## PROBABILITY AND INFERENCE CONCEPTS

A basic familiarity with the calculus of mathematical probability is assumed for this development of statistical modelling and inference. From a purely mathematical perspective, the probability calculus can be studied and developed with scant regard to the notion of just what probability means. A comprehensive development of probabilistic modelling for all problems of measuring and manipulating uncertainties requires, however, an appreciation of the existing definitions and interpretations of probability, and of their appropriateness and limitations in given contexts.

What is probability? Common answers include allusions to chances or likelihoods of events occurring, or, even vaguer, to propensities of occurrence. At one level, this is simply word substitution, but at another it is a reflection of a personal appreciation based on experience with uncertain events, and the vagueness or imprecision is an inherent feature of uncertainty assessment. One event may be judged more likely or more probable than another, having a greater chance of occurring or propensity to occur, but quantifying the difference and assigning numerical measures to the uncertainties, is somewhat unnatural. And so formal interpretation and, most important, operational definitions are required.

A probability is a numerical measurement of uncertainty. An event that may occur, or may have occurred, is uncertain to a degree dependent on the known circumstances of the event and on the individual concerned with making assessments of the uncertainty. The familiar interpretations and definitions of probability are those based on counting and relative frequency concepts. In these, the circumstances are clearly defined and articulated, as are the bases on which individuals are supposed to judge the events.

To begin with counting concepts, we have statistical sampling as the archetype paradigm for the development of probabilities. From a collection of $n$ individuals, one is selected; if the individuals are judged equally likely to be selected, the probability, or chance, or any specific individual is $1/n$. There is a still a degree of vagueness here; what does "equally likely" mean? The connotations are of fairness and randomness - the flips of fair coins, rolls of unbiased die, random numbers provided by statisticians, and so forth. Ultimately, however, the implied assumption must be deemed acceptable to the concerned individual before the $1/n$ probabilities, and resulting implications and consequences, follow.

The frequency concept of probability, critical in statistical theories, is similarly vaguely defined, though with strong appeal to what, in principle, are empirically verifiable circumstances. An event is embedded in a (conceptually) infinite sequence of similar events, and the relative frequency of past occurrence in the sequence is taken as defining a probability for the current case. If it has rained on about 32% of past days that are accepted as suitably similar to today in terms of climatic conditions, then a chance of around 0.32 is deemed appropriate, by the frequency definition, for the probability of rain today. If an industrial production process has produced less than 0.1% defective items in past operation then, under similar operating conditions the chance of future defectives should be around 0.1%. Key weaknesses of this definition relate to the difficulties, if not impossibilities, in determining appropriate embedding sequences in many problems, and the associated vagueness of classification of "similarity" of the circumstances of the current occasion to those past. But the definition is widely assumed in developing probability models. Certain applications involve natural embedding sequences of similar events and thus provide empirical foundation for extrapolating relative frequencies (the chance of the sun rising tomorrow). Still, the definition has strong subjective elements, requiring observers to make or accept the vague notion of appropriate similarity.

Counting and frequency concepts are fundamental features of basic probability theory. A wider definition is required, however, in order to extend the domain of probability to problems with no logical counting or frequency structure. To attach numerical measures of uncertainty to unique events, for instance, requires notions of chance not covered by these two concepts. Definitions of subjective probability attempt to cover such cases, and all other problems of quantifying uncertainty. The basic notion here is that probability is a numerical measure of belief of an individual, the observer, in the occurrence or non-occurrence of events. An individual who strongly believes it will rain today will have a high probability for that event. Identifying, or measuring such beliefs in numerical terms requires and operational definition, and there are several related approaches, the most easily understood and acceptable related to assessment of uncertainties by comparison with events of known and agreed chances. Thus an individual judging an event to be more likely than a 6 on the roll of a single die, but less likely than an even outcome on the roll, should be comfortable with a probability lying between 1/6 and 1/2 for that event. Further reflection and comparison with more refined, "standard" events might narrow down the assessment to a small enough range to adopt a particular value. Obviously, such definitions are operational only up to a margin; extremely precise assessments are impossible, but then, very high precision is not necessary in many problems.

There are other assessment methods, particularly some involving gambling issues (we might talk about this sometime later on). For now, two key points are evident. Firstly, probability as degree of belief is clearly subjective, individual and information dependent. Secondly, the "objective" counting and frequency concepts are subsumed; in assessing beliefs about an event with counting or frequency context, an individual should take such contexts into account in forming his or her probability judgements. Hence the statement by Bruno de Finetti that *Probability does not exist* (generally) in any absolute or "objective" sense.

## CONDITIONAL PROBABILITY

All probabilities are conditional. The probability you assign to an event depends (i.e., is conditional upon) the information and knowledge you bring to bear in assessing the probability. This is most clearly evident in connection with statistical inference where we have a fundamental focus on *how probabilities change* as new data or information is analysed. The simple diagnosis example (over the page) makes this clear. Let's take a few details from that example:

> *An individual ("you") presents for testing for the presence or absence of a rare disease. You are assumed to be drawn from the general population and it is know that the disease prevalence rate is 1% – that is, 1% of people carry the disease. On the basis of only this information, my probability that you are diseased is 0.01. That is a probability conditional on information $H$ – where $H$ stands for my background knowledge and the assumption that you are a representative/random draw from the population. Now you get a screening test; the test turns out negative, contra-indicating disease. Based on the known accuracies of the screening test – which is not perfect – this data revises the probability of 0.01 to 0.0001. – see the diagnosis example below and class discussion for the How? and Why? behind the revised number. Call the data $D$, so that $D = \{``Test\ Negative''\}$ is added to $H$ to get the revised probability: i.e.,*
>
> $$Pr(\text{Disease}|H) = 0.01 \qquad but \qquad \Pr(\text{Disease}|H, D) = 0.0001.$$
>
> *So here we see clearly that there is no such thing as THE probability of you being diseased – it depends on conditioning information and data, as it does on who does the analysis.*

The concepts and mechanisms for revising probabilistic representations of information and knowledge – illustrated in this simple (the simplest?) example – form the corner-stone of scientific inference in the face of uncertainty.

**INTRODUCTORY INFERENCE CONCEPTS: A simple diagnosis example**

*(This is closely related to the genetics example in GCSR Section 1.4)*

Individuals tested for presence of a rare disease are assumed to be drawn from a population with about 1% diseased. Absent other information, a tested individual then has initial, or prior, chance 0.01 of being diseased. The screening test is 99% sensitive, correctly diagnosing a diseased individual with probability 0.99; it is only 90% specific, however, correctly diagnosing a non-diseased individual with chance 0.9. These test characteristics may be based on trials with patients of known condition. Introduce binary indicators $\theta = 1(0)$ to denote presence (absence) of the disease for the individual, and $x = 1(0)$ to denote a positive (negative) test result indicative of disease. The information defines probabilities $p(\theta)$ and $p(x|\theta)$ for all values of $\theta$ and $x$, and hence the joint distribution through chances $p(x, \theta) = p(x|\theta)p(\theta)$. We can answer the following questions.

First, how many patients can we expect to test positive? Those testing positive are destined for further tests and possibly treatment. Decisions about allocation of resources for patient care will be based on projected numbers in this group, so health care administrators need to ask such questions. We can deduce the marginal probability $p(x = 1) = 0.1089$, or about 11%. Most of these will healthy since by far the majority of all tested individuals are healthy. Secondly, if a patient tests positive, what are the chances he or she is diseased? By Bayes' theorem, $p(\theta = 1|x = 1) = 0.091$. The odds on disease given a positive test are $o(\theta = 1|x = 1) = 0.1$, which can be obtained from $o(\theta = 1|x = 1) = o(\theta = 1)r$ where $r = p(x = 1|\theta = 1)/p(x = 1|\theta = 0)$. So the final or posterior odds $o(\theta = 1|x)$ are obtained from the prior odds $o(\theta = 1)$ through multiplication by $r = p(x|\theta = 1)/p(x|\theta = 0)$; $r$ depends only on the test characteristics, and is called the *likelihood ratio* for $\theta = 1$ versus $\theta = 0$ based on outcome $x$. For any known outcome $x = 1(0)$, $p(x|\theta)$ may be viewed as a function of the unknown $\theta = 1(0)$; it is called the *likelihood function* for $\theta$ given the observed value of $x$. One way to assess the impact that data $x$ has on uncertainty about $\theta$ is to quote the likelihood ratio, equivalently the ratio of posterior to prior odds; based on $x = 1$, $r = 9.9$, so that the posterior odds on $\theta = 1$ are 9.9 times the prior odds, whatever the prior odds may be.

The focus on updating of prior $p(\theta)$ to posterior $p(\theta|x)$ is a central inferential concept. Via Bayes' theorem

$$p(\theta|x) \propto p(\theta)p(x|\theta)$$

where the constant of proportionality can be identified to ensure that $p(\theta|x)$ is a density function. The problem of inference is one of updating probabilistic descriptions of uncertainties by conditioning on further information. The prior $p(\theta)$ is of course a conditional distribution, conditional on all previous data and information used in its assessment, and all other assumptions. This is not made explicit in notation though formally we might write $p(\theta|H)$ where $H$ denotes all such prior information. This information is also relevant to the test description, so that the chances of test outcomes might formally be written $p(x|\theta, H)$. Then Bayes' theorem is, more formally,

$$p(\theta|x, H) \propto p(\theta|H)p(x|\theta, H).$$

Typically, $H$ will be suppressed in notation, though it is important to identify in conditioning statements all quantities that are assumed to define the corresponding distributions if they may be subject to change in the course of analysis or if they are themselves to be viewed as uncertain at some point.