

How to take the guess work out of expert judgment

Judgment is the freest form of information there is ...



WORKBOOK

Presented by

Dr Chris Jackson

Reliability engineer



Copyright © 2022 Christopher Jackson

All rights reserved. No part of this publication may be reproduced, distributed or transmitted in any form or by any means, including photocopying or other electronic or mechanical methods without the prior written permission of the author (Christopher Jackson), except as permitted under Section 107 or 108 of the 1976 United States Copyright Act and certain other non-commercial uses permitted by copyright law.

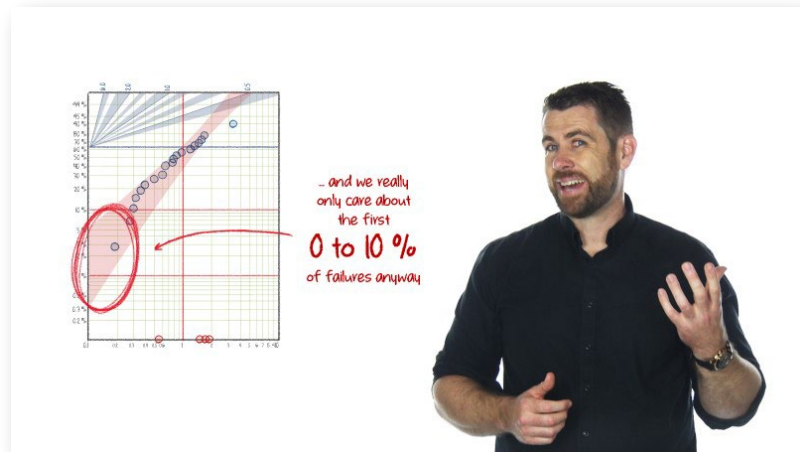
For permission requests, write to Christopher Jackson using the details on the following pages.

Limit of Liability/Disclaimer of Warranty: While the author has used his best efforts in preparing this document, he makes no representations or warranties with respect to the accuracy or completeness of the contents of this document and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. The author shall not be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Who is your teacher? ... Dr Chris Jackson

Dr Jackson holds a Ph.D. and MSc in Reliability Engineering from the University of Maryland's Center of Risk and Reliability. He also holds a BE in Mechanical Engineering, a Masters of Military Science from the Australian National University (ANU) and has substantial experience in the field of leadership, management, risk, reliability and maintainability. He is a Certified Reliability Engineer (CRE) through the American Society of Quality (ASQ) and a Chartered Professional Engineer (CPEng) through Engineers Australia.

Dr Jackson is the cofounder of IS4 online learning, was the inaugural director of the University of California, Los Angeles (UCLA's) Center of Reliability and Resilience Engineering and the founder of its Center for the Safety and Reliability of Autonomous Systems (SARAS). He has had an extensive career as a Senior Reliability Engineer in the Australian Defence Force, and is the Director of Acuitas Reliability Pty Ltd.



He has supported many complex and material systems develop their reliability performance and assisted in providing reliability management frameworks that support business outcomes (that is, make more money). He has been a reliability management consultant to many companies across industries ranging from medical devices to small satellites. He has also led initiatives in complementary fields such as systems engineering and performance-based management frameworks (PBMFs). He is the author of two reliability engineering books, co-author of another, and has authored several journal articles and conference papers.

What is INFORMATION?

There are lots of definitions of information. Information is sometimes seen as ...

... processed, organized and structured data.

This is not a good definition. Why? Because information is not data. Data can be (for example) a spreadsheet or table full of numbers (like times to failure, oxide layer thickness or anything else that we have spent a lot of time measuring and gathering). If we to organize this data (like listing the numbers from 'smallest' to 'largest') this does not simply become information.

A better view of information is based on how you need to use it. Information needs to be seen as ...

... knowledge obtained from investigation or learning.

It is very important that you don't have to get or do anything from data to gain information. You can read a book and get information. You can be taught something ... which involves imparting information.

This is very important. Gathering data can be time consuming and expensive. But reading a book or speaking to an expert is typically free (or at least, very inexpensive). So why don't we use our existing knowledge more often?

In practice ... we actually do. When we design an amazing new device, product, machine, plant or manufacturing line, we use our judgment and skill to make decisions about what our 'item' will look like. Will we use a pneumatic or hydraulic pump? Electric motors or internal combustion engines? Streamlined looks or chunk designs that look robust? Each of these decisions is made 'every minute' by young designers, engineers and technicians.

But then there are these other scenarios where some of us seem to crave information that is derived from data and nothing else. This is particularly true when it comes to reliability and understanding how long our thing is going to last. And that sometimes comes down to confidence.

Confidence is a measure of you. Not a product or item.

Decision-makers who don't have a background in how to design 'reliable' things often crave confidence generated by data and statistics. They have no idea how to assess the methods used to bake reliability into a product or device – and so the only thing they have left is numbers.

And when it comes to reliability, it can be very long and expensive to generate these numbers during test. So in many cases, it is much better to understand what reliability looks like through understanding how it happens. This can be based on things like expert judgment and knowledge. But how do we 'get' this knowledge?

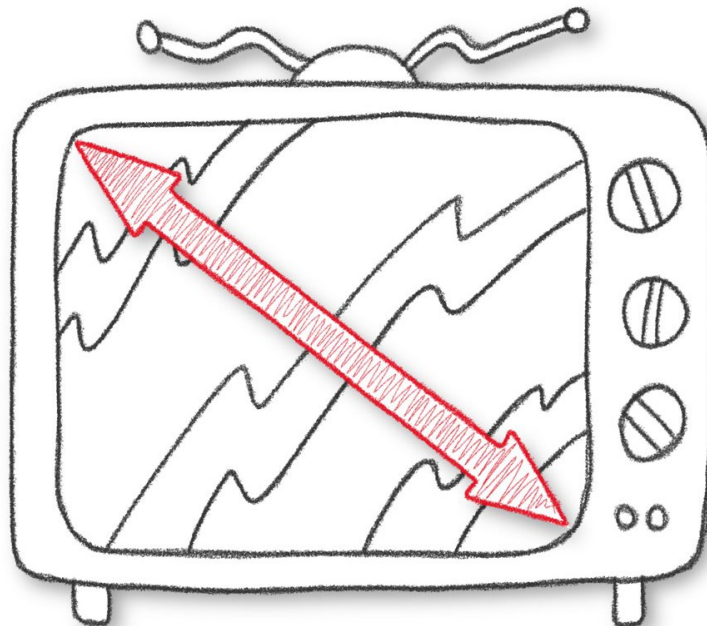
So how do we START?

Let's say we are interested in understanding the quantity of something that we either have not measured or find difficult to measure. It could be the reliability of a product at two years. It could be launch reliability of spacecraft.

When NASA investigated the 1986 Challenger Space Shuttle Disaster where the external tank exploded 79 seconds after takeoff, it found several troubling cultural issues across the agency. One interesting observation they made was that the engineers and technicians estimated that inherent space shuttle reliability would result in one catastrophic failure in every 50 to 100 launches. NASA's management team (non of whom were engineers or involved in the space shuttle design or operation) routinely estimated this frequency to be one catastrophic failure over thousands of launches. There were a total of 135 space shuttle launches, resulting in two catastrophic failures (at an observed frequency of one catastrophic failure every 67.5 launches).

So what good is it to have a safety program, modelling and analysis which convinces decision-makers that the true reliability is orders of magnitude higher than it actually is? ... especially if the engineers and technicians KNOW what it really is the whole time?

So let's look a more tangible metric. Such as the diagonal length of a large computer monitor or TV.

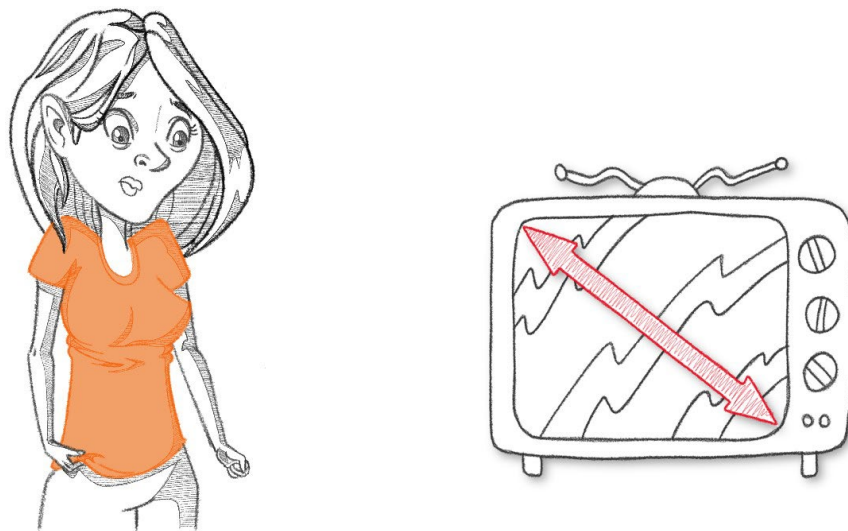


This was actually the basis of many experiments that were run in courses where the instructor would ask students what their 'knowledge' of the length of this diagonal consisted of. The classroom would have a suitably large computer monitor at the front that every student could observe.

The instructor set a challenge to the students. The students were not allowed to get out of their chairs, but they were going to work together to help estimate the length of this diagonal.

To do this, we need to characterize information stored in the minds of each student. None of the students are professional 'distance estimators.' But they have each had to estimate distances at various stages of their lives. This experience could have been accumulated while playing sports where it is important to estimate how far away goals, teammates and holes are. This experience could have been accumulated while undertaking some hobby carpentry on weekends. This experience could have been accumulated from any activity where students had practices trying to turn images their eyes detect into quantities inside their brains to base decisions on.

And this experience helped students to estimate how long the diagonal of the computer monitor was.



Quantifying HUMAN KNOWLEDGE is hard.

The simplest way to extract information from the human brain is to ask someone to simply provide their point estimate or 'best guess' of something. But there is a problem with this. This doesn't communicate all the information inside that human's brain ... and it can influence the information in another person's brain.

For example, let's say that all students were asked to come up with their 'best guess' of the computer monitor diagonal. Let's say that Clara is one of those students, and quietly guessed that the diagonal length was 57 inches or 145 cm. She sits at the back of the classroom – meaning that she will be the last person to shout out her best guess.

So the instructor asks each student in turn. The first student shouts out '34 inches or 86 cm.' The second student shouts out '37 inches or 94 cm.' The third student shouts out '33 inches or 84 cm.' The rest of the students shout out their best guesses, with the largest best guess being 40 inches or 102 cm.

Clara gets to hear all these best guesses. She realizes that her best guess is considerably larger than any of the other best guesses.

So what does she do?

When it is her turn, she shouts out '42 inches or 107 cm.' This is actually lower than her original best guess, and still the largest best guess out there. She might have felt embarrassed or not confidence about her original best guess because of what everyone else shouted out. But the information she passed on was actually a combination of *her* best guess and *all the other best guesses*.

We don't want this. This is a very human tendency to not appear 'stupid.' And the best way to not appear 'stupid' is to not appear 'different.' So there is a tendency to conform with other opinions out there. So in fact, the most influential best guess was the first best guess, as this had the ability to influence everyone else's 'shouted' best guess.

So how do we get around this? ANONIMITY. We make sure that these best guesses are written on pieces of paper and then collated by the instructor so there is no way we can attribute any best guess with any student. This not only ensures one best guess doesn't influence another ... but it also ensures no one has to feel 'stupid.'

But we haven't solved all the problems with quantifying human knowledge.

Let's say that one of the students (David) has spent a lot of time working in an electronics store. One that sells computer monitors like the one in the classroom. He is very familiar with monitors of all different sizes. He even thinks he knows its model number! His best guess for the diagonal is 38 inches.

But Clara does not have the same experience that David has. She is also slightly short-sighted.

So what does this mean? It means that David is a lot more confidence in his best guess than Clara is. And this is entirely rational given his experience and knowledge base.

So we should place a lot more 'importance' on David's best guess than Clara's.

But how? One approach is to try and understand David's and Clara's CONFIDENCE. If we can somehow measure how much more confidence David is for his best guess when compared to Clara, then we can place that much more importance on David's best guess.

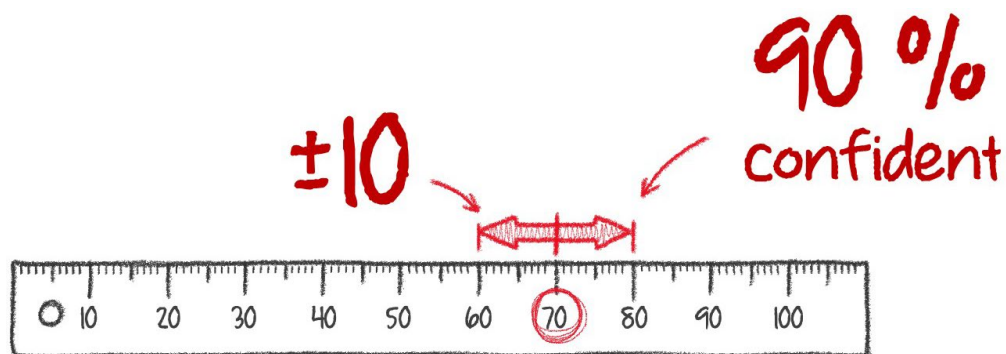
But how?

How do we QUANTIFY human knowledge?

One approach ... is to ask people! For example, let's look at another student – Sophie. Her best is 27.5 inches or 70 cm. But the instructor also asks Sophie about how accurate she thinks her best guess is. The instructor asks Sophie the following questions:

1. What is your *BEST GUESS*? ... *27.5 inches or 70 cm*
2. How *ACCURATE* do you think your best guess is in terms of deviation from your best guess ... *within 4 inches or 10 cm above/below the best guess*
3. How *CONFIDENCE* do you think that the true value is within your accuracy range? ... *90 %*

This means that Sophie is 90 % confidence that the computer monitor's diagonal is within the interval 23.5 to 31.5 inches, or 60 to 80 cm.



Now we are much closer to being able to quantify (in this case Sophie's) knowledge.

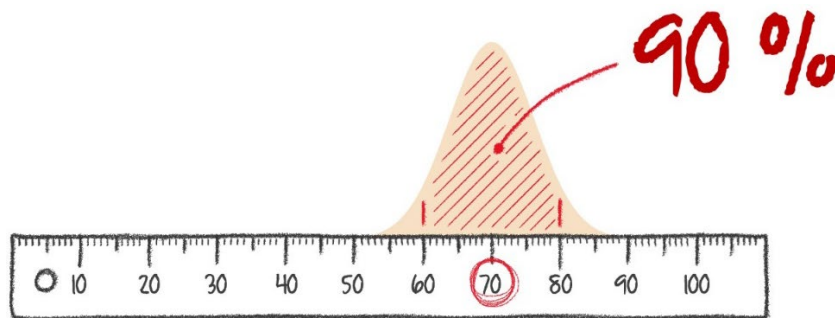
BUT THIS IS OFTEN EASIER THAN IT SOUNDS! The instructor first needed to teach his students how to quantify their knowledge like this. So he starts by asking everyone to write down their (1) best guess, (2) accuracy range and (3) confidence for the instructor's height. The reason he does this is to make sure everyone knows how to do this. There are of course many errors. Some students just write a best guess and confidence. Other use a percentage for the accuracy range. So the instructor corrects these attempts, and then focuses on the height of a door. He asks them to quantify their knowledge again. He keeps doing this until everyone knows how to quantify their understanding of distance.

Then he asks them to do the same for the diagonal of the computer monitor. He then gathers their pieces of paper up ... and is then ready to turn these numbers into information.

Here comes the BELL CURVE.

The 'bell curve' is also known as the normal (probability) distribution. It is used to describe how probability is distributed to lots of potential random variable values. The normal distribution is observed in the world around us A LOT. It just so happens that when lots of random processes are added up, the resultant number is described by the normal distribution.

So what does this mean? Well ... it means that we can use a normal distribution to characterize every student's knowledge. For Sophie (in terms of cm) ...

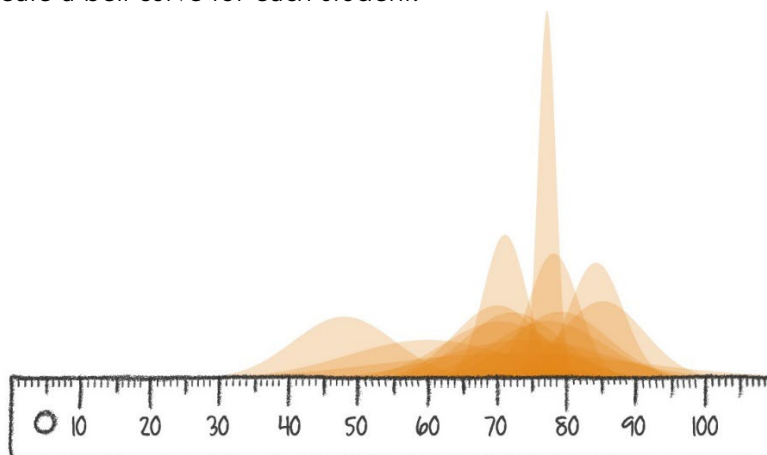


You can clearly see the 'bell' shape. Assuming a bell curve to describe human knowledge is usually a 'good assumption.' It has been shown to mimic how we characterize inaccuracies in measurements and estimates. In fact, the bell curve was first used for this purpose (albeit in astronomy).

So how do we get the bell curve we see above? Well, we need computers to help. For example, we used the following function in Microsoft Excel to get the bell curve above that characterizes Sophie's knowledge or understanding of the diagonal of the computer monitor.

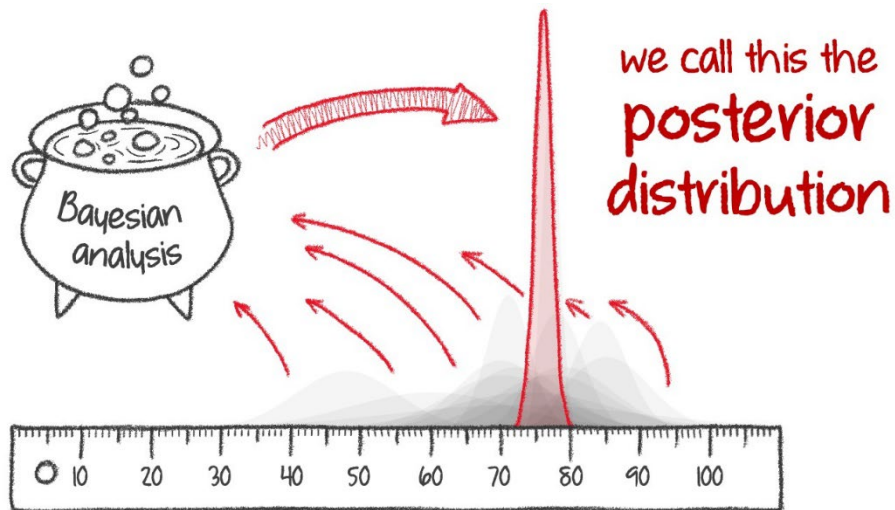
 `=NORM.DIST(x,<best guess>,
<range>/NORM.INV(1/2+(<confidence>/2),0,1),FALSE)`

This allows us to create a bell curve for each student.

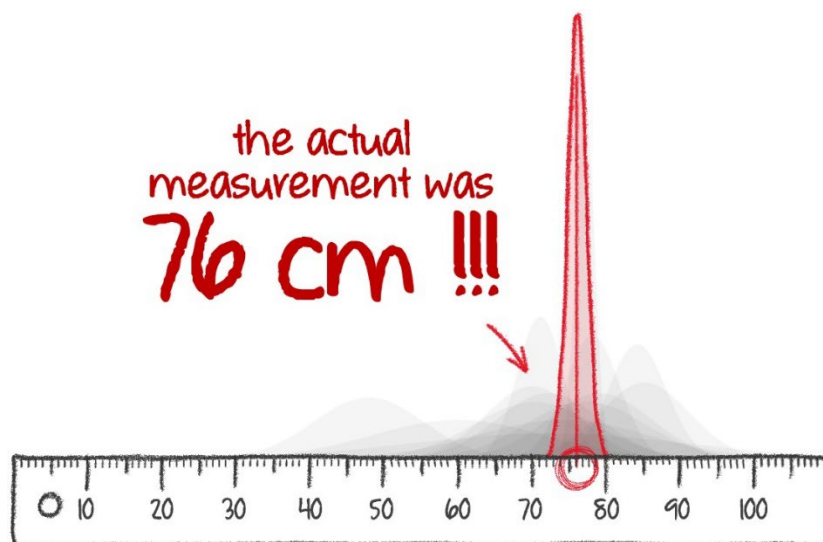


So now what can we do?

We can use this approach called Bayesian analysis. Bayesian analysis is fantastic at taking a lot of different information that we quantify with things like bell curves and turn them into what we call a posterior distribution. For our students guessing the diagonal of our computer monitor, the posterior distribution looks like this.



So why is this important? Well, the posterior distribution summarizes the combined knowledge and information across the brains of all our students. And when we combine information we get something really useful. Especially in this case, because the actual diagonal length of the computer monitor was 30 inches or 76 cm. Look at how well that compares to the posterior distribution.



So how do we get this posterior distribution?

What is BAYESIAN ANALYSIS?

Bayesian analysis is based on Bayes theorem, which is of the following form:

$$f(x|E) = \frac{L(E|x)f(x)}{\int_{\forall x} L(E|x')f(x')dx}$$

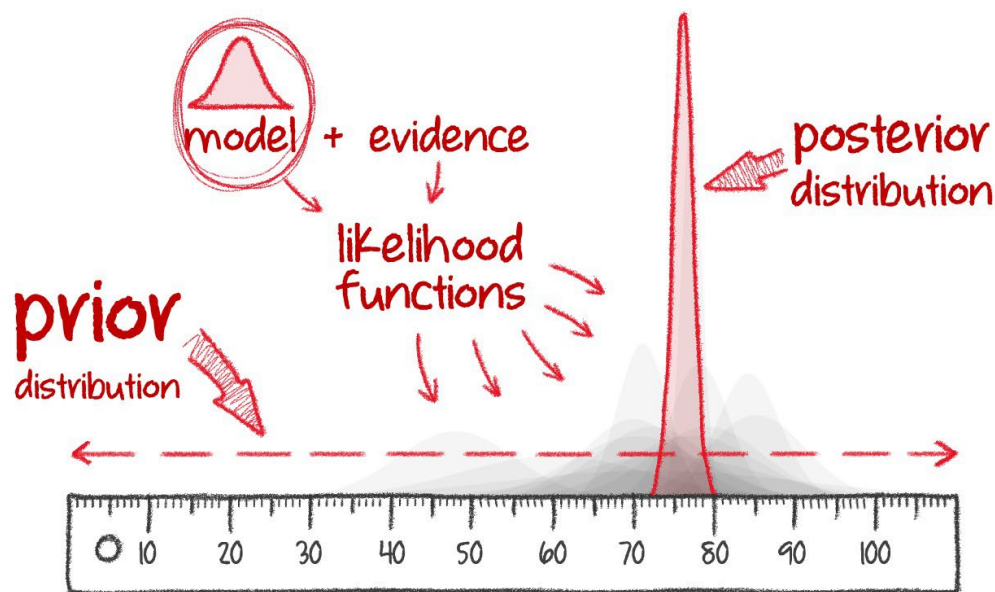
... where $L(E|x)$ is what we call the likelihood of observing evidence set E given a particular value of something we are interested in investigating, x , $f(x)$ is the prior distribution of the thing we are investigating, and $f(x|E)$ is the posterior distribution of the thing we are investigating.

For our computer monitor, x was the diagonal that we were interested in understanding better, $f(x|E)$ was the red curve on the previous page that summarized all the information from our students, and $L(E|x)$ is one of those bell curves that summarized one of our student's knowledge.

But of course, there was more than one student. So Bayes theorem looks like this ...

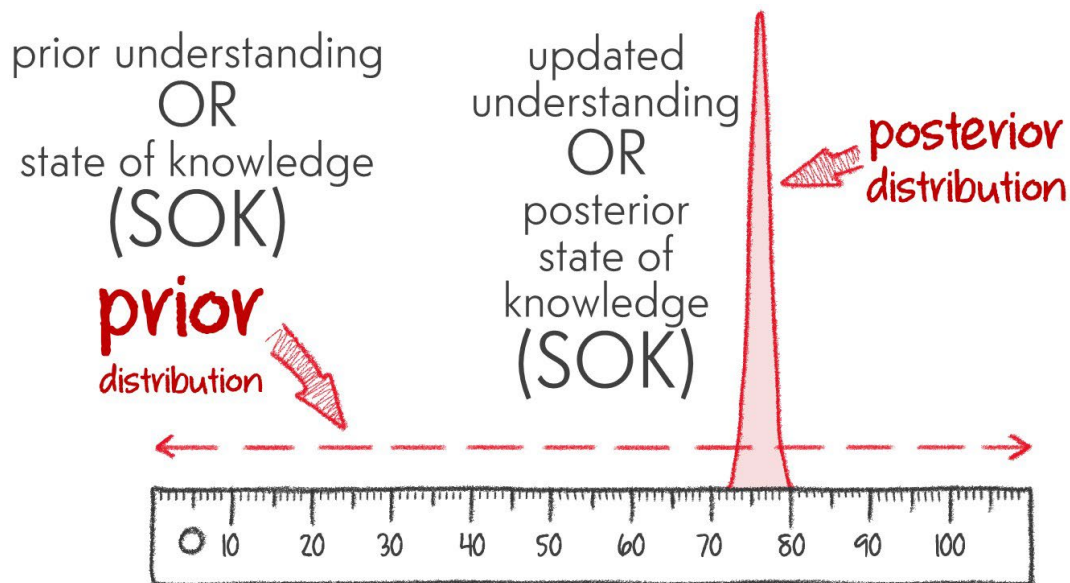
$$f(x|E_1, \dots) = \frac{L(E_1|x)L(E_2|x) \dots L(E_n|x)f(x)}{\int_{\forall x} L(E_1|x')L(E_2|x') \dots L(E_n|x')f(x')dx}$$

This of course looks really complicated. It is simpler than it looks. For our students estimate the diagonal of the computer monitor ...



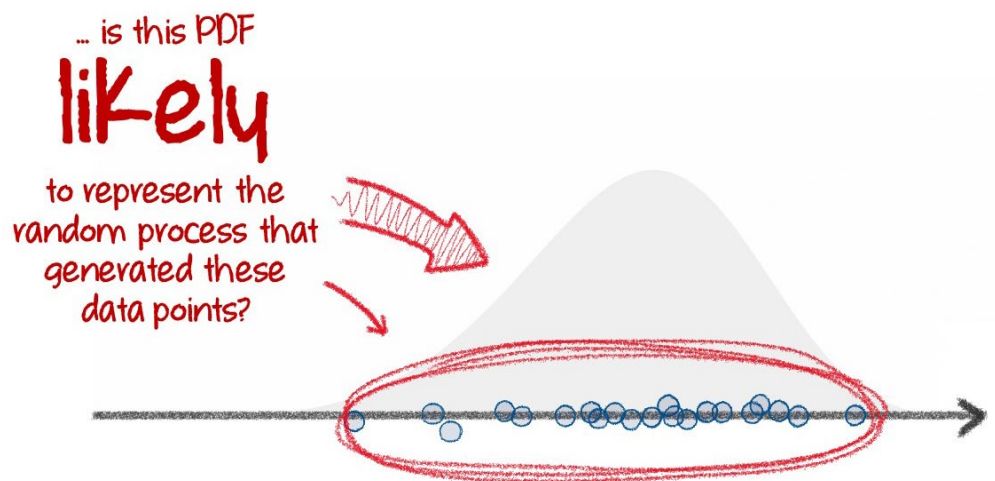
In this case, the prior distribution is 'flat' or equal to one. What is the prior distribution anyway? It is supposed to summarize our knowledge before we get additional information from evidence, textbooks, students or anything else. Many people assume a 'flat' prior distribution to try and replicate a 'lack of knowledge.' This is not perfect, but works well if you have lots of evidence.

From another perspective ...



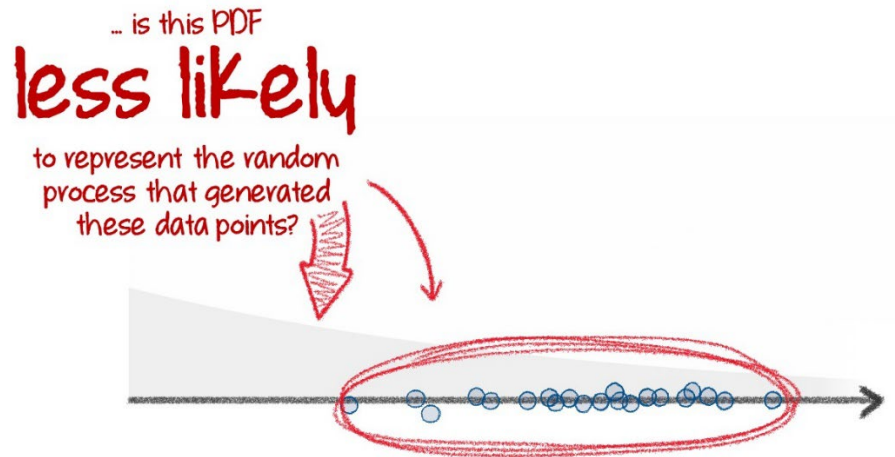
And it all starts with the LIKELIHOOD

The likelihood is used across a lot of probability and statistics. For example, let's say we run an experiment where we gather 20 data points. We know want to see if something that looks like a bell curve is a good fit.

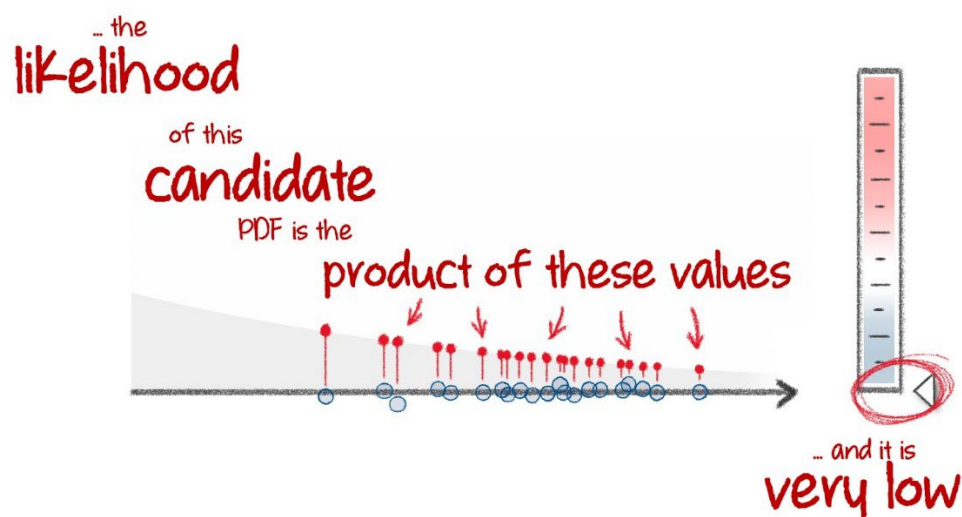
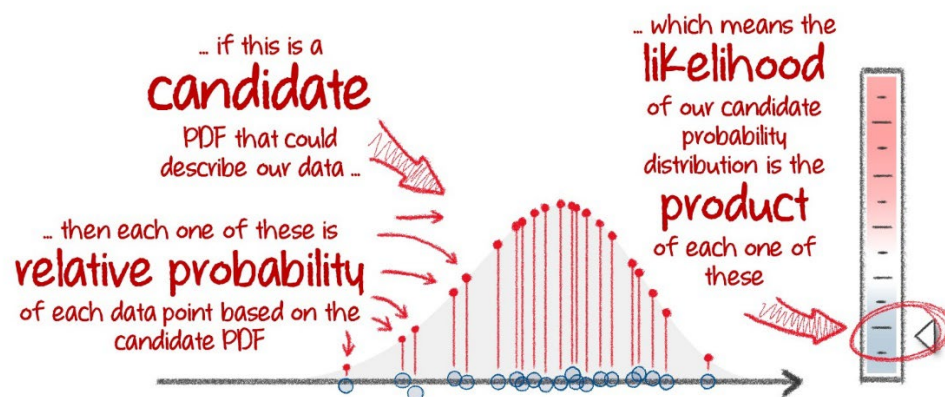


The function $f(x)$ or the probability density function (PDF) gives the relative probability that a continuous random variable is equal to a specific value. You might look at the grey PDF above and conclude that it seems to describe the relative frequency or density of the data points. So you might say that this PDF curve is 'likely'

At least compared to other PDF curves ... like the one below.

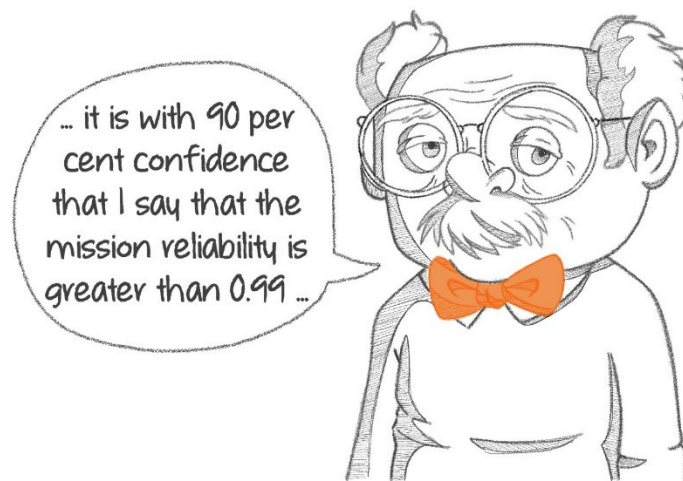


But how do we quantify how much more likely one PDF is compare to another? Well, we simply multiply all the different PDF curve heights for each data point.



But not everyone likes BAYESIAN ANALYSIS

The opposite approach to statistical analysis is what we call the 'frequentist' or 'classical' approach. A full explanation is outside the scope of this document, but this is what it means when it comes to trying to understand what 'information' is. For example, let's say we have a classical statistician (below) who says the following ...

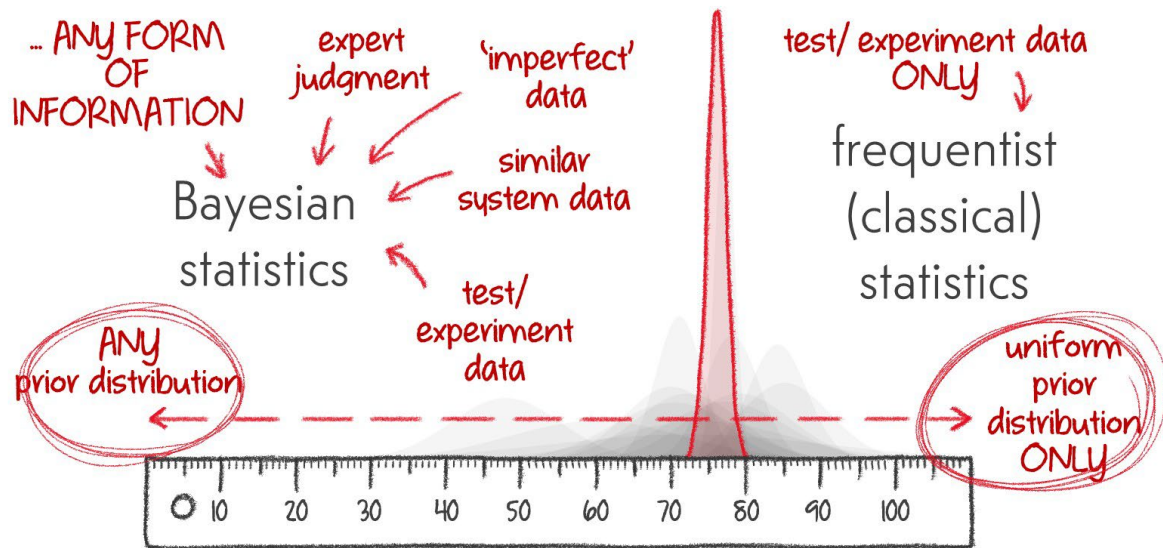


Now you might think that this means that there is a 90 % probability that the mission reliability of something is greater than 0.99 or 99 %. But it doesn't.



This might seem like semantics. But it is not. Most statistical approaches are 'classical' or 'frequentist.' But we actually use things like confidence bounds in a Bayesian way. We aren't interested in how often our statistician is right or wrong. We are interested in whether the true value of our underlying metric is within a confidence interval (or not). We can use this probability. Not how often our statistician is right or wrong.

And there is another problem with 'classical' or 'frequentist' statistics.

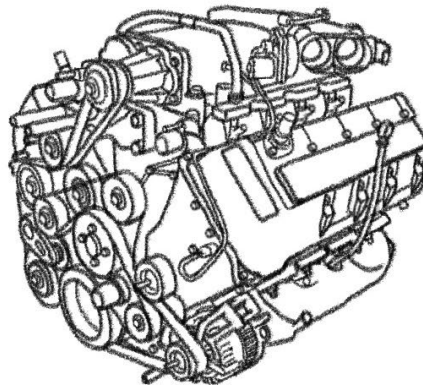


So BAYESIAN ANALYSIS can help you solve lots more problems

Look at the following scenario ...

... one test showed
1 out of 20 products
had a defect

5 %

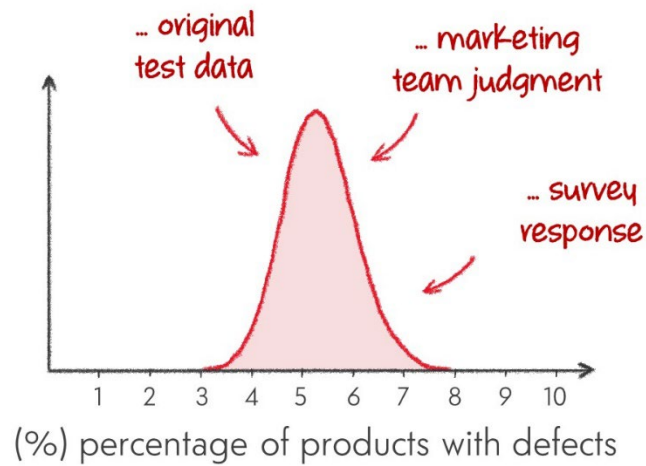


... we sent out a
quick, one question
electronic survey to
1000 customers

... we got 48
negative
responses!

So what does this information mean? Well, we know that not everyone will respond to a survey. We also know that some people are more likely to respond than others. Let's say we speak to our marketing team who believe that if a customer has a negative experience, they are 60 % likely to write a negative response. They also believe that if a customer has a positive experience, they are 1 % likely to write a positive response.

This means we can use Bayesian analysis to come up with the following posterior distribution of the percentage of products with defects ...



This is really helpful.

Classical statisticians complain that there is never enough data. Bayesian statisticians complain we never use the data we have well enough.

Which one are you?

