Dpto. de Logica y Filosofia de la Ciencia, Valencia 2007
LECTURE 1 "INTRODUCTION TO BAYESIAN INFERENCE"

Jan-Willem Romeijn, University of Groningen
Email: j.w.romeijn@rug.nl
Web: http://www.philos.rug.nl/~romeyn

1. The dual nature of probability

The first mention of probability in its modern meaning is found in a correspondence between Pascal and Fermat concerning a game of chance. This discussion concerns the estimation of the probability of events given a fixed chance setup. The probabilities are thus assigned to events in the world.

After being converted to the Catholic sect of Jansenists, Pascal devised an influential argument for believing in God, known as Pascal's wager. Among other things the wager involves the probability that God exists. The thing to note is that in this case, probability is assigned to a belief. It is epistemic.

It is this distinction between two kinds of probability that is at the heart of Bayesian inference, as it was first conceived by Thomas Bayes. Statistical inference, as it will be discussed in the next lecture, also drives on this distinction: it is about aligning epistemic probability with the probabilities 'out there'.

2. Physical probability

Fairly quickly the notion of probability was applied to other physical events than just the rolling of dice. An important early application is the calculation of mortality rates with the aim of valuating insurance contracts. Probability, it turned out, can be used for all sorts of mass phenomena.

There is a more or less continuous line from these first applications to modern times. A number of highlights over the centuries:

- Jakob Bernoulli and the law of large numbers.
- Poisson, Gauss and Daniel Bernoulli: the start of error statistics.
- Maxwell, Quetelet and the notion of a 'mean man'.
- Galton, Pearson, and Fisher: statistical methodology and the advance of science.
- Boltzmann, Einstein and quantum mechanics: using statistics at the heart of physics.

3. Epistemic probability

The basic idea of epistemic interpretations of probability is that probability is an expression of uncertain opinion. We may distinguish two main developments.

The analyses of bets by Pascal and Huygens marked the start of decision theory. An important advance is the use of a notion of utility, in Daniel Bernoulli's solution to the St Petersburg paradox. In the work of Ramsey, Savage, and Jeffrey, this idea is developed further. Savage axiomatised probability in tandem with utility, with the motivation that on itself probability does not have any empirical import.

Leibniz picked up on the revolution of the probabilists, but used it for his own main interest, legal reasoning. Unfortunately the use of probability in court remains controversial up to the present age. But the use of probability in a theory of sound reasoning with uncertainty has only gained popularity. In the following I want to concentrate on the simplest and, to my mind, the most promising of probabilistic logics: Bayesian logic.

4. Axiomatisation of probability

Before doing so, it is useful to specify the modern axiomatisation that is due to Kolmogorov. In his treatment, probability is a measure of sets A, B, etc.

- $p(\mathrm{A}) \geq 0$
- $p(\Omega) = 1$
- $p(\mathrm{A} \cup \mathrm{B}) = p(\mathrm{A}) + p(\mathrm{B})$ if $\mathrm{A} \cap \mathrm{B} = \emptyset$.

We can conveniently represent the sets by means of Venn diagrams. Their areas are a natural measure, and thus represent the probabilities assigned to the sets. The areas nicely illustrate Kolmogorov's axioms.

A collection of sets forms a so-called algebra if it is closed under a number of set theoretical operations. For Kolmogorov a probability measure is defined on such an algebra. For the applications we will consider, it is useful to think of sets as collections of possible worlds. Each set is characterised by a proposition that is true in exactly those possible worlds belonging to the set.

In this way we can associate sets with propositions, and thus assign probabilities to propositions. The beauty of Kolmogorov's axiomatisation is that it is just a formal system. It does not suggest anything towards an interpretation of the probability measure.

5. Bayesian logic

In the set-theoretical formulation of Kolmogorov it is rather easy to derive the theorem that Thomas Bayes painstakingly derived some 250 years ago: $p(\mathrm{A} \mid \mathrm{B}) = p(\mathrm{A})\, p(\mathrm{B} \mid \mathrm{A}) / p(\mathrm{B})$. This theorem is the centerpiece of Bayesian inference.

Following the work of De Finetti, Howson, Fitelson, we can view the theory of probability itself as a deductive logic. There is a striking similarity between probability as a function over an algebra, and truth values as a function over a language: probabilities can be viewed as generalised truth valuations.

In this view Kolmogorov's axioms determine what probability assignments are consistent. Usually this idea is defended by means of so-called Dutch book arguments, betting setups in which the gambler who does not comply with the axioms is guaranteed to loose money. Inferences that obey this consistency criterium on subjective probabilities is often called Bayesian.

Bayesian inference, understood in this way, dictates the probability values that must be assigned to specific propositions on the basis of certain values for other propositions. There are numerous philosophical applications of this idea, for example the Monty Hall dilemma.

6. An example: Monty Hall

There is a car behind door $i = 1, 2, 3$, denoted A_i. We choose door 1, denoted K_1. Now say that Monty opens door 2, denoted M_2. According to Bayes' formula, the probability that the car is behind door i is:

$$p(A_i \mid M_2 \ \& \ K_1) = p(A_i \mid K_1) \times \frac{p(M_2 \mid A_i \ \& \ K_1)}{p(M_2 \mid K_1)}.$$

where the expression in the denominator is

$$p(M_2 \mid K_1) = \sum_i p(A_i \mid K_1)\, p(M_2 \mid A_i \ \& \ K_1).$$

So we can derive everything from the so-called prior probabilities $p(A_i \mid K_1)$ and the so-called likelihoods, $p(M_2 \mid A_i \ \& \ K_1)$. The priors are $p(A_i \mid K_1) = ⅓$ for each i. We further know that $p(M_2 \mid A_2 \ \& \ K_1) = 0$, because Monty will not open the door with the car, in this case door 2. We also know that $p(M_2 \mid A_3 \ \& \ K_1) = 1$, Monty opens a door you did not choose, in this case M_2 or M_3. Finally, we assume $p(M_2 \mid A_1 \ \& \ K_1) = ½$.. If the car is behind door 1, Monty can choose to open either door 2 or door 3, and in those cases he chooses randomly. We can now derive that $p(M_2 \mid K_1) = ⅓×½ + ⅓×0 + ⅓×1 = ½$, and filling this in we also have $p(A_3 \mid M_2 \ \& \ K_1) = ⅓ × 1 / ½ = ⅔$, and $p(A_1 \mid M_2 \ \& \ K_1) = ⅓ × ½ / ½ = ⅓$.

7. Induction

The interest of the next lecture is in the application of logical probabilistic inference to induction. Induction is supposed to bring us from data to general conclusions on the populations from which the data is obtained. Statistics can be seen as an attempt to provide a probabilistic warrant for such inferences.

The warrant of classical statistics is based on the law of large numbers, and typically associated with the exclusive use of physical probability. Against this, Bayesian statistics makes explicit use of both epistemic and physical probability.