

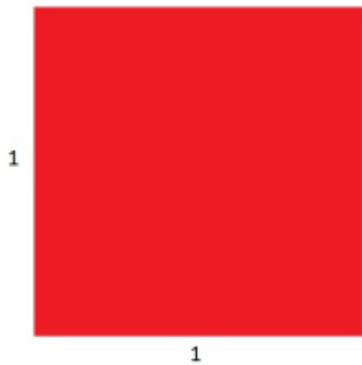
How to build an Anti Aircraft Missile: Probability, Bayes' Theorem and the Kalman Filter

Ever wondered how an Anti Aircraft Missile works? A plane can move at different speeds and altitudes, so how do you know when to fire your missile? Well you need to know two things: where the aircraft is now and where it will be a short time in the future. The reason you need to know where it will be in the future is because your missile takes time to reach the plane's altitude, so you need to aim at where the plane will be, not where it is now. Engineers use a nifty thing called the Kalman Filter in order track objects and predict where they will be. The Kalman Filter isn't just used for missiles, it also plays an integral role in GPS, self driving cars, auto pilot, AI and robotics. What is this Kalman Filter then? It's a recursive way to use Bayes' Theorem. What's Bayes' Theorem? It's a useful tool in probability. What's probability? Read on, you'll find out.

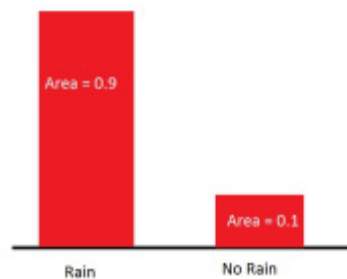
This post therefore describes some basic probability, what Bayes' Theorem is, what the Kalman Filter is and finally how it is used in an Anti Aircraft Missile. Nothing I describe here requires maths knowledge beyond A-level (or even GCSE) as i'm giving a very broad overview of the whole process. But hopefully after you read it you will have learnt something new.

Probability

Here's a box:



It has an area of 1. Imagine the area represents all your certainty about an event. What does that mean? Let's say I'm about to flip a coin. I'm 50% certain it will be heads and 50% certain it will be tails, so I give half the area to heads (as it has half of my certainty) and half the area to tails. If I'm 90% sure it will rain tomorrow, I will give 0.9 of the area to rain happening tomorrow and 0.1 to it not happening



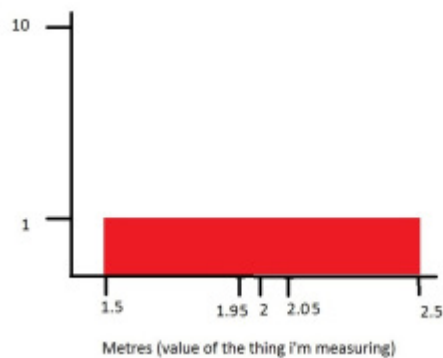
tomorrow, like this:

So I can distribute my certainty, or area in the box, how I want. If there are three outcomes I can distribute my certainty: 0.3, 0.6, 0.1 or 0.9, 0.1, 0 or if you were completely certain that the first outcome would happen 1, 0, 0. The one rule about the box is that the area must always equal 1.

When you are estimating a coin toss or rainfall you are distributing your certainty into *discrete* outcomes, this means you are separating your box into different segments of different sizes. You separate your box into two other discrete boxes, rain/no rain, heads/tails. As well as *discrete* outcomes to events, there are also *continuous* ones. When there is a continuous outcome you don't split your box of certainty into smaller boxes, you mould your box into the shape of your certainty. Sound weird? It's not. The next example will make the difference clear.

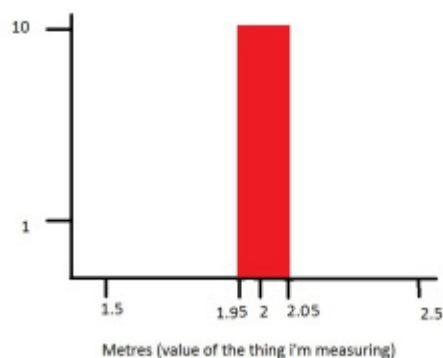
I have two sensors that measure distance, one is a good sensor and one is bad. Instead of giving an answer like '10 metres' it gives the upper and lower limit of what it thinks it is, like '9.90-1.13 metres'. I measure an object that is two metres

away with the bad sensor and it gives me the answer '1.5-2.5 metres'. I can then mould my box of certainty between 1.5 and 2.5 like



this:

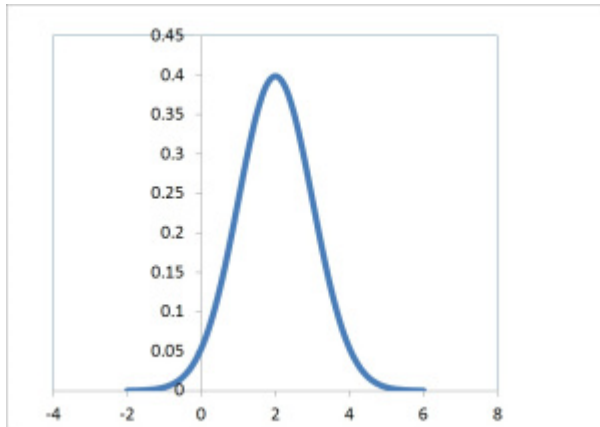
Remember the box area always has to be equal to 1. The width is $2.5 - 1.5 = 1$ so the height must be 1. Now I take a measurement with my good sensor and get the answer '1.95-2.05'. I can now mould my box like



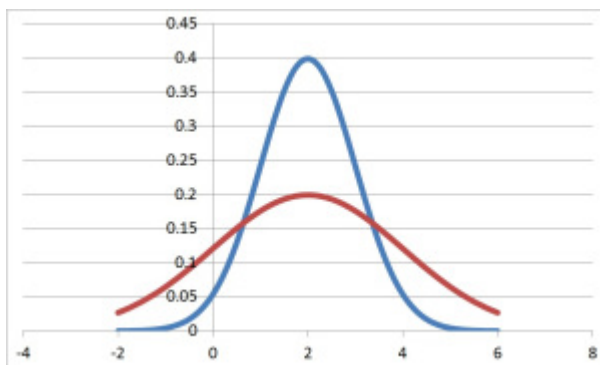
this:

See how the area is still 1? The range we're given with the better sensor is more narrow, so the box must be higher. When the outcomes were discrete, our certainty was measured by what area of the box we gave to certain outcomes. In continuous outcomes (like the graph above) we still use the area of the box to measure certainty, but we now need to measure the area in a range.

Let's say we want to measure how certain each sensor was that the object is between 2-2.05 metres away. On the good sensor the height is 10 and the width is $(2.05 - 2) = 0.05$ metres. So the area of the box is $10 \times 0.05 = 0.5$, which means it is 50% certain the object is within that range. On the bad sensor the height is 1 and the width is still 0.05, so the area of the box is $1 \times 0.05 = 0.05$. It is 5% certain the object is within that range. In the real world a sensor that measures distances doesn't give you a range, but an exact value. However some sensors are better than others, so we still need a way of quantifying how good a sensor is. So what shape accounts for these two facts and still gives us a good way to shape our probability? A bell curve!



You've probably seen this shape before. It occurs everywhere in maths and nature. The posh word for it is a Gaussian but we'll stick with bell curve. What's so useful about the bell curve is that you only need two numbers, it's mean (where the highest point is) and it's variance (how wide it is). So what do variance and mean represent? Well the mean is obvious, it is what the sensor thinks the value is. In the graph above the mean is at 2 (because that's the highest point in the bell curve) so the sensor will tell me that the object is 2 meters away. Variance isn't as obvious. The variance is a measurement of how sure the sensor is about it's reading. On the graph below the blue bell curve has a variance of 1 and the red bell curve has a variance of 4, both have the same mean of 2:



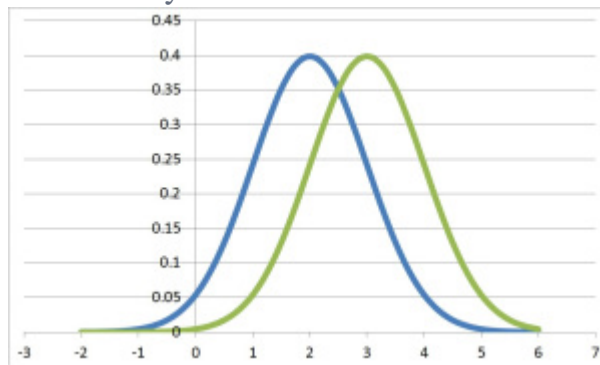
As you can see, if the variance of a bell curve is increased then it gets wider. The reason that the wider curve (the red line) is shorter than the thinner curve is because the area underneath both curves still has to equal 1. So if I make a curve fatter, by increasing the variance, then it has to get shorter in order to keep the same area. It's exactly like the sensors that gave a range, if you have a larger range the rectangle had to be shorter to keep the same area. A tall thin bell curve is more certain about it's value than a short wide bell curve. This is because all of the certainty (which remember is the area of the shape) is concentrated in a smaller range when it is thinner. The variance is a measure of how wide a bell curve is, so let's put this into a rule:

The larger the variance, the worse the measurement

Okay enough about bell curves and certainties. Let's move onto Bayes' Theorem and see why it's useful.

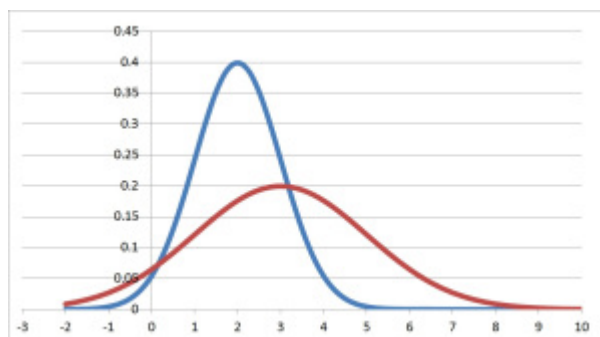
Bayes' Theorem

So we have seen how to represent probability and the most useful way to represent it: with a bell curve. The reason a bell curve is useful because all you need to define it is a mean and a variance. Bayes' Theorem is a way of getting a load of different bell curves and creating one bell curve that sums them all up. What does this mean practically? Well if I have 20 different sensors, of different quality, trying to measure the same thing, Bayes' Theorem says 'i've looked at all the measurements and it's probably this value'. Let's look at a couple of examples. I have two sensors measuring distance again and they are exactly the same quality. One tells me the object is 2 meters away and the other says it is 3 meters. They are the same quality, so this means that they have the same variance. The two sensors'



certainty will look like this:

So if we want to infer what the distance is from these two measurements, it would be 2.5 metres right? We trust each sensor exactly the same, and 2.5 metres is in the middle of the two readings given. How about if we trust one sensor more than the other? Let's say the sensor that gave us a reading of 3 metres has a large variance (i.e. it isn't very good) and the sensor that gave us 2 metres has a small variance (it is good). Our certainties would then look like this:



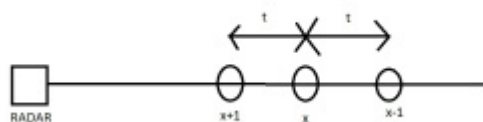
How do we infer the distance now? The blue curve is more certain, so surely our result should be closer to 2 metres than to 3 metres. Bayes' Theorem helps us calculate what the 'actual' distance is from lots of measurements from sensors of different qualities. It does this by giving us a new bell curve. This is really useful

because it's condensing information for us. We are given a load of bell curves and are given one bell curve as a result. Remember a bell curve needs two bits of information to be defined, mean and variance. The mean of the new bell curve moves around based on where the mean of the old bell curves are. Think of it like a vote. Each sensor casts a vote on what it thinks the measurement is, but the better sensors have a more influence in their vote. Bayes' Theorem counts up all the vote and takes the best guess at where the mean for the new bell curve is.

What happens to the variance? Interestingly it ALWAYS decreases, i.e. the new bell curve you get will have a smaller variance than the old bell curves you used. This means that the new bell curve you have will always be more certain of it's value than the previous bell curves. This translates to: The more information you have about something the better. You can never be given more information about something and be less certain about it's value. It seems obvious, but it has interesting corollaries (perhaps for another post).

The Kalman Filter

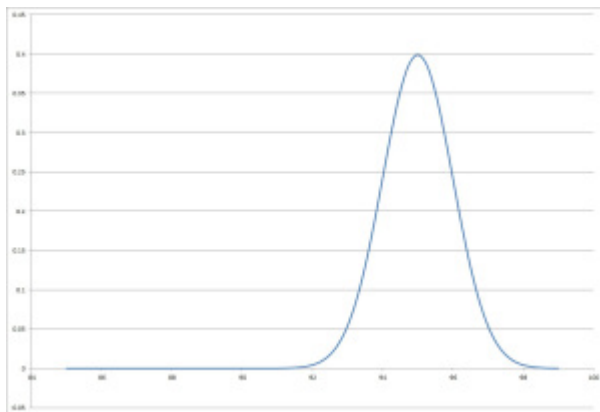
Let's take a step back and see what we know. Sensors give us measurements of objects and also how certain they are of their measurements. This is represented in a bell curve of certainty, the peak of the curve being the value of the measurement. The variance is a measure of how certain the sensor is of its prediction. Bayes' Theorem allows us to get multiple readings from different sensors and produce one bell curve which has a mean that is based on the mean of all the other sensors, and a variance that will always be less (better) than any of the other sensors. What's this Kalman Filter then? Well first of all it's not really a filter. A better name would be a Kalman Tracker, because it's main job is to track moving objects. Let's say we have a ball that is approaching a radar. The radar tells us how far away the ball is in meters and takes a new measurement every t seconds:



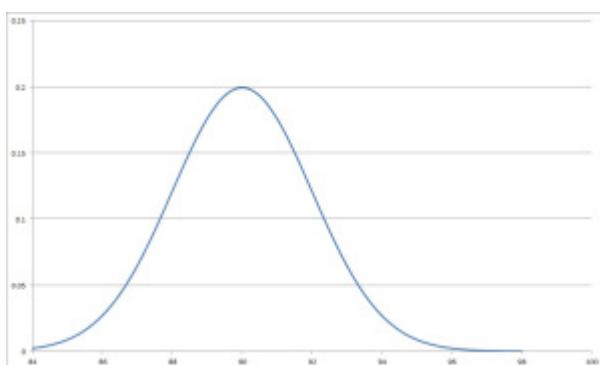
As the ball is moving towards us there will be a new value every time we take a measurement. The measurement at $x-1$ will be further away than the measurement at x because the ball is moving. So this problem is different to the one that Bayes' Theorem solves. Bayes' Theorem lets us take lots of measurements of one thing and guesses the value. So if there were 20 radars that all guessed the distance $x-1$, then we could easily use Bayes to get a good guess on what the actual value of $x-1$ is. This problem is different because we only have 1 thing doing the measuring (the radar), and the thing we're measuring is going to move around all the time.

However Bayes can still be used in a sneaky way here. The trick is that we use the past measurements of the ball to predict the future position of it. So we use the position of the ball at $x-1$ and x in order to predict where it will be at $x+1$. Here's how:

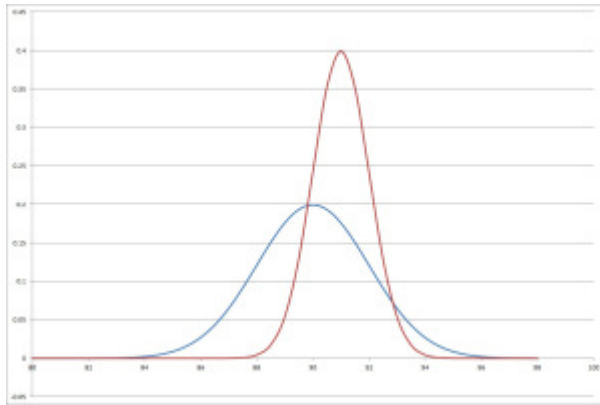
The ball is 100m away at $x-1$. The radar takes a new measurement every 0.5 seconds. When the radar takes another reading (where x is) it is 95m away. Therefore in 0.5 seconds the ball has traveled 5 meters. Speed = distance/time = $5/0.5 = 10\text{m/s}$, so the average speed of the ball is 10m/s. Now assuming that the ball is travelling at a constant speed of 10m/s, we can predict that in the 0.5 seconds it takes for the radar to take its next reading, the ball will have traveled: Distance = time x speed = $0.5 \times 10 = 5$ meters from its previous position, therefore we predict it will be $95 - 5 = 90$ meters away at $x+1$. We can now shift our bell curve in order to predict what the next response will be. When we measured x it was 95m away, so the bell curve will look like this:



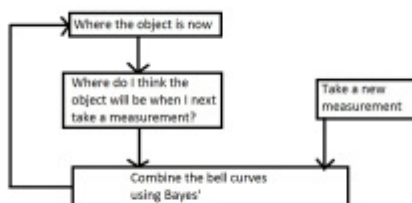
Then we predict it will be at 90m at $x+1$, so the bell curve will look like this:



Incase you can't read the scale the mean has shifted from 95m to 90m. Also notice how the variance has got wider when we shift the bell curve. That's because we are less sure about our prediction of $x+1$ than we are about our measurement of x . This is due to the fact we are assuming that the ball will be travelling at a constant speed, which increases our uncertainty, and hence the variance. So we have used the measurements of $x-1$ and x in order to predict where the ball will be at $x+1$. We then take the radar measurement of $x+1$ and see how similar they are:



The blue bell curve is our prediction of where the ball will be (90m) and the red bell curve is where the new measurement from the radar thinks the ball is (91m). Now it's time to use Bayes' Theorem. We combine our prediction and the measurement using Bayes in order to produce one bell curve that is our best guess of where the ball will be at $x+1$. We then use the measurement at $x+1$ and x in order to predict where the ball will be at $x+2$, make a prediction, take a measurement, and combine the two to get the best guess of $x+2$. The process is recursive, looking like this:



This is better than only using the measurement from the radar. Using information about past measurements lets us take an extra guess at where the ball is. Remember if you have more information about something the better. The Kalman filter is a really effective way of using the past to inform the present and predict the future.

Anti Aircraft Missiles

Lets wrap this up by giving an overview of how an Anti Aircraft Missile works. It's almost the same as the example of the ball moving towards the radar, but more tailored towards missiles. Your missile can shoot vertically up into the air. If a plane passes over your missile station, you want to fire the missile so that it will come into contact with the plane. You have a few radars that are tracking the plane and they take a new measurement every 0.1 seconds. You also know how long it will take for your missile to reach a certain altitude.

Say the plane approaching your station is flying at 30,000ft and you know that it will take your missile 10 seconds to reach that altitude. You use the Kalman Filter

on the radars in order to track where the plane is and it's speed. But the Kalman filter can also make predictions about the future position of the plane. So you not only plan 0.1 seconds in advance (the amount of time it takes for the radar to take a new measurement), but also 10 seconds in advance (the amount of time it takes for the missile to reach the appropriate altitude). When the prediction 10 seconds in advance says that the plane will be vertically above the missile, you shoot.

As the plane is already being tracked, figuring out the altitude of the plane is as simple as looking at where the Kalman Filter thinks the plane is. So if the plane is at 20,000ft and the missile takes 7 seconds to reach that altitude, the Kalman Filter will predict 7 seconds in advance. It's easy to write a program that will do this automatically without any human input. What if the plane is changing altitude? The problem becomes slightly harder, but not too hard. It's just a matter of figuring out how far to plan in advance.

Conclusion

In this post i've described what a Kalman Filter is and hopefully you've understood the potential of it. To get there i've gone through how to represent probability and Bayes' Theroem because I believe you get a deeper understanding of the whole process if you have a grasp of these.