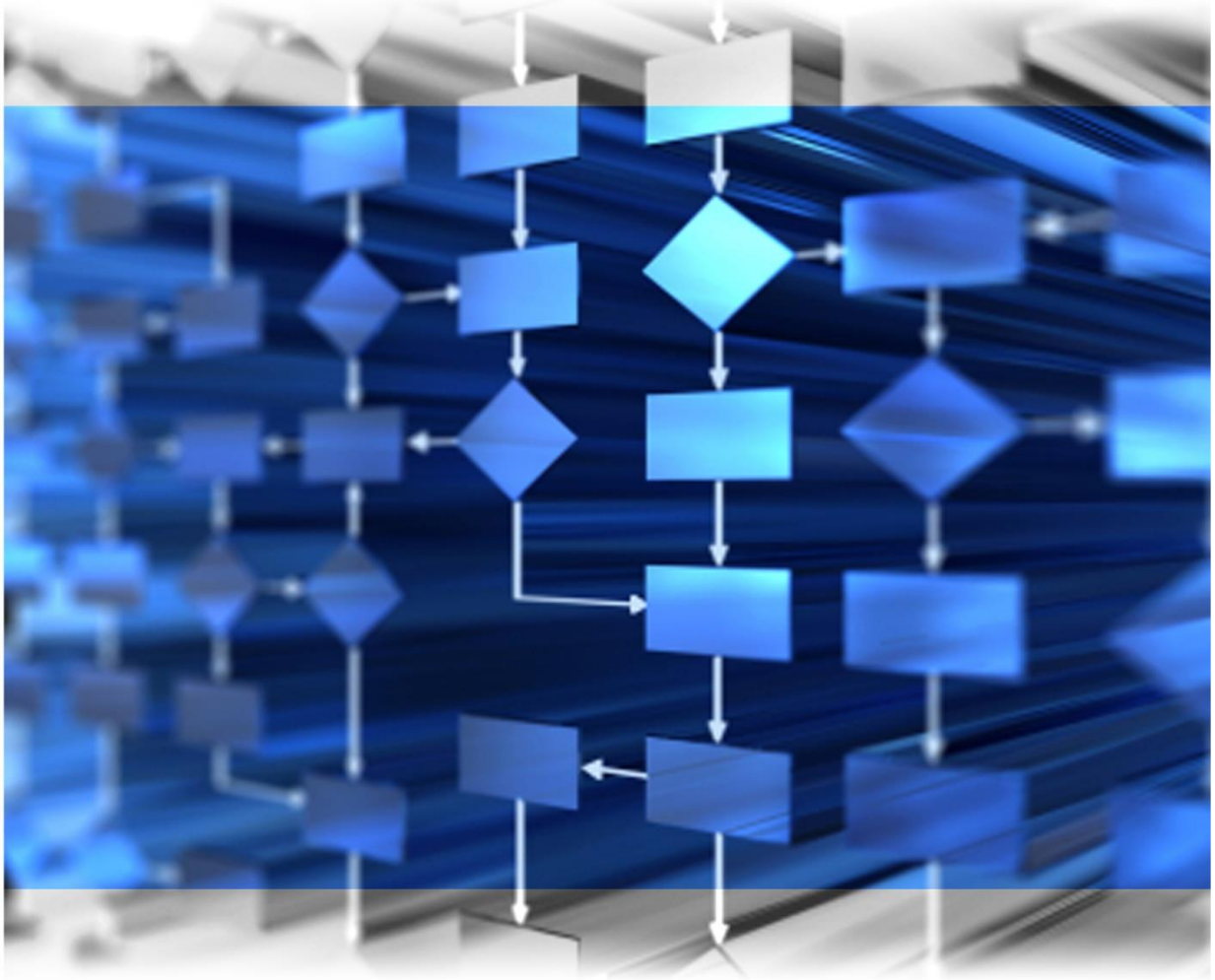
LONE STAR

THE RIGHT ANSWER



What is “Bayesian” These Days?

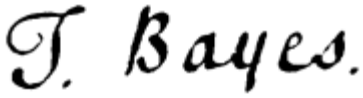
August, 2013

What is "Bayesian" These Days?

2013 marks the 250th anniversary of the Royal Society's receipt and publication of work by Thomas Bayes. His work forever changed probabilistic reasoning; nearly all probabilistic modeling and research is Bayesian now. It's the foundation of a good deal of Lone Star's work. But few people understand it.[1] Those who do know of it generally have only a limited grasp of Bayesian applications and usage. This paper provides a survey of Bayesian applications and a basic tutorial on Bayes and his work. It looks at "traditional" Bayesian thinking, as well as modern applications of Bayes found in surprising places. It describes a little about Lone Star's application of Bayes

Some History

Thomas Bayes was a mathematician, a Presbyterian minister and the son of a minister. Both father and son were from a prominent family of "Nonconformists."[2] The Royal Society published his paper, *An Introduction to the Doctrine of Fluxion, and Defense of the Mathematicians against the Objections of the Author of the Analyst* (1736). In it, he countered arguments attacking the foundation of Newton's calculus.[3] But his most famous work, on probability, was published after his death.



Bayes' Signature

[1] Sir Harold Jefferys of Cambridge University wrote, *"This theorem (due to Bayes) is to the theory of probability what Pythagoras's theorem is to geometry."* It's nearly impossible to do geometry without using the principle that the longest side length of a right triangle is the square root of the sum of the other sides squared. And, it's nearly impossible to work on problems in probability without Bayes. Humans understand tangible triangles better; Pythagoras has better recognition.

[2] Nonconformity was refusal to "conform" to the Church of England. The 1662 Act of Uniformity meant any English subject belonging to non-Anglican church was a non-conformist. Presbyterians, Baptists, and Quakers were among the first considered Nonconformists. The Act of Toleration (1689) exempted Nonconformists who had taken the oaths of allegiance and supremacy from most penalties. However, Nonconformists were ineligible for some educational and other opportunities, until the 19th century, long after Bayes' death. D.R. Bellhouse provides background on nonconformity and its relationship to Bayes as an educated member of society, a Royal Fellow, and a minister in his paper *The Reverend Thomas Bayes, FRS:* A *Biography to Celebrate the Tercentenary of His Birth,* Statistical Science 2004, Vol. 19, No. 1, 3–43

[3] Newton's *Method of Fluxions* was published in 1736. "Fluxions" is Newton's term for differential calculus. Newton's paper, published by the Royal Society, was *The Method of Fluxions and Infinite Series: With Its Application to the Geometry of Curve-Lines.* Newton was not precise in his definitions. Bayes clarified. Bayes paper might be titled, "*What Newton meant to say…"*

Little is known with certainty about his life, the only picture we have may not really be him, and few of his works survive.

We know more about him after he published the paper on Fluxions. It seems to have been his defense of Newton which attracted the attention of the Royal Society and led to thinking Bayes would be a good candidate for a Royal Fellow. One of the sponsors at the Royal Society seems to have been the 2nd Earl Stanhope, a patron of other Nonconformists including Joseph Priestley the discoverer of oxygen and other substances.

Bayes was elected Fellow of the Royal Society (F.R.S.) in 1742 and contributed to the review of the papers and publications of other Royal Fellows, but apparently tended to correspond privately with a few other members of the Royal Society, rather that publish in the Society's proceedings.

By 1755 his health began to fail. He died in 1761. He was probably about 60 years old. His relatives asked another Nonconformist minister and scholar, Richard Price, to examine papers Bayes had written on various subjects. Price found an incomplete and imperfect paper with a, "solution of one of the most difficult problems in the doctrine of chances."[4] Price edited and submitted Bayes work to the Royal Society; it was published in 1763.

Bayes had been interested in what we would call "Probability and Statistics" for several years before his death. He provided a sophisticated review/critique of work by others in this area, which is recorded in a few remaining, papers, and acknowledged by other authors and reviewers. He made insightful contributions to the work of others. But, during his life, Bayes published none of his work in probability.

Bayes' Family Tomb

What Price found in dead man's papers led to what we call "Bayes Rule" today.

Bayes didn't live to see his most important work published or to see the recognition it earned.

[4] From Morgan's Memoirs of the Life of the Rev. Richard Price, D.D., F.R.S.R., as quoted by Bellhouse

What Bayes Discovered

Bayes wanted to address topics with uncertain outcomes, either because we just don't know, or because the topic has random variation. Bayes didn't know what caused an epidemic. He also didn't know where a ball thrown randomly on a table would come to rest.

What he saw was that the alternatives could be testable hypotheses. We can think of different causes for an epidemic, we can think of different grid locations the ball can be in. With perfect information, each hypothesis is either true or false.

But what about imperfect information? Bayes saw that our idea about the probability of each hypothesis should change as we add information, even if imperfect. He showed that the relationship of how probability shifts can be described in a fairly simple way, explained below.

He showed how to compute the best estimate of probability, in spite of all kinds of uncertainty we face.

The organization of different alternatives (hypotheses) is a model. It's a representation of our understanding of a problem. Bayes showed how to take our understanding, our model of an uncertain topic and treat it with math. And, he showed it always worked, even when our knowledge isn't perfect.

He showed how to compute the best estimate of probability, in spite of all kinds of uncertainty we face.

Bayes Rule and an Example

Suppose your child says, "one of the kids got sick at school today." If the school is coed, we'd guess there's a 50% probability the sick child is a girl, and a 50% chance it's a boy. Either seems equally likely. We don't know. If you are the girl's PTA sponsor, you may want to know if it was a girl that was ill.

But, if your child then adds, "it happened in the girls' locker room," we are probably justified in thinking the odds are very high (but still not 100% certain) a girl was sick.

If we find the sick child wore a dress, had braided hair, and wore Mary Jane shoes, we are approaching 100% certainty it was a girl. Bayes took this common sense idea (we should modify our estimate of probability when we get more information) and extended it to mathematics.

Bayes Rule looks like this:

(These are gentle equations, if read slowly this will make sense, but if you have math aversion, please skip to the next section and feel no guilt)

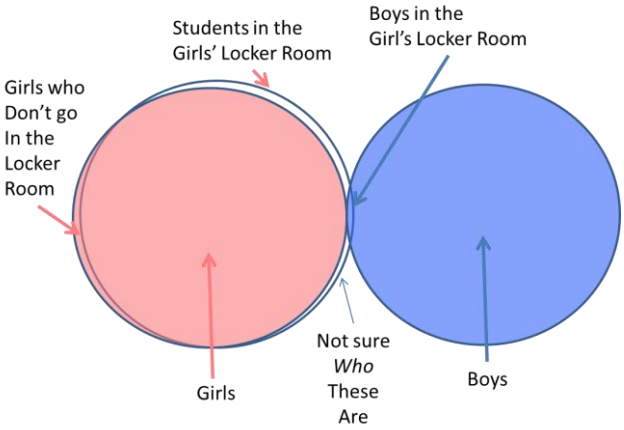
The probability the child is a boy is P (B), and P (G) is the probability the child is a girl. When we begin the conversation (and we don't know about the locker room, dress, hair, or the shoes) P (G) = 50%.

If you have math aversion, please skip to the next section and feel no guilt.

We also have to think about "conditional probability" - the odds something is true, given some other condition. Like the odds that a child goes in the girls' locker room if the child is a boy.

We write this P (L|B), or the probability of L (going in the locker room) given, B (being a boy).

Since there are forces at work to keep boys out, we assume most boys never go there. The parents of girls hope P (L|B) is a low probability, and for our example, we'll guess it's 1%. On the other hand, we might have the statistics to show P (L|G) is 95%. Nearly all girls go into their locker room for some reason at one time or another. We might know these things by watching security video.



This may not account for 100% of locker room occupants, but that's ok. In the real world, we may have some kids we didn't see very well, and we weren't sure of gender. Perhaps some kids never go in the locker room. We can account for about 95% of the cases. This is part of the uncertainty Bayes accommodates for us.

What we want to know is the probability the child was a girl, if the location of the incident was the locker room, or P (G|L).
Was the Sick Child a Girl?

What Bayes discovered was how these probability values related to each other.

P (G|L) = P (L|G) (PG)/P (L) or;

The probability the child is a girl, given the location is the locker room is equal to:

The probability of going to the locker room if a girl, multiplied times the probability a child is a girl, divided by the chances a child goes to the locker room.

Remember that we had a problem with accounting for the kids going to the girl's locker room. So all we know about P(L) is that P(L) = P(L|G) (PG) + P(L|B) P(B) which is the probability of each gender going to the girls locker, times the probability of that gender.

So, now we replace P (L) with what we know.

Now we get this equation

P (G|L) = P (L|G) (PG)/ (P (L|G) (PG) + P (L|B) P (B))

Which is 95% x 50% / ((95% x 50%) + (1%x50%) Or 47.5 %/(47.5% + 0.5%).

That's 47.5%/48%, which is 98.96%, so we are pretty sure it's a girl who was sick in that locker room, even without taking into account the dress, the braids and those cute Mary Jane shoes. And, we hope she's feeling better now.[5]

Bayes Rule in the 20th and 21st Century

Bayes won because it works.

A wonderful narrative of Bayesian progress is found in the book, *The Theory That Would Not Die; How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. This isn't a math book. It's a book about people and ideas, how things change, and how we can hope good ideas will win out.

Sharon Bertsch McGrayne wrote it. She found great story telling because of the "two centuries of controversy." A number of things ended the controversy.

First, Bayes won because it works. World War II code breakers used it to choose the most probable keys used to crack coded messages. It found a lost

[5] In Bayesian jargon the initial estimate of 50% that the sick child was a girl is a "Prior" and the revised estimate of 98.96% is a "Posterior." Depending on how we do this, we may have a single probability as either a prior or posterior, or, we can have a distribution, with the prior distribution being shifted and reshaped into a posterior distribution.

submarine in the middle of a very big ocean. And, Bayes won because some people didn't know they were using it, or simply didn't know about Bayes and the controversy. It's just math, and math is true.

Two very important applications of Bayes are machine control and signal processing. Many of the problems in these two fields are roughly the same; dealing with imperfect information.

In machine control, we find errors like the GPS doesn't tell perfectly where we are, the compass doesn't really point exactly north, and the speedometer isn't quite truthful. We have to generate an estimate of where we are, and where we are going in the face of these errors.

In signal processing, we have "noise" in our system. Most signal processing is digital these days. Everything is represented by "1" or "0." But in your computer, GPS receiver, or iPhone, the 1 and 0 are represented by a voltage. 0 might be no volts (zero volts) and 1 might be about 1 volt. But real voltages are never quite perfect. So, if we have ½ volt, does it means 1 or 0? In the case of GPS it's much worse. The GPS radio signal coming down from the satellite is about as strong as a night light bulb, and its over 10,000 miles away. We have almost no chance of reading any particular 1 or 0 in the signal stream.

They work because they use Bayes Rule (or 'obey Bayes Law'), whether their inventors knew it or not.

In machine control and signal processing, "filters" are what we use to figure out what we think is probably true. Where is north? What's the data stream of 0's and 1's for my iPhone or GPS?

These filters have many different names, with the Kalman filter being one of the most famous. R.A Kalman published his findings in the 1960's. Today Kalman filters are everywhere. They run in airplane autopilots, machine control, and other places. It turns out many of these filters are Bayesian.[6] They work because they use Bayes Rule (or 'obey Bayes Law'), whether their inventors knew it or not.[7]

Two Kinds of Bayes

Digital filters adjust estimates of truth (where am I?) or, the odds of a proposition (is that email spam… maybe). The second kind, estimating probability is easier to recognize as Bayesian.

[6] Most "optimal predictors" can be derived using Markov methods (e.g., Markov Chains). They are named for various proponents and inventors; Feynman-Kac, Moran, Kalman, etc. They are nearly all Bayesian because it's almost impossible to manipulate probability estimates without using Bayes Rule, whether the inventor knew it or not.

[7] Kalman doesn't reference Bayes in his seminal paper. He does make one reference to Laplace, and Laplace was an ardent promoter of Bayes.

Spam filters are easy to recognize as Bayesian. They estimate probabilities (or odds). An email that has a high probability of being spam is rejected.

The first time a server sees email from a Chinese philanthropist, claiming to be a distant cousin, offering you a job for $10 million a year to help give his money away, who knows? It could be true. Computers are pretty dumb, and it might not be flagged as spam. But, a second identical email to someone else seems odd, even to a computer. After 1000 identical emails it seems certain the philanthropist can't have so many cousins.

Spam, or, Not Spam – both are hypotheses. Worldwide, about two thirds of all the email in is Spam. The safe thing to do for an email filter is to reject *all* your email. Somehow that seems to be a bad idea.

Spam filters are easy to recognize as Bayesian.

So, Bayesian email filters are adjusting all the time. They change the odds they use to examining mail which comes to a server by looking for things such as every person in your company getting an invitation from a billionaire Chinese cousin. Some of these filters also adjust the basic odds that any email is spam (the prior) because we don't really care about the worldwide average odds; we care about your inbox. It turns out spam frequencies vary with where you live, what kind of work you do, and many other factors.

This kind of filtering can use many different conditional probabilities to test two or more alternative hypotheses. There are many tests we can apply to shift our guess about the odds of being spam or not.

What about the other kind of Bayesian filter?

They tend to estimate outcomes, rather than odds. If you have seen a GPS map with a point that shows your estimated location, and a circle around the point showing where the navigator is estimating your location, then you have seen this kind of estimator at work.

What is less obvious is these estimators generate many estimates at once. In a GPS navigator, we will probably find estimates of the distance between you and a dozen satellites; it may estimate the relative velocity between you and the space vehicles. If we have a good map, the navigator will use that information estimate the height of the earth's surface, whether you seem to be on the surface, or not.

So, this kind of navigator is estimating dozens of things, and then using those estimates to generate the estimate or your location, and your speed.

Preferring Outcomes over Odds

This kind of system estimator (or filter) that seeks to calculate the best assessment of "truth" doesn't look much like the classic formula in the example of the sick child. But they are just as Bayesian as the spam filter.

Why do some applications prefer this kind of estimator that generates predictions of outcomes?

There are many reasons.

First, outcome predictions are more actionable. You turn left or right depending on your location on the map. You know whether to buy a company (or not) based on an estimate that it meets certain criterion (or not). Corporate decision makers and robot controllers both want estimates not odds.

Humans are predictably bad at guessing odds.

Pigeons are better.

Executives rarely want to know a model predicts its 81% probable buying a company fits all our criteria. Rather, they want to know what our model predicts for our criteria (which might be earnings, sales, and cash flow). They want to see the numbers, the estimates.

They may also want to know the odds. For example, they may care if our models show there is a 92% chance we will make our earnings goals, a 93% of meeting sales goals, and, a 95% chance of generating or exceeding target cash flow, while there's an 81% chance of meeting all these goals at the same time.

But, they will prefer to know *predictions of outcomes* not just odds.

This approach allows us to use a large number of input parameters while estimating very diverse outcomes. At Lone Star, we've built system models that estimated market behavior, production capacity, and logistics performance all in the same model at the same time.

This is a very interesting kind of problem; it's the like real world; tangled, interconnected and messy.

Another benefit is the inputs we can use. Humans are predictably bad at guessing odds. Pigeons are better[8] It is better to ask about values for inputs. In

[8] An example is how bad we are at the "Monty Hall Problem" involving when to change the door you choose if the game show host gives you a chance. Researchers Walter Herbranson and Julia Schroeder detailed their findings of humans vs. pigeons in the February 2010 issue of the *Journal of Comparative Psychology*. Herbranson is a comparative psychologist at Whitman College in Walla Walla, Washington. His comparative cognition lab is in Maxey Hall at the Whitman campus – the name of the Hall is a delightful and improbable coincidence.

other words, we are better off to ask about how much gas you'd expect to find in a fuel storage site, rather than ask about the probability we'll find a certain amount. Mathematically, these are equivalent, but they aren't equivalent terms of human cognition.

Summary

It's been 250 years since Reverend Price sent the Royal Society the work of his deceased friend Bayes. Now, powered by computers with power beyond what Bayes or Price could have imagined, we can use the power of Bayes Rule to generate estimates and predictions of astonishing accuracy.

What is Bayesian now? Nearly everything interesting dealing with uncertainty and probability. And, modern models estimating outcomes (not just probability) are found all around us and may be the most common kinds of Bayesian tools, whether we recognize them or not.

About Lone Star

Headquartered in Dallas, Texas, Lone Star provides business and technical and advisory services that address clients' most complex, mission critical challenges. By utilizing processes based on best practices and tools that enable effective outcomes, Lone Star delivers the value clients expect.

For more information the Lone Star home page is www.Lone-Star.com