Introduction and background

Scientific method requires classification and quantification. If scientific methods are to be applied to the interpretation of religious texts – or any texts for that matter – it is necessary that there should be a link between meaning and what is measured. We speak of scientific method because we wish to emphasise that inferential procedures are carried out according to rules of rationality. This is the great strength of the scientific approach to religious texts. It ensures that different individuals, proceeding from the same pre-suppositions, will be led to the same conclusions. This stands in marked contrast to some of the methods employed by traditional biblical scholars. It is true that there are many situations where it can be truthfully asserted that biblical scholars have reached a consensus. Indeed in some situations the evidence is so overwhelming that one does not need to be too explicit about how the conclusions were reached. Unfortunately, it is also true that there are wide divergences among scholars on such matters as the historical character and the doctrinal implications of, for example, such matters as the reported resurrection of Jesus. Scientific method is concerned not only with the numbers themselves but what is done with them. It is as important to process the numbers correctly as to create them in the first place. For this reason we need to pay particular attention to inference as well as description.

We shall identify two distinct areas where quantitative methods have been applied to biblical texts. The first is associated with what is often called stylometrics, which aims to learn something about the structure or authorship of texts by analysing their statistical characteristics. The second is more fundamental in the sense that it is concerned with the interpretation, in particular with the historical accuracy, of biblical texts. It is in this context that Bayes’ theorem is often invoked, and we shall make a major digression, at the appropriate point, to explain what the theorem can and cannot do.

Stylometrics

It may be far from obvious how one might learn anything useful from the statistical analysis of the numerical properties of words and sentences, say. However, a very simple example will illustrate
what may be possible. The London *Times* normally prints three leading articles each day on topics of current interest. The first leader, on 15 April 2011, was about immigration and consisted of 27 sentences ranging in length from four to 41 words. The second leader, about the stock market launch of a large corporation, contained 37 sentences ranging from five to 38 words per sentence. At first sight, the two appear rather similar until we note that there were two sentences of length 38 words in the second case and the next longest sentence had only 28 words. This difference is brought out very clearly if we look at the averages. In the first leader the average sentence length is 23.15 words and in the second it is 16.30. The difference is quite striking and immediately raises the question of why this should be so. The type of subject matter was not greatly different, so the most obvious explanation is that they were written by different people, the latter of whom preferred shorter sentences. There could, of course, be other reasons but the only point we wish to make here is that the quantitative evidence of the word count requires explanation. If that explanation lies, for example, in different authorship then we would have established a link between a statistical characteristic of the text and a matter of potential substantive interest.

Scientific study, as we have already noted, is concerned not only with quantification but with what is done with the numbers at which we arrive. It is as important to process the numbers correctly as to create them in the first place. Thus we are concerned with inference. In this section we shall give two examples of how the quantification of written text poses questions about, and suggests answer to, questions of substantive interest. The first relates to the authorship of religious texts; the second to their mode of composition.

The earliest major account of the statistical approach to literary studies was by Yule (1944), but this was not immediately followed up. Another pioneering study which attracted much statistical attention was that of Mosteller and Wallace on the Federalist Papers (Mosteller and Wallace 1964). Their object was to identify the authorship of several papers whose authorship had previously been disputed. At about the same time Morton (1965) (see also Morton 1978) used a stylometric approach to the New Testament epistles attributed to St Paul; he also went on to consider the disputed authorship of the Johannine material. A somewhat broader sweep was taken by A. Kenney, whose book *A Stylometric Study of the New Testament* appeared in 1986. These are reviewed in Bartholomew (1996).

Before moving on to discuss particular examples it is necessary to specify how documents of any kind can yield quantitative measures for stylometric analysis.

**Statistical characteristics of religious texts**

There are many things one can measure in any text. Allowance may have to be made for the characteristics of the original language in which the text was written, of course, but for present purposes we can leave such matters on one side.

**Words**

Words are the basic units of any text. We may look at the statistical characteristics of the set of words which the author uses, such as the frequency distribution of their length and its various summary measures. There is also the selection of words – the author’s vocabulary. When comparing two texts, we may compare the size of the vocabulary or the particular choice of word. Some authors may have a particular predilection for certain words which ‘gives away’ their authorship. All of these things, and many others, are included in reference to the richness of the vocabulary.
Sentences

Sentences may differ in length (i.e., the number of words) or in their structure. The number of words in a sentence is easily measured and the properties of its frequency distribution determined. The structure is more difficult to quantify but, at its most elementary, it would involve such things as word order.

Multivariate characteristics

Pieces of writing which we wish to compare will usually differ in many respects. At the most basic level it may be sufficient to compare average sentence length, say, but it will usually be desirable to make a number of comparisons simultaneously. We then move into the realm of multivariate statistics. To get the idea, consider how we might compare the ‘size’ of members of different races. One way would be to compare height because ‘larger’ people tend to be taller. Equally they also tend to be heavier so we might wish to bring weight into the comparison. Height and weight together should give a better picture of group differences. One way of looking at this would be to plot height and weight in a scatter plot with height along one axis and weight on the other. Any differences in size would then be reflected in the separation of points. If we found, for example, that the points for one group clustered together in one part of the plane and those for another elsewhere, we would have a picture of the difference in which both dimensions played a part.

In a similar way we might measure several further characteristics of size and see whether different authors (or whatever the subject of investigation might be) clustered in different parts of the plane. This would only work visually, of course, in two dimensions, but conceptually there is no reason for not using many indicators of style and ‘looking’ for a clustering effect. We could continue to use the geometrical terminology, and speak of points, without being able to visualize things as in two dimensions.

Richness of vocabulary

The size of an author’s vocabulary and the way in which words are used can be an important indicator of the origin and structure of a text. A number of attempts to measure ‘richness’ have been made and these have been used in stylometrics.

A single index may be useful, but no single index can capture the subtlety of a written text, so attempts have been made to use a collection of such measures. A notable example is provided by Holmes (1992), which we use as our first example.

First example: disputed authorship of the Mormon scriptures

We shall illustrate some of the principles involved in stylometric analysis by taking a study carried out by Holmes on the Mormon scriptures (Holmes 1992). His study uses a multivariate idea, the richness of vocabulary, as the main tool for discriminating among possible authors.

Holmes used five indices, each of which captures part of what we mean by ‘richness’. A particular interest was in words which are used only once in a text – known as hapax legomena. An author with a rich vocabulary will tend to use a great many different words and it is the rarer words which are particularly indicative of authorship. The number of such words will depend, of course, on the length of the text so the actual number has to be standardised in some way to make the index independent of text length. A second, similar, index is the number of
words which are used exactly twice – *hapax dislegomena* – as they are called. A third index which Holmes used was first introduced by Yule (1944). This index is based on the supposition that different words occur in a random fashion according to a Poisson distribution, and the index is based on the variability of the frequency distribution of the number of times each different word in the text is used. It, thus, goes beyond the first two indices by using all occurrences and not just the first two. Strictly speaking it does not require the occurrences to be distributed exactly according to the Poisson law, but, if they are, it turns out that the index does not depend on the size of the text.

His fourth and fifth indices are derived from something called the *Sichel* distribution. This specifies the proportion of times that any word type (e.g., noun, preposition, etc.) is expected to occur in a text of given length. The distribution appears to fit actual texts well and it depends on two unknown parameters. What Holmes did was to devise a method of estimating these two parameters, and it is these estimates which he used as the final two indices.

Holmes wished to compare the richness of the vocabulary for a set of texts with a view to identifying their origins. In particular he wished to compare the writings known to be by Joseph Smith, the founder, with the text of The Book of Mormon and other writings to which Joseph Smith may have contributed or even authored. The methodological question is how to compare sets of five indicators. Holmes used two methods: cluster analysis and principal components analysis. Cluster analysis, as its name suggests, aims to see whether the various writings cluster together in groups, each one of which might indicate a common authorship or, at any rate, a comparable degree of richness in the vocabulary used. Holmes also included analyses from the King James translation of the Bible, in particular, parts of Isaiah, because that had often figured in earlier discussions. The relevance of this comparison is that it is claimed that the book of Mormon was written long before the King James translation was made, and this would make it surprising if it shared a common style with the King James Bible.

The interpretation of Holmes’ analysis is complex and the interested reader must consult the original paper to obtain a full account. Broadly speaking Holmes’ findings do not support the understanding of the origin of Mormon writings on which Mormon doctrine depends. In particular, the findings suggest that the Book of Mormon, and other writings believed to have been given to Joseph Smith by revelation, were in fact composed by Joseph Smith himself writing in what is termed his ‘prophetic mode’.

The important thing for our purposes is the status of the results obtained by cluster analysis and principal components analysis. As used in these studies, they are essentially descriptive methods. They throw into relief the differences between the various documents which may have literary significance but this must not be confused with statistical significance. Holmes’ results do not, therefore, settle in an objective way the questions of the authorship of the Mormon texts. They do, however, make it that much more difficult to argue the contrary case.

**Second example: the synoptic problem**

Many biblical documents are supposed to have been composed by drawing on a variety of sources. In retrospect, it is then of interest to identify the origins of extant passages in those earlier documents. One major study of this kind was carried out on the book of Genesis by Radday et al. (1985). It was supposed that there were three strands running through Genesis and it was the purpose of the study to disentangle them. A more recent field of study of this kind relates to the manner of formation of the synoptic gospels.

The synoptic gospels are those attributed to Matthew, Mark, and Luke. The so-called synoptic problem arises because some of the material is common to all three gospels, some is
common to two of them and some only occurs in one. The problem is to discover why this has arisen and to explain the occurrence of the overlaps. This problem is closely linked to the dating of the gospels because, if the overlaps have occurred because one writer had access to one or more of the others, this could only happen if some had been composed before others. For example, Luke could only have used material from Mark if Mark’s gospel was already available. The synoptic problem has been investigated by comparing the material which is common to pairs of gospels and making deductions about how this might have arisen.

There is an enormous literature on this problem, mainly written by biblical scholars but, more recently, attempts have been made to study the problem statistically. A useful review of this work has been made by Poirier (2008). Very recent, and sophisticated, studies in this tradition are by Mealand (2011) and Abakuks (2012).

One can imagine many ways in which the gospel writers may have operated. For example, and not very realistically perhaps, they may have worked independently, but all drawing on a common source of oral, or possibly written, material, now lost. Or Luke, for example, may have used Mark as one of his sources, drawing some of the rest of his material from elsewhere. One of the best known hypotheses of this kind, known as the two-source hypothesis, is that both Matthew and Luke drew on Mark and on another document, often known as Q. The supposed source Q is entirely hypothetical since no such document has been discovered, and Abakuks (2006) has used probability arguments to show that the assumption of its existence may not even be necessary. This uncertainty has not prevented many biblical scholars, for example Crossan (see below), from treating Q as if it really existed. It is not possible to settle the question statistically because there is only a limited amount of text available, and the best we can hope for is to shift the probability in favour of one hypothesis or the other. Mealand’s analysis, using a variety of multivariate methods, moves the odds somewhat in favour of the two-source hypothesis.

The number of possible hypotheses is clearly very large, and the object of stylometric research in this field has been to find which of the possibilities are the most plausible in the light of the data. Literary scholars will judge plausibility with reference to the canons of literary criticism, whereas statisticians will proceed by constructing models of the composition process and choosing amongst them using criteria of closeness of fit. The two approaches are not necessarily in conflict, and one may hope, instead, that they will be mutually reinforcing.

All approaches must begin by specifying the units of text which will form the basis of the comparisons. This is not a straightforward matter, but it is natural to begin with words as the most elementary units of which text is composed. However, meaning is derived from context, as well as the individual words, so it is important to ensure that, if one chooses to use words as the elements of analysis, one also checks that they are being used in the same sense in the various locations in which they occur. Although these questions are of fundamental importance, they are not primarily statistical and therefore will not be pursued further.

The central idea behind any model is a specification of how the synoptic authors are presumed to have worked. It is easy to see that they did not copy directly from one another, or some other source in a purely mechanical manner, because if they had, the fact would be immediately obvious. It is evident that if only one source had been used by any one author it must have been used in a more unpredictable fashion. The best way to capture this unpredictability is by introducing probabilities into the specification of models. This does not imply, of course, that authors drew lots in deciding what to write next, merely that the unpredictability in their behaviour can be described in probability terms. In other words, they behaved as if they were drawing lots. The proof of this particular pudding, of course, is in the eating!
Ideas of statistical inference applied to the interpretation of texts

The methods of statistical inference have sometimes been seen as providing a tool to help in answering questions posed by Biblical texts. There are certainly some similarities between the two approaches, so before proceeding further, we outline some of these and comment on their relevance.

In a typical statistical problem there are some data, usually numerical, and one or more hypotheses to explain them. The aim is to find a hypothesis which satisfactorily explains the data. There are various procedures of inference for judging how closely the data conform to the hypothesis. The data will usually consist of a collection of numbers (often a random sample) from some population. Next there is a set of hypotheses which, it is hoped, will explain the data. The problem of statistical inference manifests itself in a variety of forms according to the objectives of the analysis. For example, one of the hypotheses may be special in some sense, and the objective will then be to test whether the data are consistent with that hypothesis. Or one may simply wish to select that hypothesis which provides the best explanation of the data— in some clearly defined sense.

In the case of biblical texts the written texts comprise the data, and the hypotheses purport to explain what the texts mean or who wrote them. In the statistical problem, as we have seen, the data usually consist of numbers derived from some measuring process. In textual analysis, the data will seldom be numerical and the position is obviously more complicated even if it is of essentially the same form. The words may be parts of narrative accounts, and they do not necessarily have a single unambiguous meaning. Or they may be poetry or other literary forms which do not have a simple factual interpretation. The task of the biblical scholar is to decide what it all means and to express the conclusions in terms of one of the hypotheses which are being entertained as possible explanations. Another, important difference is that in the scientific case the size of the sample is often at choice. If, initially, the result is unclear we can usually go back and increase the sample size. That option is rarely available in the analysis of texts because what we have may be the total corpus, or, even if it is not, what we have is hardly likely to be anything like a random sample of the corpus.

However, in spite of the obvious differences, there are certain common features and one of them is particularly relevant here. It is often true that there is more than one statistical hypothesis which accounts for the data. In such cases the hypothesis is said to be poorly determined, or underdetermined, by the data. Subsequent work may then be directed to collecting more data which will help to resolve the uncertainty. In biblical studies this situation is the norm. What is written in the gospels, for example, is held to support many different and sometimes seemingly contradictory interpretations. For example, according to one hypothesis, the texts are to be treated as recording actual events, whereas, in another, they may be understood as symbolic or as fictitious in some sense. The most that the biblical scholar can expect to do is to construct a coherent account of what he finds—that is, to find a hypothesis which can account for everything we know under certain interpretations of the data. Such interpretations often depend on hidden assumptions which it is assumed, or hoped, the reader will share with the author. A common example of this is to suppose that God (if he exists) is not involved in any way with what happens in the world. If this assumption is shared by the reader a naturalistic hypothesis such as this may appear convincing. The point being made here is not that such assumptions may be true or false but that any uncertainty about them bears on the final conclusion.

Before moving on to Bayesian methods, we shall critically examine attempts by a number of biblical scholars and philosophers to use quantitative methods in the interpretation of biblical texts. To do this we shall examine their approaches to a single, critical event reported in the
Christian gospels: namely, the bodily resurrection of Jesus Christ. All of them focus on the historicity of what is reported, but in this example, especially, interpretation and historicity are inextricably bound up together.

We begin with J.D. Crossan, who has published many books relevant to the subject (for example, Crossan 1991, 1994). Crossan is a leading figure in the Jesus Seminar, which classifies the purported records of the life of Jesus according to their authenticity by a process of collective decision-making which involves voting by members of the seminar. The data used by members of the seminar consist of a set of sayings gleaned from many sources of which the Gospel of Thomas and the alleged Q Sayings of Jesus figure prominently. This approach is scientific to the extent that more weight is given to multiply attested sayings (i.e., those that occur in several sources, assumed to be independent). However, little attention appears to be given to the basic authenticity of the sources. For example, there seems to be no recognition that the Q source is hypothetical and that there are some scientifically based studies (such as that of Abakaks 2006) which show that Q need not be assumed to exist in order to account for what we find in the gospels. In fact what Crossan appears to be doing, with a great show of imaginative erudition, is to construct a naturalistic hypothesis which accounts for all of the evidence. One of its most questionable elements, picked on by N.T. Wright (2003), is that Crossan introduces the idea of retrojection. That is, in order to account for the resurrection stories in the gospels, he supposes that the events did not really happen but were written back into the record by later Christian communities in order to justify their beliefs and practices. No evidence is given for believing that this is possible (for example, no instances are given of this having certainly happened, elsewhere). The success of Crossan’s efforts must be judged by whether the reconstruction represents a possible and credible explanation. Crossan himself offers no guidance on how one should choose between possible explanations (supposing there to be more than one!). He uses no probability statements, formal or informal, to express his own judgements. There is therefore no justification for describing his method as scientific.

N.T. Wright is a major New Testament scholar who treats the resurrection in The Resurrection of the Son of God (Wright 2003). He claims to have shown that the bodily resurrection of Jesus from the dead provides the best explanation of what we find in the biblical record. To focus on the main point we wish to make, we begin with his conclusion, which is: ‘We are left with the secure historical conclusion: the tomb was empty, and various “meetings” took place not only between Jesus and his followers (including at least one initial sceptic) but also, in at least one case (that of Paul; possibly too, that of James), between Jesus and people who had not been among his followers. I regard this conclusion as coming in the same sort of category, of historical probability so high as to be virtually certain, as the death of Augustus in AD 14 or the fall of Jerusalem in AD 70’ (2003: 710).

Although the final conclusion is expressed in probability terms the logic employed owes nothing to probability theory. Instead, he draws on the idea of necessary and sufficient conditions as they occur in logic and mathematics to establish the truth of a proposition. For example, he says: ‘The empty tomb and the “meetings” with Jesus, when combined, present us with not only a sufficient condition for the rise of early Christian belief, but also, it seems, a necessary condition.’ He adds, ‘Nothing else historians have been able to come up with has the power’. Nevertheless, Wright acknowledges that this does not deliver the certainty to be expected from formal logic (because it is concerned with what the record says, rather than with what actually may have happened) by saying: ‘This remains, of course, unprovable in logical or mathematical terms. The historian is never in a position to do what Pythagoras did: not content with drawing more and more right-angled triangles and demonstrating that the square on the hypotenuse always does in fact equal the sum of the squares on the other two sides, he constructed a
theorem to prove that this must always be the case, with history it is not like that. Almost nothing is ever ruled out absolutely; history, after all, is mostly the study of the unusual and unrepeatable. What we are after is high probability; and this is to be attained by examining all the possibilities, all the suggestions, and asking how well they explain the phenomena.

Before moving on to the Bayesian approach, we shall look at a third, and different, approach. Licona (2010), by contrast, is very clear that, ideally, Bayesian inference is appropriate in this case, and he quotes some attempts to proceed in this fashion and he discusses it on p. 116ff. However, Licona does not find the conditions applicable, so he introduces an alternative, non-probabilistic, method to find the hypothesis which provides the best explanation of the data. It is impossible to do justice to Licona’s careful and thorough treatment in a short space but the essence is as follows.

Licona begins with what he regards as the irrefutable facts, which he calls the ‘historical bedrock’, which has to be explained by any hypothesis. These facts are universally agreed by virtually all scholars. He then considers the chief ‘naturalistic’ hypotheses which purport to explain these facts and then rates all these competing hypotheses according to how good an explanation they provide. He concludes that the hypothesis of actual bodily resurrection provides the best explanation.

The Achilles heel of Licona’s treatment, for our purposes, lies in what he calls the ‘horizons’ of the authors of the various hypotheses. These, essentially, are their prior suppositions about the nature of reality. He acknowledges that agreement among historians is unobtainable because each uses a different horizon. In particular, any historian who rules out supernatural actions, a priori, is precluded from giving credence to the resurrection. The great merit of Licona’s treatment is that it is totally transparent and makes it clear where those who disagree with him must aim their criticism.

Bayes’ theorem

There are several approaches to uncertain inference, but one which has captured the imagination of many biblical researchers is the use of Bayes’ theorem or, more generally, Bayesian inference. The reason for this is that it seems to provide exactly what literary (and other non-quantitative) researchers require. It appears to enable us to calculate the probability of a hypothesis. Thus, for example, if we were able to objectively calculate the probability that St Paul wrote the letter to the Ephesians that bears his name, we would have a scientifically defensible measure of something which, hitherto, may seem to have depended on subjective – and fallible – human judgement. Unfortunately as we shall see, this is rarely the case and most such hopes will lead to disappointment. In order to explain why this is so, we must digress to give a necessarily technical account of what Bayes’ theorem actually is.

Background

In order to avoid as much mathematics as possible, we shall state the theorem in an incomplete form that nevertheless enables us to make the essential points. For this purpose, we need, first, to explain the notation. The theorem is about probabilities, and, hence, we need to be able to designate a probability in a way which provides full information about what is involved. ‘Pr’ is short for ‘probability’. The thing whose probability we are talking about is described in brackets after ‘Pr’. Thus

\[ \text{Pr(it will rain tomorrow)} \]
means ‘the probability that it will rain tomorrow’. Almost all probability statements are conditional; that is, the probability depends on something else being the case. Thus we might be interested in the probability that it will rain tomorrow given that the previous year was the wettest on record. The latter circumstance might possibly make a difference and this possibility needs to be acknowledged in determining the probability. This is done using the symbol ‘|’ as a shorthand for the word *given*. The foregoing probability might therefore be more fully written

\[ \Pr(\text{it will rain tomorrow} \mid \text{the previous year was the wettest on record}) \]

In practice we often use single symbols to replace the rather cumbersome forms of words used in the foregoing illustrations. Thus we might encounter \( \Pr(H \mid C) \) where \( H \) refers to some hypothesis of interest (such as that it will rain tomorrow) and where \( C \) refers to the conditions under which the probability has been calculated. Typically, we shall be interested in probabilities of the form

\[ \Pr(H \mid E) \]

where \( H \) might be the hypothesis that Paul wrote the letter to the Ephesians and \( E \) would be the relevant evidence – which would, principally, include all the relevant information provided by the so-called Pauline corpus. Bayes’ theorem now tells us that

\[ \Pr(H \mid E) \text{ is proportional to } \Pr(H) \times \Pr(E \mid H). \]

The terms in this expression have names. \( \Pr(H) \) is called the ‘prior probability’ – because it is the probability before we consider the evidence \( E \). \( \Pr(H \mid E) \) is called the ‘posterior probability’ of \( H \) because it is the probability after the evidence \( E \) has been taken into account. The remaining term, \( \Pr(E \mid H) \), is called the ‘likelihood’. It is important to notice that the word ‘likelihood’ is being used as a technical term in a sense which is somewhat more limited than its everyday meaning. In particular it is not a synonym for ‘probability’. The reason for this is that we shall only be interested in how it varies with \( H \) (and not how it varies with \( E \)). The context in which it is used is one in which we are comparing different hypotheses on the same data.

It is now immediately clear where one problem in the use of Bayes’ theorem lies. Bayes’ theorem is essentially about how probabilities are changed by evidence. In the example given above, we could only find the probability that Paul wrote Ephesians after the evidence is available, if we already know what the probability was before the evidence became available. In other words we must know the prior probability before we can determine the posterior probability. The theorem says nothing at all about how the prior probability should be determined. In most cases this is a personal matter, and there is no requirement that different people should agree. It is this essential subjectivity which has made the theorem so controversial. It must be emphasised that the theorem itself is not controversial. The theorem is the necessary consequence of a few axioms of rational choice which are quite unexceptional. The controversy resides in the applicability of the theorem in practice. There are some scientific contexts, in genetics, for example, where there is no problem at all. There are many others where the matter is much more questionable. Anyone who accepts the axioms must, if they are to act coherently, accept their consequences.

### The Bayes factor

Bayes’ theorem is often used to compare two hypotheses, and we shall see it used in this way below. Suppose \( H_1 \) and \( H_2 \) are two competing hypotheses. We could compare their posterior
probabilities by computing the ratio \( \Pr(H_1 \mid \text{evidence}) / \Pr(H_2 \mid \text{evidence}) \). According to Bayes’ theorem this may be expressed as

\[
\left( \frac{\Pr(H_1 \mid \text{evidence})}{\Pr(H_2 \mid \text{evidence})} \right) = \frac{\Pr(H_1)}{\Pr(H_2)}, \quad \frac{\Pr(\text{evidence} \mid H_2)}{\Pr(\text{evidence} \mid H_1)}
\]

The second factor \( \Pr(\text{evidence} \mid H_2) / \Pr(\text{evidence} \mid H_1) \) is called the Bayes factor. If it is large, it means that \( H_2 \) gives a much higher probability to the evidence than \( H_1 \). If the Bayes factor is very large indeed it might swamp the prior probability ratio so that, even if we do not know the latter, its actual value hardly matters. This means that progress can sometimes be made if we are very uncertain about the prior probabilities because whatever the prior probability ratio is, the Bayes factor will be ‘large’.

There is a long history of using this way of bypassing the need for prior probabilities. In statistics it is known as ‘likelihood inference’ and it has been in use for nearly a century. But it has achieved wider currency more recently under the title of ‘inference to the best explanation’. As proposed by Lipton (2004), ‘best’ is interpreted to mean ‘that which gives what we observe the greatest probability’. Other users of the term have interpreted ‘best’ in other senses, as we have seen with Licona above. When used probabilistically, this mode of inference is essentially Bayesian analysis but without the prior probability.

There are three further important points to note about the application of Bayes’ theorem. First, it can be used in stages to build up a posterior probability. Suppose the evidence \( E \) becomes available in two stages. After stage 1, the posterior probability is \( \Pr(H \mid E_1) \), say. If further independent evidence \( E_2 \) is added, the posterior probability becomes \( \Pr(H \mid E_1 \text{ and } E_2) \). It is straightforward to show that \( \Pr(H \mid E_1 \text{ and } E_2) \) is proportional to \( \Pr(H \mid E_1) \times \Pr(E_2 \mid H) \). In words this says that the prior probability at the second stage is the posterior probability from the first stage. So we can proceed, sequentially, to build up evidence with the prior probability at any stage being the posterior probability from the previous stage. Second, it is essentially a means by which an individual can ensure that the set of their beliefs is coherent. It does not lead to agreement among individuals unless, of course, they happen to agree on their prior probabilities and any other subjective inputs. Third, a point which is mathematically trivial but of fundamental importance in practice, if a person chooses their prior probability to be zero, no amount of evidence can produce a posterior probability of anything other than zero.

Two applications of Bayes’ theorem to texts

We shall discuss two applications of Bayes’ theorem – both by philosophers – to the question of whether or not there was a bodily resurrection of Jesus Christ. Taken together they illustrate the pitfalls which await the unwary.

Swinburne

We begin with Swinburne (2003). He uses an approach similar to that employed in his earlier work on the existence of God (Swinburne 1979/1991). The greater part of the book is taken up with a highly detailed and thorough, but qualitative argument. In an appendix, Swinburne clothes the qualitative argument in algebraic terms and assigns some numerical values to some of the probabilities involved. Swinburne requires, of course, a prior probability – in this case the
probability that God exists. The process of building up a posterior probability starts, according to him, with a prior probability conditional on what he calls *tautological evidence* (that is, the evidence provided by natural theology). He also speaks of intrinsic probabilities in the same sense. It is not clear why he regards this as the starting point, because natural theology is based on what nature tells us about the existence of God and this evidence is necessarily provided by the natural world. We need to go back further to the point before there is any evidence derived from the senses at all. In those circumstances, as I argued in Bartholomew (1996: 167), any prior probability is strictly indeterminate and hence there can be no truly objective posterior probability of any hypothesis about the world dependent on it. This applies to Swinburne’s argument, and hence all of the probabilities which appear in his discussions must be subjective (including his prior probability for God’s existence of 0.5). Swinburne recognises this fact but does not sufficiently emphasise that, as a result, all his probabilities are subjective and are relevant to no one but himself. His claim that the resurrection is highly probable expresses his own belief, systematically arrived at, but places no obligation on others to agree. It does, however, place an obligation on anyone who claims to be forming their beliefs rationally to arrive at them using Bayes’ theorem.

**McGrew and McGrew**

The second paper is by McGrew and McGrew (2009). This is a major study set in the context of a wider discussion of Hume’s famous argument about miracles. It avoids the need to specify prior probabilities by looking only at the Bayes factor. McGrew and McGrew aim to show that the Bayes factor in favour of the resurrection must be very large, so that the prior probability ratio required to outweigh it would have to be very small indeed. They conclude, using an approximation and making other assumptions, that only if the prior probability of God’s existence were judged to be less than 1/10^40 (a value which Licona quotes in a footnote on p. 117) would it be credible to doubt the resurrection. A probability of 1/10^40 is unimaginably small – so small that it is virtually indistinguishable from zero. Hence, for all practical purposes, the resurrection, according to the McGrews, is virtually certain.

This is a very surprising result which, if truly objective, would move the subject beyond further argument. It is therefore important to examine the argument very closely. The McGrews start from the biblical evidence and explain in a section entitled ‘Textual Assumptions’ (p. 597) that they will assume ‘that we have a substantially accurate text of the four Gospels’. Intuition would suggest that this builds in a substantial advantage at the outset to hypotheses which are consistent with that assumption (it rules out, for example, Crossan’s various hypotheses). But intuition is not necessarily a sure guide. The McGrews identify many events recorded in the gospels which all testify to the reality of physical resurrection. It turns out that the relevant Bayes factor requires the determination of the probability of the conjunction of many (about 40) events. This probability cannot be determined because the events concerned are not independent. The McGrews recognise this and therefore propose an approximation which yields the same Bayes factor as if they were independent. Their justification for doing this is that any departures from independence would tend to make the Bayes factor even smaller and so they are erring on the safe side by treating the events as if they were independent. Their argument for doing this is unconvincing and entirely qualitative. The McGrews also devote considerable space to a defence of the general historical accuracy of the gospels without, apparently, noticing that their conclusion on that seriously compromises the validity of their Bayesian analysis. Their analysis does show how dependent conclusions may be on the assumptions made at the outset – or imported during the analysis.
Conclusions

Religious texts concern, directly or indirectly, the nature of ultimate reality. They are often used to support various contentions about the nature of that reality. We may approach these questions using one of two assumptions: either that the reality as we experience it through the senses is the totality of all that exists, or that this is not the case and there is something beyond what we experience with the senses.

The methods described in the first part of this chapter were descriptive in the sense that they enable us to see the issues more rationally and to contribute more sensibly to debates about such things as the authorship and structure of religious documents. They did not (and cannot) decide issues with certainty. Indeed their chief value may be in insisting that such things must always be uncertain to some extent.

Our discussion of inferential methods in the second part of this chapter showed that neither Bayes’ theorem nor any other method of inference can lead us to certainty about the nature of ultimate reality. Any hypothesis which ultimately depends for its truth on which stance we take about the nature of ultimate reality involves a prior probability, and there is no way to rationally assign such a probability – it is essentially indeterminate. In particular, if we were to assign a probability of zero to God’s existence – as some seem to do – no amount of subsequent evidence could shift us from that position. Conversely, if the reverse is true, we could, in principle at least, be convinced of the truth of any proposition for which there is sufficient evidence. However, there is no way that the beliefs derived from the two opposed starting points can ever be reconciled rationally.

Acknowledgement

In preparing this chapter I have greatly benefited from discussions with Dr Andris Abakuks.