Topic 1: Introduction to Probability and Statistics

1.1 Introduction to Probability and Statistics

So why should you care about probability and statistics? Basically, it's because this is a very powerful tool for dealing with uncertainty.

First of all, navigation and search engine that I showed you, those are very advanced problems, as is the life insurance market.

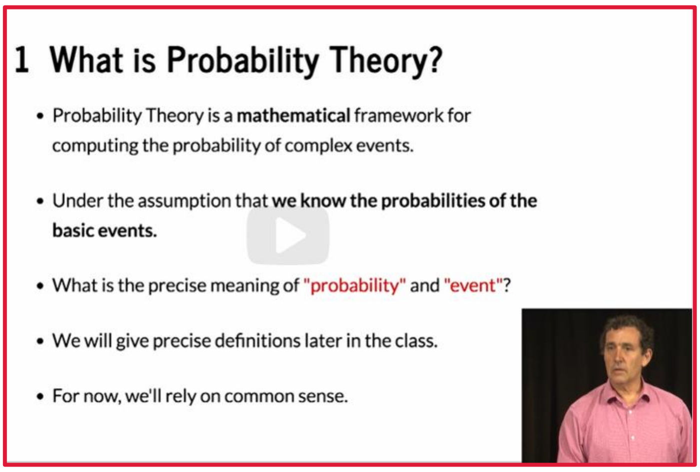
So to summarize, what we have all around us when we are doing anything in the world, is uncertainty. And probability and statistics provide a rational way to deal with uncertainty.

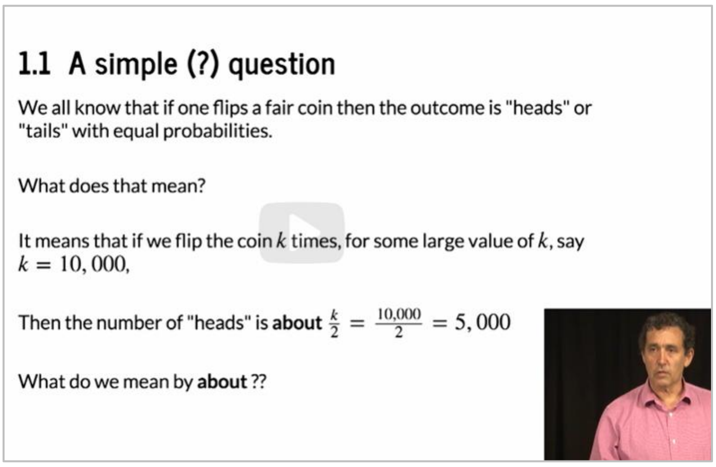
Probability and statistics provide mathematical tools for estimating the likelihood of random events.

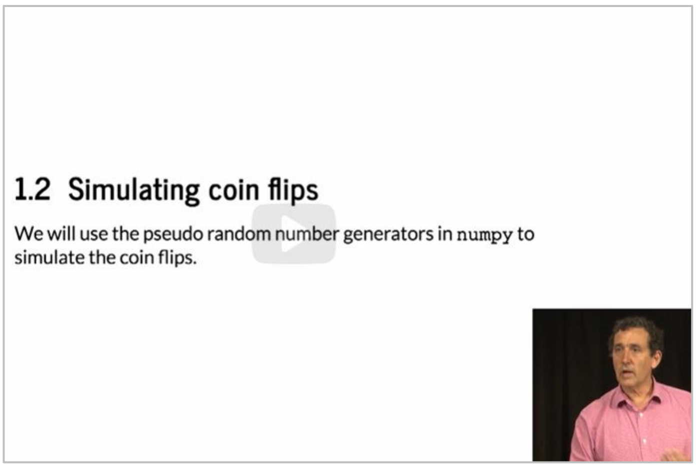
What is probability and statistics useful for?
- To quantify uncertainty
- To make predictions about future events

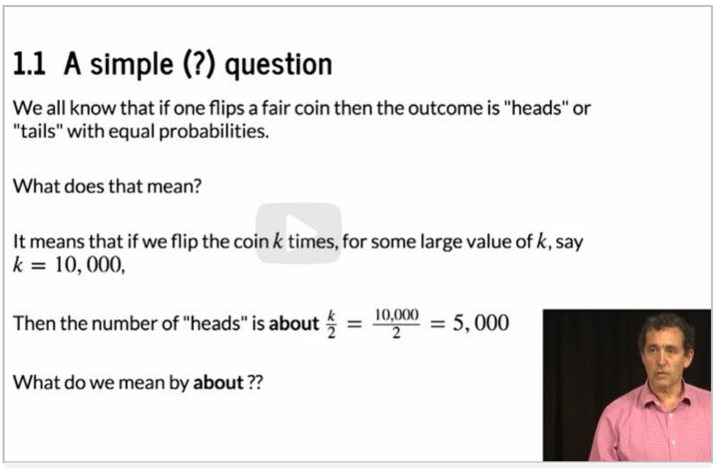
1.2 What is Probability Theory

There are two main subjects that we will study. One is probability and the other is statistics.

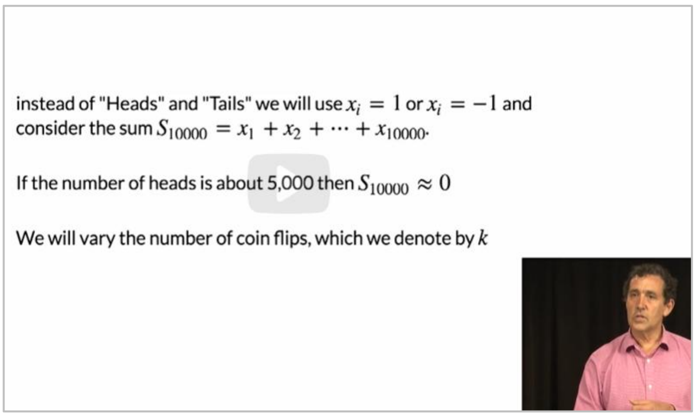


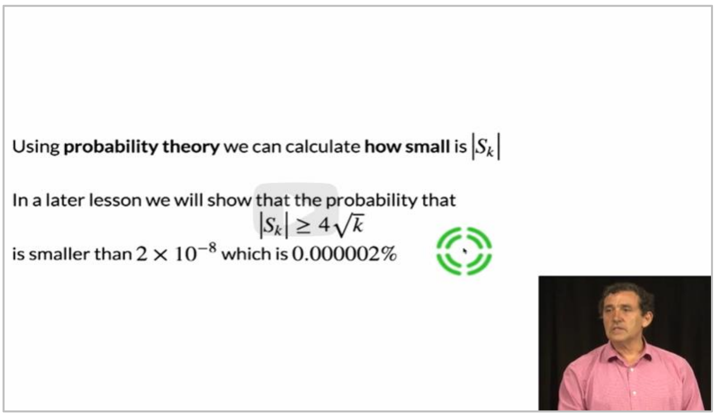




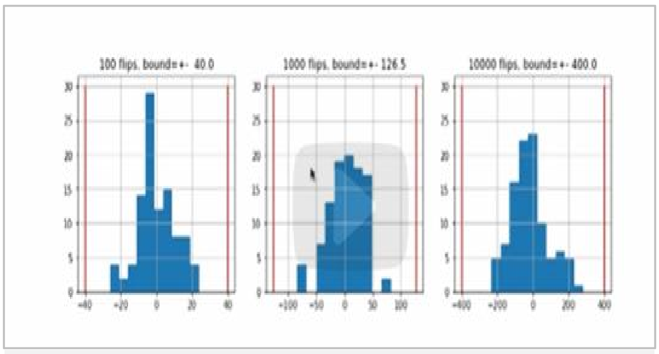


what do we mean by about, how can we express this notion of about in a better way because we might be actually interested in knowing how far from that we are.

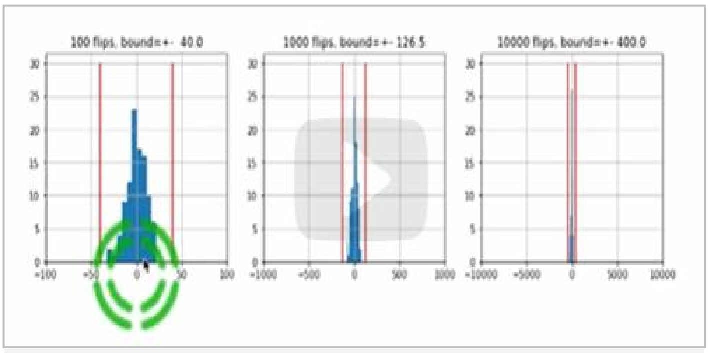




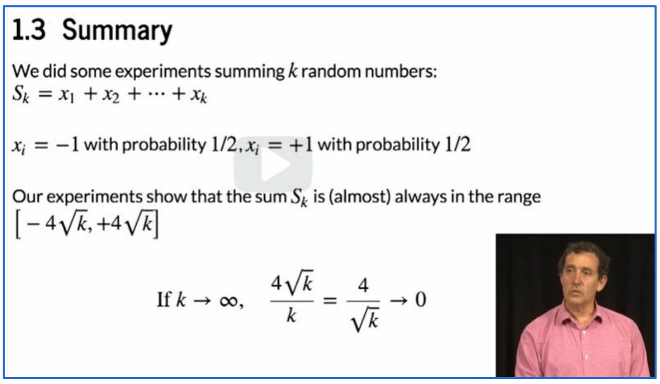
总结:
- Probability Theory is a **mathematical** framework for computing the probability of complex events.
- Under the assumption that we know **the probabilities of the basic events**.

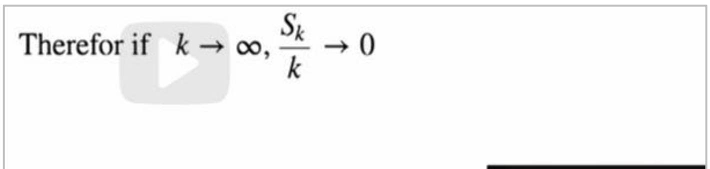


Okay, so here is our simulation. What we see is, here, we have one hundred coin flips.Here, one thousand coin flips, and here ten thousand coin flips,
and the red line mark what probability theory says is the boundary in which it is very, very likely that the total number of coin flips resides (存在，在于，定居).



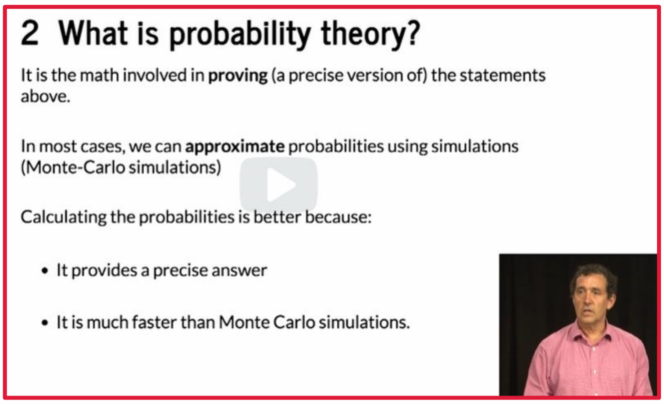
If we scale, if we plot the full scale of these coin flips, what we see is the following.
We see something like this, so when we plot the whole scale from minus one hundred to one hundred, for hundred coins, and from minus ten thousand to ten thousand for ten thousand coins,
then we see that the distribution becomes more and more concentrated around zero, relative to this scale.





And so what we can say is that Sk, relative to k,

so the ratio of the number, of the difference between heads and tails, divided by k, that goes to zero.
And that's basically what we mean by the probability is being half and half.



In most cases, we can approximate the output, these probabilities, using simulations. These are called Monte-Carlo simulations,
and that's essentially what we did in this little experiment that we did.

So why isn't that enough?
Because, first of all, calculating the probability gives you a precise answer, and doing Monte-Carlo simulations just gives you an approximation and you need to run the experiment longer and longer to get more and more accurate answers.
And the second is that it is much faster than Monte-Carlo simulations, essentially for the same reasons.

POLL

If we flip a coin 100,000 times, and denote $S_n = x_1 + x_2 + \ldots + x_n$ and $x_i = -1$ = tails, $x_i = 1$ = heads. What is the smallest value we can expect this sum to not exceed?

RESULTS

400	30%
1265	53%
4000	14%
6580	3%

Submit

Results gathered from 1185 respondents.

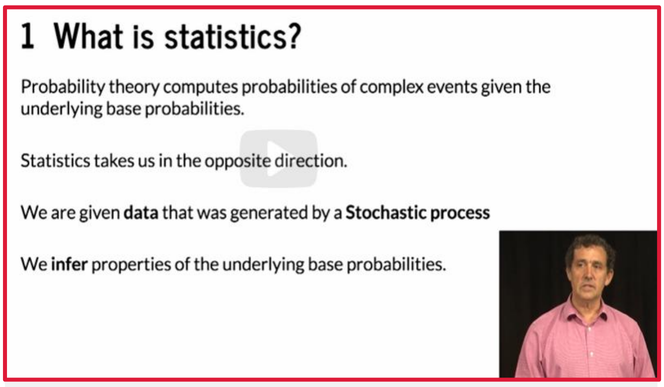
总结:

① Probability theory is the math involved in **proving** (a precise version of) the statements above.

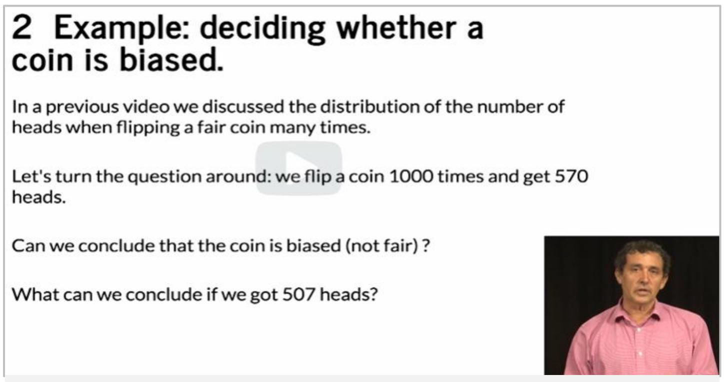
② In most cases, we can **approximate** probabilities using simulations (Monte-Carlo simulations). Calculating the probabilities is better because:

- It provides a precise answer.
- It is much faster than Monte Carlo simulations.

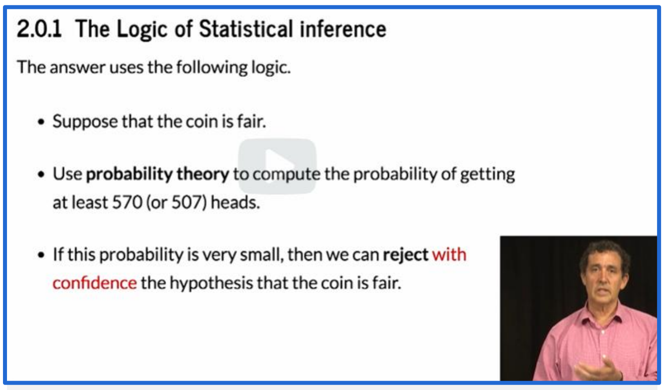
1.3 What is Statistics



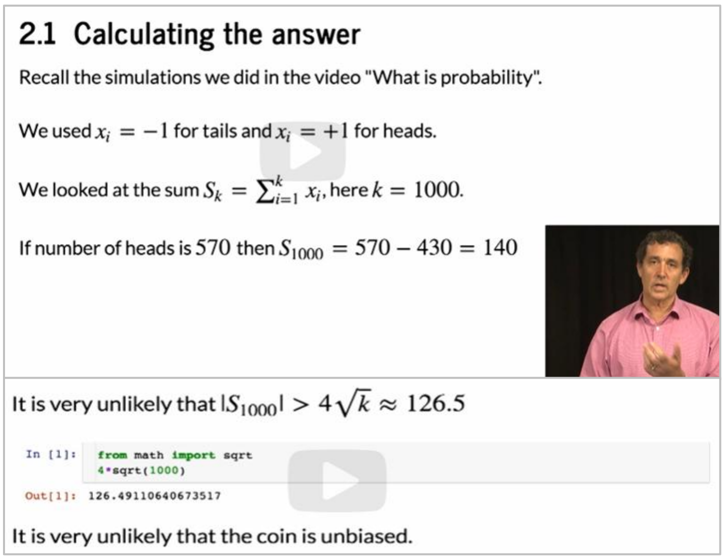
We're given data that was generated by some Stochastic process, or some random process, and from that data we infer properties of this Stochastic process.

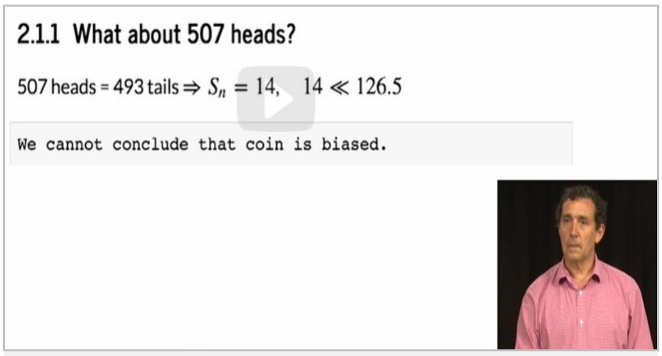


Let's go back to our coin. And, we believe that it is an unbiased coin.
So it gives us exactly half heads and half tails. But how can we be sure?
So, previous time we saw that, we talked about the distribution, and now we wanna turn the question around.
If we flip the coin 1000 times, and we get 570 heads, then can we conclude that the coin is biased? So, the coin is not a fair coin?
What about, can we conclude that the coin is biased? Not a fair coin?



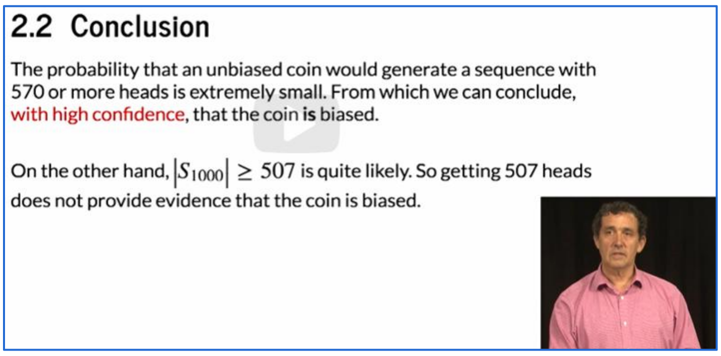
And then we can calculate what is the probability that the coin will give us 570 coins.
If this probability is extremely small, then we can reject with confidence the hypothesis that the coin is fair. We can say it is very unlikely that a fair coin would generate this sequence, and therefore it is not a fair coin.





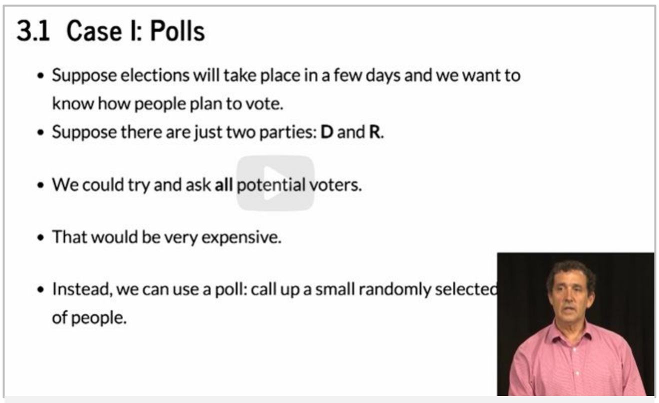
What about 507 heads? Well, if you have 507 heads, 493 tails, then Sn is going to be 14, and 14 is much, much smaller than 126.5.

So at least according to the rule we just had, we cannot say that the probability is very small, and as we'll see, the probability is actually quite reasonable, that it's quite large, that a fair coin would generate 507 heads.
So, we cannot conclude that the coin is biased. It might still be biased, but we might have to flip the coin many, many, many more times, before we can deduce that.

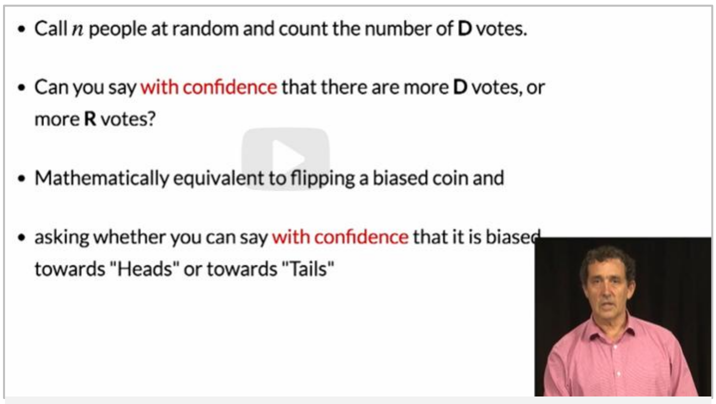


So statistics, unlike probability which is a part of math, statistics is really about problems in the real world.
And so you might ask, "Why should I care about the coin being biased? That's not a problem I kind of face many times." And that's a very valid critique.
And we will now give some examples for real problems where the statistics is very closely related to whether or not a coin is biased.

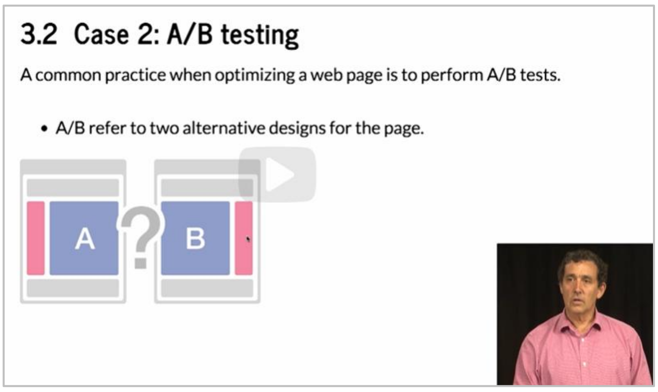
总结:
Probability theory computes probabilities of complex events given the underlying base probabilities.
Statistics takes us in the opposite direction.
- We are given **data** that was generated by a **Stochastic process**.
- We **infer** properties of the underlying base probabilities.



We call a small randomly selected set of people, ask them of their opinions, and then we extrapolate from that what do people think in general.



Mathematically, this is exactly equivalent to flipping a biased coin, and asking whether heads is more likely than tails. Or tails more likely than heads. It's the exact same question, and the same math holds for it.

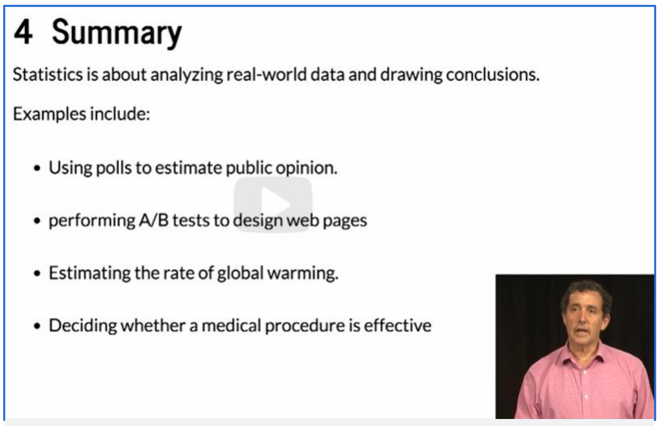


This is called A/B testing, which is a very common practice on developing web interfaces. You basically think about two alternative designs for your web page, one is A, one is B.

To see which design users prefer, we randomly present the design A or design B, when people visit our website.

We measure how long the user stays on the page, or whether the user clicked on an advertisement, or any other indication that the user likes one of the designs more than the other.

We want to decide **with some confidence**, which of the two designs is better. Again, this is very similar to making the decision with confidence on whether head is more probable than tails, or vice versa.



so to summarize, statistics is about taking data from some real-world process, and drawing conclusions about this process from the data you collected.

POLL

If we flip a coin 1,000 times and get 507 heads, can we conclude that the coin is unbiased?

RESULTS

Yes	50%
No	50%

Submit

Results gathered from 1070 respondents.

2

1/1 point (graded)

Suppose a coin is tossed 1000 times and comes up heads 610 times. Is the coin biased?

- (•) Yes, with very high probability
- ( ) Unclear

✔

备注:

1.4 The three card puzzle.

Suppose we have three cards in a hat:

- **RB** - One card is painted blue on one side and red on the other.
- **BB** - One card is painted blue on both sides.
- **RR** - One card is painted red on both sides.

So we have these cards in the hat, and we mix them.

2 The setup

- I pick one of the three cards at random, flip it to a random side, and place it on the table.
- U be the color of the side of the card facing up. (**B** or **R**)

3 Do you want to bet?

- If the other side of the card has a different I pay you $1,
- If the other side has the same color you pay me $1.

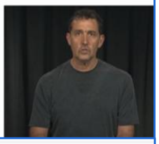
4 Why is this a fair bet ?

- Suppose U is **R**.
- Then the card is either **RR** or **RB**.
- Therefor the other side can be either **R** or **B**
- Therefor in this case the odds are equal.
- A similar argument holds for the case where U is **B**

5 Lets use a monte-carlo simulation

The code below selects one of the three cards at random and selects a random side to be "up".

It then prints the card and indicates if the two sides have the same or different colors.



```
BB same      RB different BB same      RR same      RR same
BB same      BR different RB different BB same      BB same
BB same      RR same      BB same      BB same      BR different
BR different BB same      BB same      RR same      BB same
BB same      RR same      BR different BB same      BR different
BB same      RR same      BB same      RR same      RB different
BR different RR same      RB different BB same      BB same
RB different RB different RR same      BB same      BR different
RB different BB same      BB same      RR same      BR different
RR same      BR different RB different BB same      BB same

{'different': 17, 'same': 33}
```

But if you look at the numbers that come out down here, it shows that different is 17 times, and same is 33 times.
So clearly, different happens much fewer times than same, okay?
So I basically have here a game that is, even as simple as it is, it's going to, it's unfair to you and I'm going to gain money on average from playing it.

6 The simulation does not agree with the argument

- In Simulation: the two sides have the same color about **twice** the number of times that they have different color.
- you are twice as likely to lose as you are to win.
- On average you lose 33 cents per iteration:
  $\$1 \times (2/3) - \$1 \times (1/3)$

7 Alternative argument

If we pick a card at random 2/3 of the time we pick a card where the two sides have the same color, and only 1/3 where the color is different.

8 How can we be sure?

- The original argument also sounds convincing, but is wrong.
- To be sure that our argument is correct, we need to define some concepts, including **outcome** and **event**. Which we will do next week.

So supposedly now we understand the game, but the problem is the original argument also sounds convincing, but it is wrong.
So how can we distinguish between this argument and another argument and say which one is right without doing a simulation?

Okay, simulation is fine, but sometimes, as you saw already, running a simulation to answer a probability question is hard, you have to run for a long, long time, and you get only an approximate result, okay?

POLL

Why was the assumption incorrect that both players have a 50% chance of winning a game?

RESULTS

We never accounted for a blue side showing.	5%
We cannot know the probability until many tests are performed.	10%
We only have a 2/3 chance of drawing a card with a red side showing.	21%
We didn't consider the probability of drawing a card with two of the same colored sides.	64%

Submit

Results gathered from 884 respondents.

If we repeat an experiment many times, the long term frequencies of the outcomes converge to the probabilities.

- (x) True
- ( ) False

✔

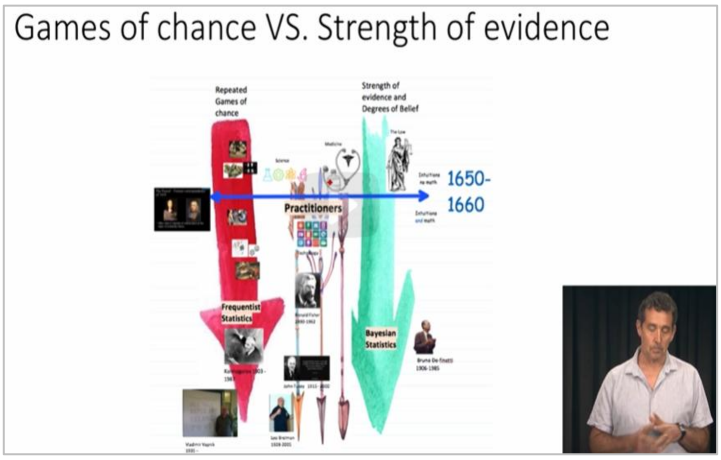
Answer

Correct: Video: History of Probability and Statistics

备注:

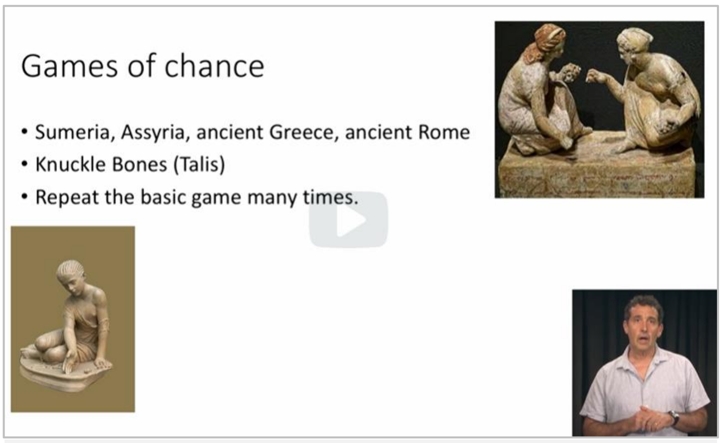
1.5 History of Probability and Statistics

Today I want to tell you a little bit about the history of probability and statistics, this is not because this is a course about history, but simply it would really help you, before we start working on the details and the mathematics of doing probability, it would help you to know what is the general framework and what are the kind of questions, that people ask in statistics and it turns out that it's a very diverse set of questions.

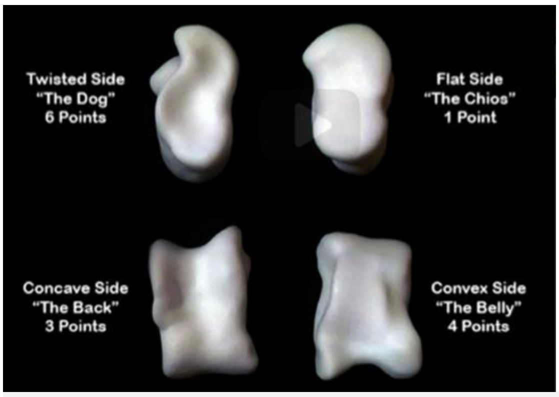


So in general, you can say that statistics and probability in its modern forms started around 1650, but the types of questions, that have been asked fall into two groups in general

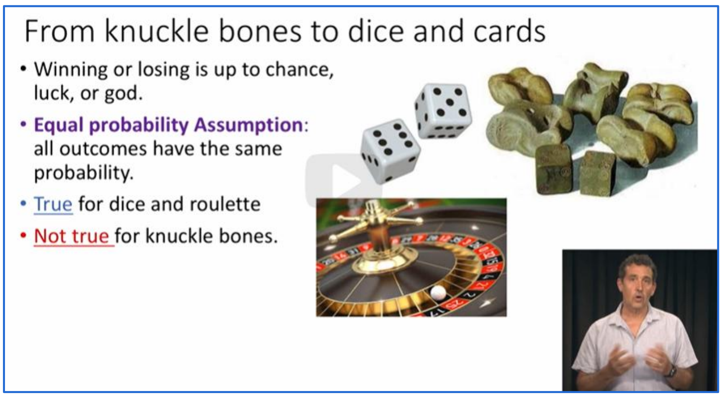
- and one is repeated games of chance
- and the other is strength of evidence and degrees of belief,



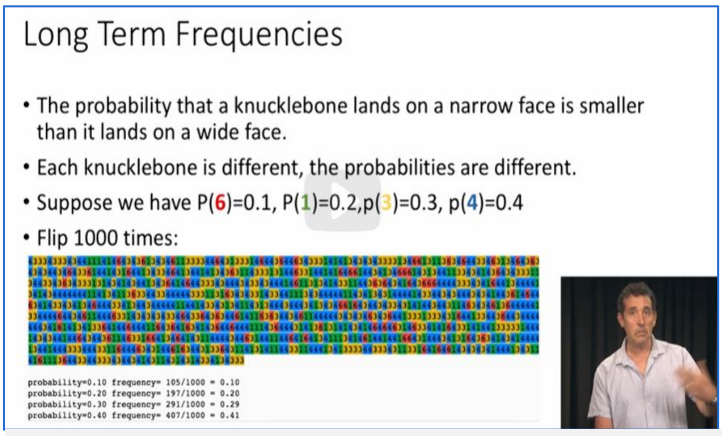
Okay, so games of chance, games of chance have existed for a very, very long time, much before the 1600s, probably from before Christ and the earliest kind of games were based on what was called Knuckle Bones or Talis, so those are bones that are taken from goats.

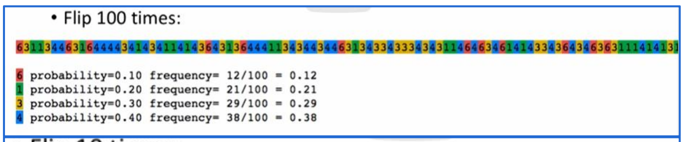


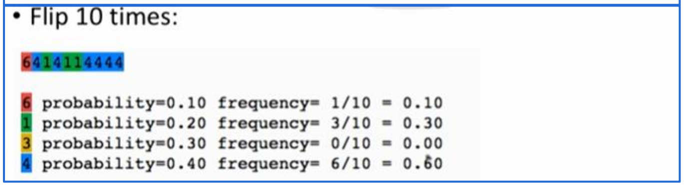
Here is a picture of what the basic knuckle bone game is, this is the strange shape of the knuckle bone
and depending, you throw it in the air, it falls,
and depending on the side on which it falls, you get different amounts of points,
okay, so this is the ancient game of knuckle bones,



so it's true for dice and roulette, but it is actually not true for knuckle bones, because knuckle bones are asymmetric and so the probability that they fall on different sides is not the same.



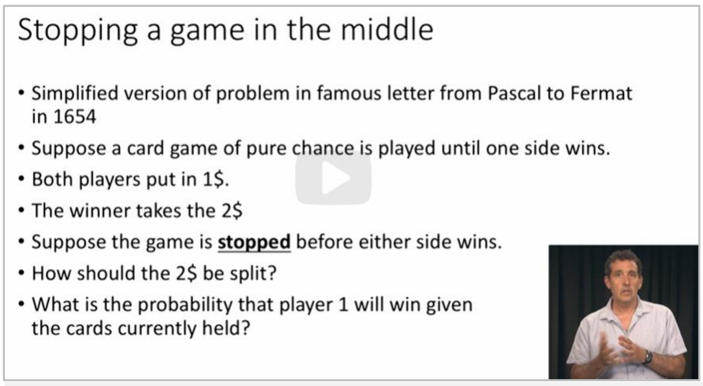


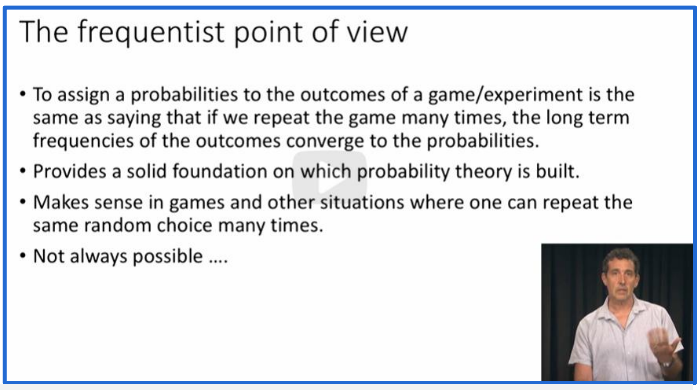


Long term frequencies are basically the assumption that when you throw something many, many, many times.

so if you repeat this kind of game just a small number of times, you don't get anything related to the probabilities, but on the long term, you get the long term probabilities.

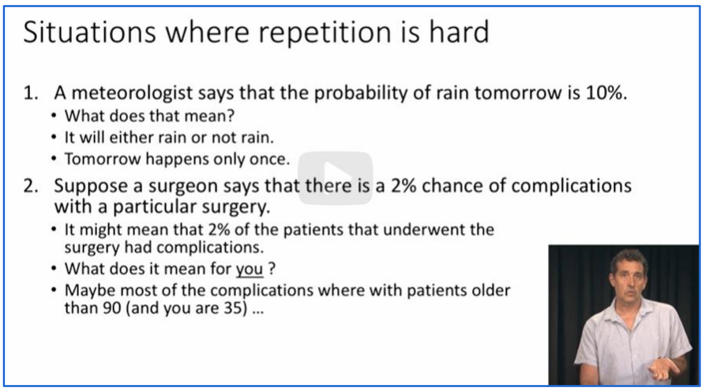
备注:





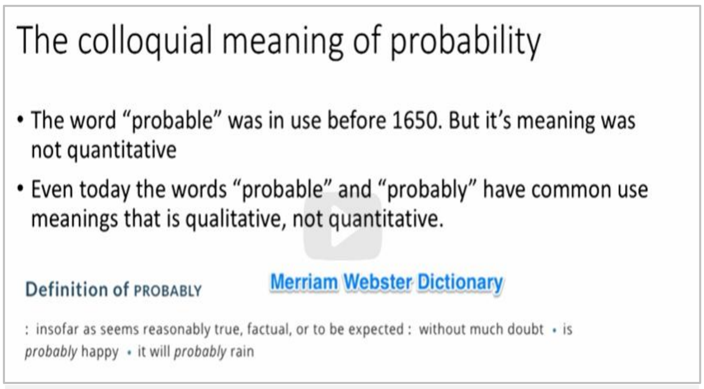
Okay, so that was the question that Pascal solved for some cases and sent in a letter to Fermat and that is a classical question, that is about games of chance, frequencies and probabilities.

So that's **the frequentist point of view** and it basically says the only thing that probability means is that when you repeat the same game or the same trial many, many times on many, many people or many, many cases, then what you get converges to the probabilities, that's what probabilities mean and this gives you a foundation on which most of the mathematics of probability theory can be built and it makes sense in games and other situations like polling, but it doesn't always make sense, right.

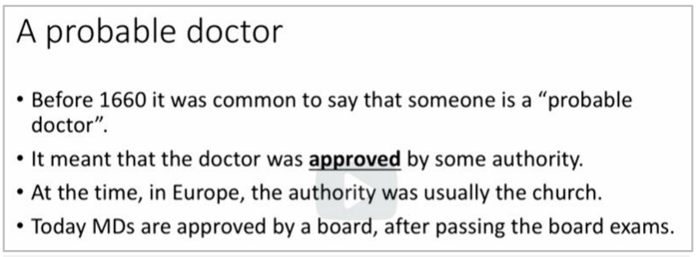


Sometimes it really does not make sense to think about probability as something that is the result of repeating the same game many times, so here are some examples.

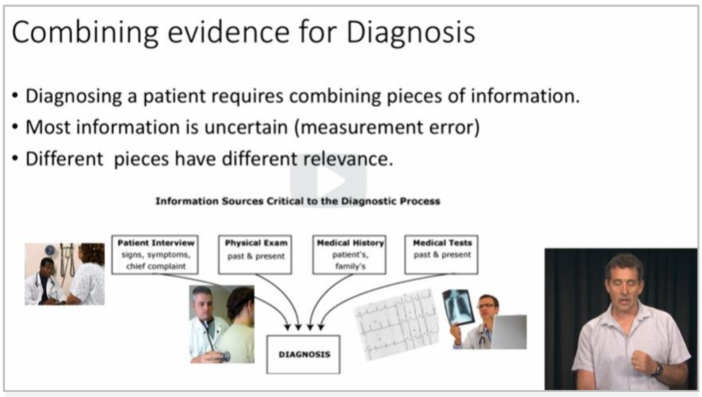
And this leads us to the other type of probability that has to do with confidence, with measuring evidence and basically quantifying opinions.
.



So all of these things, they're just basically saying probably means very likely, it is very likely to happen or I'm pretty sure it will happen, right, so that's a very common but unrelated to mathematics, unrelated to games of chance.

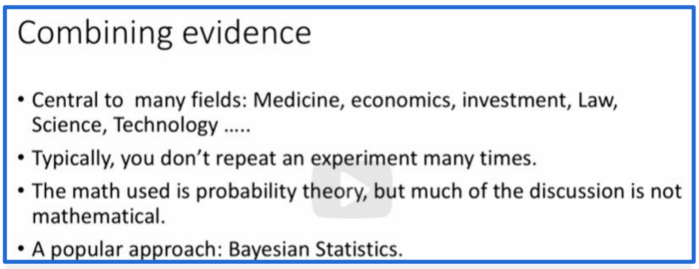


When you say that someone is a probable doctor, it meant that this doctor was approved by some authority, okay, so probable just means approved in some way and approved meant that at that time in Europe, it means simply the church, the church said that this is an approved doctor, right, so that meaning of probably was really very different than what we intuitively think about today.
Today MDs are approved also, they are approved by a board and after they do the board exam, then they're approved and they're a good doctor or they're approved doctor or the board certified doctor.



Okay, so let's think about the kind of problems, that are faced by a doctor. So if you want to diagnose a patient, it requires putting together many different pieces of information, okay, so this is a little diagram, that gives you different types of information, that goes into the diagnosis, the patient interview, physical exam, medical history, medical tests and most of the information is uncertain, right,
you don't really know exactly what is the blood content of a person, you know for a particular sample, it might change from day to day and also different pieces of information have different levels of relevance, right,
it might not be very relevant, the results of the blood test or it might, the x-ray might reveal a lot or the MRI might reveal very little.

So when you're in this situation, you have information coming from different places and with each piece of information, you somehow want to associate the confidence that you have, now the doctor does this by and large intuitively.



This kind of problems appear all over and not just in medicine, so when we talk about combining evidence, we are talking about things that are central to medicine, to economics, investment, law, science, technology, many, many different things, that rely on basically combining evidence and if they want to do it quantitatively,
then they use probability and statistics.
And typically you don't repeat an experiment many times, so in some very particular situations you do, but in most of the important decisions that you're faced,or a doctor is faced, they have this one patient that they need to treat, so there's no meaning to saying really, what's the probability of something, okay.

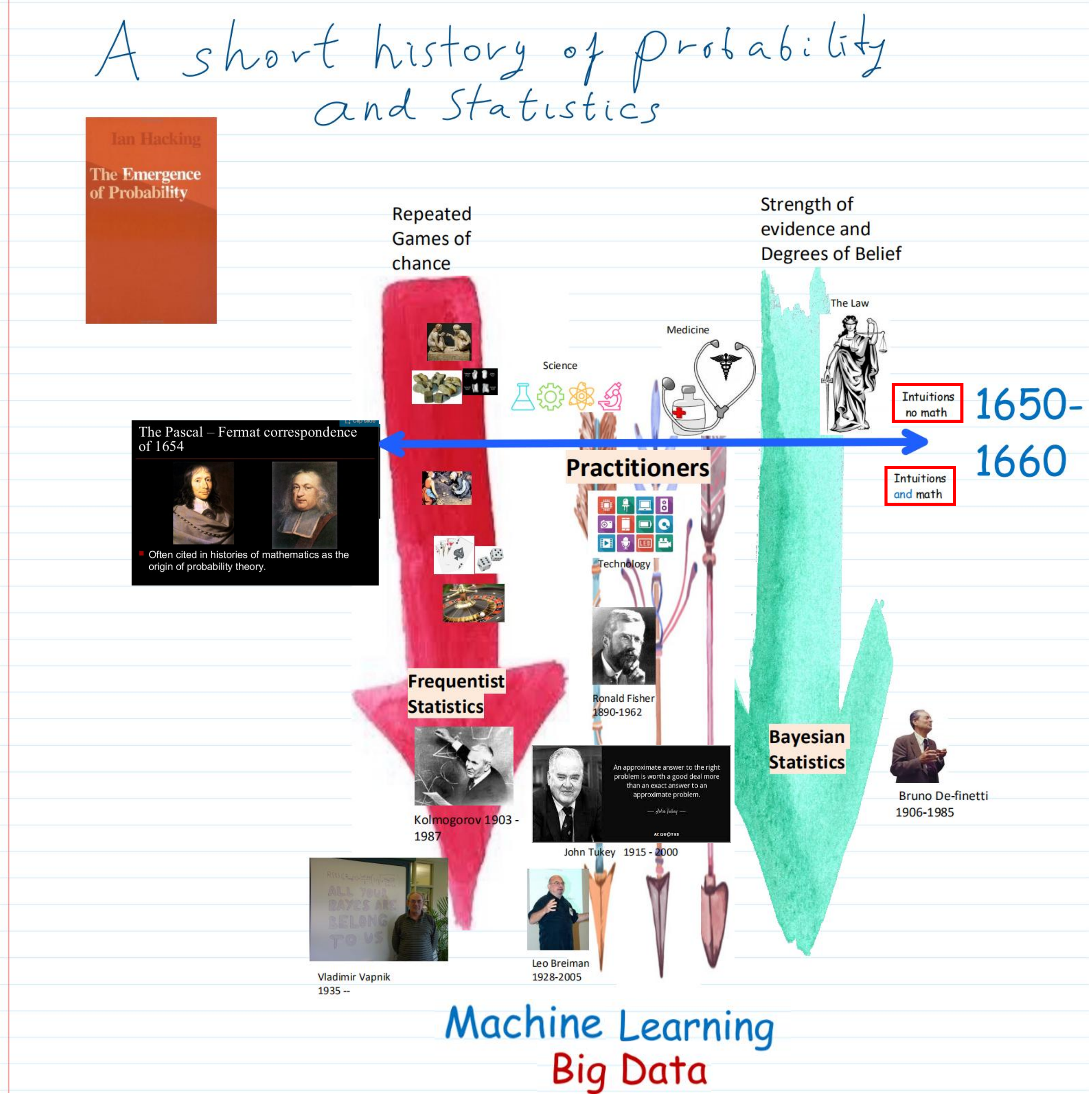
so the math of this is provided by probability theory, but much of the discussion around it is not mathematical, so it is really discussion that has to do with convincing nd with how do you compare this evidence to another evidence, some of it is math and much of it is discussion and that's the type of things that you get in statistics, you don't get really cut and dry outcomes.
A popular approach to putting all of this in a very common framework is **Bayesian Statistics**. so Bayesian Statistics puts as its main thing how to evaluate evidence and how to combine evidence and it is not, it doesn't take the same approach as frequentist, even though fundamentally they use the same math.

A Poster

History 1

Thursday, August 17, 2017 2:13 PM

Hi, last video we talked a little bit about the history of probability and statistics. And I gave you some pointers to the main ideas that were there. Now, I would like to put these things together with the timeline so that you can have a better sense about how probability and statistics developed over the years. So let's look at this poster. This is a short history of probability and statistics and much of it is based on this book by Ian Hacking, 《The Emergence of Probability》. So if you're interested in more you can go read this book. It's an excellent book.



There are two parts to probability and statistics. The two threads are repeated games of chance, on the one hand, and the strength of evidence and degrees of belief on the other. That's the two arrows, the red one going down and the green one going down, and time goes in this direction, so this is as time progresses.

And statistics and probability in its modern form is pretty much agreed to be starting at the time of Pascal and Fermat that were two mathematicians in 1654. But the main interesting thing is around 1650 is when modern mathematical probability and statistics started to be developed. Okay, so that's the timeline that we have with the blue line. All right, so let's look a little bit from what happened before that point.

- Before that point, you had repeated games of chance, so those were games played with different things, like knucklebones and dice and cards. And those raised questions of the type of what is the right way to split the money when you stop a game early? So that is the part of statistics that we will actually deal with quite a lot in the beginning, which has to do with games of chance.
- The other part that is much less well-defined, but probably even more important is what do we do when we have a state of uncertainty? We have some evidence towards some conclusions but we are not sure how to weigh different evidences that might be contradictory. So these kinds of things come up in law. So here is the law. It comes up in medicine and it comes up in science and later on, technology.

Basically, in modern science and technology, probability and statistics are a necessary part. Now, in public policy, it's also a necessary part. So those things existed from before and in these correspondences, Pascal and Fermat also related to them. But it's important to remember that these two things are quite different from each other. One is about evidence and about how people think about evidence and the other is much more mechanical. So it has to do with rolling dice and so on.

So of course, the rolling of dicedid not stop at that point. We have casinos, also, now. And so these questions are natural, and these questions give rise to the **frequentist** approach to probability and statistics that was described in the other video.

- And the best known champion of Frequentist Statistics is Andre Kolmogorov, one of the great mathematicians from Russia, and he invented what's called the axiom of probability. So he was central to this view.
- And in the more recent, current, still alive is Vladimir Vapnik, who has developed some of the foundations for machine learning.

Okay, so this is about the frequentist.（以下剪切过来的）So, just to add a little bit,there's this area that we will also talk about called hypothesis testing and P-values, which is the Frequentist approach to arguing about degrees of belief. So it's an interesting contrast and this actual approach is now very, very commonly used in science to accept or reject papers according to the strength of the evidence that they have.

Now, in the other direction, in the side of evidence and degrees of belief, there was a different line of development, which is called **Bayesian Statistics**, and we will talk also about that in a later time, in which you take your belief before you see the evidence and you update them when you see the evidence

- and the champion of that was Bruno De-finetti.
- Okay, so you have on the one side this Bayesian Statistics approach and on the other side, the Frequentist approach and there's definitely a tension between the two. So this is a pretty famous picture by now of Vapnik standing around next to a board and in the board it says, "All your Bayes "are belong to us." So this is a clear slight of Bayesian Statistics.

All right, but then what develops over time is **people that are statistics practitioners**, people that actually use probability and statistics in order to solve real world problems and I draw them in the middle here, the practitioners, because they are, in general, not dogmatic to one side or the other. They would use Bayesian Statistics when it's appropriate, Frequentist Statistics when it's appropriate and other heuristics when that's appropriate.

- Okay, so the father of those methods is Ronald Fisher, who has brought statisticsto the sciences and also to the social sciences.
- And then more recent ones are John Tukey and even more recent is Leo Breiman, the inventor of bagging and cart and other important algorithms.

And as I said before, these middle ones are the practitioners, I drew different colored arrows in all kinds of ways because they're each unique. They take various ideas from probability and statistics and they apply it to real problems in their own unique way.

So the modern version of these practitioners today is Machine Learning and even more recently, Big Data, when we try to apply Machine Learning methods to Big Data. So let's zoom out and to see the complete picture and what was important for me to show you here is that while the methods that people are more familiar today with are like Machine Learning, Big Data, and Neural Networks are very popular, there is actually a very long history and in this long history people developed a lot of very important methods that are worthwhile knowing about.

So I hope that that gives you a perspective that will be useful for you for the rest of the course. Thank you very much.

POLL

Bayesian Statistics is useful for situations where an experiment is repeated many times.

RESULTS

True	29%
False	71%

Submit

Results gathered from 782 respondents.

备注：