The Bayesian View of Extreme Events

Allan McRobie, Cambridge University Engineering Dept
Henderson Colloquium, Magdalene College, Cambridge, July 2004

July 5, 2004

Abstract

A key step in "Designing for the Consequences of Hazard" is choosing which hazards to design for. This paper argues that the Bayesian perspective is the rational and prudent framework.

1 Motivation

Loosely speaking, structural engineering involves "loads versus strength". It is a failure of our system of educating structural engineers that the strength side of this dichotomy is usually given so much pre-eminence. It would be a caricature to suggest that designers expect to look up loads in a code of practice, and then invest many man-hours in detailed finite element analysis of the structural behaviour under those loads. Sometimes, though, this caricature does not seem to be so far from the truth. Structural behaviour is often calculated to assumed accuracies of a few percent. Although codes may dictate specific figures for what a structure "should" be designed for, the reality of what a structure may subsequently experience is far more intangible. Ultimately, structural design is a question of the management of uncertainty, and the greatest uncertainties are usually in the loads. If structural engineers wish to be involved in the design process as decision-makers rather than as mere service-providers, then they need to engage with this process of assessing the possibilities that may occur to their structure. There is a three-stage hierarchy: 1. How does a structure behave, given a load environment?, 2. What is the nature and physics of that load environment? and 3. What is the chance that such a load environment will occur? Improved modelling only addresses the first two stages: fully-coupled fluid-structure interaction CFD simulations of long-span bridge aeroelasticity; explicit finite-element simulations of aircraft impact events; computational models of crowd behaviour; - none of these address the question "which storm/aeroplane/crowd/flood/earthquake/terrorist incident/accident/tsunami/volcanic event etc. Again, to caricature the process, someone will pick a scenario, and the structural engineer will then try to design a structure that just satisfies that criterion, usually to a remarkably inappropriate degree of accuracy.

This paper argues that the Bayesian perspective is the appropriate framework in which to address the third stage, but there are at least two obstacles that inhibit its adoption. Firstly, attitudes may be coloured by educational background. Often, the more deterministic and scientific the training, the greater the reluctance to accept the Bayesian perspective. Secondly, many treatises on Bayesian theory are so thick with Greek letters, integrals and formulae that the power of the underlying philosophy is lost in the haze of symbols. As a consequence, hard-headed managers may consider that anything so mathematically ornate must be inapplicable to the noisy subtleties of everyday decision-making. This is unfortunate, for in a broad sense, the way which many managers currently make decisions (on the basis of a combination of experience and analysis) can be readily assimilated into the Bayesian framework.

Perhaps the primary contribution that Bayesianism can make to the debate about extreme possibilities is its clarification of the language of and thinking about risk. It provides a clear, coherent and unambiguous framework in which discussion can take place. Many discussions of extreme possibilities are riddled with sentences which, to a Bayesian, are incorrect and indicative of underlying misunderstandings.

Rather than describing the detailed mathematical techniques of Bayesian statistical analysis and decision theory, this paper will give a broad outline of the underlying philosophical framework. There are numerous versions of this, and I have chosen to follow the one expounded in the recent book by Jaynes: *Probability Theory: the Logic of Science.*

2 Bayesian Fundamentals

2.1 Bayes' Formula or Bayes' Theorem

Bayes proved a theorem about conditional probability. It states that, given fairly broad conditions,

$$p(AB) = p(A|B)p(B) = p(B|A)p(A) \text{ whence } p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{1}$$

Here, $p(AB)$ means the probability of both A and B, and $p(A|B)$ means the probability of A given B.

The theorem is not disputed (and is much used) by frequentists. The dispute lies in the interpretation of the various terms in the formula, and when and how the formula can be used. Most engineers learn the theorem at A level or in their first year at University. Unfortunately it is usually taught in a frequentist setting (assessing the probability of drawing coloured balls at random out of a bag, for example) such that the philosophical differences between Bayesianism and frequentism are avoided.

2.2 Probability as belief

The disagreement between Bayesians and frequentists concerns the interpretation of what "the probability of A" means. For frequentists A is some repeatable event and $p(A)$ is its relative frequency of occurrence. For Bayesians, A is a logical proposition, and $p(A)$, the probability of A, is a number representing of ones degree of belief in the truth of that proposition. Examples of such propositions are

1. A = " The next card drawn will be an ace"
2. A = " The next concrete cube measurement will lie between σ and $\sigma +$ dσ"
3. A = " The Earth's average surface temperature in 2100 will be more than 7°C higher than in 2000"
4. A = "My Old Nag will win the 3.30 at Epsom tomorrow"
5. A = "My opponent's hand can beat my pair of aces"
6. A = "More than 25,000 people will cross the bridge on its opening day"
7. A = "An aircraft will crash into the building in the next 50 years, and the aircraft's design has not yet left the drawing board"

For frequentists, the first two propositions can be ascribed probabilities if some assumptions are made about the underlying distribution of samples. The others, though, have **no meaning** in terms of probability for frequentists. For Bayesians, all statements can be ascribed a probability. For the probability of drawing an ace, a Bayesian would include a subjective assessment of the chance that the deck is rigged. Remarkably, the theory of statistics and probability that is taught on engineering courses can say little or nothing about most everyday matters of chance. More importantly, the issues faced by engineers, of assessing the likelihood of future events, **cannot** be expressed in frequentist language. Why then do we persist in teaching engineers any theory of probability other than Bayesian?

2.3 Updating

A central pillar of Bayesianism is the interpretation of Bayes' Formula, which states that

$$p(H|D) \quad \alpha \quad p(D|H)p(H) \tag{2}$$

where H is a proposition called the hypothesis, and D is the data. On the right hand side, the term $p(H)$ is called the **prior**: ones prior belief in the possible truth of the hypothesis (before looking at the data). The term $p(D|H)$ is the **likelihood** of obtaining the data D given the truth of the hypothesis H. The product gives the left hand side $p(H|D)$, the **posterior** on H, and is the updated degree of belief in H after having seen the data. There is thus a temporal/causal dimension to Bayesian theory, with clear 'before' and 'after' concepts which are lacking in the pure theorem.

Essentially, one has some beliefs, some new information arrives, and one takes cognisance of this and updates ones beliefs accordingly. Bayes' simple formula has thus become a model for learning. A rigorous Bayesian is considered to be conceived/born with something fairly close to a blank slate and the experiences of life gradually update this in the fashion described by the above formula.

2.4 Subjectivity

A primary source of objection to the Bayesian framework comes from the scientific community. This is a consequence of history. Science blossomed in The Enlightenment. The great scientists, Newton, Hooke and so forth, felt they were uncovering an underlying objective reality, free from subjective opinion. Subjectivity was the realm of mumbo-jumbo. To some scientists, Bayesianism, with its emphasis on subjectivity, seems to give equal legitimacy to the views of cranks and nutcases as it gives to careful scientists.

2.5 The difference between science and engineering

A primary distinction between science and engineering is the temporal aspect. For scientists, there is an underlying reality which they are attempting to uncover. It does not really matter how long this takes. For structural engineers, the problem is that of anticipating what may happen in the future. The rigorous scientific approach to answering the question "What is the probability that an aeroplane will crash into the power station in the next 25 years" is to observe the power station for the next 25 years, and answer either 0 or 1 as the case may be. Engineers do not have that luxury of simply gathering more and more data. They need to make decisions about the future. The future is inherently unknowable and unpredictable. The only rational and coherent way to formulate such questions about the likelihood of future events is the Bayesian perspective.

2.6 Prediction versus Estimation

It is unfortunate that many texts on statistics focus on the problem of parameter estimation. Again this has its historical roots in scientists attempting to use available observations to obtain the best possible estimates of the numbers describing the world around them (the speed of light, the diameter of the Sun, the genetic make-up of a batch of peas, etc.). It is even more unfortunate that many texts on Bayesian theory have taken the same path. For engineers, the perennial problem of assessing future possibilities are not questions of estimating parameters. As a consequence many numerate engineers apply methods such as Maximum Likelihood to help them make decisions, in the belief that this is scientific best practice. Unfortunately, such engineers are rarely embarking upon a scientific question but are instead charged with trying to anticipate future possibilities before they happen. Design engineers need to face up to **The Prediction Problem**.

The estimation problem, in Bayesian terms, focuses upon the posterior distribution

$$p(\theta|\mathbf{X}) \; \alpha \; l(\mathbf{X}|\theta)p(\theta) \tag{3}$$

where $\mathbf{X}$ is the historical data, and $l(\mathbf{X}|\theta)$ is the likelihood of obtaining that data, given the parameters θ. Estimation concerns itself with updating beliefs about the underlying parameters θ.

The Bayesian formulation of The Prediction Problem extends its consideration to the predictive distribution

$$p(x_{n+1}) = \int_{\forall\theta} p(x_{n+1}|\theta)p(\theta|\mathbf{X}) \; d\theta \tag{4}$$

This focuses directly on the probability of the next value x_{n+1}. It is the probability of any x_{n+1} GIVEN a parameter value θ, combined with the probability that the parameter value actually is θ (as described by the posterior distribution).

This second step, from estimation to prediction, is often omitted in simple explanations of Bayesianism. This means that the debate occurs on the estimation territory for which frequentism has been designed. I am often asked "Do Bayesian methods give more precise estimates?". The question misses the whole point. The answer is "Bayesianism is a more rational and prudent approach to prediction".

2.7 The myth of objectivity

There is an underlying objective reality. However, anticipating and assessing the likelihood of what might happen next is a subjective issue. There is no objective assessment of which horse will win the 3.30 at Wincanton tomorrow, or what the earth's average surface temperature will be 100 years from now. Until engineers face up to the inevitable conclusion that assessing the future is subjective, and can only be approached via a Bayesian stance, they will continue to make ill-considered, erroneous and imprudent decisions no matter how much they back them up with purportedly-objective techniques of engineering science and frequentist statistics.

2.8 Extremes

If there is plenty of data, then it will tend to dominate the prior beliefs. For estimation questions within the span of the data (such as what is a likely value for the location parameter μ of an underlying normal distribution) different analysts will then draw generally similar inferences, and all will be very close to the frequentist estimates. If, however, there is little data, or if prediction questions are asked which lie well beyond the span of the data, then the prior beliefs will have a more significant effect. Typical of such questions is the famous case of the flood protection in Holland against catastrophic inundation from North Sea storm surge. The Dutch have set a design criterion that the flood protection have an annual exceedance probability of 10^{-4}, and yet only around 10^2 years of historical data is available on which to base such an assessment [2, 3]. This is one example of the spectrum of possibilities. In this case, there is a significant amount of high quality, numerical data, but inferences are required well beyond the span of this. For such cases, the techniques of Bayesian extreme value theory may be applicable (see for example Coles [4, 5]). A different form of example would be material sampling, where only one or two readings are taken (core samples perhaps), and yet inferences are required about characteristic strengths. A third form of example would be more abstract, such as predicting crowd densities on the opening days of a bridge, or assessing the possible forms and severity of terrorist attacks. There may be copious historical information, but the assessment involves predicting the actions of other humans and the possibility that they may do something beyond anything that has been previously seen. However, the common theme in all such examples is that the prior beliefs will have a large effect on the inferences, such that comparable care should be taken in properly constructing prior belief distributions as in assessing the available data.

2.9 There is no such thing as "No Priors"

It is common for engineers to consider Bayesian theory to be "another method". Given their unfamiliarity with its deeper underlying interpretations and their training in so-called objective engineering science, they are reluctant to ascribe priors. They thus often apply the methods in the form "Bayes without priors". There is, however, no such thing. Consider for example the case of a random variable where the form of the underlying distribution is known, but it depends upon an unknown parameter t say. Rather than ascribing no prior, the "no priors" argument would typically ascribe uniform prior $p(t) = 1$ over all possible values of t, in the hope of denoting the ideas that "we know nothing about t beforehand", "we are trying to be objective" and "the function $p(t) = 1$ is so structureless that we are letting the data speak for themselves". However, rather than being "no prior", this procedure inadvertently incorporates massive amounts of prior information. Even an analyst who claimed to have complete prior ignorance would be inclined to reject the state of prior belief that is unconsciously applied by this approach.

For example, $p(t) = 1$ can represent a state of prior belief which holds that there is over a trillion times the chance that t is greater than a million, than that it is less than a million. Note that the choice of the parameter t with which to parameterise the family of distributions is completely arbitrary. The family could equally well be parameterised by $\beta = 1/t$. The "no-prior" argument for t implies there is much prior knowledge about β: it is almost definitely less than one, and, in fact, is probably smaller than any positive number you can name. The function $p(t) = 1$ may appear structureless but the implied prior belief distribution on $\beta = 1/t$ is a very spiky function. Engineers should refrain from claiming to apply Bayesian techniques with "no priors", for they are unwittingly applying very outlandish priors.

2.10 Redefining the language of probability

Many commonplace and supposedly-understood terms need to be re-interpreted from the Bayesian perspective. Foremost, as discussed, is the word 'probability'. Much used "confidence intervals" need to be replaced by "highest posterior density regions", although both concern estimation and so have less relevance for prediction. For prediction, the simple notion of return period is central, and it has a different interpretation in Bayesianism: the 1 in 10,000 level event is that event beyond which lies 1/10000th of the predictive distribution for the value of the next event. Significantly, this incorporates uncertainty directly. The more uncertain we are about the underlying parameters and model, the fatter the tail of the predictive distribution and the larger the 1 in 10,000 level event. Rather than seeking a maximum likelihood estimate of the 1-in-10000 quantile and assessing uncertainty by means of confidence levels, the Bayesian view considers the proposition $A =$ "The next event exceeds Q" and seeks the value of Q for which $p(A) = 0.0001$. That value of Q is the 1 in 10,000 level event. As a concrete example, consider the extrapolation of storm surge levels on the UK East Coast undertaken in

Coles [5]. Note in particular the way that Bayesian predictions diverge away from the frequentist estimates for large return periods, reflecting the increased uncertainty in extrapolating beyond the span of the data.

2.11 Who decides?

The obvious adjunct to the inevitability of subjectivity is the question "Who decides?". This is simple: the decision-maker decides. Whosoever is going to take the final responsibility for the successful performance of the structure over its lifetime is the person who should ensure that it is their own subjective opinions that are reflected in the formulation of the prior beliefs.

2.12 Constructing priors

There is a large literature on how to formulate priors, particularly in the face of vague prior information. A few pointers are included here.

One should avoid the use of "obstinate" priors which move only slowly in the face of substantial information. Normal distributions are often used as priors as they are simple and familiar. However, they have light tails, and can thus be quite obstinate. More heavily-tailed priors are generally to be preferred.

Conjugate priors (meaning the resulting posterior will be from the same mathematical family as the prior) are much used, but they have no rational basis: analytical elegance is not reason enough for their use in prudent decision making.

Ignorance priors and probability-matching priors provide half-way houses between Bayesian and frequentist approaches and may be worthy of further investigation by analysts determined to simulate the "no prior" information case.

The rigorous Bayesian view, however, is that priors should be true representations of ones beliefs before looking at the data, and that they can be constructed by a sequence of gedanken experiments and careful self-examination. One can consider many "relative likelihood" possibilities, ("is it as plausible that this parameter is above 1 as below 1 ?", "is it equally plausible the parameter lies between 1 and 10 as between 10 and 100 ?", "the parameter cannot possibly be negative (and so it cannot be normally distributed)", etc.).

Jaynes envisages the use of gedanken betting games, such as "Would you prefer the chance of receiving N pounds if A turns out to be true to the chance of receiving N pounds if the single roll of a fair dice returns a six?". It may sound a little convoluted, but there is no fundamental impediment to the careful construction of distributions that are reasonable representations of ones prior beliefs.

2.13 The easy excuse

When something extreme happens, it is common to hear: "We didn't expect that", or "The chances of that happening were billions-to-one against". Such statements can be a smokescreen, and are hard to argue against since despite their irrelevance they can often be true. Given the lack of a coherent language for debate, ambiguities arise through the lack of specifying what "that" means. One trick is to refer to a very specific set of circumstances so precise and restricted that of course it appears extremely improbable that it could have been anticipated. The statement "ten passengers were killed when a freight train collided with a passenger train that had been derailed by a Land-Rover pulling a trailer whose driver had fallen asleep at the wheel and had careered off an overpass and down an embankment onto the track, and who was trying to ring for help as the first train approached" is very highly specified. The probability of A = "a car will cause the death of more than five rail passengers in the UK in the coming decade" may however be quite appreciable. After-the-fact statements about "billions-to-one" are fatuous. For any A = "this specific event occurs", then **after** the event occurs, $p(A) = p(A|A) = 1$ (it is a certainty - it happened). A statement such as "We did not expect 100,000 people to cross on opening day" is overly specific. A Bayesian interviewer would ask "Before the opening day, what degree of belief would you have placed in the truth of the proposition A = "More than 50,000 people will cross on opening day"? ".

The excuse "We didn't expect that" is too easy. It would be even more untenable if it were found that designers harbour such an excuse at the back of their minds as ready defence in case something later goes awry.

2.14 The imprudence and irrationality of existing statistical methodologies

Consider the archetypal design scenario where a designer knows the magnitudes of the n previous events, and knows the form (but not the parameter values) of the distribution from which they are drawn and needs to design a structure such that there is a small probability against failure for future events. A standard methodology would involve using the data to estimate the parameters using (bias-corrected) Maximum Likelihood, selecting a quantile with a high return period from that distribution, and then augmenting that with confidence intervals. The designer could claim to have acted in accordance with accepted best practice if they then design the structure to withstand an event which has a high confidence level above the estimated high return period event.

To prove that this is irrational and imprudent, consider, for demonstration purposes, the simple case where events are drawn from a uniform distribution $U(0, a)$ with unknown scale parameter a. Consider also the case where the data consists of a single previous observation. The designer may design for the 1 in a 1,000,000 return event, at a 99.99999% confidence level. If the designer habitually repeats this process, then the next event will exceed the design value for around **1 in 8** of those structures [6]. Talk of 1-in-a million and high confidence is thus inappropriate. The actual performance can be considerably less than desired. Perhaps when some designers say "We did not expect that to happen" it reflects genuine surprise that the results of their frequentist techniques were so easily overcome.

There are current statements in the scientific literature that the mean global temperature of the planet in 2100 will be between 1.3 and 6.7°C hotter than it is today, with an implication that for $A =$ "The earth will be more than 7°C hotter in 2100", we have $p(A) = 0$. Most climate **scientists** refuse to ascribe explicit probabilities to propositions such as A, but continue to do so implicitly. Climate prediction and its inherent uncertainties are not only the domain of scientists: engineers are currently designing against extreme wind and flood possibilities, and the Thames Barrier protects Central London from North Sea storm surge. Surprising things may happen in the future, but future analysts may look back and ask whether current generations of scientists and engineers really should have been so very surprised by the way events subsequently turned out. If Central London is devastated by a storm surge in the coming century, or by a horrendous terrorist outrage, will future generations be so accepting of the easy excuse?

2.15 Against Maximum Likelihood

The statistical method known as Maximum Likelihood (including its Bayesian variant - maximum posterior) is one of the most elegant and sensible methods for estimating unknown parameters. However, it has little place in prediction and in the prudent management of uncertainty. One should not consider the distribution with the most likely parameters, but the integral over all possible parameters and their respective probabilities. This is the predictive distribution.

The problem with Maximum Likelihood extends beyond this mathematics: the concept creeps into wider thinking. The phrase itself sounds good - considering what is "most likely" sounds sensible. When considering extremes, however, it is not sensible. In many spheres of risk assessment, scenarios are described, and some engineer will argue at some point that some unknown parameter should be set to its most likely value. One hears "The fuel tanks are likely to be empty", "the driver is likely to slow down", "the Earth's temperature is most likely to rise by around 3°C", "it is most likely that only one train (log/truck/plane/ship/...) will be involved", "the crowd is likely to walk in an orderly fashion", etc. There is an obvious contradiction in considering extreme possibilities whilst arguing for the "most likely". Proper assessment of extreme possibilities requires that the phrase "most likely" be recognised as irrelevant and imprudent. Different, but similar, and worse, is the following: if $A =$ "crowds can synchronise their footfalls to vertical oscillations", then since A is unproven it may be claimed that one does not need to consider the possibility further. However, lack of conclusive evidence for A does not imply $p(A) = 0$.

3 Summary

Current debates about the risk of extreme events are riddled with dubious statements about 'chance'. Many patronising documents have been written about "the public perception of risk" and yet most members of the public have little problem in making subjective probability assessments, as evidenced by recent pre-match betting on propositions such as $A =$"Scholes will score the first goal". I contend that many of the people who least understand risk and probability are the very scientists and engineers who claim to be expert practitioners of its theory.

This paper proposes

- that structural engineers be educated in Bayesian theory and its underlying philosophy, and about the centrality to the prediction problem of the predictive distribution (as opposed to that distribution with parameters of highest posterior density);

- that quasi-scientific frequentist methodologies currently applied by engineers be subject to scrutiny;
- that design-for-robustness offers a more prudent strategy against the unknown future than investment in frequentist-based “reliability” studies.

Until such time as the Bayesian perspective is the norm, designs will be built which await not-to-be-quite-so-unexpected surprises; engineers will have the easy excuse at the ready; clear thinking and coherent debate will be absent, and the rationale behind engineering decisions will be obscured by the myth of objectivity and the frequentist language inherited from historical scientists who had very different aims.

Frequentist methodologies are the wrong approach to the decisions that engineers need to make, decisions that involve assessments of abstract future possibilities based on incomplete and abstract information. Is the Bayesian framework more coherent and more prudent for this task? You decide.

References

[1] Jaynes, E. T., *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

[2] van den Brink, H., Konnen, G., and Opsteegh, T., “Uncertainties in extreme surge level estimates from observational records”, *Phil. Trans. Roy. Soc. Lond., Theme Issue “The Big Flood”*,(to appear 2004).

[3] Deltacommissie, *Rapport Deltacommissie*. SDU uitgevers, Den Haag (in Dutch), 1960.

[4] Coles, S., *An Introduction to Statistical Modelling of Extreme Values*, Springer Series in Statistics, Berlin, 2001.

[5] Coles, S., and Tawn, J., “Bayesian modelling of extreme surges on the UK East Coast”,*Phil. Trans. Roy. Soc. Lond., Theme Issue “The Big Flood”*,(to appear 2004).

[6] McRobie, F.A., *Exact Exceedance Estimators and the* $1/\sigma$ *reference prior*, Cambridge University Engineering Dept Technical Report, CUED/D-STRUCT/TR213, 2004.