdots$$

$$= p + pq^2 + pq^4 + pq^6 + \cdots$$

$$= p(1 + q^2 + q^4 + q^6 + \cdots)$$

$$= p\frac{1}{1-q^2}$$

$$= \frac{1}{6} \cdot \frac{1}{1-(\frac{5}{6})^2}$$

$$= \frac{1}{6(1-\frac{25}{36})} = \frac{1}{\cancel{6}(\frac{11}{\cancel{36}_{6}})} = \frac{6}{11} \approx 54.55\%$$

Geometric series

$$\sum_{i=0}^{n} q^i = \frac{1}{1-q} \quad (q<1)$$

$$\sum_{i=0}^{n} (q^2)^i = \frac{1}{1-q^2}$$

$$\frac{36}{36} - \frac{25}{36} = \frac{11}{36}$$

45.44%

36 Discrete moments

What can we say about the expected value and, more specifically, the moments of a discrete distribution. The expected value comes naturally

$$E(X) = \sum_x xP(X = x) \quad \neq \infty$$

but only if this sum is well-defined! So, we say that the expected value is $E(X)$ (as defined above) *if* the series is absolutely convergent:

$$\sum_x |x| P(X=x) \quad < \infty \text{ Converges}$$

Why absolute convergence and not normal convergence? This comes from the Riemann series theorem which states that we can only rearrange the terms in our series if the series is absolutely convergent. Note that all of our discrete distributions *are* convergent if and only if they are absolutely convergent. Why? Because we stated that $P(X = x) = p_x \geq 0$ and we said that our sample space is all non-negative integers! So a non-negative integer times a non-negative integer is always non-negative! We only need to care about absolute convergence if we allow negative terms in our sample space.

$1+2+3+4+\cdots \neq -\frac{1}{12}$

What if we look at an arbitrary function instead of just X for the expected value? Things like the moments $E(X^2)$, $E(X^3)$ or just any other function. Supposing that our function is $f(X)$ then

$$\boxed{E(f(X)) = \sum_x f(x) P(X = x)}$$

is defined if the right-hand side is absolutely convergent.

Example 36.1 Let's go back to the geometric distribution we've been working with most of this week. What is the expected value for when the first six will show up when rolling a fair six-sided die? In other words, how many times do we expect to roll the die before the first six shows up?

X is the roll where we get a 6 for the first time

$$E(X) = \sum_{x=1}^{\infty} x \underbrace{P(X=x)}_{q^{x-1}p} = \sum_{x=1}^{\infty} x q^{x-1} p = p \sum_{x=1}^{\infty} x q^{x-1}$$

$$p\sum_{x=1}^{\infty} x q^{x-1}$$

$$q^0 + q^1 + q^2 + q^3 + \cdots = \sum_{i=1}^{\infty} q^{i-1} = \frac{1}{1-q}$$

$$\left(\sum_{i=1}^{\infty} i q^{i-1}\right) = 1\cdot q^0 + 2\cdot q^1 + 3\cdot q^2 + 4\cdot q^3 + \cdots$$

$$-\left(q\sum_{i=1}^{\infty} i q^{i-1}\right) = 1q^1 + 2q^2 + 3q^3 + \cdots$$

$$= 1\cdot q^0 + 1\cdot q^1 + 1\cdot q^2 + 1\cdot q^3 + \cdots = \sum_{i=1}^{\infty} q^{i-1}$$

$$(1-q)\sum_{i=1}^{\infty} i q^{i-1} = \sum_{i=1}^{\infty} q^{i-1} = \frac{1}{1-q} \Big/ (1-q)$$

$$\Rightarrow \sum_{i=1}^{\infty} i q^{i-1} = \frac{1}{(1-q)^2}$$

$$E(x) = p\sum_{x=1}^{\infty} x q^{x-1} = \frac{p}{(\underbrace{1-q}_{1-q=p})^2} = \frac{p}{p^2} = \boxed{\frac{1}{p}}$$

$$p = 1/6 \Rightarrow E(x) = \frac{1}{1/6} = \boxed{6}$$

$E(X) = 1/p$

Calculate the standard deviation. $SD(X) = \sqrt{Var(X)} = \sqrt{E(X^2) - E(X)^2}$

$$E(X^2) = \sum_{x=1}^{\infty} x^2 P(X=x) = p\underbrace{\sum_{x=1}^{\infty} x^2 q^{x-1}}$$

$$\left(\sum_{i=1}^{\infty} i^2 q^{i-1}\right) = 1q^0 + 4q^1 + 9q^2 + 16q^3 + \cdots$$

$$-\left(q\sum_{i=1}^{\infty} i^2 q^{i-1}\right) = 1q + 4q^2 + 9q^3 + \cdots$$

$$(1-q)\left(\sum_{i=1}^{\infty} i^2 q^{i-1}\right) = 1q^0 + 3q^1 + 5q^2 + 7q^3 + \cdots$$

$$\left(2\sum i q^{i-1}\right) = 2q^0 + 4q^1 + 6q^2 + 8q^3 + \cdots$$

$$-\left(\sum q^{i-1}\right) = q^0 + q^1 + q^2 + q^3 + \cdots$$

$$= q^0 + 3q^1 + 5q^2 + 7q^3 + \cdots$$

$$\underbrace{(1-q)}_{p}\sum_{i=1}^{\infty} i^2 q^{i-1} = 2\sum_{i=1}^{\infty} i q^{i-1} - \sum_{i=1}^{\infty} q^{i-1}$$

$$= 2\cdot\frac{1}{p^2} - \frac{1}{p}$$

div. by p.

$$\sum_{i=1}^{\infty} i^2 q^{i-1} = \frac{2}{p^3} - \frac{1}{p^2}$$

$$= \frac{2-p}{p^3}$$

$$E(X^2) = p\cdot\frac{2-p}{p^3} = \frac{2-p}{p^2}$$

$$SD(X) = \sqrt{\left(\frac{2-p}{p^2}\right) - \left(\frac{1}{p}\right)^2} = \sqrt{\frac{2-p-1}{p^2}} = \boxed{\frac{1}{p}\sqrt{1-p}}$$

$p = 1/6$

$SD(X) = 6\sqrt{5/6}$

$= 5.477$

Example 36.2 Let's look at a second example since these things can get a little confusing. Say that McDonald's has a new collection of toys for their happy meals and you've decided you want every single toy! Suppose that the chance of getting each toy is independent and there are n toys in total. If each purchase gives you a toy at random, how many happy meals are you expected to buy before you get every toy?

$E(X)$ where X is # happy meals to get every toy.

1^{st} toy = 1^{st} happy meal $= 1$

$\Rightarrow$ if $n=1$ $E(X) = 1 = E(X_1)$

X_i = chance of getting ith toy after having gotten $i-1^{st}$ toy.

2^{nd} toy $\to$ every happy meal has $\underbrace{\frac{n-1}{n}}$ $= p$, w/ geo.

if $n=2$ $E(X) = E(X_1) + E(X_2)$

$= 1 + \frac{n}{n-1}$

3^{rd} toy $\to$ every happy meal has $\frac{n-2}{n} = p$, w/ geo

if $n=3$ $E(X) = E(X_1) + E(X_2) + E(X_3)$

$= \underbrace{1}_{\frac{n}{n}} + \frac{n}{n-1} + \frac{n}{n-2}$

any n $E(X) = \sum_{i=1}^{n} E(X_i)$

$= \sum_{i=1}^{n} \frac{n}{n-(i-1)} = n \sum_{i=1}^{n} \frac{1}{n-(i-1)}$

$= n \underbrace{\sum_{i=1}^{n} \frac{1}{i}}_{\text{Harmonic series}}$

$\frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1}$

$n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{1}\right)$

$n\left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}\right)$

if $n=5$ $\to$ $E(X) = 5\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right)$

$= \frac{137}{12} = 11.42$

37 Poisson Distribution

We're now gonna take a step back and look at a distribution from way back in the day: the Poisson distribution. In Week 3 we had talked about the binomial distribution where we have some trial which is successful with probability p. We then repeat this trial n times and we ask what is the probability that exactly X were successful. There were two ways we could approximate this: the normal distribution for when p is close to $\frac{1}{2}$ and the Poisson distribution for when p is sufficiently small. Recall that the Poisson approximation was given by:

$$\boxed{P(X = k) \approx e^{-\mu}\frac{\mu^k}{k!}}$$

Binomial

where $\mu = E(X) = np$.

The *Poisson distribution with parameter μ* is the distribution:

Wikipedia: Poisson distribution

$$\boxed{P(X = k) = e^{-\mu}\frac{\mu^k}{k!}}$$

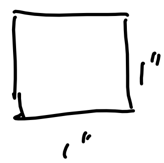
$k \in \mathbb{Z}_{\geq 0}$

The super nice thing about the Poisson distribution is that it has a super easy expected value and standard deviation. Supposing that X has a Poisson distribution with parameter μ then

$$\boxed{E(X) = \mu} \text{ and } \boxed{\mathrm{SD}(X) = \sqrt{\mu}}$$

Example 37.1 It is known that the number of raindrops that fall on a particular square inch of roof in a one-second interval of time is given by the Poisson distribution. Supposing that it has Poisson distribution with parameter 2, what is the variation for the number of raindrops (that fall on a particular square inch of roof in a one-second interval)?

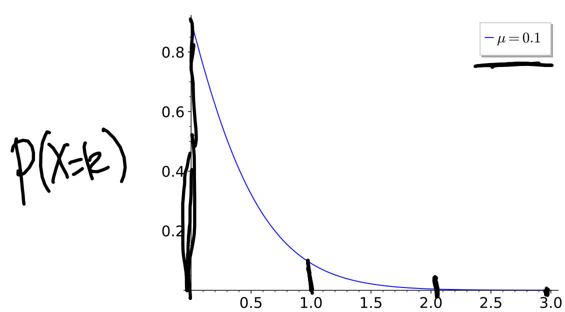
$\mu = 2$



$X = \#$ raindrops/sec.

$$\mathrm{Var}(x) = \mathrm{SD}(X)^2 = \sqrt{\mu}^2 = \mu = 2$$

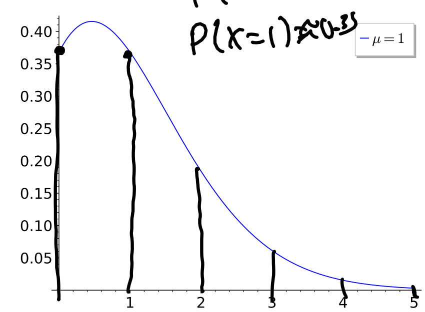
Let's look at a couple of examples of the Poisson distribution with different values of μ.

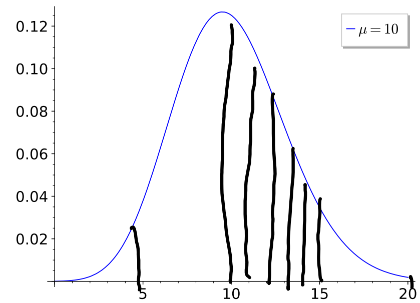


$P(X=0) \approx 0.36$

$P(X=1) \approx 0.35$

$p(X=k)$





Notice how, as we plug-in higher μ, we get closer to having a normal distribution. What this means is that we can use our normal approximation for Poisson distributions as well! So if we have a Poisson distribution on a random variable X then

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b+\frac{1}{2}-\mu}{\sqrt{\mu}}\right) - \Phi\left(\frac{a-\frac{1}{2}-\mu}{\sqrt{\mu}}\right)$$

And the skew-normal approximation is given by

$z = \frac{b+\frac{1}{2}-\mu}{\sqrt{\mu}}$

$$P(X \leq b) \approx \Phi(z) - \frac{1}{6\sqrt{\mu}}(z^2-1)\phi(z)$$

Another cool thing about the Poisson distribution is that they're relatively easy to sum up. Suppose we have n independent Poisson variables $X_1, X_2, \ldots, X_n$ with parameters $\mu_1, \mu_2, \ldots, \mu_n$ respectively. Then we know that $X_1 + X_2 + \cdots + X_n$ is a Poisson random variable with parameter $\mu_1 + \mu_2 + \cdots + \mu_n$.

Example 37.2 Suppose that we're watching a game of bowling where four players are on a team together. The first player has a $\frac{1}{15}$ chance of hitting a strike, the second player has a $\frac{1}{20}$ chance of hitting a strike, the third play has a $\frac{1}{70}$ chance and the final player has a $\frac{1}{10}$ chance. They each will throw the bowling ball 10 times and see how many strikes they get. What is the probability distribution for how many strikes the team will get as a whole? (Use Poisson distribution for each player and assume that each throwing of a ball is independent.)

Poisson for each roll

$\mu + \mu + \mu + \mu + \mu + \cdots + \mu = 10 \cdot \mu$

$\mu_1 = \quad = 10 \cdot \mu = \frac{10}{15}$

$\mu_2 = \frac{10}{20}$

$\mu_3 = \frac{10}{70}$

$\mu_4 = \frac{10}{10}$

$S = X_1 + X_2 + X_3 + X_4$

$\mu = \mu_1 + \mu_2 + \mu_3 + \mu_4$

$= \frac{10}{15} + \frac{10}{20} + \frac{10}{70} + \frac{10}{10} = \frac{97}{42}$

$\Rightarrow S$ is Poisson

$\boxed{\text{Dist. is a Poisson distribution w/ parameter } 97/42}$

38 Random Scatters

We're now going to look into a particular way of looking at the Poisson distribution in order to help us tackle a common question. We've stated the problem before, but we'll state it again.

Suppose you're trying to measure how many times a particle hits some surface. You can either think of this as raindrops hitting some sheet, or other particles like dust, molecules, photons, etc. In biology you can think of this as the positions of different cells or organisms on a microscope slide. In astronomy you can think of this as positions of stars in the night sky. In baking, you can think of this as positions of chocolate in a chocolate chip cookie. As you can see, this idea can be thought of in a lot of different ways!

This entire model is dependent on the fact that the number of points in a given area has the Poisson distribution. The same idea holds if we change our dimension from 2 to 3 to 4 or higher! We'll still have the Poisson distribution if looking at the number of points in a given volume.

In order to do this, we'll make two key assumptions:

(1) α No point is in the exact same location.

(2) β There's an equal chance for the point being anywhere. (Aka, it's random)

The frequency of how many points will occur will be estimated by a constant λ. This gives us the Poisson scatter theorem.

Wikipedia: Poisson Scatter Theorem

$\alpha, \beta \Rightarrow$ ① ②

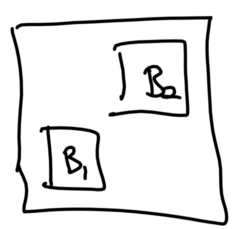
Theorem 38.1 (Poisson Scatter Theorem) *Suppose we have a square sheet where there are no points in the exact same location and all the points are randomly placed.*

α β

Let B be a subset of the square and X_B be the number of hits in B.

(1) Then X_B is a Poisson random variable with parameter $\lambda \times area(B)$.

(2) If B_1 and B_2 are disjoint subsets, then X_{B_1} and X_{B_2} are independent events.



This random scatter is called the Poisson scatter with intensity λ. The intensity λ is the expected number of hits per unit area.

Conversely, if the above two properties hold, then there are no points in the exact same location and all the points are randomly placed.

①,② ⇒ α, β

Example 38.2 Suppose we live in a fancy house that has a pool, but we're not rich enough to clean the pool during the winter. When summer hits we look at our pool and find a bunch of bacteria growing! In a volume of 1000 drops we find roughly 2000 bacteria (which are all separate from each other and randomly located throughout the water). Loving biology, we take out a single drop and we smear it (uniformly) over the surface of a dish. Since we're totally enamoured with bacteria, we put some food for the bacteria and leave them alone. We come back after a few days and find nice healthy colonies of bacteria on our dish! What is the distribution of the number of colonies over the whole plate and over an area of half the plate? (A colony is basically a large group of bacteria that huddles together after being birthed from the same great great grandbacteria.)

$\frac{2000 \text{ bacteria}}{1000 \text{ drops}} \sim \frac{2 \text{ bacteria}}{\text{drop}}$

$\boxed{\mu = 2} \to$ # of colonies on whole plate

1/2 plate $\Rightarrow \boxed{\frac{\mu}{2} = 1}$

Now suppose that there's a probability of p that each bacterium will die (independently) for whatever reason. What's the distribution of the number of colonies on the whole dish?

$\lambda = 2$ is our intensity as that's our scatter.

Let B be small enough that it contains exactly 1 colony

$P(1 \text{ bacteria in } B) \approx 2 \cdot \text{area}(B)$

$P(2 \text{ bacterian in } B) \approx 0$

$P(1 \text{ colony in } B) = P(1 \text{ bacteria in } B \text{ \& colony forms}) + P(2 \text{ bacteri \& colony forms})$

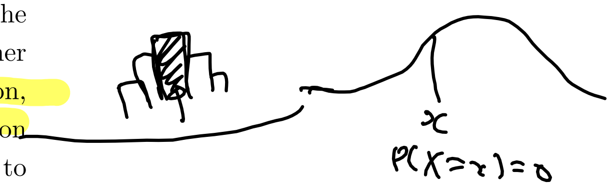
$= 2 \cdot \text{area}(B) \cdot p + 0 \cdot p$

if area$(B)=1$ $\Rightarrow$ the scatter of colonies has intensity $2p$ per unit area
$\Rightarrow$ the total # of colonies on the whole plate has a Poisson dist. w/ parameter $2p$.

This example is actually what is known as the "thinning of a Poisson scatter". If we have a Poisson scatter with intensity λ and each point has a probability of p of staying, then the scatter of points that are kept is a Poisson scatter with intensity λp.

39 Continuous Distributions

We've seen both continuous and discrete distributions so far, but have mainly focused on the discrete. In the discrete case we learned how to handle the expected value and the standard deviation, but now we want to do the same thing for continuous distributions. Remember that the main difference between discrete and continuous distributions is whether we can count them or not. For example, with the binomial distribution, we can go through the cases one at a time. With the normal distribution we couldn't do that! In order to figure out the distribution we had to take the area under the curve. And so now, we get to the part where integration will finally be used.



Let's go back and remember how the probabilities of a normal distribution were determined. We had some function, let's say $f(x)$, and this gave us some bell curve. This bell curve came from looking like the histogram of the binomial distribution. To get the probability at a certain point, the binomial distribution was easy since it was all rectangles. So calculating the area was simple. Calculating the probability at a certain point for the normal distribution was more difficult since we couldn't look at the distribution at a given point since the width was basically zero! So we said we just need to take the integral:

$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

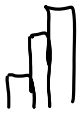
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

That's it! As long as the function $f(x)$ is continuous, we can take this integral, which is why these distributions are called *continuous distributions*. The function $f(x)$ is called the *probability density*. Other than being continuous, the only other thing we require of a probability density is that the area under the curve must be equal to 1 (or else it's not a distribution). This allows for

Wikipedia: Continuous Distribution
Wikipedia: Probability Density

$$P(\text{everything}) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

At this point we might ask how all of the terminology from the



$\int = S$ Gauß

discrete world moves into the continuous world. It turns out, it's almost the same! For example, in the discrete case we had the following formula for expected value:

$$E(X) = \sum_{\text{all } x} x \, P(X=x)$$

If you recall, by making the boxes smaller and smaller, we can turn this into a Riemann sum which leads into integration! So for the continuous case we have:

$$E(X) = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

To calculate the variance, we have $\text{Var}(X) = \underline{E(X^2)} - E(X)^2$ and so we also need:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \, dx$$

which also gives us the standard deviation $\text{SD}(X) = \sqrt{\text{Var}(X)}$.

Ok, so what about independence? Remember for the discrete case we had the following definition for independence

$$\underbrace{P(X = x, Y = y) = P(X \overset{=0}{=} x)P(Y \overset{=0}{=} y)}$$

Notice how for the continuous case this formula doesn't help! Since the probability at any point is equal to 0 we have that both sides always equal 0. In other words, if we used this formula, every continuous distribution would be independent, which doesn't make sense. We *won't* be covering how to define independence in the continuous case. If we have time at the end of the year, I'll cover it as a bonus, but for now we will skip it. The main thing to know is that *if* two continuous random variables are independent then all other independence rules hold (like we still have $\underbrace{E(XY) = E(X)E(Y)}$ if X and Y are independent and both $E(X)$ and $E(Y)$ are defined and finite).