Thomas Bayes (1702-1761)

Thomas Bayes was born in 1702 in London, England. His father was one of the first six Nonconformist ministers to be ordained in England. Thomas Bayes’ parents had their son privately educated, and some believe that he was taught by de Moivre, who was doing private tuition in London during this time.

Bayes went on to be ordained a Nonconformist minister. He first assisted his father in Holborn, England. In the late 1720’s, Bayes took the position of minister at the Presbyterian Chapel in Tunbridge Wells near London. Bayes continued his work as a minister until he retired in 1752, but continued to live in Tunbridge Wells till his death on April 17, 1761.

Throughout his life, Bayes was very interested in the field of mathematics also, more specifically, the area of probability and statistics. Bayes is thought to be the first to use probability inductively. He also established a mathematical basis for probability inference. Probability inference is the means of calculating, from the frequency with which an event has occurred in prior trials, the probability that this event will occur in the future. According to this Bayesian view, all quantities are one of two kinds: known and unknown to the person making the inference. Known quantities are obviously defined by their known values. Unknown quantities are described by a joint probability distribution. Bayesian inference is seen not as a branch of statistics, but instead as a new way of looking at the complete view of statistics. Bayes’ mathematical contributions may not look very impressive, but his discovery of this original (Bayesian) approach to statistics itself is what is to be considered very important.

Bayes wrote a number of papers that discussed his work. However, the only ones known to have been published while he was still living are: Divine Providence and Government Is the Happiness of His Creatures (1731) and An Introduction to the Doctrine of Fluxions, and a Defense of the Analyst (1736). The latter paper is an attack on Bishop Berkeley for his attack on the logical foundations of Newton’s Calculus. Even though Bayes was not highly recognized for his mathematical work during his life, he was elected a Fellow of the Royal Society in 1742.

Perhaps Bayes’ most well-known paper is his Essay Towards Solving a Problem in the Doctrine of Chances (parts of which are reproduced in the Classics section of this issue of Resonance). This paper was published in the Philosophical Transactions of the Royal Society of London in 1763. This paper described Bayes’ statistical technique known as Bayesian estimation. This technique based the probability of an event that has to happen in a given circumstance on a prior estimate of its probability under these circumstances. This paper was sent to the Royal Society by Bayes’ friend Richard Price. Price had found it among Bayes’ papers after he died. Bayes’ findings were accepted by Laplace in a 1781 memoir. They were later rediscovered by Condorcet, and remained unchallenged. Debate did not arise until Boole discovered Bayes’ work. In his composition the Laws of Thought, Boole questioned the Bayesian techniques.
Boole’s questions began a controversy over Bayes’ conclusions that still continues today. In the 19th century, Laplace, Gauss, and others took a great deal of interest in this debate. However, in the early 20th century, this work was ignored or opposed by most statisticians. Outside the area of statistics, Bayes continued to have support from certain prominent figures. Both Harold Jeffreys, a physicist, and Arthur Bowley, an econometrician, continued to argue on behalf of Bayesian ideas. The efforts of these men received help from the field of statistics beginning around 1950. Many statistical researchers, such as L J Savage, Bruno de Finetti, Dennis Lindley, Morris DeGroot and D Basu, began advocating Bayesian methods as a solution for specific deficiencies in the standard system.

A specific contribution Thomas Bayes made to the fields of probability and statistics is known as Bayes theorem. It was first published in 1763, two years after his death. It states ($P$ denoting probability and ‘|’ denoting ‘given’):

$$P(H|E, C) = \frac{P(H|C)P(E|H, C)}{P(E|C)}.$$  \hspace{1cm} (1)

It uses probability theory as logic and serves as a starting point for inference problems. The left hand side of the equation is known as the posterior probability. It represents the probability of a hypothesis $H$ when given the effect of $E$ in the context of $C$. The term $P(H|C)$ is called the prior probability of $H$ given the context of $C$ by itself. The term $P(E|H, C)$ is known as the likelihood. The likelihood is the probability of $E$ assuming that $H$ and $C$ are true. Lastly, the term $1/P(E|C)$ is independent of $H$ and can be seen as a scaling constant.

Bayes theorem can be derived from the ‘product rule’ of probability which states


We also have the ‘sum rule’ given by

$$P(A|I) = P(A, B|I) + P(A, B^c|I),$$

where $B^c$ denotes the complement of $B$ (‘not $B$’), having the probability $P(B^c|I) = 1 - P(B|I)$. Combining these two we get

$$P(A|I) = P(A|B, I)P(B|I) + P(A|B^c, I)(1 - P(B|I)).$$

It follows from these that, if $P(A|B, I)$, $P(A|B^c, I)$ and $P(B|I)$ are given to us, we can obtain $P(B|A, I)$ using the inversion:

A corresponding expression can be given for the posterior probability density function when the unknown quantity (say $\theta$) is assumed to be a continuous variable. Then, $\pi(\theta)$ denoting the continuous prior density of $\theta$ and $f(x|\theta)$ denoting the likelihood function (or the model density of data $x$ given $\theta$), we have the following expression for the posterior probability density function of $\theta$ given the data $x$:

$$\pi(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int \pi(u) f(x|u) du}.$$ 

It has been realized that in many complicated problems, especially involving complex models, the Bayesian approach is the most natural one. Statistical problems involving image analysis, spatial modelling, microarrays and gene expressions are some such areas. Further, the advent of powerful computers and sophisticated statistical computing tools have made easy the difficult calculations involving posterior probabilities which had been once considered an impossible task, fit to be avoided. Some of these new computing techniques are the EM algorithm and the Markov Chain Monte Carlo (MCMC). We have already seen some articles related to these in the past issues of Resonance, a few of which are listed in the Suggested Reading. A new series article on MCMC is beginning with the present issue.

In what follows, we attempt to explain in modern language what Bayes established in his essay. Proposition 9 in his essay is the most important result as far as Bayesian inference is concerned. In this proposition Bayes considers estimating the probability $\theta$ of the Binomial($n$, $\theta$) distribution. i.e., if a coin with an unknown probability $\theta$ of coming up heads on any toss is tossed $n$ times independently and it comes up heads $x$ times, how is $\theta$ to be estimated without the help of any other extraneous information? This will be known today as the Bayes estimation of the parameter $\theta$ using the non-informative prior distribution of Uniform on the interval $(0,1)$. Let us recall Example 1 from [2]. As stated there, if we let $X$ denote the number of ‘Yes’ responses in a sample of $n$ randomly chosen individuals, then $X$ can be modelled as a binomial random variable with the probabilities

$$P(X = x \mid \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \ldots, n.$$ (2)

As explained earlier, in Bayesian inference, the parameters are also regarded as random variables, and thus the above probability is regarded as conditional probability of observing $x$ ‘Yes’ responses when the random variable $\Theta$ takes the value $\theta$. The purpose of statistics is then to invert this conditional probability so that the probability distribution for $\theta$ conditional on the observed data $(x)$ can be derived. For this, one uses the the likelihood function: the sample density rewritten as a function of $\theta$ for the observed value of sample data $x$:

$$\ell(\theta|x) = f(x|\theta),$$

to quantify the sample information. In the example above, the likelihood function of the proportion $\theta$ is simply

$$\ell(\theta|x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}.$$ (3)
We also need a (prior) probability distribution $\pi(\theta)$ on the unknown parameter. Once we have both of these, the Bayes theorem (1) can readily provide us a post-data (post-experimental or simply posterior) distribution of the unknown parameters conditional on the observed sample data. Suppose that, as assumed by Bayes, we have no special information available on the unknown proportion $\theta$ (apart from what we hope to get from sample data $x$). Then we may assume that $\theta$ is uniformly distributed on the interval $(0,1)$. i.e., the prior density is $\pi(\theta) = 1$, for $0 < \theta < 1$. In the Example, then we readily see that the posterior density of $\theta$ given $x$ is given by

$$\pi(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int \pi(u) f(x|u) \, du} = \frac{(n+1)!}{x!(n-x)!} \theta^x (1-\theta)^{n-x}, \quad 0 < \theta < 1.$$ 

Notice that this is just the density of the Beta distribution with parameters $x + 1$ and $n - x + 1$. (Note that the probability density function of the Beta distribution with parameters $\alpha$ and $\gamma$ is given by $f(y) = \Gamma(\alpha+\gamma)/\{\Gamma(\alpha)\Gamma(\gamma)\} y^{\alpha-1}(1-y)^{\gamma-1}$ for $0 < y < 1$, with $\Gamma$ denoting the gamma function: $\Gamma(a) = \int_0^\infty \exp(-u)u^{a-1} \, du$, for $a > 0$.)

Bayes’ essay is mostly concerned with the derivation of this posterior Beta distribution and subsequent calculations of posterior probabilities of the form,

$$P(a < \theta < b|x) = \frac{\int_a^b \pi(\theta|x) \, d\theta}{\int_0^1 \pi(\theta|x) \, d\theta} = \frac{\int_a^b \theta^x (1-\theta)^{n-x} \, d\theta}{\int_0^1 \theta^x (1-\theta)^{n-x} \, d\theta}.$$

Since these probability calculations involve incomplete Beta integrals, he provides approximations using geometric arguments.

**Suggested Reading**


Mohan Delampady
Statistics and Mathematics Unit
Indian Statistical Institute, RVCE Post, Bangalore 560 059, India.