$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

$$P(E_1 \mid E_2) = \frac{P(E_1)P(E_2 \mid E_1)}{P(E_2)}$$

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{1/8}{3/8} = \frac{1}{3}$$

Bayesian, Likelihood, and Frequentist Approaches to Statistics

A comparison of methods

Stephen Senn

The Italian mathematician, actuary, and Bayesian, Bruno de Finetti (1906–1985), once estimated that it would take until the year 2020 for the Bayesian view of statistics to completely prevail.[1] Whether or not his prediction comes true, there is no question that Bayesian statistics is gaining ground. In drug regulation, however, the alternative frequentist view continues to dominate, although there are areas (for example the regulation of medical devices) where the Bayesian approach is being applied. Many readers of *Applied Clinical Trials* will have heard of Bayesian statistics, and some will have wondered what it is. If de Finetti is right, those who have not wondered yet will have to reason to do so in future.

Statistics versus probability

Before explaining the difference between Bayesian and frequentist statistics (and a third alternative, the likelihood approach, which has some features of both), it is useful to draw another distinction—probabilists and statisticians. Probabilists are mathematicians and, like others of that breed, are involved in a formal game. The game they play is subjunctive, which is to say that it is a matter of if and then. If such and such are true then such and such follow. "If the die is fair, then there is one chance in six that I will roll a one," is a trivial example of the sort of question probabilists deal in. However, if you ask the probabilist, "Is the die fair?" then you will probably receive the reply, "That's not my department." Enquiry as to whose department it is leads to the statistician. The statistician cannot restrict life to subjunctive matters. Statisticians deal not just with if and then, but also with whether and what. Whether the die is fair or not, and, if not, what exactly the bias is, are the sorts

Dawn of a Bayesian era? Although other statistical methods are more popular, the Bayesian method is quickly gaining ground.

of questions that statisticians are supposed to try and answer and their answer is supposed to rely on data.

In my book, *Dicing with Death*, I have described this difference between probability theory and statistics as the difference between the divine and the human.[2] Probability theory is a divine theory because it works from known initial conditions to consequences using universal laws. These initial conditions are declared by the probabilist in a fiat to begin with—"let there be theta." (Theta is a popular choice of symbol to represent a probability.) Thus the probabilist acts as creator of his or her own universe.

Statistics on the other hand is a human science. The state of nature has been declared and given, but we don't know what it is. All we can do is observe the consequences and try and divine the rules that govern our existence.

The distinction between probability and statistics is also sometimes made in terms of direct and inverse probability. A question in direct probability might be, "In 100 tosses of a fair coin, what is the probability of having exactly 40 heads?" A question in inverse probability might be, "In 100 tosses of a coin, 40 showed heads. What is the probability that the coin is fair?" The former question is thus the province of probability theory and the latter the province of statistics. The second sort of question is much harder to answer than the first. In fact, it is so hard that mathematicians, scientist, philosophers, and, of course, probabilists and statisticians can't agree how it should be answered. The difficulty can be illustrated with a simple example.

An example
Suppose that I have two urns, urn A and urn B, each containing four balls. Urn A contains three white balls and one black ball, and urn B contains two black balls and two white balls. I am informed that an urn has been chosen and that a ball has then been drawn at random from the urn chosen. The ball is black. What is the probability that it came from urn A?

One simple answer might be as follows. Before the ball was chosen the urns contained three black balls between them: one ball in urn A and two balls in urn B. If any ball in either of the urns is equally likely to be chosen, it is twice as likely that the black ball chosen was from urn B as that it came from urn A. Hence, the probability that it came from A is 1/3.

This answer can be formally justified by a theorem in probability, Bayes theorem, named after Thomas Bayes (1701–1761), an English Presbyterian minister, and which was communicated posthumously by his friend Richard Price and read to the Royal Society of London in 1763. In words, Bayes theorem states that the probability of an event E_1 given another event E_2 is the joint probability of both events divided by the probability of event E_2. In symbols we would write this as Equation 1:

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

Here P( ) means "probability of," | means "given," and ∩ means "and." Because the probability of the joint event $E_1 \cap E_2$ is $P(E_1 \cap E_2) = P(E_1)P(E_2 \mid E_1)$ (which, expressed in words means that the probability of "E_1 and E_2 "is the probability of E_1 multiplied by the probability of E_2 given E_1), an alternative representation of (Equation 1) is Equation 2:

$$P(E_1 \mid E_2) = \frac{P(E_1)P(E_2 \mid E_1)}{P(E_2)}$$

Suppose, in our example, that event E_1 is "choose urn A" and E_2 is "choose black ball." Then if each urn is equally likely a priori to have been chosen we have $P(E_1)=1/2$. Furthermore, if each urn is equally likely to be chosen, because both urns contain the same number of balls, each ball is equally likely to be chosen.

Out of the eight balls in total, one is the black ball in urn A so that the probability of "urn A and black ball" is $P(E_1 \cap E_2)=1/8$. On the other hand, three out of eight balls in total are black. Hence we have $P(E_2)=3/8$. Now applying Bayes theorem we can substitute these values in the right hand side of the equation given by (Equation 1) to obtain

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$$

which is the answer we had before.

Some difficulties
This is all very well and may even seem trivial, but there is a difficulty with this answer. In formulating the question in the first place I did not say that the decision from which urn to withdraw a ball was made at random, with each urn being given an equal chance of being chosen. I did specify that the ball was chosen from the urn at random. The net result of this is that although some of the probabilities for this problem are well defined, for example the probability of choosing a black ball if urn A was chosen in the first place, one important probability is not, that of choosing urn A.

It is the case that many of the problems we encounter in science have probability elements that can be divided into two sorts. One sort can be fairly well defined. We assume that a given theory is true and then calculate the probability of the consequences. For example, we might assume that the probability, Θ, that a patient will be cured if given a particular drug is 0.3. We can then calculate very precisely, for example, given this assumed value what the probability is that exactly 40 patients in a sample of 100 will be cured. In fact, given that we have a sample of 100 patients 40 of whom have been cured we can calculate the probability of this event as a function of the probability Θ substituting all sorts of values, not just 0.3. This type of probability, where the event is fixed and the hypothesis changes, were called likelihoods by the great statistician, geneticist and evolutionary biologist R.A. Fisher (1890–1962) and play a central part in statistical inference. Suppose in our urn-sampling problem that we had drawn a white ball. The probability of sampling a white ball is 3/4 if A is chosen and 1/2 if B is chosen. These are the so-called likeli-

> Many of the problems we encounter in science have probability elements that can be divided into two sorts—one sort well defined, the other sort not well defined.

hoods. Note that they do not add up to one and there is no general requirement for likelihoods, unlike conventional probabilities, to do so. This is because the event (white ball) is fixed and the hypothesis (urn A or urn B) is allowed to change. For conventional probability we fix the hypothesis (for example urn A) and vary the outcome (black or white ball).

The second kind of probability element is not well defined. This is the probability of a given hypothesis being true in the first place. For example the hypothesis, in advance of running a trial in 100 persons, as to the probability of a cure. It turns out, however, that to issue inverse probability statements, it is necessary to assume such prior probabilities. Because we lack an objective basis for them, this can only be done subjectively. In attempting to solve the urns and ball problems you have to assume a prior probability that urn A was chosen, even though this was not specified in the problem.

This brings us to the heart of the problem. In order to use Bayes theorem to allow us to say something about the probability of a scientific hypothesis H being true given some evidence e, we would have to use (Equation 2) to write something like Equation 3:

$$P(H|e) \quad \frac{P(H)P(e|H)}{P(e)}$$

Here $P(H|e)$ is sometimes referred to as the posterior probability of the hypothesis—the probability after seeing the evidence. The difficulty is that of the three terms on the right-hand side of (Equation 3), we can usually only find objective values for $P(e|H)$, the probability of the evidence given the hypothesis. However, the prior probability of the hypothesis, $P(H)$ is needed for the solution, as is the probability of the evidence $P(e)$. The latter is particularly awkward to obtain, because many different hypotheses would give rise to e (albeit with differing probabilities). Thus you need to know the prior probability of every single such hypothesis in order to calculate it.

Odds, Bayes, and likelihood
However, by reformulating our objectives slightly, the difficulty of having to estimate $P(e)$ can be finessed. Suppose that we wish to compare the posterior probabilities of two hypotheses, H_A and H_B in

terms of their ratios, or odds. We can use (Equation 3) to write

$$P(H_A|e) \quad \frac{P(H_A)P(e|H_A)}{P(e)}$$

and also

$$P(H_B|e) \quad \frac{P(H_B)P(e|H_B)}{P(e)}$$

The ratio of these two expressions gives us what we require and, fortunately, the awkward term $P(e)$ cancels out so that we are left with Equation 4:

$$\frac{P(H_A|e)}{P(H_B|e)} \quad \frac{P(H_A)P(e\mid H_A)}{P(H_B)P(e\mid H_B)} \quad \left\{\frac{P(H_A)}{P(H_B)}\right\}\left\{\frac{P(e\mid H_A)}{P(e\mid H_B)}\right\}$$

This is the odds ratio form of Bayes theorem, promoted by the British mathematician and statistician George Barnard (1915–2002). It states that the posterior odds of one hypothesis compared to another is the product of the prior odds (the first of the two terms in curly brackets) and the ratio of likelihoods (the second of the two terms in curly brackets). This still leaves us with the problem of estimating the prior odds. There are three common "solutions."

The first is the Bayesian one of stating that there is nothing inherently problematic about subjective probabilities, because probabilities anyway are nothing more or less than a statement of belief. The difficulty with using Bayes theorem only arises because of the myth of objective probabilities, which is part of the myth of objective knowledge. Indeed, de Finetti himself referred contemptuously to the, "inveterate tendency of savages to objectivize and mythologize everything."[1] What the Bayesian says is, "abandon your pretensions of objectivity, embrace subjectivity, and recognize that you need to include personal belief as part of the solution of any problem." Thus introspection is the key to the solution. It is personal belief that provides the final (otherwise missing) ingredient to the calculation of posterior probabilities.

The second solution is to go halfway

only. This is to say that of the two terms on the right-hand side of (Equation 4), one (the ratio of likelihoods) is well defined and may attract a fair degree of assent as to its value but the other (the prior odds) is not. For example, in my urn and ball problem, because I did not define the mechanism by which the urns were chosen then

$$\left\{\frac{P(H_A)}{P(H_B)}\right\}$$

is completely speculative and not worth including in the problem. However, the second term,

$$\left\{\frac{P(e|H_A)}{P(e|H_B)}\right\}$$

is defined by the problem. Indeed, it is equal to $(1/4)/(2/4)=1/2$. The ratio of likelihoods is thus one to two comparing urn A to urn B or two to one in favor of urn B. This quantity is then perfectly objective. The Bayesian will counter that this may well be so but it still fails to capture an important element of the problem, namely the prior odds. Furthermore, it turns out that for more complex cases it is not always possible to calculate such simple ratios of likelihoods.

The frequentist approach
The third solution is the frequentist one. This is to abandon all pretence of saying anything about hypotheses at all. Effec-

<div style="text-align:center; color:#2a6ebb;">
The difficulty with using Bayes theorem only arises because of the myth of objective probabilities, which is part of the myth of objective knowledge.
</div>

tively, inverse probability is rejected altogether and one tries to work with direct probabilities only. For example one could adopt the following rule of behavior. If a black ball is chosen I shall act as if it came from urn B. If a white ball is chosen I shall act as if it came from urn A. We can then calculate the probabilities of making two types of error. If urn A was the urn from which the ball is chosen, then there is a one in four chance of choosing a black ball. Thus there is a one in four chance of being wrong. On the other hand, if urn B is the urn from which the ball is chosen,

then there are two chances out of four of choosing a white ball; thus there is a 50% chance of being wrong. This is referred to a "hypothesis testing" and is an approach that was developed at University College London in the later 1920s and early 1930s by the Polish mathematician Jerzy Neyman (1894–1981) and the British statistician Egon Pearson(1895–1980). Neyman later emigrated to the United States and founded an extremely influential and vigorous school of statistics at Berkeley. Note, however, that these error rates are subjunctive. The probability statements are of the "if/then" form. They do not correspond to probabilities that the hypotheses are true and, indeed, Neyman would

deny that any such statement has meaning. A hypothesis either is or is not true and hence does not have a probability of being true.

The Neyman-Pearson system is the one that appears to be the one used in drug regulation. We refer to type I error rates, to null and alternative hypotheses, to power of tests, and so forth. All of these are concepts that play an important part in that system. Nevertheless, the way in which the system is applied in practice reflects elements of a slightly older and similar system, much developed by RA Fisher. A problem in applying the Neyman-Pearson system in practice is that often the probability of the evidence is often only well defined under a so-called null hypothesis. In a controlled clinical trial such a hypothesis might be, "there is no difference between the treatments." Given such a null hypothesis the probability of observing a result as extreme or more extreme than the observed difference between treatments, the so-called p-value, may be calculated. This may be

compared to a standard level of significance, for example 5%. If the p-value is less than this standard, the hypothesis is considered rejected. Such a procedure can be employed to guarantee a given type I error rate, as in the Neyman-Pearson system, but does not employ any specific reference to alternative hypotheses, which can be difficult to characterize. For example, the logical alternative to, "the treatments are the same," is the, "treatments are different," but because there are infinitely many ways in which treatments can differ, this does not yield a unique way of calculating probabilities.

It would be simplistic to conclude that the difference between frequentist and Bayesian statistics is that the former is objective and the latter subjective. It would be more accurate to say that the former is subjunctive (if the null hypothesis is true I shall make an error with this probability) and the latter is subjective (my personal

belief of the truth of this statement is such and such). Bayesians would claim that frequentist methods give an illusion of objectivity. Frequentists deny any place for subjective probability, but, in fact, the interpretations that result from any application of frequentist statistics depend very much on personal actions. For example, the decision to inspect a trial during its running with the possibility of stopping the trial early may impact on the reported p-value. Thus, it is not only the data that affect the conclusion but the trialist's intentions also. Such behavior has no direct impact on Bayesian calculations, which are not affected by the number of times one looks at a trial and so from this point of view can claim to be more objective.

Where does this leave us?
In my view, it is too early to say. It may be that de Finetti's prediction will come true and we shall move to a consensus that Bayesian methods are those we should use. Perhaps, on the other hand, drug regulation will continue much as before. Per-

sonally, I like the advice of George Barnard. Starting with a key paper in 1949, Barnard produced many trenchant criticisms of the then dominant frequentist school but never accepted that Bayesian methods alone would be sufficient for the applied statistician.[3] Towards the end of his life he suggested that every statistician ought to have basic familiarity with the four major systems of inference—de Finetti's fully subjective Bayesian approach, a less extreme version pioneered by the British geophysicist Harold Jeffreys (1891–1989) (which has not been discussed here), the Neyman-Pearson system and Fisher's mix of significance tests and likelihood.[4]

Of course this can be regarded as a rather unsatisfactory situation. We have to have four systems rather than one. Is statistics not complicated enough as it is? As already explained, however, statistics is a human subject not a divine one. The difficulties it attempts to overcome are genuine and our human powers are limited. In making sense of clinical trials we have no choice but to count and measure; and to make sense of our counting and measuring we have no choice but to use statistics. As in other areas of human struggle, pragmatic compromise, although not perfect, may avoid the disasters to which fanatic single-mindedness can tend.

References
1. B.D. de Finetti, Theory of Probability, Volume 1. (Wiley, Chichester, 1974).
2. S.J. Senn, Dicing with Death. (Cambridge University Press, Cambridge, 2003).
3. G.A. Barnard, "Statistical Inference (with discussion)," Journal of the Royal Statistical Society, Series B 11, 115–149 (1949).
4. G.A. Barnard, "Fragments of a statistical autobiography," Student 1, 257–268 (1996).

Stephen Senn, PhD, CStat, is Professor of Pharmaceutical and Health Statistics at University College London, Department of Statistical Science, 1-19 Torrington Place, London WC1E 6BT, UK, +44 20 7679 1698, fax +44 87 0052 3357, email: stephen@senns.demon.co.uk. He is a member of the Applied Clinical Trials editorial board. His book, Dicing with Death (2003), a popular account of medical statistics, is published by Cambridge University Press.

> Barnard produced many trenchant criticisms of the frequentist school, but never accepted that Bayesian methods alone would be sufficient for the applied statistician.