



Bayesian Statistics

Bayesian techniques are an alternative to classical or frequentist methods. Although frequentist statistical techniques are more commonly used, the use of Bayesian methods is increasingly widespread. Bayesian techniques are used in a broad range of situations such as radar tracking, inertial navigation, spotting credit card fraud and analysing clinical trials.

Bayesian methods are increasingly popular due to the advantages they offer over frequentist statistical methods:

- A Bayesian analysis provides direct statements about the quantities of interest, providing more intuitive results and feeding naturally into a decision making process.
- Bayesian statistics allows us to obtain some useful information even from a single piece of evidence, whereas many more samples would be required for a frequentist statistical result.
- The Bayesian approach allows all evidence to be taken into account in an explicit way. Different forms of evidence can be combined in the overall probability model or included via the prior belief. Analysing the data using different priors allows the data to be interpreted from different points of view, such as regulatory or business.
- The Bayesian framework more naturally allows for the modelling of biases or systematic error, and for modelling any hierarchical structure in the data or in the problem. For instance in a clinical trial with 200 subjects at 5 different hospitals, we can model not just the overall effect of the treatment studied, but also variations in outcome between different hospitals.
- Bayesian statistics provides greater flexibility. It offers a natural way to adapt an experiment in progress in the light of results collected so far, or to halt the experiment early if the result is clearer than expected and no more data is needed to achieve the desired degree of certainty.

What is Bayesian Statistics?

The essence of Bayesian methods is a mathematical rule explaining how you should change your existing beliefs in the light of new evidence. It allows people to combine new data with their existing knowledge or expertise.

In Bayesian statistics, probability represents an individual's degree of belief that a particular event will occur. Thus the Bayesian approach is based on personal or subjective probabilities. Central to Bayesian methods is the use of Bayes' Theorem, which states that if A and B are events and P(B), the probability of event B, is greater than zero, then:

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

The theorem is most commonly interpreted such that A represents a model that we think might be true (in frequentist terms, a hypothesis) and B represents our observations. In this case:

- The prior estimate of probability, P(A), is our initial belief about the probability of A being true.
- The posterior estimate P(A|B) is the probability of A being true given that B has been observed.
- The likelihood factor, P(B|A) is the probability of event B occurring if A is true.
- Finally we consider a range of possible A's so we can calculate P(B), the total probability of B happening for any A. Since B did happen, we need to divide by P(B) to normalize the answer.

In frequentist inference, tests of significance are performed by supposing that one particular hypothesis, the null hypothesis, is true, and then computing the probability of observing a statistic at least as extreme as the one actually observed during hypothetical future repeated trials (this is the P-value). In other words, frequentist statistics examine the probability of the data given a model (hypothesis). By contrast, Bayesian statistics allows us to examine the probability that a possible model is true, given the available data. It also allows us to compare the different possible models and assess their relative probabilities of being true.

Should Statistics be Objective?

From the Bayesian point of view, inferences are always conditional, conditioned on prior information, i.e., what (we believe) we already know. The same data can lead to different conclusions depending upon other available information.

Objections are often raised because Bayesian results are subjective. A Bayesian statistician would answer that in order to use frequentist statistics, subjective decisions have to be taken and that Bayesian statistics merely makes this explicit.

As an example, suppose we toss a coin 10 times and get 10 heads. Should we conclude that the coin has two heads?

A frequentist would calculate a P-value of 0.001 for the hypothesis that the coin is normal. This does not mean that we are 99.9% sure that we have a two-headed coin. It merely says that given a normal coin, what happened is highly unlikely. This is an objective fact but is not an answer to the original question. We now need to decide how small the P-value must be in order to convince us that we have a two-headed coin rather than just being lucky. This is where the frequentist analysis becomes subjective.

The Bayesian approach requires a prior for how likely two-headed coins are. Say we believe that, never having seen one, only one coin in a million has two heads. Then Bayes' Theorem tells us that despite seeing 10 heads in a row, the probability of this being a two-headed coin is still only 0.001. If on the other hand the coin is very old and we think two-headed coins were more common when it was minted, this changed prior belief, through Bayes' Theorem, would lead to a higher posterior probability of the coin having two heads.

Thus the Bayesian paradigm has given a clear answer to the original question but the answer is subjective, depending on our beliefs about two-headed coins. However, since this subjectivity is included explicitly, we can easily test how important the choice of prior was by trying different priors.

This can lead people new to the Bayesian approach to wonder "what, or whose, prior should I use?" or to think that they can prove anything because they can set the prior to whatever they want. But the answer is simple: you need to use the prior of the person you have to convince of the result.

Bayesian Networks

An important subset of Bayesian statistics is Bayesian Networks, or Bayesian Belief Networks (BBN).

A Bayesian Network models a decision problem by mapping out cause-and-effect relationships among key variables and assigning to them probabilities that represent the extent to which one variable is likely to affect another. Each variable in the problem domain is assigned to a node. Each node can take different states, and the probability of the node taking each state, given the states of its parent nodes, is defined using a Conditional Probability Table.

A simple example is a network that models the effect of a possible train strike on whether two workers are late (see diagram below).

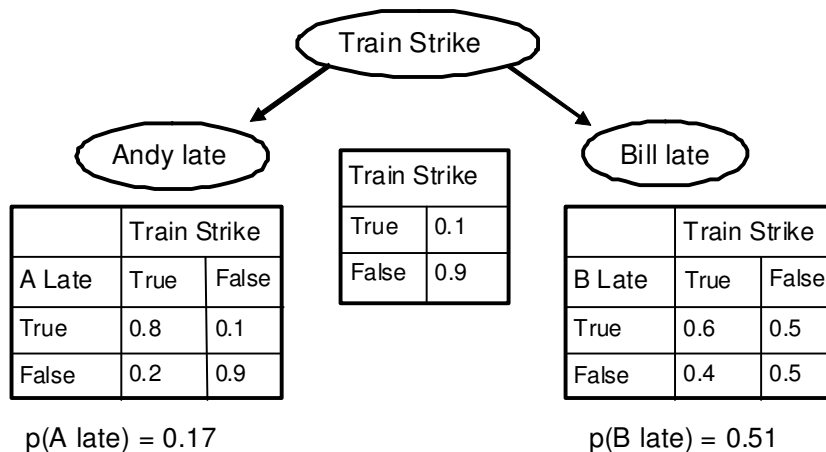
This network has been used to calculate the overall probability of each worker being late. But the main advantage of this method is the capacity to update the probabilities given new information – for example, if Andy turns up late, then it is possible to calculate the probability that there is a train strike, and thus calculate the new probability that Bill will be late, given that Andy has already arrived late.

Bayesian Networks can have hundreds of individual nodes with probabilities that can be updated every time new information is received. This makes them ideal for assessing complex situations, and they are used as tools to support decision-making processes. They are suitable for problems where uncertainty is unavoidable, since they express such uncertainty in an explicit way. They can be used to generate optimal predictions or decisions even when key pieces of information are missing.

In a decision-making process, Bayesian Networks can also be used to optimize costs or benefits. Outcomes are associated with probabilities, but actions that might lead to those outcomes can also be associated with risks or rewards. The Bayesian Network can then be used to calculate the route that is most likely to give the best return, for example the largest monetary profit.

Adaptive Testing

Suppose we are trying to conduct an experiment that consists of a number of tests (or measurements) to determine some parameters for a model – for instance the dose response curve of a drug or the performance of a car engine with different tunings. Whereas the frequentist approach tends to yield fixed experiments with a pre-determined set of tests, the Bayesian approach allows us to modify the tests to optimize what we learn during the experiment. We do this by “updating” our prior after each test with the results of that test, to get a new prior for the next test. We can then simulate running this next test with a computer program to determine which set of test conditions (which dose to give the next subject, which values of tuning parameters to next run the engine with) will tell us most about the outcome we are interested in. Such sets of tests are known as “sequential adaptive” experiments. They allow us to focus on the range of values for the parameters that prove most promising, allowing the experiment to use fewer tests, be completed quicker and yield more information than a traditional frequentist design.



Uses of Bayesian Statistics

Bayesian statistics has a wide variety of possible uses, some of which are described below.

Spacecraft Pointing Errors

Satellites have complex systems to control their orientation – the direction the satellite is pointing in. This is one of the most crucial aspects of satellite control. A communications satellite, for example, must remain pointing towards the earth at all times, while the Herschel telescope must keep its field of view fixed during its entire observation period. The orientation of a satellite has many possible sources of error due to movement of the satellite and uncertainty in the sensor measurements. All these error sources, together with others such as payload thermal distortion and misalignment, contribute to the overall error in the spacecraft instruments.

The European Space Agency (ESA) developed a rigorous set of statistical algorithms to capture individual sources of error and combine them to give overall performance. This approach is encapsulated in ESA's *Pointing Error Handbook*, originally written by staff now at Tessella, that can be used by companies developing spacecraft to construct detailed pointing error budgets. Central to the methodology is the use of Bayesian methods to update the probability distributions of the individual errors when new information is received, in order to make optimal use of available information. These probability distributions can then be recombined to obtain a new estimate of the distribution of the overall error.

ESA also commissioned Tessella's Analytic Pointing Performance software tool, which fully implements the Bayesian methodology from the *Pointing Error Handbook* to allow pointing performance to be calculated.

Kalman Filters

A Kalman filter is an example of a Bayesian estimator: an estimator or decision rule that calculates the optimal result based on the given criteria.

A Kalman filter estimates the state of a dynamic system from a series of incomplete and noisy measurements. One example application is inertial tracking of people. Appropriate sensors are attached to a person and the data gathered is used to track their position as they move around, using a Kalman

filter algorithm. At each timestep the filter uses a model of the dynamics of the object or person to predict its current position and velocity. It then corrects this prediction, based on the current set of measurements. The estimate obtained by this method is the one that minimizes the mean squared error over all the properties being measured.

The Kalman filter naturally takes into account the relative uncertainty of the measurements and the model, relying more on the measurements than the model if the measurements are believed to be more accurate. It also tracks the error in the estimates over time. Furthermore, often it can be set up to track and continually recalibrate sensor biases over time.

There are also similar techniques such as the Extended Kalman Filter (for non-linear problems), the Unscented Kalman Filter and the Particle Filter. These techniques have been used in many different applications including radar tracking (see below), inertial navigation, spacecraft and aircraft attitude determination, and numerical weather forecasting.

Banking and Finance

Bayesian techniques have wide-ranging uses within the financial sector. Bayes' Theorem is commonly exploited within hedging and within asset and portfolio management. Banks and credit card companies also use Bayesian methods to determine whether or not to grant customers credit and to what limit.

Bayes' Theorem is extremely effective within banking, where the objective is to assess risk and make forecasts, predictions and inferences based on this assessment. Risk assessment is inherently subjective and a central idea within Bayesian statistics is that of subjective probability.

Security and Fraud Detection

Bayesian methods provide a method of detecting credit card fraud. The way in which a stolen credit card is used will be different to its normal use, but spending patterns can also change for legitimate reasons. Bayes' Theorem is crucial in assessing the likelihood of fraud given the spending patterns. Bayesian methods are used to find patterns of behaviour, rather than individual anomalies that could lead to incorrect results.

AT&T developed a similar system to detect telephone fraud. Again, patterns of behaviour are sought, in this case calls to a previously

uncalled country, calls of unusually long duration, etc. The system also considers the total financial loss being incurred by a possibly fraudulent series of calls.

Medical Diagnostics

In medicine, Bayes' Theorem has assisted in diagnosis – identifying a patient's ailments as a particular disorder – and in prognosis – forecasting the natural course of disease. It is also integral to some models being used to determine optimal treatments for various disorders for individual patients.

In the early 1980s, researchers attempted to build a rule-based system for lymph-node pathology diagnosis, but the system was not reliable enough to meet the needs of pathologists. They then built another system using simple Bayesian methods. The system uses inference models and details of previous cases to create a differential diagnosis of plausible diseases based on the patient's symptoms. The user can also ask the system to identify the features that would be most useful for distinguishing between competing diagnoses, considering the costs and benefits of each observation or test. The assessments from this system were much more reliable than the rule-based method, and analyses took only one fifth of the time. The system was commercialised as Intellipath.

Spam Filtering

Bayesian systems are widely used in the continuing fight against spam, the unsolicited marketing and other junk email that deluges most email systems. Numerous Bayesian anti-spam email filters have been developed. Many are open source and can be downloaded from the internet. They make use of a simple Bayesian component that "learns" how to recognise spam and differentiate it from non-spam. This is achieved by training the software – telling it which of the emails you receive are acceptable, and which are spam. The software analyses the words in the message, and builds up a model of the indicators that a message might be spam. After training, the system can then accurately determine which messages are spam. When first introduced, these systems were capable of spotting 99% of all spam emails. Unfortunately, a combination of Bayesian and non-Bayesian methods is required for modern spam filters, since many anti-Bayesian spam generators have been developed in response to the Bayesian filters.

Other uses

Bayesian methods have been used in a diverse range of other software systems.

The US Navy have developed real-time software for determining the performance of various ship self-defence weapon systems against varying types and ranges of incoming attack weapons. Traditional techniques to solve the problem had been unsuccessful, but an approach that involved the use of Bayesian Networks led to a solution that was both effective and efficient.

NASA's Mission Control Center in Houston use Bayesian Networks to interpret live telemetry and provide advice on the likelihood of failures of the space shuttle's propulsion systems. It also considers time criticality and recommends actions of the highest expected utility.

Intel uses Bayesian Networks for diagnosis of faults in processor chips. Given end-of-line tests on semiconductor chips, a statistical process can be used to infer possible processing problems.

Nokia Networks uses the Hugin Decision Engine, a commercial tool making use of Bayesian Networks, in a prototype tool for efficient diagnosis of problems with mobile networks. By having an automated tool that reads network performance data and from that estimates and monitors network problems ranked by probability, the network operator gets an efficient troubleshooting procedure saving both expensive expert resources and downtime of the network.

Case Studies – Bayesian Statistics at Tessella

Tessella has worked on a number of projects that used Bayesian statistics. Details of some of these projects are given below.

Clinical Trials

Tessella has developed a Clinical Trial support system for a large pharmaceutical company. This customer wanted to run a Phase II trial testing 15 different doses of a particular drug in order to find the optimal one. In a conventional trial this would have meant that a large number of patients would need to be assigned to each dose in order to achieve statistically meaningful results. With the study drug being expensive to produce and the high cost of subjects in clinical trials, this would have made the study prohibitively expensive.

The customer asked Tessella to produce a system to provide centralized adaptive dose allocation for this clinical trial. The Bayesian learning and dose allocation component was an application developed by Don Berry and Peter Mueller, then of Duke University. The “prior” information and basis for the model in the system was provided by a database of several hundred patients in a past study of patient responses who were not given any treatment for their condition.

The Study Management team decided to use an adaptive randomization scheme, based on Bayesian statistics. The technique allowed the system to learn about the responses of patients at each of the study doses, discovering which doses it believed were ineffective, compared to placebo, and so allocate more patients to the effective doses as the study progressed. This method is much more ethical than a standard trial: in a standard trial patients may be assigned to a dose even after results suggest that that dose is not working; in the adaptive trial, the doses are chosen to be the best for the trial as a whole, but are penalized so that the patient is not affected badly. This has the effect of reducing the cost and the number of patients because the patients are “used” a lot better and not wasted on doses that are known to be too toxic, or not effective enough. A further benefit was that the system could provide decision making support on whether to stop the trial early, if the dose response was predicted to be below that required.

The use of this adaptive dose allocation system in running this clinical trial has been

estimated to have saved approximately 25% of operational costs of a multi-million dollar trial, compared to running the trial using a standard randomization protocol.

Another benefit of the adaptive method is that you can optimise the criteria for choosing the best parameters and best dose based on what you want at the end of the study. The output from the trial is the last set of model parameters and this final model can be used to calculate (for example) the minimum effective dose, or the optimal successful (i.e. effective but not toxic) dose. The function used for the parameter updates and dose finding can be changed so that the model “focuses” on one of these values.

Tessella has worked on a number of Bayesian clinical trial algorithms for pharmaceutical companies and the MD Anderson Cancer Center. This has involved implementing adaptive dose allocation algorithms. Tessella has worked on techniques for combination trials in which doses of two drugs are allocated dynamically and the two dimensional dose response surface is constructed without having to investigate every possible dose combination. Tessella has also worked on algorithms that choose the optimal dose by considering both therapeutic benefit and side effects. This work has involved implementing published algorithms, collaborating with leading statistical experts and developing novel techniques.

2D Gel Mapping

Tessella has worked on a project for a leading biotech company that specializes in proteomics, the comprehensive study of proteins, as a foundation for the discovery of new drugs and ways of diagnosing disease.

This company compares the proteins from different subjects, looking for those that differ between healthy and unhealthy subjects. One of the key stages in this investigation is the production of a 2D “map” of the protein peptides, so that the peptides are separated according to mass and charge by gel electrophoresis. The maps from different samples must then be accurately compared with each other. Unfortunately, the production of the maps is an inexact art, and so the correct aligning of the map images, and the correct matching of peptides from one map to another, is a complex problem, which previously had to be performed manually. This involved many man-hours of painstaking work, and so an automated approach was required.

The solution involved state-of-the-art image processing techniques, developed in collaboration with Professor Wilson of Warwick University, and new statistical methods formulated by Tessella.

The image analysis involved pre-processing of the map images, and then a multi-resolution approach, which aligns the maps at successively increasing levels of detail. Once the image maps are aligned, a statistical algorithm is used to match the peptides from one map to the corresponding peptides in the other map, using a Bayesian model which draws upon prior assumptions from theoretical modelling and an extensive database of manually matched peptides.

The accuracy of the system, and the reduction in the amount of manual work required have drastically improved the efficiency of identifying potential drug candidates. This has led to a much higher throughput of studies and, therefore, an increase in the rate at which drug candidates can be identified and subsequently brought to market.

Radar tracking

Tessella have experience in creating target tracking algorithms for a number of radar systems including SAMPSON, a phased-array Multi-Function Radar being built for the new Type 45 Destroyer by BAE Systems Insyte.

A modern radar system takes in a huge mass of data that must be processed effectively. In a fast-moving tactical situation the tracking subsystem needs to automatically detect, recognise and track planes, missiles and ships, in an environment which can also include radar returns from cliffs, rainclouds and flocks of birds. There are a number of applications of Bayesian statistics in the algorithms that Tessella have developed to solve the tracking problem:

- When a new radar signal is received it may be a target, an object that should be tracked such as a plane, or it may be “clutter”, objects that need not be tracked such as birds or reflections from the sea surface. Information about the signals received on subsequent radar scans is combined using Bayesian statistics to update the estimated probability that a target is present. When this probability rises above a threshold, a track is initiated.
- In order to accurately track targets it is important to know the likelihood of receiving radar signals that are due to

clutter. This is likely to vary as a function of range and the direction in which the radar is looking, and so the radar maintains maps of clutter density. The search area of the radar is divided into a number of cells and parameters describing the probability density function for the number of clutter detections are assigned to each cell. These parameters are updated using a Bayesian model every time the radar scans a particular cell. This is an example of an adaptive method, where the results of the last measurement are used as the prior values for the next calculation.

- As measurements of an object’s position and velocity are received they can be combined using the technique of an Extended Kalman Filter to provide an estimate of object state.
- Bayesian techniques have been used for estimating the mean radar cross section of each target based on the amount of energy received each time the target is seen. This can vary due to the inherent variability of the target (changes in radar reflectivity depending on its orientation) and due to the changing direction of the radar beam relative to the target (which means both the estimated target position and its uncertainty must be considered).

The application of these techniques to the radar tracking problem has allowed robust algorithms to be developed that make maximum use of the information available to the radar and increase tracking performance.

Other radar tracking algorithms developed by Tessella include a Bayesian technique for data fusion of Electronic Support Measure (ESM) bearing measurements and radar returns. A radar with ESM can detect platforms (ships, aeroplanes and other objects) using radar signals, and can also detect other emitters. The ESM provides an approximate bearing measurement for each of these emitters, but these measurements are far less accurate than the radar position fixes. The radar must combine the two (possibly incomplete) sets of data to determine which platforms have emitters. The algorithm uses Bayes’ Theorem to derive the posterior estimates, or new association probabilities, from the new data and the prior estimates (the association probabilities before the new data was received).

Other Case Studies

Tessella was commissioned by a consortium of water companies to implement a Bayesian asset model designed by a leading statistician. This model allowed them to manage their assets with less frequent inspections that would be required by a classical statistical method, a significant advantage given the large number of assets, and the inaccessibility of many of them. The system can also be used to assess and predict asset value for financial purposes.

Tessella worked with the financial department of a large multinational company to develop an overnight investment decision support system. They wanted a system to track performance of the various decisions and to create a decision support tool that would help the investment staff move the funds at the appropriate time to the appropriate market.

Conclusions

Bayesian statistics offers a range of advantages over conventional methods. It has a better foundation, offers greater power and flexibility and provides results in a more natural and intuitive form. Where data has a complex structure or where the parameters of an experiment are adjusted dynamically, Bayesian statistics deals naturally with the difficulties that would confound a frequentist approach. Bayesian Networks offer the ability to tackle complex decision problems, which have resisted conventional approaches. The benefits of Bayesian methods can be seen in situations where there is only “vague” prior knowledge.

Tessella has experience in developing a range of algorithms and systems that make use of Bayesian techniques, and we can choose the techniques that are most appropriate for each individual problem and find the most efficient solution.

References

Many books are available that explain the principles of Bayesian statistics and Bayesian inference. A few examples are given here:

- P. M. Lee, *Bayesian Statistics: An Introduction*, 3rd edition (Hodder Education, 2004)
- A. Gelman, *Bayesian Data Analysis* (Chapman & Hall / CRC Press, 2003)
- A. O'Hagan and J. Forster, *Kendall's Advanced Theory of Statistics*, Vol. 2B Bayesian Inference, 2nd edition (Wiley, 2004)

Many more references to Bayesian statistics can be found on the Internet, including the following:

- For a good introduction to Bayesian statistics, without too much mathematics, see “A Brief Introduction to Bayesian Statistics” by Floyd Bullard at <http://www.ncssm.edu/courses/math/Talks/PDFS/BullardNCTM2001.pdf>
- An essay on the relative merits of Bayesian and Classical statistics in High Energy Physics, by G. D'Agostini, at http://nedwww.ipac.caltech.edu/level5/March01/Dagostini/Dagostini_contents.html
- For a more detailed critique of Classical vs. Bayesian statistics in a clinical arena, see Toward Evidence-Based Medical Statistics. 1: The *P* Value Fallacy” by S. Goodman at <http://www.annals.org/cgi/content/full/130/12/995> and “Toward Evidence-Based Medical Statistics. 2: The Bayes Factor” by S. Goodman at <http://www.annals.org/cgi/content/full/130/12/1005>

Tessella plc 26 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire OX14 3YS, UK
T: +44 (0)1235 555511 | F: +44 (0)1235 553301 | E: info@tessella.com

Tessella Inc 233 Needham Street, Suite 300, Newton, MA 02464, USA
T: 1 617 454 1220 | F: 1 617 454 1001 | E: info@tessella.com

Tessella – successfully delivering IT and consulting services to world leaders in R&D, science and engineering.

For decades, Tessella has been successfully delivering IT and consulting services to world leaders in R&D, science, and engineering. Through the application of scientific methods and rigorous quality procedures, we enable clients in life sciences, energy, the public sector, and consumer industries to achieve a wide range of objectives, including, forecasting floods, developing fusion power, enhancing military sensor capability, increasing drug discovery and development efficiency, and reducing risk to health and the environment in the extraction and production of oil and gas. With offices in Europe and North America, global companies rely on Tessella for business critical assignments.

Copyright © Tessella plc 2009, all trademarks acknowledged. Issue: V2.R1.M0 | June 2009

