The Opportunities and Advantages of Using Bayesian Statistics in Health Diagnostics

Bayesian Statistics focuses on the key question: "How does this new piece of evidence change what we believe?" For instance, what do we believe the effect of this drug or treatment is? What is the probability the patient has this disease, given this new piece of diagnostic evidence?

The Bayesian approach allows all evidence to be taken into account in an explicit way, and in different ways. Different forms of evidence can be combined in the overall probability model or included via the prior belief. Analyzing the data using different priors allows the data to be interpreted from different points of view – regulatory, therapeutic and business.

The Bayesian framework more naturally allows for modelling biases, systematic error and any hierarchical structure of the data.

The result of a Bayesian analysis is a direct statements about the quantities of interest, (likely therapeutic effect, probability of toxicity etc.) providing more intuitive results and feeding naturally into a formal decision making process.

These general benefits of Bayesian Statistics lead to some specific opportunities in health care.

Diagnosis

Bayesian statistics lend themselves simply and intuitively to medical diagnosis. Following a positive diagnostics test, for HIV for example, a clinician's estimate of the likelihood that the patient actually has HIV will be influenced by their prior expectation of that outcome given the patient's lifestyle and other symptoms.

It is easy to see how this updating of the estimate of a diagnosis can be carried out over a sequence of diagnostics tests. Where the estimate after one test is used as the prior before the next and so on.

For a full example see the Merck Manual of Diagnosis and Therapy: http://www.merck.com/mrkshared/mmanual/section21/chapter295/295e.jsp

Where the diagnosis is more complicated, where the clinician has to decide between a number of different possible causes to explain a given set of symptoms, a Bayesian Belief Network can be used to capture the effect on the final diagnosis of various bits of information and results of diagnostic test. (Examples: MSN > Health and Fitness > Children's Symptoms Checker, and the Net Library at www.Norsys.com).

The Bayesian Belief Networks can be enhanced with 'utilities' and 'costs' to become Decision Networks. These can take into account the effect of data known already, the cost of gathering additional data (taking a MRI scan for instance) and the effect of knowing this additional data to recommend whether to gather it or not.

Imaging

An active area of research is the application of Bayesian inference to image recognition and analysis. Example applications are:

- Noise removal.
- Quantitative image analysis – size of tumour or infarct (stroke), for instance.
- Automated identification of brain structures in MRI scans.
- Three-dimensional density reconstruction from PET data.
- Organ / vessel tracking in two- and three-dimensional MRI data as a function of time.

Typically the advantage of applying Bayesian inference to these problems is its rigorous, formal framework for starting from prior belief and updating it based on the data gathered.

Pharmacogenetics

Pharmacogenetics is an exciting new area of research in drug development, whereby sub-groups of the population are identified as having a genetic pre-disposition to react very differently to a particular drug. For instance the gene CYP2C9 is linked to the ability to metabolize Warfarin, people with certain mutations of this gene metabolize it much more slowly and so need much lower doses of Warfarin for effective treatment and suffer toxic side effects if given the normal dosage.

In determining an association between a particular gene and a drug, DNA microarrays are used to measure gene expression and these are then related to drug exposure and other biological outcomes. This gives rise to a 'Large p, Small n' problem where there are many thousands of potential explanatory variables (the number of genes tested) and relatively few test results. This problem has recently been successfully tackled using Bayesian Statistics.

http://ftp.isds.duke.edu/WorkingPapers/00-15.html

Biomarkers

One of the biggest problems facing those developing new drugs or treatments is that the 'clinical outcomes' used for measuring the effectiveness of treatment are often very subjective (e.g. measures of pain or depression) or coarse grain (e.g. the Modified Rankin Scale) or take a long time to collect (e.g. lack of progression in Alzheimer's). The Pharmaceutical industry is keen to develop biomarkers that can act as surrogates for these clinical endpoints. Biomarkers would be objective biological measures e.g. FMRI scans to show changed mental state, brain scans to measure infarct size in Stroke or progression of Alzheimer's. These would be more reliable measures and available earlier. Problems of image analysis and of establishing the correspondence between the Biomarker and the clinical endpoint are all area where Bayesian Statistics could be applied to great advantage.

Embedding Intelligence in Devices

Bayesian inference can be used to build a rigorous decision model based on training data. This can be used for example, for distinguishing the pattern pulses in an ECG monitor that is likely to indicate to a cardiac arrhythmia. Unlike some methods, such as neural networks, the decision algorithm produced by Bayesian Statistics can be completely interpreted with respect to the problem domain – all the nodes or steps in the decision making correspond to a decision an expert would understand. This makes it much easier to validate and be confident in the algorithm.

Bayesian Belief Networks can be used to intelligently and adaptively structure the workflow in the devices interfaces, as in Ricoh Printer's Troubleshooter.

http://www.ricoh.co.jp/about/business_overview/report/23/pdf/099_105.pdf

Why Bayes, Why Now?

With all these benefits why isn't Bayesian Statistics more widely deployed? As Bayesian Statistics is the only formal, coherent calculus of statistical inference, why is there any other form of statistics? The simple answer is that most Bayesian models are not analytically evaluable, and we need to use numerical methods that require computing power that is commonplace today, but has only been readily available since the late 1980s.

Following a blossoming of fundamental research and tool building Bayesian Statistics only became mainstream in the mid 1990s. So there is not yet an established body of procedures and guidance as has been built up for conventional statistics. For instance there are still no established standards for designing, analyzing and reporting Bayesian Clinical Trials. Using Bayesian Statistics thus both permits and requires more innovation than conventional methods.

Conventional statistics work, are widely accepted and for a lot of problems have well documented approaches to tackling them. Bayesian Statistics on the other hand has a better foundation, offers greater power and flexibility and provides results in a more natural and intuitive form. We expect its use to become well-established in areas not well covered by existing conventional statistics and in the long term become the dominant method for doing statistics.

The advances in areas of complex data analysis and decision-making, offer potentially great benefits in the area of health care in particular.

Tessella's Experience

Tessella has been at the forefront of the application of Bayesian Statistical research to real commercial problems for over 10 years. We have successfully implemented Bayesian statistics and Bayesian Belief Networks in a wide range of projects.

For Pfizer, Tessella worked with leading academics from the US to enable an ambitious Bayesian model to be used to run the ASTIN Stroke trial. This allowed Pfizer to explore the dose response to 16 different doses of a drug rather than the usual 2 or 3, without requiring a larger trial than usual.

For a biotech company Tessella worked with an imaging expert to develop a Bayesian algorithm to first align and then analyze the differences in populations of 2D protein peptide maps, resulting from a 2D chromatography process. This allowed the automation of what had previously been an extremely labour intensive process.

Tessella worked with a large multinational to develop an overnight investment decision support system. The system tracked the performance of the various decisions and provided a decision support tool that would help the investment staff move the funds at the appropriate time to the appropriate market.

A significant issue for a water company is the condition of a large number of fixed assets, many of which are in remote locations, or are underground. A consortium of companies commissioned Tessella and a leading statistician to design and implement a Bayesian asset model that could help them maintain the condition of the assets using less monitoring than a conventional statistical model would have required.

Recommended reading:

- "Statistics: A Bayesian Perspective", Donald A Berry. Duxbury Press."Bayesian Data Analysis 2nd
- Ed" Gelman, Carlin, Stern & Rubin. Pub: Chapman and Hall

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