CENTRE FOR RESEARCH ON INTRODUCED MARINE PESTS

TECHNICAL REPORT NO. 17

BAYESIAN STATISTICAL INFERENCE IN ECOLOGICAL RISK ASSESSMENT

KEITH R. HAYES

C S I R O

MARINE RESEARCH

SUMMARY

This report is Volume III of a three volume series on ballast water risk assessment. The report examines the use of Bayesian statistical inference in quantified ecological risk assessment. The purpose of the report is to determine whether or not Bayesian statistical methods should be used in the ballast water risk assessment being developed by the Centre for Research on Introduced Marine Pests (CRIMP), for the Australian Quarantine and Inspection Service (AQIS).

Bayesian inference techniques for discrete and continuous variables are illustrated with reference to a ballast water sampling problem (how many samples must be taken to ensure confidence in a negative result), and a journey survival problem (what is the probability that a given species will survive a journey of a specified duration).

The use of Bayesian (as opposed to classical) inference techniques in ecological risk assessment has recently attracted considerable attention. The debate for and against each statistical approach has largely revolved around the following issues:

1. what does probability mean when applied to ecological systems and how should it be calculated;

2. the assumptions that each approach makes when analysing uncertainty about physical and biological parameters;

3. providing ecological risk estimates with complete descriptions of uncertainty;

4. the extent to which each approach is able to deal with complex, multi-parameter models in data sparse situations; and,

5. can the results of each approach be clearly understood by stakeholders who may not be well versed in statistical science.

On reflection there is very little to choose between the two approaches on these grounds. Having said this, it is possible to identify two advantages that Bayesian statistical techniques have over classical approaches:

1. they are able to employ subjective interpretations of probability. This is important to ballast water risk assessment because there is currently no historical database detailing species assemblages under specific ballasting conditions, together with the timing and frequency of successful introductions, that would allow empirical, deductive, risk assessment methods. Analysts seeking to quantify the risks associated with ballast water introductions will have to use inductive risk assessment methods and may therefore need to use subjective interpretations of probability;

2. they immediately direct the analyst to the full distributional qualities of parameter uncertainty, through the posterior distribution function. Furthermore once the posterior distribution has been derived it can be quickly updated as more information becomes available, without having to repeat the assessment from start. Bayesian techniques are therefore well suited to the iterative development of quantitative risk assessment whereby risk estimates are made and then continually updated in light of additional information.

Potential problems remain, however, when specifying the prior probability distribution in data sparse situations. In these situations the likelihood function may be quite diffuse and therefore the prior, even a non-informative one, is likely to exert considerable influence on the shape of the posterior distribution function. Under these circumstances Bayesian risk assessments may be less repeatable than assessments using classical inference techniques.

It is difficult therefore to define when and where Bayesian approaches might be better suited to quantitative ecological risk assessment than more traditional approaches. On balance each case is probably best approached on its merits bearing in mind that:

1. both classical and Bayesian risk assessments require important subjective decisions of the analyst – the extent to which these decisions dominate a Bayesian analysis, however, is dependant on the availability and quality of data;

2. classical inference techniques cannot be used with subjective interpretations of probability, and since this is a valid component of the risk analyst's tool box, Bayesian techniques form an important alternative approach to quantified ecological risk assessment; and,

3. Bayesian statistical inference is well suited to the iterative development of quantitative risk assessment and quickly emphasises the full distributional qualities of uncertain parameters. The results of a Bayesian analysis, however, should be judged in light of the data that was available to the analyst, and the extent to which an independent analyst might arrive at the same (or similar) conclusions.

ACKNOWLEDGMENTS

CONTENTS

1 INTRODUCTION

1.1 Background and objectives

Quantitative Risk Assessment (QRA) can be broadly classified as either deductive or inductive. Deductive risk assessments are usually easier to do: they are based on historical data and employ a well-accepted (frequentist) interpretation of probability. For example there is a plenty of data on road accidents involving pedestrians. The Fatal Accident Frequency Rate (FAFR) can be easily calculated from this data, and the risks to pedestrians calculated accordingly[1].

Risk assessments are harder to conduct when there is no relevant historical record, for example when assessing the risks associated with a new technology. Under these circumstances the analyst must use inductive assessment methods, and may also be forced to seek alternative (eg subjectivist) interpretations of probability. The first risk assessments conducted for nuclear power stations faced exactly this problem. Rasmussen (1981) summarises the problem as follows:

> *"the use of probabilistic risk assessment in large accidents of low probability must employ the logic of the subjectivist (or Bayesian) approach since rarely will enough actual data exist to use the frequentists' definition."*

There are some interesting parallels here with quantitative ecological risk assessment, particularly for ballast water introductions. Ecological risks are usually characterised by one-off events with little, if any, historical precedence. For example there is currently no historical database detailing species assemblages under specific ballasting conditions, together with the timing and frequency of successful introductions, that would allow an accident-frequency approach to ballast water risk assessment. Ecological risks may also be associated with low probability/high consequence events, and again ballast water introductions are no exception:

1. the small number of successful introductions, relative to the many millions of tonnes of ballast water released each day in ports around the world, indicates a very low event frequency. In the United States, for example, the total volume of foreign ballast water entering US waters is approximately 220,100 tonnes per day. The total number of foreign marine species probably introduced through ballast water is 57 (Carlton *et al*, 1995). If we conservatively assume that all introductions occurred within the last ten years, and that ballast water discharges have remained approximately constant in this time, then the event frequency becomes 7.19×10^{-8} per tonne ballast water;

2. non-native species can have dramatic economic impacts. The zebra mussel *Dreissena polymorpha*, for example, is estimated to cost as much as US $300 million a year due to physical obstruction of cooling water intakes and perhaps a total of US $500 million a year in total nuisance costs (Weathers and Reeves, 1996);

---

[1] The simplest model would assume the FAFR to be constant, and make risk estimates on this basis, eg the annual risk of a pedestrian being killed in a road accident is 4.2×10^{-5} (Wilson and Crouch, 1985).

3. non-native species can have equally dramatic ecological impacts. They may prey upon native species, or compete for food and habitat, leading to the displacement and possible extinction of native populations (see for example McKaye *et al* 1985, Iongh and Van Zon 1993, and Welcomme 1992). Non-native species may also cause more subtle ecosystem changes, the full consequences of which may not be apparent for many years.

It is not easy to conduct quantitative risk assessments under these conditions. Deductive approaches are unsuitable, and the frequentist interpretation of probability they employ questionable. Alternative (Bayesian) approaches may therefore be recommended.

The objective of this document is to critically examine the use of Bayesian statistical techniques in quantitative ecological risk assessment, and to comment on their potential use in the ballast water risk assessment being developed by the CSIRO's Centre for Research on Introduced Marine Pests (CRIMP).

The impetus for this analysis dates back to September 1996 when CRIMP began to develop a quantitative risk assessment framework for ballast water introductions on behalf of the Australian Quarantine and Inspection Service (AQIS). The first stage of the project required a comprehensive review of ecological risk assessment methods employed within various disciplines, but particularly those relating to biological introductions and invasions. The risk assessment review was completed in April 1997, and subsequently published as a CRIMP technical report (Hayes, 1997). This report made a number of recommendations, including:

1. the ballast water risk assessment framework should be modelled on the Quantitative Risk Assessment (QRA) paradigm used by the chemical and nuclear process industries, and the import risk assessment framework advocated by the Office International des Epizooties;

2. QRA is an iterative process that improves with use. The ballast water risk assessment framework should emulate this and provide increasingly accurate risk estimates as more data are made available to the analyst; and,

3. Bayes theorem is a statistical way of updating estimates of uncertainty in light of new information. It should therefore be investigated as an alternative statistical approach to ballast water risk assessment.

The development of the risk assessment framework took place subsequent to this review. The framework documentation is divided into three volumes. Volume I summarises the approach, the analysis and the data requirements of the risk assessment, but only mentions Bayes theorem in passing. Volume I was published as a CRIMP technical report in March 1998 (Hayes and Hewitt, 1998). Volume II provides a detailed description of the modules that are built into the framework, culminating in a demonstration model for ballast water risk assessment. Its development is currently on going. Volume III (this document) completes the series by examining Bayesian statistical inference techniques, and the extent to which they should be adopted within the risk assessment framework.

1.2 Scope and structure of the report

The purpose of this document is to examine the use of Bayesian statistical inference within ecological risk assessment. It does not provide an in-depth mathematical treatise of Bayesian methods. Accordingly the mathematical treatment of Bayes' theorem has been kept deliberately simple, restricted in the main part to univariate analysis. The reader is referred to the numerous textbooks on Bayesian methods if he or she wishes to take the statistical theory further than that presented here.

Chapter 2 discusses probability and the role it plays in describing uncertainty within the risk assessment process. Bayesian methods adopt a fundamentally different approach to statistical inference and encourage alternative interpretations of probability. Chapter 2 explores these interpretations, and contrasts Bayesian statistical methods with some classical approaches to statistical inference. A variation on the basic Bayesian approach – Empirical Bayes - is introduced here.

Chapter 3 describes Bayes' theorem and its application to discrete and continuous data sets. This is illustrated with reference to a ballast sampling problem and a simple journey survival model. This chapter also examines the role played by conditional probability and the likelihood function in Bayes theorem, and discusses the derivation of conjugate, non-conjugate, informative and non-informative prior distributions.

Chapter 4 explains why Bayesian methods might be attractive to risk analysts. The use of Bayes' theorem in ecological risk assessment is examined, together with the current debate between advocates of Bayesian statistical inference and those that adhere to the classical paradigm. This chapter does not detail the finer statistical points of this debate, but rather examines the basic premises which underline Bayesian methods, the arguments for and against these, and their practical implications from a risk assessment perspective.

Chapter 5 summarises the preceding discussion, highlighting the case for and against the use of Bayesian statistical inference in ballast water risk assessment.

Much of the mathematical and statistical detail is omitted from the main body of the text but included in the Appendix.

1.3 Notation

Random variables are represented by capitals, such as X or Y. Values taken by these variables (the data) are represented by x or y. A vector of n observations is represented as $\mathbf{y} = (y_1, y_2, y_3......y_n)$, where the subscript denotes individual observations.

P( ) denotes the probability of a particular outcome or event. A letter will be used in the parenthesis to refer to the outcome or event in question. If this probability is conditional upon a second event or outcome, then this is denoted P(/). The letter after the slash denotes the conditioning event. For example $P(x/H_0)$ denotes the probability of the data x, given a null hypothesis H_0.

The probability mass or density function that assigns probability to values of a discrete or continuous variable is denoted $p(x)$. In both cases $F(x)$ signifies the cumulative distribution function. The joint probability distribution of two or more variables is denoted $p(x, y)$. The terms 'density' and 'distribution' are used interchangeably. Prior probability distributions are denoted by the subscript 0, for example $p_0(x)$.

The parameter(s) that characterise a probability mass or density function are generically denoted by θ. It is common therefore to write $p(y/\theta)$ to signify that the probability function is conditional on the parameters of the distribution. The probability of the parameter given the data is written $p(\theta/y)$. When considered as a function in θ, this distribution is called the likelihood function, written $l(\theta/y)$.

2   PROBABILITY, UNCERTAINTY & RISK ASSESSMENT

2.1 The nature of probability

Defining probability

Most quantitative expressions of uncertainty and risk involve the use of probability. Unfortunately probability is difficult to define beyond what one might intuitively understand of the term 'probable'. The three most commonly accepted definitions are:

1. the classical, equal-likelihood definition - if an event can occur in n mutually exclusive and equally likely ways, and if n_A of these have the attribute A, then the probability of A is the fraction n_A/n. This definition usually derives from a physical understanding of the system in question. For example the probability of selecting any of the hearts from a deck of cards is 0.25. This is verified by recognising that there are 13 hearts in a deck of 52 cards, each of which is (assumed) equally likely to be selected.

2. the frequentist approach defines the probability of event A as the number of times A occurs (n_A) in the total number of repetitions (n) of an experiment or trial. Under certain conditions[2], the relative frequency n_A/n tends to a limiting value called the probability of A. This view is empirical; it derives from past experience rather than a physical understanding of the system in question. For example a works manager may estimate the probability of a production line item being defective as 0.01, because he has observed that 1 in every 100 items produced by the line have been faulty in the past, rather than through a detailed understanding of the way in which the production line works.

3. the subjective definition, commonly associated with Bayesian methods, interprets probability as a rational expression of an individual's degree of belief. This definition is more flexible than the others because it does not rely on the notion of equal-likelihood, or the repeated iteration of trials. For example a businessman may estimate the probability that his next venture will be a success, purely on the basis of his experience (Hampton *et al*, 1973). For the very same reasons, however, this notion of probability is often viewed with a certain degree of scientific scepticism[3].

There are, however, numerous other approaches to probability. For example Good (1959) identifies three other types of probability and distinguishes physical probability – the probability of a success given the experimental set-up – from the attempted proof of this in terms of long-run frequency. This distinction seems useful in light of the Mendelian theory of gene frequency, in which the probability of a genetic trait is an inherent property of the system (McPherson, 1990).

---

[2] The conditions are that each iteration of the experiment is performed in an identical manner, and the outcomes of the experiment are physically independent of one another.
[3] Proponents of this approach, however, point to certain consistency criteria that ensure mathematical rigour in the derivation and application of subjective probabilities (see Freund, 1971 p. 39).

Anderson (1998) recognises at least four other probability concepts (including knowledge, confidence, control and plausibility) and emphasises the distinction between probability statements that are internal to the observer, such as the degree of confidence in a hypothesis, and those that are external to the observer, such as the chance or frequency of an outcome in a random process. Edwards (1992) maintains that this distinction is fundamental, refusing to accept any definition of probability that is not external to the observer, ie does not involve a random choice from a defined population or the generation of events by a chance set-up.

In practise the mathematical theory of probability (and statistics) relies upon the agreed axioms of probability (refer to Appendix A), and not on philosophical arguments regarding the meaning of the term. However, questions of meaning are often used to discredit the use the of one statistical paradigm in favour of another, particularly in ecological science. For example Power (1996) suggests that the classical definition of probability has no practical meaning to ecological risk assessment because the analyst is rarely able to numerate all the possible outcomes of an event within complex ecological systems. The frequentist definition may also be inapplicable because it is rare to find repeated "experiments" of an ecological event. Hampton et al (1973), Smith (1984), Holdway (1997) and Suter (1993) express similar sentiments.

By contrast, the subjective interpretation of probability is inherently flexible, and places no restriction on its application. It does, however, bring into question the objective value of assessments made on this basis (but see below). Power (1996) warns that once a degree-of-belief interpretation is allowed in situations where there is very little external evidence (such as the incidence of premature deaths caused by failures at a nuclear power station), then one must expect different conclusions from person to person. This raises questions of scientific validity and has obvious implications for the practical application of ecological risk assessment.

In summary it is only safe to say that a risk analyst may interpret probability in a variety of ways. None of these, however, should necessarily be viewed as mutually exclusive (Cordue and Francis, 1994). Ideally the problem at hand will allow a classical or frequentist interpretation or, at the very least, the conception of a "chance" probability model that is external to the observer. In cases where this is not possible, however, the practising risk analyst may have to use a subjective interpretation, but at the same time be aware of its implications. The analyst should also note, however, that classical statistical inference techniques cannot use subjective interpretations of probability.

Conditional probability

Irrespective of which definition the analyst employs, it is important to realise that probability cannot be isolated from the circumstances under which it is derived. It is absolutely dependent on the data available to the analyst and the assumptions that he or she makes. In other words all probability is conditional upon the underlying conditions.

The probability of an event A, given the conditions of the experiment H, or the occurrence of an event B, is written as $P(A/H)$ or $P(A/B)$ respectively. Many Bayesian authors (see for example Good 1959, Lindley 1965, or Schmitt 1969) go to great lengths to emphasise the conditional nature of probability, insisting that the conditioning events be explicitly recognised on each occasion. This is largely because of the key role played by conditional probability within Bayesian methods. In standard statistical texts, however, it is common practise to abbreviate the notation to $P(A)$ when the conditional circumstances are clearly understood.

To elaborate on the concept of conditional probability consider Table 2.1. This summarises accident statistics, occurring over the period of a year, on two hypothetical oil and gas platforms (H and F) operating on Australia's Northwest shelf.

Table 2.1   Accident statistics for two hypothetical offshore oil and gas platforms

Accident type	Platform H	Platform F	Total
Serious*	2	8	10
Minor	17	43	60
Total	19	51	70

*A serious accident might be defined as one requiring more than 3 days off work

Suppose an accident investigator is to select one of these incidents at random for further investigation. It can be seen that the probability of selecting a serious accident, denoted P(S), is 0.143 (10 serious accidents out of a total of 70). If the investigator is sent to platform H then the probability of selecting a serious accident, denoted P(S/H), becomes 0.105 (2 serious accidents out of a total of 19). Note how, in this instance, the sample space shrinks from a total of 70 accidents on both platforms, to only 19 - those occurring on platform H.

Generalising from this observation allows the following definition: The conditional probability of an event S given the condition H is given by

$$P(S\,/\,H) = \frac{P(S \cap H)}{P(H)} \qquad\qquad P(H) \neq 0 \qquad\qquad (2.1.1)$$

where; P(S ∩ H) = the function of both S and H occurring
        P(H)   = the marginal probability of H.

In this example P(S ∩ H) is 0.029 (2 out of 70), and P(H) is 0.271 (19 out of 70), giving a conditional probability P(S/H) of 0.105, ignoring differences due to rounding error.

Note how division by the marginal probability P(H) acts as a normalising factor to ensure that the proportion of serious and minor accidents on any given platform sums to unity, thereby ensuring consistency with the second axiom of probability.

Probability and objectivity

When faced with uncertainty, most scientists turn to classical statistical techniques for an 'objective' solution The idea that subjective probability should be the key to the rational treatment of uncertainty is often decidedly unpalatable (Smith, 1984). By implication scientists often distrust Bayesian methods because of their extensive use of subjective probability.

In reality most practical applications of probability entail some form of subjective input, and this is usually carried over into the statistical analysis. The classical and frequentist notions of probability require a subjective choice of null hypothesis and significance levels (Ludwig, 1996), 'plausible symmetries' and 'repeated iterations' (Smith, 1984). Even the simplest of classical hypothesis tests involve quite fundamental subjective choices in the design and termination of experiments (Berger and Berry, 1988). More generally all probability-based inferences rely on a statistical model, the choice of which is largely subjective.

Proponents of subjective probability see it as a means for introducing the beliefs of the analyst in a transparent manner. Bayes theorem provides a vehicle for adjusting these beliefs in light of the evidence (data), in a manner that is both logical and rigorous. Lindley (1965) therefore believes that if two scientists differ in a question of science, rather than taste, then it is because they have accumulated different evidence. Bayesian methods would allow them to pool their evidence and come to some mutual agreement. Box and Tiao (1973) demonstrate this point, showing how two physicists in disagreement over the value of some physical constant, would come closer in belief following a single (presumably jointly planned and executed) experiment, using Bayesian inference techniques.

None of this, however, provides particularly compelling evidence for or against the use of Bayesian methods in science or risk assessment. Probability and statistical methods are usually no more objective than the analyst that employs them. Quantitative Risk Assessment is therefore no more objective than qualitative risk assessment for having used probability and statistical techniques (classical or otherwise). The strength of quantitative risk assessment, as in science, lies not in its objectivity but rather in the way it exposes subjective input.

2.2 The role of probability in risk assessment

Risk and risk assessment

Risk is traditionally characterised by the occurrence of accidental events that have undesired effects. Risk assessment is the means by which the frequency and consequences of these events are determined, such that

$$Risk = \frac{Event}{Time} \times \frac{Consequences}{Event} = \frac{Consequences}{Time}$$

This traditional, engineering approach to risk assessment, leads to risk functions that describe accidental events in terms of the frequency of consequences. The consequences in this approach are often expressed in terms of human injuries or fatalities (the risk assessment endpoints).

Kaplan (1997) adopts a more flexible definition, emphasising that risk is defined not as a number, a curve, or a vector, but by three questions: What can happen? How likely is that to happen? If it does happen, what are the consequences? The answer to these questions constitutes a triplet [S_i, L_i, X_i] where S_i denotes individual risk scenarios, L_i denotes the likelihood of the ith scenario and X_i the consequences of this scenario.

Uncertainty regarding the likelihood of risk scenarios, together with uncertainty regarding the type or magnitude of their consequences, means that these components should be expressed in probabilistic terms, denoted [S_i, $p(\varphi_i)$, $p(X_i)$]. The definition of risk is completed by identifying the complete set of possible risk scenarios such that

$$Risk = \left\{ S_i, p(\varphi_i), p(X_i) \right\}_c .$$ (2.2.1)

This approach encourages a broader interpretation of risk and is therefore better suited to ecological risk assessment, where the events in question may not be 'accidental' in any sense, nor the endpoints restricted to human fatality or injury.

Uncertainty and probability

Uncertainty plays a critical role in risk assessment. The distinguishing feature of ecological risk assessment is that it acknowledges uncertainty in the type, likelihood and magnitude of environmental impacts. By contrast environmental impact assessment simply identifies what components of the environment might be affected by the activity in question. It does not quantify the likelihood of an impact, nor the assessors uncertainty regarding the magnitude of impact.

Uncertainty comes in many forms. Not all of these are reducible (Faber *et al*, 1992) or equally amenable to analysis. Figure 2.1 summarises the types of uncertainty that occur in ecological risk assessment. This document is primarily concerned with uncertainty regarding empirical quantities. This is because it is concerned with the types of uncertainty that can be expressed in probabilistic terms, and probability is only appropriate to uncertainty about empirical quantities (Morgan and Henrion, 1990), and model credibility. There are other types of uncertainty within risk assessments, but it is inappropriate to examine these in probabilistic terms.

Empirical quantities represent properties of ecological systems. To be empirical these quantities must be measurable, at least in principle. In other words they can be said to have a correct value, as opposed to an appropriate or good value (Morgan and Henrion, 1990).

Ecological risk assessment uses models of ecological systems. Empirical quantities within these systems are represented by model variables. Uncertainty regarding these empirical quantities, and the variables used to model them, occurs because:

1.   ecological systems are variable; and,

2.   random and systematic errors occur when the properties of these systems are measured.

Almost all empirical (ecological) quantities are variable; the flow in a river, for example, will vary from hour to hour and from day to day. The analyst will therefore be uncertain about the dilution of contaminants discharged into this river, hence there is a risk that the concentration of these contaminants will exceed some specified level - a typical risk assessment problem.

Uncertainty regarding empirical quantities can be described by collecting data on the quantity, and using this to define an empirical distribution function, or to fit a theoretical probability distribution using statistical inference techniques. Fitting an appropriate theoretical distribution to ecological data may be more difficult than defining an empirical distribution. Theoretical distributions, however, are generally better because they smooth out the data, allow the generation of values outside the observed range and require less computer memory (Law and Kelton, 1991).

Figure 2.1   Uncertainty in risk assessment and the role of probability
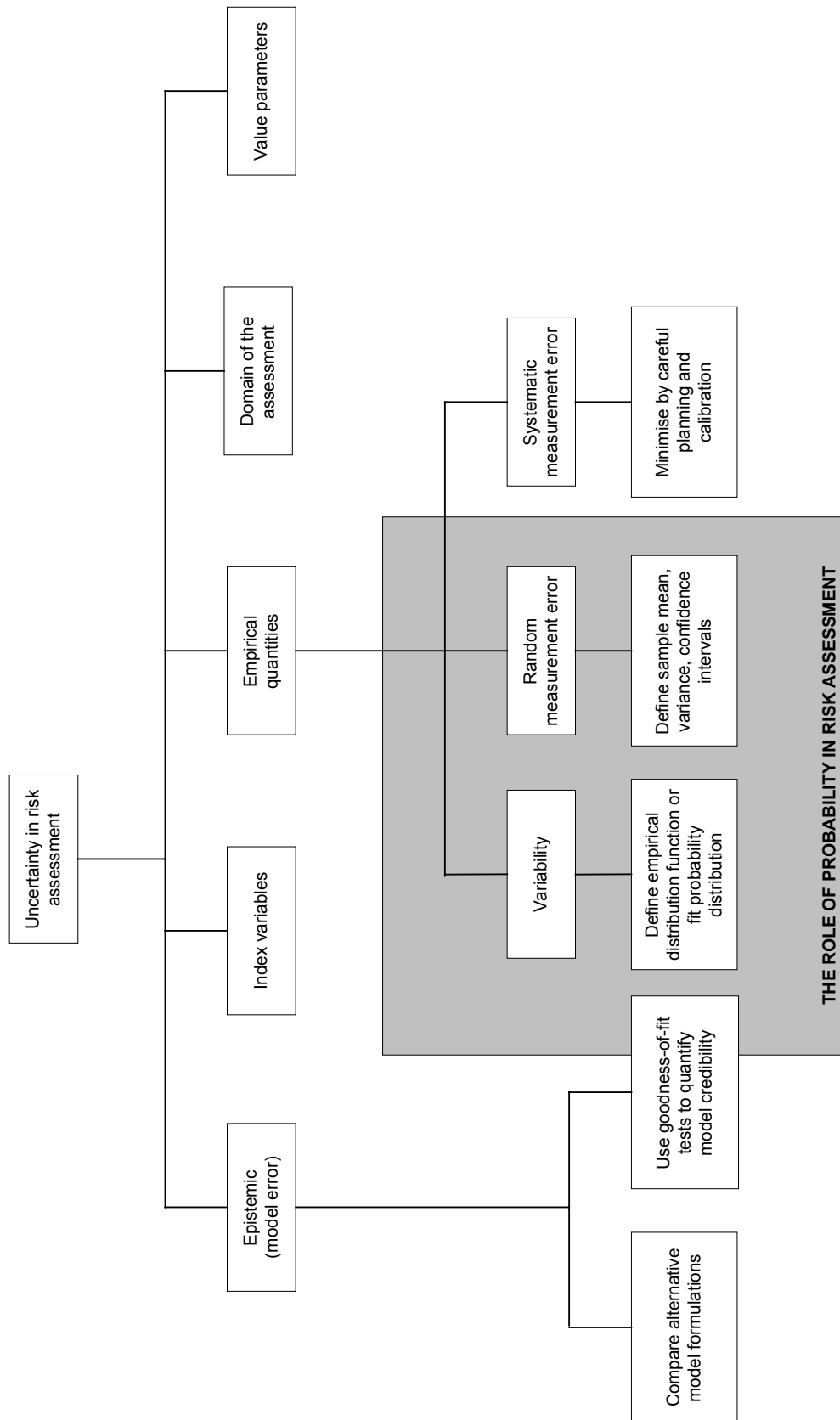
The other source of uncertainty regarding empirical quantities is measurement error. Measurement error occurs because measurements of empirical quantities are not exact: the measuring instrument and the observer are inevitably imperfect. Measurement error has a random component and a systematic component. The random component depends on the variations between observations and the number of observations, and is usually expressed in terms of a sample variance or confidence intervals around the sample mean. This component of measurement error can usually be minimised by taking additional measurements.

The systematic component is defined as the difference between the correct value of the quantity of interest, and the value to which the mean of the measurements converges as more measurements are taken (Morgan and Henrion, 1990). Systematic error can be minimised by careful experimental design and instrument calibration, but cannot be reduced by taking additional measurements. Estimating the magnitude of systematic error may therefore involve some degree of subjective judgement.

Another important source of uncertainty in ecological risk assessment is our incomplete understanding of the biological, physical or anthropogenic systems that are being modelled. This has been termed fundamental or epistemic uncertainty (Pate-Cornell, 1996). Epistemic uncertainty expresses itself as model error - the most intractable of all the potential errors in a risk assessment. Model error can occur in the boundaries, structure and components of a model, and also in the types of probability distributions used to represent uncertain empirical quantities. Analysts are generally aware, before the fact, that models are approximations of reality. The error this causes, however, is only apparent after the fact, and cannot be addressed in a predictive manner. Choosing a suitable model can be further complicated by contradictory data sets (Schnute and Hilborn, 1993) or when several models fit the data reasonably well but yield very different predictions (Moore 1996, Kot *et al* 1996).

The accuracy of a model is usually determined by comparing the model's predictions with reality (ground-truthing) or by comparing the results of alternative models that use different methods and/or assumptions (Suter 1993). Reckhow and Chapra (1983) propose three model validation rules:

1. test the model against data that reflect conditions other than those under which the model was calibrated;

2. employ a statistical goodness-of-fit test to quantify the extent to which the model's predictions are corroborated by observation;

3. prepare alternative model formulations and then base the model selection on the basis of their performance in the tests above, and their consistency with theoretical system behaviour.

Probability is used in the goodness-of-fit tests advocated by these authors, but only to test model accuracy. Probability should not be seen as an attribute of a model; every model is necessarily an approximation of reality, and therefore definitely false (Morgan and Henrion 1990). It is meaningful to say that one model is better than another, in that it produces more accurate predictions. It is misleading, however, to say that model Y has a probability X (Anderson, 1998).

There are three other potential sources of uncertainty in ecological risk assessment:

1. index variables – used to describe spatial or temporal components of a model, such as a particular location, month or year;

2. model domain parameters – used to define the scope, resolution and boundaries of a model, usually by specifying the range and increments of index variables; and,

3. value parameters – used to represent the preferences of decision makers, stakeholders or the general public. For example acceptance criteria for risk estimates.

None of these, however, are amenable to analysis using probability.

To summarise, probability plays two main roles in ecological risk assessment: describing the uncertainty in variable, empirical (ecological) quantities, and describing the random measurement error that occurs when analysts attempt to measure these quantities. Probability may also be used to determine the goodness-of-fit between model predictions and actual measurements, and thereby the credibility of alternative model forms.

This report will examine the Bayesian approach to the selection of probability distributions to represent variable quantities and the selection of parameters that characterise these distributions.

2.3 Contrasting classical and Bayesian statistical inference

Classical approach

The classical approach to statistical inference considers the parameters[4] of a study to be unknown constants. The value of a parameter is estimated in light of the data, which is considered variable (the outcome of a random process). Statistical inferences are then made by comparing this estimate with sampling distributions which express how likely this estimate is for different values of the parameter. The classical approach typically involves the following steps[5]:

1. collect the data x from a random process;

2. build a hypothetical model of the random process $p(X/\theta)$;

3. use the data to build a test statistic $s(x) = \hat{\theta}$ to estimate θ;

4. the model $p(x/\hat{\theta})$ is used to estimate hypothetical sample-space properties (the sampling distribution) of the statistic s(x), such as how variable the statistic would be under repetitions of the random process;

5. use the model to test how extreme the test statistic is (using confidence intervals or hypothesis tests), and therefore how good an estimate of θ it is.

---

[4] Parameters are usually understood to be the central objective of any statistical analysis. For the purposes of this discussion they can be thought of as the parameters that characterise a probability distribution (the moments of the distribution), such as the mean or variance.
[5] Refer to Appendix B for an example of this approach.

This approach follows from the classical and frequentist definitions of probability adopted by the classical school of statistical inference. To a classical statistician it is meaningless to speak about 'the probability of a parameter value'. Any particular value either is or is not the true one, and probability statements are restricted to how likely an estimate of this parameter is, given the observed data (Walters and Ludwig, 1994).

An important outcome of this approach is that confidence intervals for parameter should be interpreted with care. Confidence intervals are sometimes mistaken as probability statements about the value a parameter can take. The correct interpretation, however, is that if the analyst were to repeat the experiment many times, each time calculating a $1-\alpha$ confidence interval, then the intervals generated are expected to include the true parameter value $(1-\alpha)$ percent of the time.

The results of a classical hypothesis test should also be interpreted carefully. Typically a classical hypothesis test calculates a significance probability under the assumption that the null hypothesis is true (refer to Appendix B). The significance probability is the probability of observing the data conditional on the null hypothesis, ie $P(x/H_0)$. It does not, however, describe how probable the null hypothesis is - $P(H_0/x)$, or how likely an alternative hypothesis might be – $P(H_1/x)$, or necessarily imply that the alternative hypothesis produced the observed data (Ellison, 1996).

Bayesian approach

The Bayesian approach is fundamentally different: the parameters of the study are considered as random variables and the observed data are fixed. Bayes theory is then used to adjust the analysts prior belief in the value of a parameter in light of the data. Statistical inferences are then made on the basis of the analyst's subsequent (posterior) belief in the value of the parameter.

This approach is possible because the Bayesian school of statistical inference permits the subjective definition of probability as a measure of an individual's degree of belief. It is therefore perfectly acceptable to talk about the probability of a particular parameter value as being the correct one, provided we are careful about how this probability is calculated in light of the observed data (Walters and Ludwig, 1994). Note, however, that the Bayesian approach does not preclude classical or frequentist interpretations of probability.

The Bayesian approach can be regarded as a process through which the analyst updates his prior beliefs in the value of a parameter in light of the evidence provided by the data (Pascual and Kareiva, 1996). Instead of single estimates of the parameter value, the Bayesian approach assigns probability to a wide range of parameter values through the posterior probability distribution (see below). As a result credibility intervals, the Bayesian equivalent to confidence intervals, have the type of interpretation that is commonly, but mistakenly, associated with confidence intervals; a $1-\alpha$ credibility interval states that there is a $(1-\alpha)$ probability that parameter value falls within the boundaries of the interval.

Table 2.2 summarises some of the fundamental differences in the way classical and Bayesian techniques interpret probability and use statistical techniques.

Table 2.2   Fundamental differences between classical and Bayesian approaches

Concept or term	Classical interpretation	Bayesian interpretation
Probability	Proportion of equally-likely components of sample space with a specified attribute, or the result of an infinite series of trials conducted under identical conditions	The observer's degree of belief, or organised appraisal in light of the evidence (data)
Data	Random sample from an underlying population	Reflection of a random world
$(1-\alpha)\%$ CI	This interval will include the true value of a given parameter in $(1-\alpha)\%$ of all possible samples	Probability that parameter value falls in this interval is $(1-\alpha)$
Hypothesis test	$P(x/H_0)$	$P(H/x)$

(adapted after Ellison, 1986)

Empirical Bayesian approach

The Empirical Bayesian (EB) approach is a variant of the Bayesian approach - it uses the Bayesian mathematical framework, but dips into the classical school of statistical inference in order to estimate the parameters of the analyst's prior belief.

The analyst's prior belief about the value of a parameter is described through his or her prior probability distribution for that parameter. This distribution, however, is characterised by its own parameters. The analyst must specify the parameters of his prior distribution in order to proceed. The essence of the empirical Bayes approach is that the analysts prior distribution is given a (classical) sampling interpretation and estimated from data (Solow and Gaines, 1995).

Empirical Bayes techniques become possible when data are generated by repeated executions of the same type of survey, trial or random experiment, which are similar to, and precede, the one for which inferences are to be drawn. These 'prior experiments' provide the basis for a classical estimation of the parameters of the prior probability distribution subsequently employed in a Bayesian approach. Johnson D. H.(1989) for example demonstrates how historical survey data (of wildfowl populations) can be used in an empirical Bayes analysis to improve estimations of abundance based on the most recent survey result.

Strictly speaking the empirical Bayes procedure is restricted to situations in which there is substantial prior information in the form of 'prior experiments'. Otherwise the procedure violates the assumption of conditional probability upon which Bayes theorem is based, namely that the prior distribution depend only on its parameters and not upon the data (Press, 1989). Empirical Bayesian techniques are therefore only relevant to ecological risk assessment when there is substantial prior information.

The classical, Bayesian and empirical Bayesian approaches to statistical inference are summarised in Figure 2.2

Figure 2.2   The classical, Bayesian and empirical Bayesian approaches to statistical inference
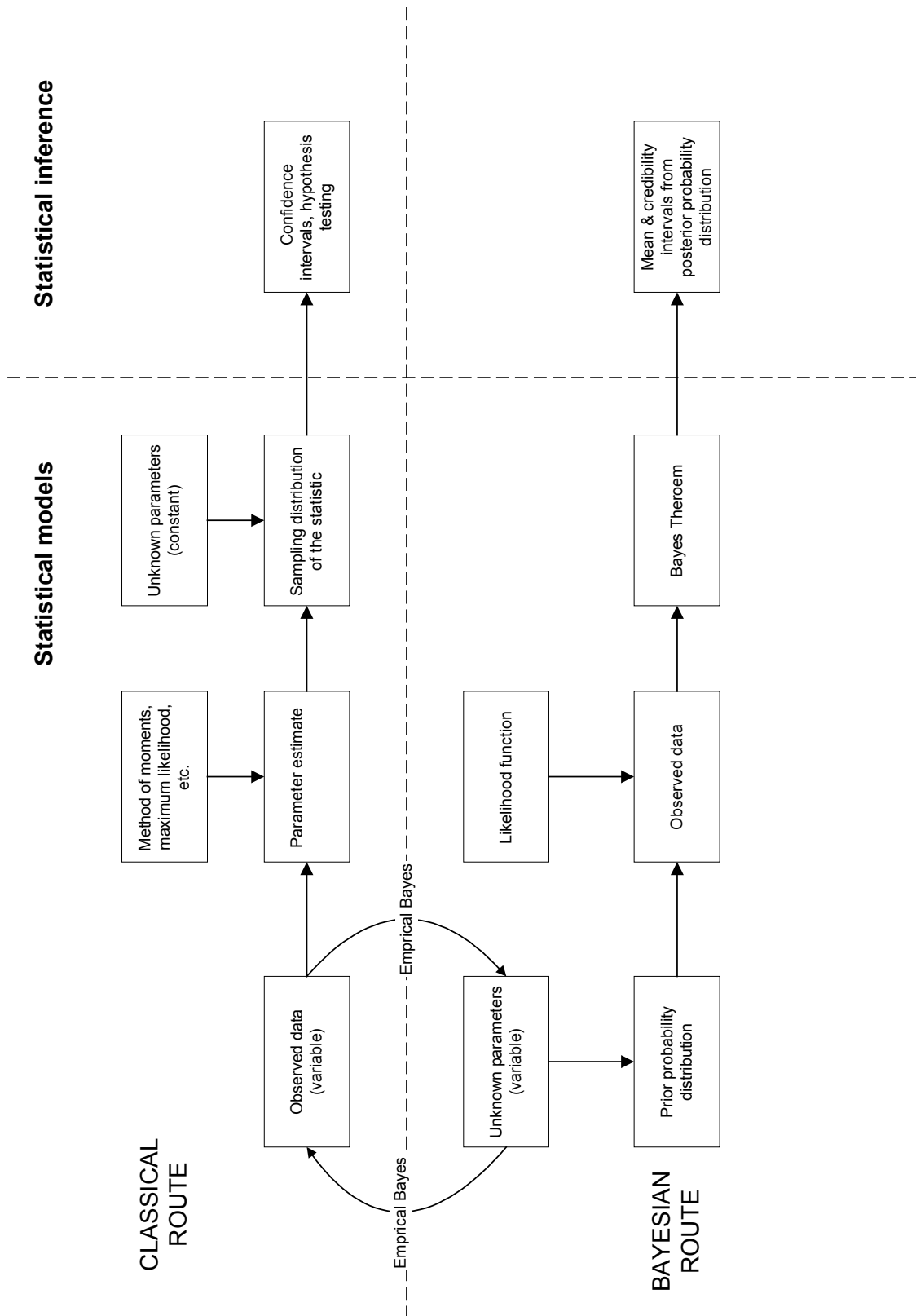
3   BAYESIAN STATISTICAL INFERENCE

3.1 Bayes theorem

Bayes theorem describes the way an analyst should update his or her prior beliefs in the value of a parameter, in light of the data. The theorem can be stated in words as

$$
\begin{array}{c}
\textit{Probability of a} \\
\textit{parameter value} \\
\textit{given the data}
\end{array}
\;=\;
\dfrac{
\begin{array}{c}
\textit{Probability of the} \\
\textit{data given the} \\
\textit{parameter value}
\end{array}
\;\;\textit{x}\;\;
\begin{array}{c}
\textit{Prior probability of} \\
\textit{the parameter} \\
\textit{value}
\end{array}
}{
\begin{array}{c}
\textit{Total probability of} \\
\textit{the data}
\end{array}
}
$$

The probability of a parameter value given the data is referred to as the posterior probability. This distinguishes it from the prior probability held be the analyst prior to the collection and analysis of the data.

Discrete variables

The more formal definition of Bayes theorem follows from the definition of conditional probability given by (2.1.1), allowing for a slight change in notation

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad \text{and} \qquad P(B/A) = \frac{P(B \cap A)}{P(A)} \qquad .$$

Since $P(A \cap B) = P(B \cap A)$

$$P(A \cap B) = P(A/B)P(B) = P(B \cap A) \qquad .$$

Bayes theorem follows easily

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \qquad , \tag{3.1.1}$$

where; $P(B/A)$ = posterior probability of B given A
$\quad\quad\;\; P(A/B)$ = the conditional probability of A given B
$\quad\quad\;\; P(B)$ = the prior probability of B
$\quad\quad\;\; P(A)$ = the marginal (total) probability of A.

We can also write

$$P(A) = \sum P(A/B)P(B) \qquad ,$$

where the summation is over all permissible values of B.

The marginal probability of A is a 'normalising' constant which ensures that the posterior probability integrates or sums to one (in exactly the same way as 2.1.1). Omitting this constant yields the unnormalised posterior probability

$$P(B/A) \propto P(A/B)P(B) \quad . \tag{3.1.2}$$

A key aspect of Bayesian analysis is the ease with which previous knowledge may be updated as new information becomes available. Given a prior probability P(B), and an initial observation A_1, Bayes theorem states that

$$P(B/A_1) \propto P(A_1/B)P(B) \quad . \tag{3.1.3}$$

If a second observation A_2 is made independently of the first then

$$\begin{aligned} P(B/A_2, A_1) &\propto P(A_2/B)P(A_1/B)P(B) \\ &\propto P(A_2/B)P(B/A_1) \quad . \end{aligned} \tag{3.1.4}$$

The expression (3.1.3) is the same as (3.1.4) except that $P(B/A_1)$, the posterior probability for B given A_1, plays the role of the prior distribution P(B) for the second sample. This process can be repeated any number of times, with the posterior probability playing the role of the prior for the next set of calculations - such that today's posterior is tomorrow's prior.

An alternative approach is to treat a set of multiple observations as one 'super-observation', define the likelihood function for this set, and proceed normally (Schmitt, 1969). In this way Bayes theorem is capable of utilising a frequentist interpretation of probability derived from a number of independent and identical trials.

A simple ballast water example

Ballast water management suffers from a recurrent problem: it is difficult to obtain representative samples of ballast water in order to test for the presence or absence of a particular species. Access to a ballast tank is usually restricted, and in most cases samples can only be taken from a limited number of locations, such as a sounding pipe or open deck cover.

Access restrictions of this type determine the type of sampling equipment that can be used. Plankton nets, for example, can be used from a deck cover but not a sounding pipe. Furthermore there is evidence to suggest that some sampling techniques are better at detecting certain species than others. For example a recent review of ballast sampling methods (Sutton *et al*, 1998) concluded that mobile zooplankton (such as crab zoea) are under-sampled by methods that rely on low flow rate pumps. As a result some ballast water samples may suffer from a large proportion of false negative readings; the sample indicates that the species is not present in the ballast water when in fact it is. For the purposes of this example assume that a particular sampling procedure gives a false negative reading for a particular species[6], on 30% of occasions.

In some instances it is also difficult to correctly identify the species found in a sample of ballast water. This is because some taxa, such as bivalve larvae, are taxonomically indistinguishable unless reared to a juvenile or adult stage. Therefore there is a finite possibility of a false

---

[6] False negative readings occur for many reasons, including species behaviour (phototaxis for example), and are therefore likely to be species specific.

positive result; the sample is incorrectly identified as containing a particular species, when in fact it does not. For the purposes of this example assume that the identification of a sample gives a false positive reading on 5% of occasions.

Figure 3.1 summarises this sampling problem with a simple event tree. First the ballast tank is either infected with the species concerned (I) or it is not (U). Second the sample identification procedure may give a positive result (Pos), indicating that that the species is present, or a negative result (Neg) indicating that it is absent. Table C1 (Appendix C) summarises the conditional probabilities associated with this example.

From a risk assessment perspective, it is important to know the probability that the ballast tank is infected, given that the sample has produced a negative reading. The analyst's prior belief regarding the infection status of this vessel may be based on:

1. no previous knowledge, in which case he or she may only be willing to say that there is a 50:50 (equi-probable) chance that the vessel is infected;

2. the past frequency of infection, assuming this vessel has been sampled on a number of previous occasions (under similar conditions);

3. a quantified assessment of infection risk.

In any case, Bayes theorem states

$$P(I\,/\,Neg) = \frac{P(Neg\,/\,I)P(I)}{P(Neg\,/\,I).P(I) + P(Neg\,/\,U)P(U)} \qquad . \tag{3.1.5}$$

Table C2 (Appendix C) illustrates the calculation with a prior infection probability of 0.5. This shows that considerable uncertainty regarding the vessel's true infection status remains after having taken one sample. This is largely due to the poor power of the sampling procedure, reflected by the high proportion of false negatives.

This uncertainty can be reduced by taking additional (independent) samples, in each case using the posterior probability generated by the previous sample as the prior probability for the next (Table C3). Table 3.1 summarises the reduction in uncertainty that follows. Under the conditions of this example, four negative samples are needed to be 99% sure that the vessel is uninfected. In practise, however, determining the probability of a false negative and/or false positive reading is likely to be more difficult than implied here. This approach also assumes that the probability of a false negative sample remains the same between samples. This may not always be the case. A negatively phototactic species for example, is less likely to be captured in later samples once the ballast tank deck cover has been removed.

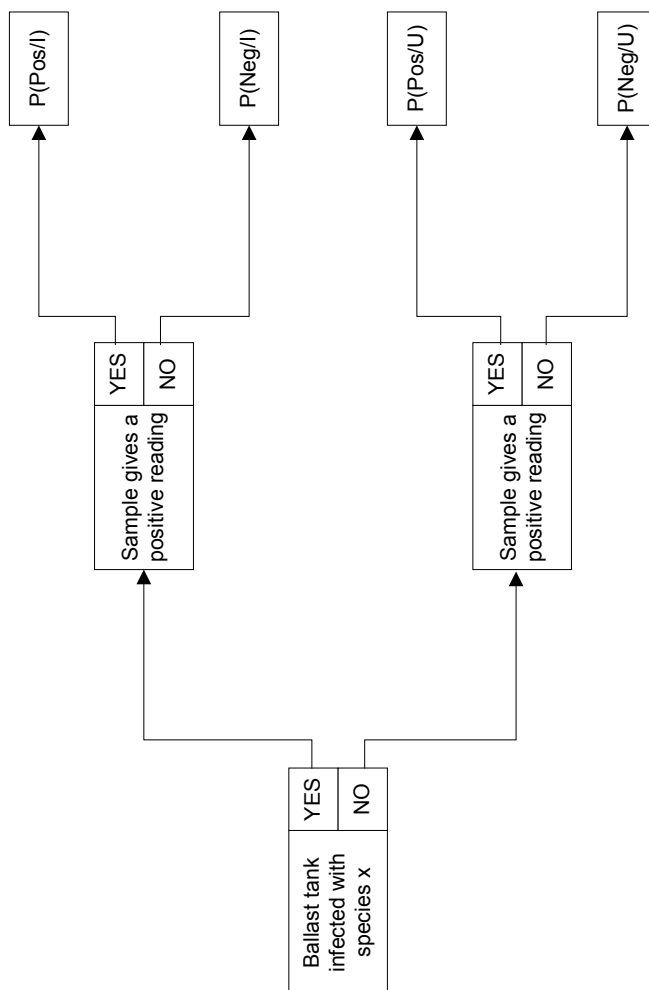Figure 3.1   Event tree for ballast water sampling problem

Table 3.1   Reducing uncertainty by collecting more samples.

Number of samples	Probabilities	
	I (vessel infected)	U (vessel uninfected)
0	0.500	0.500
1	0.240	0.760
2	0.091	0.909
3	0.031	0.969
4	0.010	0.990

Continuous variables

Probability density functions describe the random behaviour of continuous variables. Bayes theorem is applied in the same way

$$p(\theta / y) = \frac{l(y / \theta) p(\theta)}{p(y)} \quad , \tag{3.1.6}$$

where; $p(\theta/y)$ = posterior probability distribution of θ given y
     $l(y/\theta)$ = the likelihood function of y given θ
     $p_0(\theta)$ = the prior probability of θ
     $p(y)$ = the marginal probability of y.

We can also write

$$p(y) = \int l(y / \theta) p(\theta) d\theta \quad , \tag{3.1.7}$$

where the integration is over all permissible values of θ.

Note how the conditional probability $p(y/\theta)$ used in the discrete case (3.1.1), is replaced by the likelihood function $l(y/\theta)$ in the continuous case (3.1.6). This does not imply that likelihood is simply another term for probability - the two are in fact quite separate concepts (refer to section 3.2). Rather the likelihood function is proportional to the conditional probability, but the constant of proportionality is subsumed within the normalising constant p(y).

By omitting the normalisation constant, the posterior density is simply proportional to the likelihood function multiplied by the prior probability distribution for θ

$$p(\theta / y) \propto l(y / \theta) p(\theta) \quad . \tag{3.1.8}$$

The resulting posterior distribution can also be used as a new prior distribution in exactly the same way illustrated for discrete variables.

The mathematics required by Bayes' theorem is more difficult for continuous variables than for discrete variables. In particular the integral in (3.1.7) only has an analytical solution in certain cases (refer to section 3.2). This difficulty has led some risk analysts to eschew Bayesian methods on the grounds that they are laborious and cumbersome (Vose, 1996).

Fortunately this difficulty can be removed with the use of computers and numerical integration techniques. These techniques broaden the practical application of Bayes beyond models dictated by mathematical convenience, and are seen by many as the key to widespread adoption of Bayesian methods (Smith, 1984).

Journey survival example

The ability of organisms to survive in the ballast tank is a key component of the ballast water introduction cycle. Survivorship can be investigated by sampling the ballast tanks of a vessel at regular intervals over its entire journey. This is suitable as an initial research approach but is not a practical long term strategy. An alternative approach is to sample vessels at the end of the ballast leg of their journey, in order to confirm the presence or absence of target species in the ballast water. The objective here is to investigate a species' life expectancy on the basis of presence/absence and ballast water age.

The length of time T that a population[7] is expected to survive in a ballast tank can be viewed as a random variable. Survival patterns between different species are known to be quite varied, but for some species, abundance in the ballast tank has been observed to decline exponentially with journey duration (see for example Wonham *et al* 1996, Murphy 1997). This evidence suggests that an exponential distribution may be an appropriate way to model this situation. According to this model, the random variable T is said to follow an exponential distribution with parameter μ, where μ denotes the expected life expectancy, such that the probability of observing a survival time t is given by

$$p(t) = \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right) \quad . \tag{3.1.9}$$

If individuals are still alive at the end of the vessel's journey, the appropriate probability is provided by the right tail of the exponential distribution; the probability that the survival time is at least some value x

$$P(t \rangle x) = 1 - F(x)$$

$$= 1 - \left[1 - \exp\left(-\frac{x}{\mu}\right)\right]$$

$$= \exp\left(-\frac{x}{\mu}\right) \quad .$$

The results of four hypothetical surveys, measuring the abundance of a single species over time, are summarised in Table 3.2. Two types of data are recorded here; those that record the species as dead (or absent) after time t, and those which indicate that individuals were still alive after time x.

---

[7] In this example population refers to the ballast tank inoculum.

Table 3.2   Some results from hypothetical ballast water surveys

Survey	Fate of the species	t (days)	x (days)
#1	No live specimens found after 14 days	14	
#2	No live specimens found after 32 days	32	
#3	Individuals still alive after 5 days		5
#4	Individuals still alive after 8 days		8

The variable of interest here is the life expectancy of the species (μ) in the ballast tank environment. The purpose of this example is to illustrate how Bayesian techniques can be used to quantify the uncertainty regarding this variable.

Let us make the following assumptions:

1.   we have no prior information regarding μ for this species;

2.   the change in the abundance of this species with time can be modelled reasonably well with an exponential distribution;

3.   the death rate of the species was the same for each survey, remains constant over the duration of the journey and is proportional to the size of the population;

4.   the survivorship surveys were independent of each other.

This example will investigate four different approaches to this problem. In the first instance a uniform prior distribution (over suitably defined limits) and a non-informative prior distribution[8] will be used for μ. Both of these distributions are commonly used to reflect little or no prior information regarding the variable of interest.

Next a fully-bounded and partially-bounded exponential distribution will be used as the underlying statistical model. The 'natural' domain of an exponential random variable T is partially bounded, ie $0 \le t \le \infty$. Biologically, however, this is not very realistic - one would expect a finite limit to the survival time. This can be achieved by bounding the distribution at some pre-defined upper limit M such that

$$p(t \, / \, \mu) = \begin{cases} \dfrac{1}{\mu} \exp\left(-\dfrac{t}{\mu}\right) & 0 \le t \le M \\[2ex] 0 & \text{otherwise} \end{cases} \qquad . \qquad (3.1.10)$$

Table 3.3 summarises the prior distribution function, likelihood function, normalising constant and resulting posterior distribution for each of the four approaches. The mathematical details of the derivation (and notation) are provided in Appendix D.

---

[8] Section 3.2 explains the use and derivation of non-informative priors.

Table 3.3   Four Bayesian models of journey survival

Prior $p_0(\mu)$	Likelihood $l(\mu/\mathbf{t}, \mathbf{x})$	Normalising constant	Posterior $p(\mu/\mathbf{t}, \mathbf{x})$
$= \mathrm{U}(0, \infty)$	$\left(\dfrac{1}{\mu}\right)^n \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)$	$\dfrac{\left(\sum_{j=1}^m x_j + \sum_{i=1}^n t_i\right)^{n-1}}{(n-2)!}$	$\dfrac{\left(\sum_{j=1}^m x_j + \sum_{i=1}^n t_i\right)^{n-1}}{(n-2)!}\left(\dfrac{1}{\mu}\right)^n \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)$
$= \mathrm{U}(0, M)$	$\left(\dfrac{1}{\mu}\right)^n \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)$	No closed form solution – integrate numerically	$\dfrac{\left(\dfrac{1}{\mu}\right)^n \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)}{\int_0^M \left(\dfrac{1}{\mu}\right)^n \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)}$
$\propto (\mu)^{-1}$ $0 \le \mu < \infty$	$\left(\dfrac{1}{\mu}\right)^{n+1} \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)$	$\dfrac{\left(\sum_{i=1}^n t_i + \sum_{j=1}^m x_j\right)^n}{(n-1)!}$	$\dfrac{\left(\sum_{j=1}^m x_j + \sum_{i=1}^n t_i\right)^n}{(n-1)!}\left(\dfrac{1}{\mu}\right)^{n+1} \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)$
$\propto (\mu)^{-1}$ $0 \le \mu < M$	$\left(\dfrac{1}{\mu}\right)^{n+1} \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)$	No closed form solution – integrate numerically	$\dfrac{\left(\dfrac{1}{\mu}\right)^{n+1} \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)}{\int_0^M \left(\dfrac{1}{\mu}\right)^{n+1} \exp\left(-\dfrac{\sum_{j=1}^m x_j + \sum_{i=1}^n t_i}{\mu}\right)}$

Figures 3.2 to 3.5 plot the posterior and prior distribution functions for each of these models, using the data presented in Table 3.2. The problem caused by a partially bounded model is apparent in Figures 3.2 and 3.4 – appreciable probability of an unrealistically high life expectancy. The uniform prior exacerbates this problem. At first a uniform prior seems a sensible way to reflect no prior knowledge of life expectancy – allocating equal chance to all possible values of the parameter. In this context, however, it performs badly because it allocates even more probability to the tail of the posterior distribution. This leads to a very artificial posterior distribution when the model is bounded (Figure 3.3). The non-informative prior with a bounded model (Figure 3.5) provides a better reflection of what might be expected in reality.

Figure 3.2   Journey survival model # 1: Uniform prior on support $0 \le \mu < \infty$
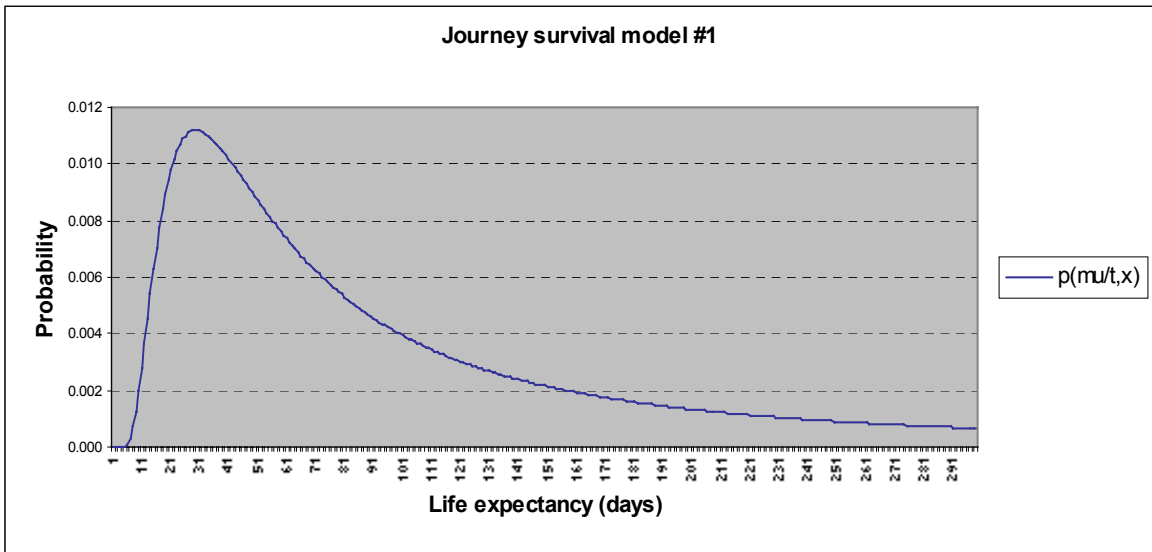

Figure 3.3   Journey survival model # 2: Uniform prior on support $0 \le \mu < 100$
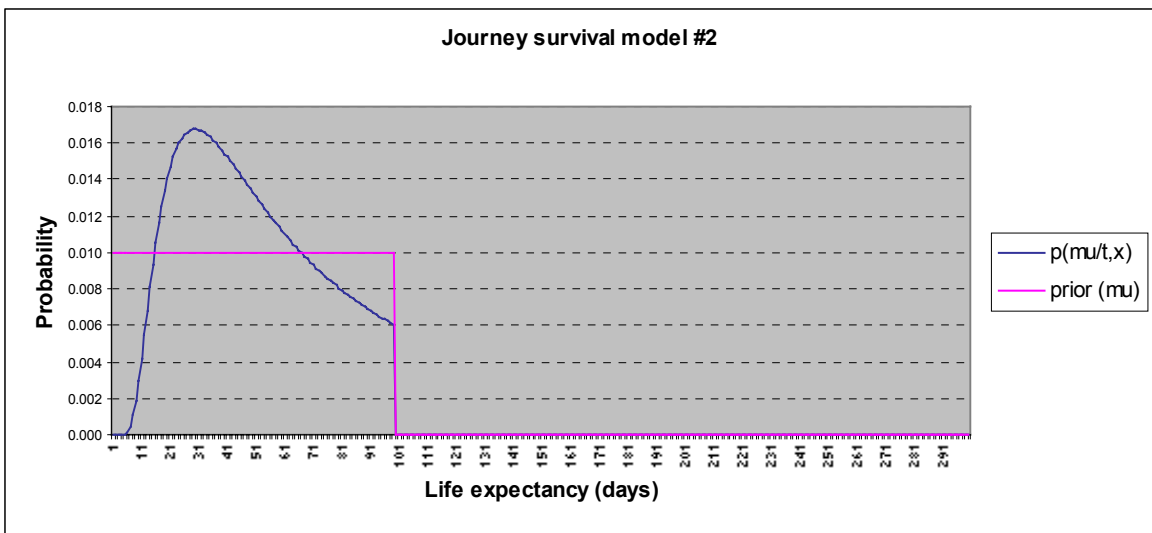
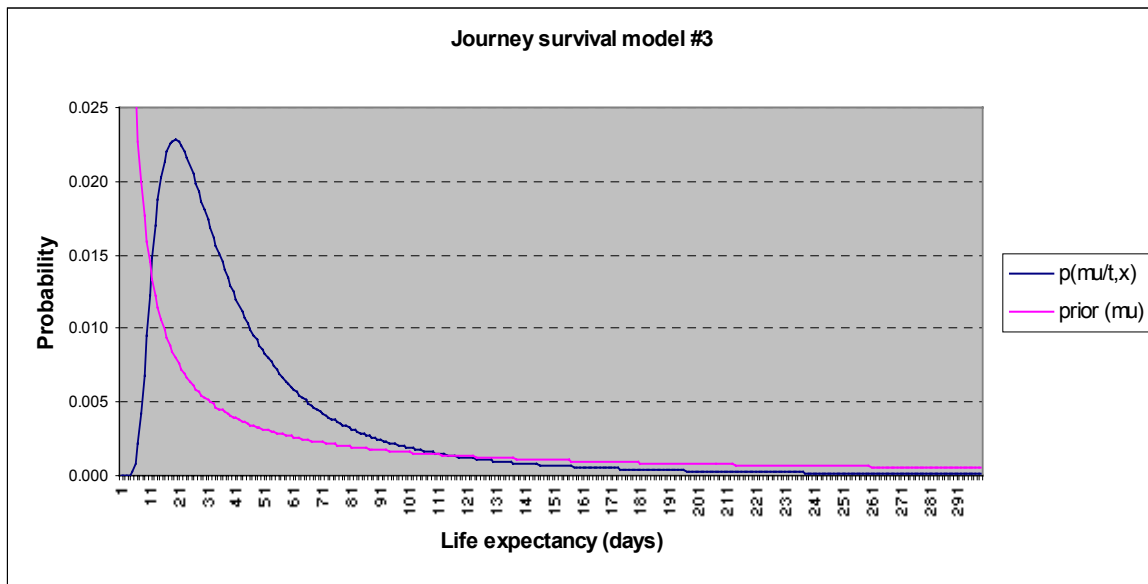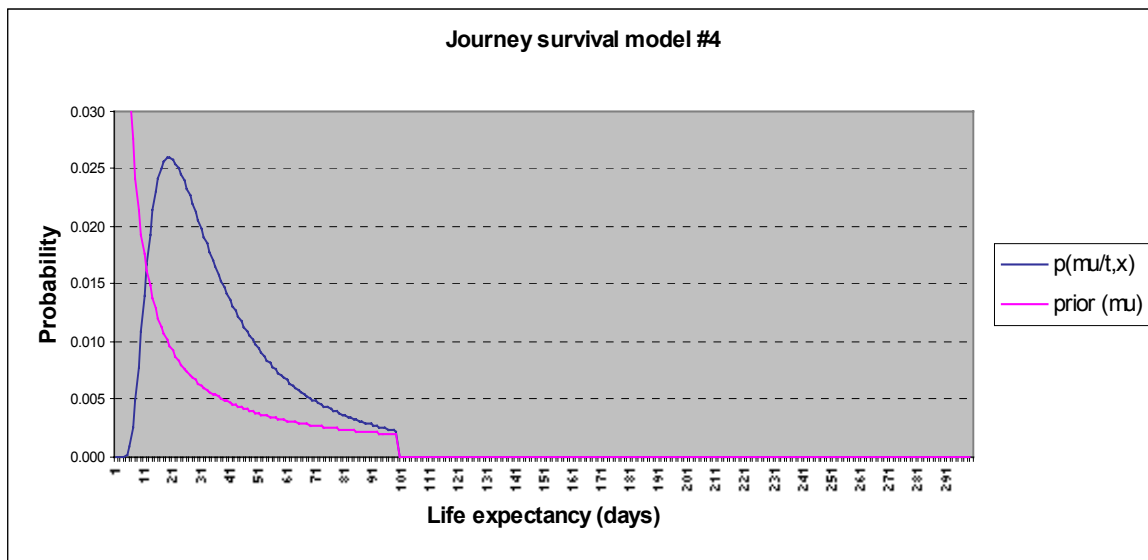Figure 3.4   Journey survival model #3: Non-informative prior on support $0 \le \mu < \infty$

Figure 3.5   Journey survival model #4: Non-informative prior on support $0 \le \mu < 99$

None of the posterior distributions above, however, are very compelling in light of the original evidence (Table 3.2). In particular survey # 2 has the effect of drawing the mode of the distribution away from the earlier survey results, which indicate that the species life expectancy lies somewhere between 8 and 14 days. This indicates two problems with the current model:

1. ballast age may not be a good surrogate for life expectancy when sampling ballast tanks at the end of the vessel's journey. In other words the absence of the species in the tank at the end of the vessel's ballast leg indicates that the life expectancy is some (undetermined) value less than the age of the ballast water;

2. the small sample size (n = 2) leads to diffuse posterior distribution which is very sensitive to 'outliers'. This also exacerbates the effect of the uniform prior distribution – put simply there is insufficient evidence to overwhelm the influence of the prior.

This last point is an important one. The posterior distribution represents a compromise between the prior information and the data, but is increasingly controlled by the data as the amount of data increases. In data sparse situations, as is often the case with ecological risk assessment, the prior retains considerable influence on the shape of the posterior (refer to section 3.2).

It is very easy, however, to incorporate new information into a Bayesian analysis - the 'old' posterior distribution function can be combined with the likelihood function for any 'new' data to produce a 'new' posterior. For example, assume that another 3 surveys were conducted to investigate the life expectancy of the same species above, each recording the absence of the species after a period of time y (Table 3.3).

Table 3.3   Further hypothetical results from ballast water surveys

Survey	Fate of the species	y (days)
#5	No live specimens found after 24 days	24
#6	No live specimens found after 28 days	28
#7	No live specimens found after 21 days	21

The 'new' posterior distribution function becomes

$$p(\mu/t, x, y) \propto p(\mu/t, x)l(\mu/y) \quad , \qquad\qquad (3.1.11)$$

where; $p(\mu/t, x, y)$ = the new posterior distribution function for μ given t, x and y
   $p(\mu/t, x)$ = the old posterior distribution function for μ given t and x
   $l(\mu/y)$ = the likelihood function for the new data y

The updated posterior distribution, using journey survival model #4, is illustrated in Figure 3.6. Notice how the additional information has reduced the uncertainty regarding the species life expectancy, in particular concentrating much more probability in the region of 21 to 41 days. For comparative purposes the updated posterior distribution, using journey survival model #2, is illustrated in Figure 3.7.

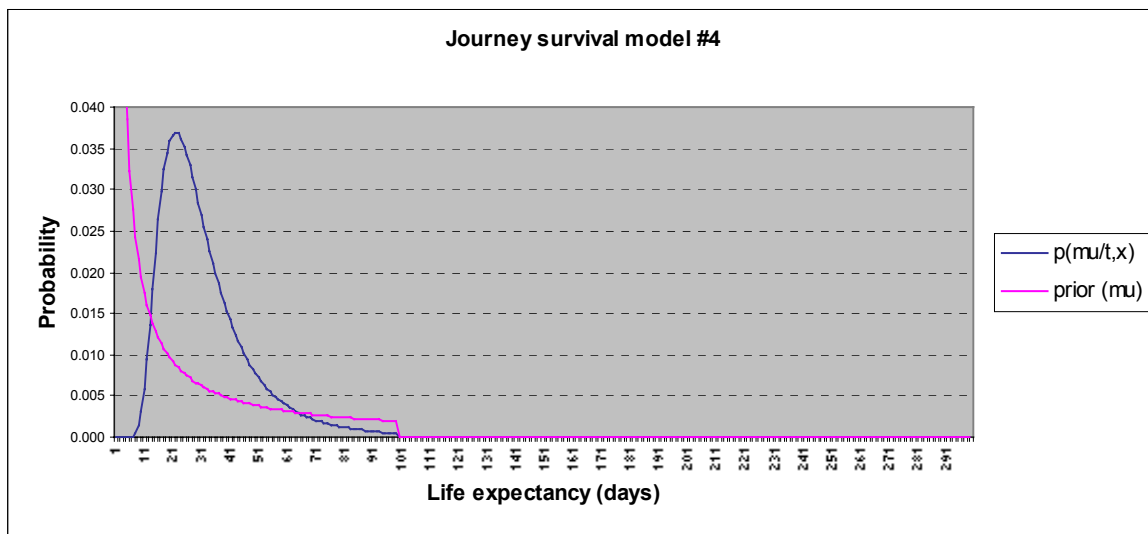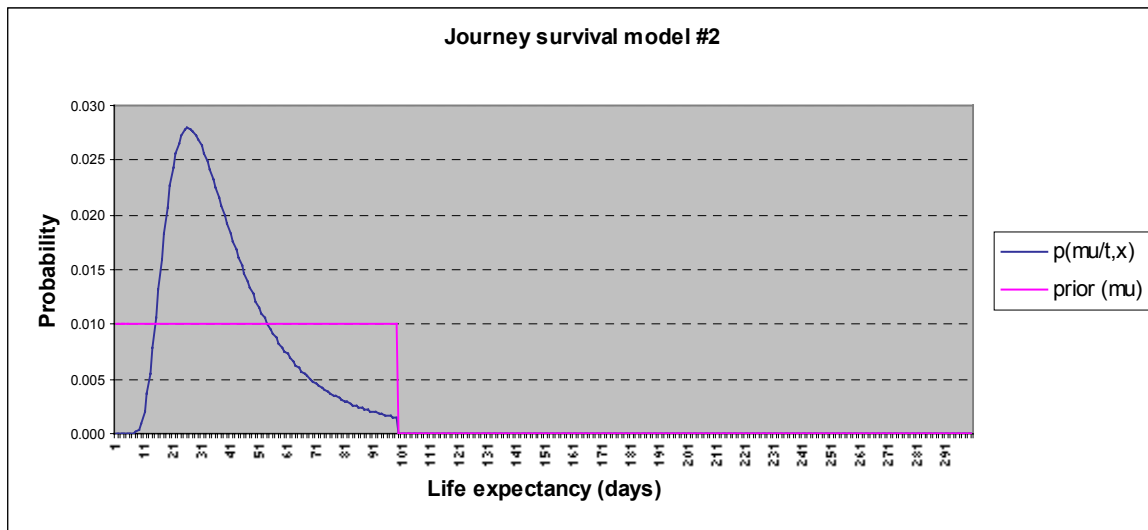Figure 3.6   Updated posterior distribution function using the additional data in Table 3.3

Figure 3.7   Updated posterior distribution function using the additional data in Table 3.3

3.2   Components of the theorem

The technical core of Bayes theorem is captured in equation (3.1.8). It consists of the prior probability $p_0(\theta)$, the likelihood function $l(y/\theta)$ and the posterior distribution $p(\theta/y)$.

The likelihood function l(θ/y)

The conditional distribution function $p(y/\theta)$ describes the probability of the data y, given the parameter θ. When viewed as a function of θ, for fixed y, this distribution is referred to as the likelihood function $l(\theta/y)$.

Likelihood is a subtle concept. Its full implications are not clear from the usual textbook definitions (such as that above), and are often missed on first reading. The conditional distribution function derives from a deductive process of assigning probabilities to different events A_i given the same hypothesis θ. The likelihood function derives from the inductive process of assigning probabilities to an event A, given different hypotheses θ_i. Therefore in $p(y/\theta)$ the event A = y is the variable, whereas in $l(\theta/y)$ the parameter θ is the variable.

As an example consider a sequence of Bernoulli trials in which the ith observation is either a success, $A_i = 1$, or a failure, $A_i = 0$, in which the probability of success, θ, is the same for each trial (i = 1...n). The data in this example form a sequence (of length n) of ones and zeroes, **y** = (1, 1, 0, 0, 1, 0, ...). The probability of any particular sequence is the product of terms θ and (1-θ), with a θ for each one and a (1-θ) for each zero, such that

$$p(y/\theta) = \theta^y (1-\theta)^{n-y} \quad . \tag{3.2.1}$$

where y is the number of ones in the sequence.

In the continuous version of Bayes theorem (3.1.6), the likelihood function replaces the conditional probability $p(y/\theta)$ in the discrete version of the theorem (3.1.1). This does not imply that likelihood is a probability function. The conditional probability $p(y/\theta)$ describes the probability of the data y on a fixed parameter θ. When considered as a function of y it defines a probability distribution which, when integrated over all possible values of y, equals one. By contrast the likelihood function $l(\theta/y)$ is predicated on fixed data, does not give rise to a statistical distribution, and need not equal one when integrated over all possible parameter values (Edwards, 1992). The likelihood $l(\theta/y)$ is, however, proportional to $p(y/\theta)$ such that

$$l(\theta/y) = k\theta^y (1-\theta)^{n-y} \tag{3.2.2}$$

in the example above, and can therefore be used in Bayes theorem. The constant of proportionality k is arbitrary, and becomes subsumed within the normalising constant p(y) when used in this way.

The inductive nature of the likelihood function underlies the central role that it plays in Bayes' theorem: it is the only way in which the data y can modify the analyst's prior knowledge of θ. It also determines how well the parameter θ allows the model to describe the data (Dilks *et al*, 1992).

In practice, however, the concept of likelihood is only reasonable within the framework of the model or family of models adopted for a particular analysis. In other words it can only be applied to situations where the functional form of the conditional distribution function $p(y/\theta)$ is specified, but the parameters $\theta = (\theta_1...\theta_n)$ are unknown.

Thus in order to use Bayes' theorem the analyst must choose the probability model which best suits the problem at hand. For many ecological risk assessment problems, however, it is not immediately obvious which probability model is the most appropriate, and rarely can the analyst be confident that the chosen model is the correct one (Gelman *et al,* 1995).

Choosing the functional form of the likelihood distribution is therefore the most crucial (and arguably the most difficult) part of any Bayesian analysis. It is surprising then to discover that this part of the analysis is often given the least attention. Much of the debate over the use of Bayesian methods in ecological risk assessment is concerned with problems of the prior distribution function (see section 4.2), and the subjective process of choosing priors.

By contrast many authors spend very little time justifying their choice of likelihood function, and rarely (it seems) is this choice questioned by reviewers, despite the fact that this can have as much impact on the results of the analysis as the choice of the prior distribution.

Cox and Hinkley (1974) offer the following guidelines for choosing a probability model:

1.   the model should establish a link with any theoretical knowledge about the system in question, and with any previous experimental work;

2.   the limiting behaviour of the model and the system should be consistent;

3.   the parameters of the model should individually have clear cut interpretations;

4.   the model should be the simplest possible consistent with the guidelines above.

In practise the analyst may need to consider a number of different models to find the most appropriate one. This point is particularly relevant to Bayesian statistical inference, as emphasised by Gelman *et al* (1995), who view an applied Bayesian statistician as one who is willing to apply Bayes' rule with a variety of different likelihood models.

The prior distribution $p_0(\theta)$

The prior probability distribution $p_0(\theta)$ usually represents the analysts knowledge regarding the parameter θ prior to the analysis and collection of data. Gelman *et al*, (1995), suggest that the prior distribution can be given two interpretations: a population interpretation in which the prior represents a population of possible parameter values from which θ must be drawn; or a state of knowledge interpretation in which the prior simply represents the analysts uncertainty regarding θ. Cox and Hinkley (1974) suggest similar interpretations:

1.   a frequency distribution whose parameters reflect analysis and synthesis of existing data;

2.   an "objective" statement of what is rational to believe about the distribution given initial ignorance of the parameter; and,

3.   a subjective measure of what the analyst actually believes before commencing the assessment.

A population or frequency distribution interpretation implies the availability of substantial prior information, in particular the family of population distributions to which the prior should belong. In these circumstances it is often possible to generate a specific prior distribution from this family, known as the conjugate prior, which has convenient mathematical properties. (It is not compulsory to use a conjugate prior, merely convenient).

Conjugate priors are constructed by interchanging the roles of the random variable and the parameter in the likelihood function, and then 'enriching' the parameters by making their values general and not dependant upon the current data set (Press, 1989). By way of example suppose the likelihood function is binomial such that

$$l(\theta / y) \propto \theta^y (1 - \theta)^{n-y} \quad . \tag{3.2.3}$$

If θ is considered to be the random variable, $l(\theta/y)$ forms the kernel of a beta distribution. The distribution is generalised by replacing the parameters y and n-y (which depend upon the data) with the arbitrary parameters α and β. The density is normalised by dividing by an appropriate constant, here the beta function $B(\alpha, \beta)$, such that the conjugate prior distribution becomes

$$p_0(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1 - \theta)^{\beta-1} \quad . \tag{3.2.4}$$

The mathematical convenience arises because many conjugate prior distributions (including all members of the exponential family[9]) give rise to posterior distributions of the same parametric form. For example employing (3.2.2) and (3.2.3) in Bayes theorem leads to the posterior distribution (refer to Appendix E)

$$p(\theta / y) = \frac{1}{B(y + \alpha, n - y + \beta)} \theta^{y+\alpha-1}(1 - \theta)^{n-y+\beta-1} \quad , \tag{3.2.5}$$

which is also a beta distribution. The property that the posterior distribution follows the same parametric form as the prior distribution is called conjugacy. Hence the beta distribution is the conjugate prior of the binomial likelihood. Appendix E summarises the conjugate prior and posterior distributions of some common statistical models.

In practise the analyst is still required to evaluate the parameters of the prior distribution, α and β in the above example. Typically these parameters would be unknown (unless the prior distribution was the posterior distribution from a previous analysis - as in the journey survival example), and would have to be estimated using prior information (data), in an empirical Bayes approach for example.

If, however, the prior information suggests a distribution which does not belong to the conjugate family then this should be used. This causes no loss of application, but is computationally more intensive. In particular the integral of the posterior distribution (the normalising constant) may not have an analytical solution, in which case it must be evaluated using numerical methods. For example it is common in fisheries risk assessment to draw prior distributions from similar biological populations, a process known as meta-analysis (Punt and Hilborn, 1997). These may not be conjugate but if they represent the best available prior information, then they should be used.

---

[9] The exponential family consists of the Binomial, Negative binomial, Poisson, Exponential and Normal distributions.

A state of knowledge ('objective', subjective or otherwise) interpretation admits a much broader class of prior distributions, allowing the analysts to specify their uncertainty with any type of probability distribution. It is commonly used when there is no population precedent to draw upon but does not exclude a population perspective. When prior distributions have no population basis, however, they can be difficult to construct. The Bayesian analyst is forced to fit some form of distribution to his prior beliefs but, as is often the case with ecological risk assessment, there may be very little evidence to support these beliefs. The analyst's judgement may therefore be subjective, vague and open to debate.

In these circumstances a prudent Bayesian may seek a prior distribution which plays a small role in the posterior distribution, and is quickly overwhelmed by the data. In this way the analysts subjective opinion ceases to be of concern because it is quickly swamped as soon as much data becomes available (Edwards *et al*, 1963). Mathematically this requires a prior distribution which is relatively uniform in the vicinity of the likelihood function (the data), and which does not peak sharply outside of this region. Prior distributions with these characteristics are called non-informative.

Box and Tiao (1973) emphasise that non-informative priors need not represent the analysts prior state of uncertainty, but can be employed as a point of reference to allow unprejudiced inference from the data. Non-informative priors should therefore express the idea that little is known about θ, prior to what the data is going to tell us, whilst not unduly influencing what the data has to say. In other words they should 'let the data speak for themselves' (Gelman *et al*, 1995).

In Bayes' theorem the data influence the posterior distribution only through the likelihood function $l(\theta/y)$. Box and Tiao (1973) describe a technique for obtaining non-informative priors based on the metric $\phi(\theta)$ under which the likelihood function is data translated - that is the likelihood curve for $\phi(\theta)$ is completely determined *a priori* except for its location, which depends on the data yet to be observed. Figure 3.4 demonstrates this principle for the binomial and Poisson models. Under the original metric θ the spread, shape and location of the likelihood curve varies with the data y. Under a suitable transformation $\phi(\theta)$, however, the likelihood curves are approximately constant except for location.

By taking $\phi(\theta)$ to be locally uniform (in the vicinity of the likelihood curve) the resulting prior distribution for θ satisfies all of the required criteria for a non-informative prior. Box and Tiao (1973) go on to demonstrate that the prior distribution for a single parameter θ is therefore approximately non-informative if it is taken proportional to the square root of Fisher's information measure[10]. That is

$$p_0(\theta) \propto J^{\frac{1}{2}}(\theta) \qquad ,$$

where $J(\theta)$, Fisher's measure of information about θ, is defined as

$$J(\theta) = \mathop{E}_{y/\theta}\left(-\frac{\partial^2 \log l(\theta/y)}{\partial \theta^2} \right) \quad .$$

---

[10] This is Jeffrey's' rule, after Jeffreys (1961)

Figure 3.8   Data translated likelihood curves for the Binomial and Poisson means

y = 3   y = 12   y = 21

0.1   0.5   0.9 $\pi \rightarrow$

a) the binomial mean π

y = 3   y = 12   y = 21

20°   45°   70° $\phi = \sin^{-1}(\pi)^{1/2} \rightarrow$
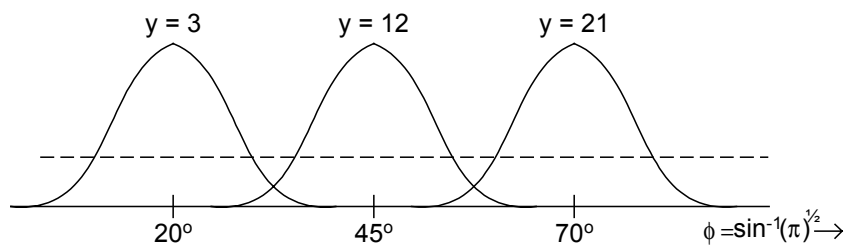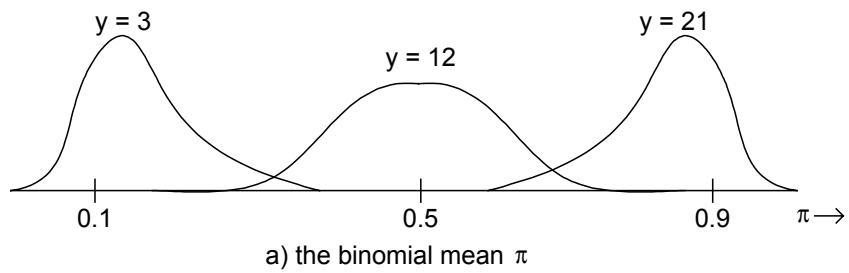
b) the transformed binomial mean $\phi = \sin^{-1}(\pi)^{1/2}$
(solid curves are standardised likelihood, broken
line is the non-informative prior)

y = 1

y = 4

y = 9

4   8   12   16 $\lambda \rightarrow$

c) the Poisson mean λ

y = 1   y = 4   y = 9

1.0   2.0   3.0 $\phi = \lambda^{1/2} \rightarrow$
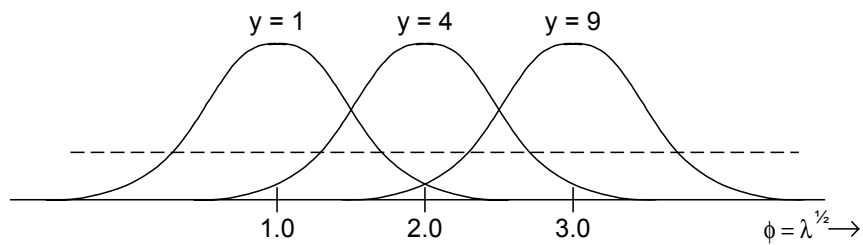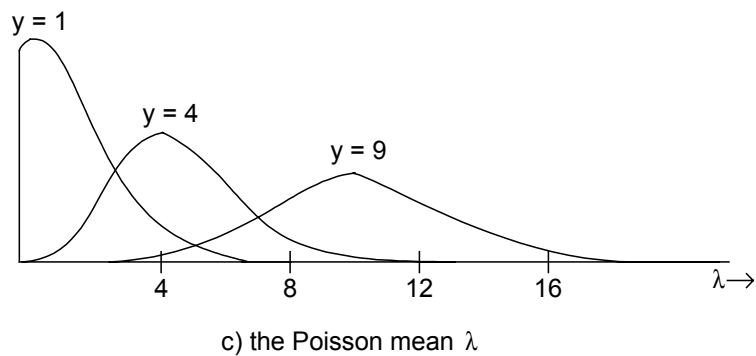
d) the transformed Poisson mean $\phi = \lambda^{1/2}$
(solid curves are standardised likelihood, broken
line is the non-informative prior)

(Source; Box and Tiao, 1973)

Non-informative prior distributions imply that the analyst is able to make a start in uncertain situations, safe in the knowledge that his or her prior convictions will not unduly bias the results of the assessment[11]. This may not always be true however. In certain circumstances they can lead to improper posterior distributions (Gelman *et al*, 1995). They are also sensitive to the measurement scale (Punt and Hilborn, 1997) although it may not always be obvious which scale is the most appropriate.

Edwards *et al* (1963) describe a number of other situations where a non-informative prior may be inappropriate:

1. the analyst has a strong prior conviction that the value of θ lies in a region where the likelihood has a very small value, and is unpersuaded by the evidence;[12]

2. the data produce a likelihood curve which is very diffuse; and,

3. observations are so expensive that the analyst cannot afford to collect enough data to ensure that the likelihood function is sufficiently well defined.

In defence Box and Tiao (1973) point out that scientific investigation is not usually undertaken unless the data supplied by the investigation is likely to be considerably more precise than the currently available information. They do concede, however, that each case should be considered on its merits - a point also taken by Gelman *et al* (1995) who warn against the automatic use of 'reference' non-informative priors, advocating a case by case approach.

The posterior distribution p(θ/y)

Bayesian inference allows the analyst to pass from the prior distribution to the posterior distribution by way of the data (represented by the likelihood function). The posterior distribution $p(\theta/y)$ represents the total uncertainty regarding θ in light of the available data and the analysts prior beliefs.

The variance of the posterior distribution (a measure of uncertainty) should be less than that of the prior because it incorporates the additional information provided by the data. This is well illustrated by the binomial model; with a uniform prior distribution the prior variance is 1/8, whilst the posterior variance is given by

$$\frac{y(n-y)+\frac{n}{2}+\frac{1}{4}}{(n+1)^2(n+2)} \quad . \qquad (3.2.6)$$

The value of (3.2.6) quickly diminishes as the data (represented by the number of trials - n) increases, underlining the increasingly dominant role played by data in the posterior distribution. In certain situations the variance of the posterior can be similar to or even larger than the prior, although this usually indicates an error in either the sampling model or the prior distribution (Gelman *et al*, 1995).

---

[11] Unfortunately the same cannot be said for the choice of the likelihood model.
[12] The experiment which produced the likelihood curve may be in error for example.

In an ecological risk assessment the posterior distribution function can be used in one of two
ways:

1. to provide a direct estimate of risk, where the endpoint of the assessment is some critical
    value of θ; or,

2. to represent uncertainty regarding an important model parameter, that is subsequently
    translated into an expression of risk through the model output.

The first approach is illustrated in Figure 3.9 which shows the posterior distribution of the
Normal mean $p(\theta/\sigma, \mathbf{y})$ when a sample of 16 observations has been taken whose average value
is 10. The distribution was derived with a locally uniform, non-informative prior (refer to
Appendix E) and therefore represents the knowledge of an analyst who, prior to the collection
of data, was indifferent to the value of θ in the relevant range. If the endpoint of the assessment
was some critical value of θ, for example 12, then the posterior distribution provides a direct
estimate of the risk that this value will be exceeded, in this case 0.159.

The risk estimate in Figure 3.9 was made using standard tables of the Normal probability
integral. Probability estimates from the posterior distribution, however, are not always so
readily derived. Figure 3.10, for example, shows the probability of a species' life expectancy
exceeding 40 days, based on the journey survival model (#4) developed in section 3.1. Here the
posterior distribution function $p(\mu/\mathbf{t}, \mathbf{x}, \mathbf{y})$ has the kernal of an inverted Gamma distribution, but
does not posses a closed integral form on limits other than $(0, \infty)$. The probability of μ
exceeding any specified value must therefore be calculated using numerical integration
techniques - in this instance Simpson's Rule was used.

The second approach use Monte Carlo simulation techniques to return values of a parameter to a
risk assessment model. This procedure is repeated many times; on each iteration a value is
randomly selected from the posterior distribution and returned to the risk assessment model.
Tabulation of the model output for each iteration allows the effect of parameter uncertainty on
the risk estimate to be investigated.

A simple Monte Carlo simulation of the journey survival problem (section 3.1) is illustrated in
Figures 3.11 and 3.12. Figure 3.11 tabulates the results of 1000 samples selected at random
from the posterior distribution function of journey survival model #4. Because the posterior
distribution is bounded such that $\mu \le 100$ the simulation rejects all samples from the parent
distribution (inverse gamma) which result in value of μ greater than 100 (refer to Appendix F).
Each of the subsequent samples was then used in a simple model of ballast tank survivability

$$P_t = P_0 \exp(-t/\mu) \quad , \qquad\qquad\qquad\qquad (3.2.7)$$

where P_t represents the ballast tank population after a journey of t days, P_0 represents the initial
inoculum and μ is the life expectancy of the species. Figure 3.12 tabulates the model output for
each iteration based on a journey duration of 40 days and initial inoculum of 1000. An estimate
of risk can be quickly drawn from this size-frequency distribution. For example the probability
of there being at least 200 individuals (20% of in the initial inoculum) still alive at the end of the
journey, is given by the number of iteration on which this occurred divided by the total number
of iterations; in this example approximately $623/1000 = 0.62$.

Figure 3.9   The posterior distribution function p(μ/**y**)

**Posterior distribution for the Normal mean
(non-informative prior)**

p(theta > 12) = 0.1032

p(pheta/y)

theta
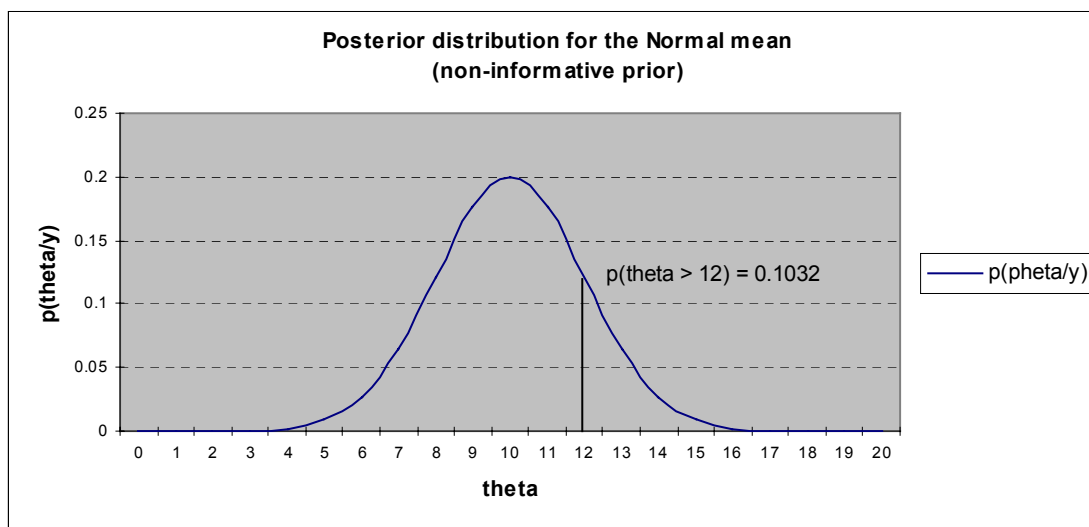
Figure 3.10   Risk assessment using the posterior distribution function p(μ/**t, x, y**)

Journey survival model #4

P(mu > 40) = 0.220

p(mu/t,x)
prior (mu)

Probability

Life expectancy (days)
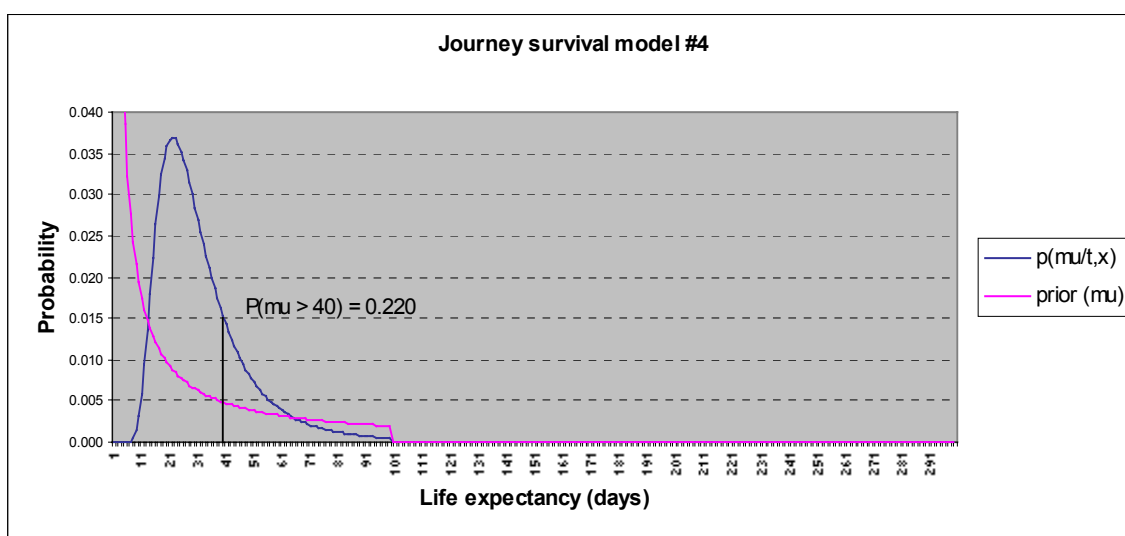
Figure 3.11 Monte Carlo simulation of p(μ/**t, x, y**) on the support $0 < \mu < 100$
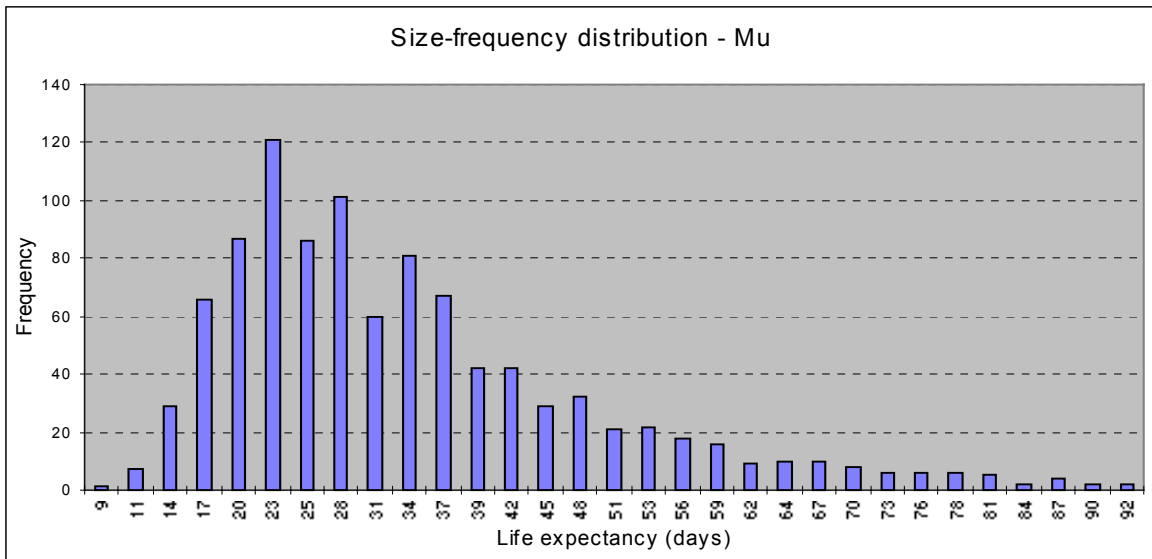
Figure 3.12 Monte Carlo simulation of the ballast tank population after a 40 day journey
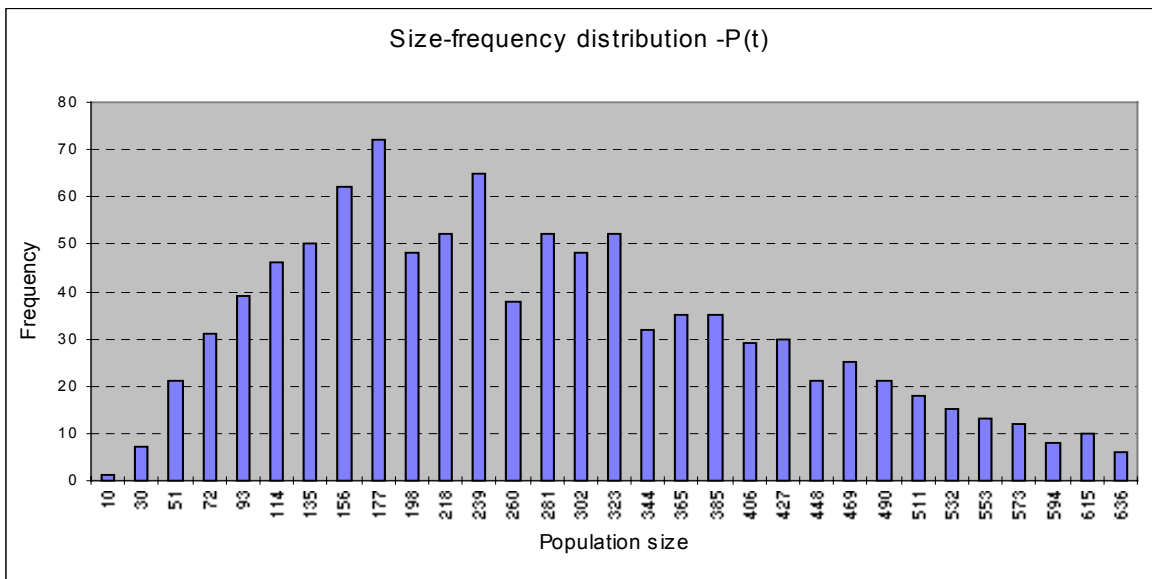
Notice how Figure 3.11, the size frequency distribution of μ, is similar to Figure 3.6, the original posterior distribution function. This occurs because Monte Carlo simulation faithfully reproduces the distribution from which it samples given enough iterations[13]. This allows the analyst to propagate variable uncertainty through a deterministic model such as (3.2.7) in order describe a range of possible outcomes (Figure 3.12) and thereby an estimate of risk.

The accuracy of the simulation can be demonstrated by comparing $P(\mu > 40)$ calculated by numerical integration techniques with the simulation size-frequency distribution. The latter gives $P(\mu > 40)$ = 252/1000 = 0.252 which is comparable to the 0.220 calculated using Simpson's Rule in Figure 3.10. The accuracy of the simulation can be improved by increasing the number of iterations at the cost of increased computer time. Note, however, that if the simulated values of the parameter are bounded (as in Figure 3.11) the resulting size-frequency distribution must be used cautiously. For example the moments of the simulated distribution may not accurately reflect the moments of the parent distribution - the variance in particular is likely to be reduced.

---

[13] A more efficient variant of the Monte Carlo approach is Latin Hypercube Sampling (LHS).

4   ECOLOGICAL RISK ASSESSMENT & BAYES THEOREM

4.1 Why use Bayes in ecological risk assessment?

Bayes theorem has recently witnessed something of a revival. This is probably due to the growing availability of computing resources and efficient numerical integration procedures, as predicted by Smith (1984). Classical statistical inference has attracted a great deal of criticism in the growing literature that has followed this revival, particularly in its application to ecological science. When following this literature it is easy to forget how spectacularly successful classical statistical practise is (Anderson, 1998), and how well it has served ecologists in the past. The question arises then, why use Bayesian approaches at all?

Ecological risk assessment presents a number of practical and philosophical challenges to statistical inference:

1. the interpretation of probability in the context of ecological science;

2. the 'true' value of biological parameters;

3. modelling complex, often poorly defined systems, with little data;

4. drawing inferences from the multi-parameter models used to describe these systems;

5. providing risk estimates with a proper description of uncertainty;

6. describing the results of the assessment to stakeholder forums who may not be well versed in statistical science; and,

7. the iterative development of risk assessment models and the predictions they make.

Bayesian statistical inference offers an alternative, and sometimes better, approach to these challenges. It is not, however, a mutually exclusive alternative – ultimately the statistical approach adopted by the analyst should be dictated by the problem and data at hand.

Ecological risk analysts may be faced with problems over the interpretation of probability (see section 2.1). Frequentist and classical interpretations of probability are often criticised because they have little intuitive meaning in ecological science. For example when attempting to assess the possible consequences of an event such as chemical discharge to the environment, the analyst can neither enumerate a complete set of mutually exclusive outcomes, nor reproduce the event enough times to estimate the frequency of each outcome.

Ideally probability should be conceptualised as characteristic of a random process, external to the observer. The process may involve repeated iterations of a trial, random selection from a defined population, or a sample space divide into n mutually exclusive outcomes. The analogy need not be direct, however. If the problem can be conceptualised in these terms, then that is sufficient rationale to undertake a simulation analysis using frequentist or classical interpretations of probability. This rationale is behind many of the simulation models used in ecological risk assessment, although this may not be explicitly acknowledged.

 Classical statical inference does, however, exclude a subjective interpretation of probability. This interpretation is very flexible and it seems foolish to exclude it from the risk analyst's toolkit. Power *et al* (1994) suggest that subjective probability is intuitively appealing to analysts seeking to make inferences about complex ecosystems, but dismiss its role in ecological risk assessment on the grounds that objective facts no longer provide a means of sorting one assessment from another. But this is precisely what Bayes theorem achieves by objectively adjusting the analysts prior convictions in light of the evidence gathered to date.

Furthermore Bayes theorem allows the analyst to make probability statements about unique and singular systems (Crome *et al*, 1996), confident that these can be readily adjusted as and when more information becomes available. The Bayesian approach is therefore eminently suited to the iterative nature of ecological risk assessment wherein risk estimates are made, but then tested, and continually updated, against the results of monitoring strategies. As this process continues each successive iteration produces a more refined risk estimate (Jaykus, 1996). Such iterations are easily performed within a Bayesian framework because of the ease with which the posterior distribution can be updated (as demonstrated by the journey survival example).

Perhaps a more compelling critique of classical statistical techniques in ecological science lies in their approach to parameters estimation. Classical estimators typically assume that there is a true fixed value for each parameter of interest and, for example, the expected value of this parameter is the average value obtained by random sampling repeated ad infinitum. The physical and biological parameters of ecological systems, however, are usually variable – no two organisms are exactly alike, and even if they were, evolution implies that their offspring would be measurably different (Ellison, 1996).

By contrast Bayes theorem recognises that the parameters of a study are variable and attempts to emphasise this through the posterior distribution. Note, however, that whenever a bayesian uses independent and identically distributed likelihoods there is an implicit assumption that the parameter(s) of concern take a single value. What the posterior distribution reflects is the uncertainty in the data when making inferences about this value (*pers comm* Andre Punt). On reflection the distinction between classical and Bayesian approaches on this point may not be as large as some authors imply.

The posterior distribution does, however, provide a full description of possible parameter values, and all inferences are based on this (Pascual and Kareiva, 1996). This is particularly important to risk assessors because scalar probability statements (point estimates) rarely provide adequate descriptions of risk. Power *et al* (1994), for example, strongly advocate complete distributional descriptions of risk rather than point estimates because:

1. knowledge of the variability surrounding an estimate of risk will change perceptions about its acceptability; and,

2. empirically produced data does not allow the analyst to assign probability outside the range of the data, and since many risk assessments involve extreme events, it is particularly important to have a complete description of the data in the form of an appropriately fitted probability model.

Via the posterior distribution function, Bayes theorem immediately directs the analyst to the full distributional qualities of the parameter(s) in question, and is therefore well suited to the risk assessment process.

This treatment of parameter uncertainty is all the more important to ecological risk assessment because of the heavy emphasis it places on modelling. Most environmental systems are in some sense 'badly defined' (Young, 1983) and if so, are best modelled within a probabilistic framework. The parameters of these models will be inherently uncertain and should only be defined in terms of statistical probability distributions.

Deterministic models can only provide adequate descriptions of ecological systems in the simplest of cases. This is because the expected value of a function E[f(A)], with a set of A of randomly varying parameters, is only equal to the value of the function using the expected value of each parameter, when f(A) is a simple linear function of the A terms (Gardner and O'Neill, 1983). The Bayesian emphasis on the full distributional qualities of parameter uncertainty is well suited in this context. The analyst is able to draw samples from the posterior distribution of each parameter and thereby propagate the full range of uncertainty through the analysis rather than use point estimates. Of course all of this is possible with classical statistical techniques, the advantage of Bayes is merely one of emphasis and the ease with which the posterior distributions can be updated.

Another potential advantage of Bayesian inference, which is often cited in the literature, is the ease with which incidental or 'nuisance' parameters are dealt with. Most ecological models involve more than one unknown parameter, but in many cases the analyst is only interested in a sub-set of these, possibly just one. Similarly in ecological risk assessment the analyst is only concerned with the distributional qualities of the endpoint but is usually faced with other (uncertain) parameters that have some bearing on this.

A large number of nuisance parameters make model calibration difficult (Stow *et al*, 1997). Furthermore inferences regarding the parameter of interest can only be made with classical techniques if the sampling distribution of the parameter(s) of interest are independent of the nuisance parameters; or sufficient statistics exist for all the parameters concerned (Box and Tiao, 1973).

A Bayesian does not worry about the sampling distribution of point estimates because inferences regarding the parameter of interest θ_1 are based entirely on its posterior distribution. This can be obtained from the joint posterior distribution of θ_1 and a nuisance parameter θ_2 by simply 'integrating out' the latter, thus

$$p(\theta_1 / y) = \int_{R_2} p(\theta_1, \theta_2 / y) d\theta_2 \quad , \tag{4.1.1}$$

where R_2 denotes the appropriate range of θ_2[14]. Alternatively the joint posterior distribution function can be factored to yield

$$p(\theta_1 / y) = \int_{R_2} p(\theta_1 / \theta_2, y) p(\theta_2 / y) d\theta_2 \quad . \tag{4.1.2}$$

Note also that the integral in (4.1.2) need not be evaluated explicitly: inferences regarding θ_1 can be made by first drawing θ_2 from its marginal distribution, and then θ_1 from its conditional

---

[14] Appendix G provides an example of this approach for a normal random variate whose mean and variance are unknown.

posterior distribution, given the value of θ_2, thereby performing the integration indirectly (Gelman *et al*, 1995).

Applying Bayesian techniques to complex, multi-parameter models, however, is not entirely straight forward. In the first instance the analysts must specify a prior distribution function for each of the parameters in the model (with all the attendant problems that this can entail) and must also account for any correlation between these distributions (*pers comm* Andre Punt). Thus Bayesian techniques may not be any easier to apply than classical inference techniques in this context.

Bayesian techniques are often quoted as being superior to classical techniques when data are scarce. Good (1959) notes that maximum likelihood estimation can give absurd results with small samples. Similar problems occur in mark and recapture experiments with low recapture numbers (Gazey and Staley, 1986), with data-poor linear model analysis (Reckhow, 1996), and with animal surveys (Johnson 1977, 1989). In each of these cases the authors suggest that Bayesian approaches are demonstrably better than their classical counterparts. In data scare situations, however, the prior distribution function exerts a strong influence on the shape of the posterior distribution (see for example the journey survival model in section 3.1). The efficacy of the analysis is therefore largely dependent on the prior distribution. This may not be a problem if there is sufficient evidence to support one prior over another. In situations where data are truly scare, however, this may not be the case, and the results of the assessment should be assessed accordingly[15].

Finally, some authors suggest that the posterior distribution function provides a common sense interpretation of the uncertainty surrounding a parameter or risk estimate, and thus the results of a Bayesian analysis are easier to interpret by the laymen. This assertion is usually supported by reference to the cumbersome definition of confidence intervals derived under classical techniques (Ellison, 1996). Indeed the requirement for 'plain English' results is reason enough for some to adopt Bayesian techniques (see for example Crome *et al*, 1996). This is an important point because ecological risk assessment is often conducted for decision makers and stakeholders who may not be familiar with the finer points of statistical science. However, the extent to which Bayesian results are easier to interpret than classical results is debateable. Classical statistician might argue that a confidence interval that is stable under a variety of different models, is no harder to interpret than a number of posterior distributions (*pers comm* Glen McPherson).

4.2   Some thoughts on the current debate

The current debate between the classical and Bayesian school of statistical inference revolves largely around two elements (Dennis, 1996):

1. the quantification of prior beliefs in the form of a prior probability distribution, and the incorporation of those beliefs into a data analysis; and,

2. the key role played by likelihood function in a Bayesian analysis, at the expense of alternative techniques.

---

[15] It is interesting to note that all but one of the authors cited here uses empirical bayesian techniques in their analysis. This suggests that data might not be as 'scare' as the authors imply.

At first sight this debate only seems relevant because ecological risk assessment utilises the concept of probability. There is, however, a much more important reason: ecological risk assessment is becoming increasingly popular because it is seen as a means of producing scientifically objective assessments (Rodier and Zeeman, 1994), divorced from subjective judgement (Holdway, 1997).

These assertions are, of course, nonsense. Furthermore classical statistical techniques are no more objective than Bayesian ones. However, the extent to which ecological risk assessment represents 'good science', and a transparent and robust decision making framework is intimately linked to the statistical procedures employed by the analyst.

The use of prior distributions

The prior distribution is arguably the most controversial aspect of Bayesian inference; a point long recognised by Bayesians (Reckhow 1996, Edwards *et al* 1963). Many classical statisticians distrust Bayes' theorem because it requires a prior distribution and they are suspicious of injecting subjective prior belief into the process of statistical inference.

Some of this controversy stems from the different interpretation of probability held by the two schools of statistical inference. Statisticians that assert probability as a true, real and objective phenomena are likely to object to the use of a prior probability distribution, particularly one based purely on subjective belief (see for example Edwards, 1972). Analysts that allow broader definitions of probability may have no problem accepting prior probabilities, particularly where there is good evidence to support them.

By the same token, however, classical and frequentist interpretations of probability are not without their detractors. Holdway (1997), for example, points out that projects which require ecological risk assessment cannot be replicated because individual ecosystems are unique, thus the frequentist requirement for repeatability is inevitably compromised. Subjective, or 'evidence based' (Kaplan, 1997), interpretations of probability are therefore more appropriate to ecological risk assessment.

Whilst these arguments are not without merit, there is little evidence to suggest that risk analysts choose one probability model over another because of problems regarding interpretation. Ecological risk analysts in particular are usually faced with much more pressing problems than the theoretical foundation of the probability concepts they employ. In practise any workable probability model is a useful one.

That is not to say that the analysts interpretation of probability should be ignored, rather if the interpretation allows a workable model to be generated, then this is usually a good start. Having said this, there are two reasons why prior probabilities should be treated cautiously by ecological risk analysts:

1. the psychological evidence suggests that humans are quite bad at estimating probability (Hampton *et al*, 1973). This is particularly pertinent to risk assessment because an expert's judgement of risk is as fallible as the public's; both are strongly influenced by knowledge and dread (Slovic, 1987)[16]; and,

---

[16] Substantive research has shown that 'experts' and laymen significantly overestimate the risks associated with unfamiliar activities, particularly if they are afraid of them (for example nuclear power

2. the parameters of the prior distribution (particularly the variance) can have an important
    bearing on the results of the analysis, even in the presence of strong evidence. Edwards
    (1996) demonstrates how over-confident prior information, that contradicts the evidence,
    can give rise to misleading results - in this instance even more confidence in the posterior
    distribution.

Additional controversy surrounding the use of prior distributions often occurs when there is
very little information on which to base prior beliefs. The analyst may be left to guess at
important parameters such as the variance of the distribution, which of itself has little intuitive
appeal (see for example Dilks *et al*, 1992). The problem can be exacerbated when prior
distributions are specified for model parameters which are difficult (or impossible) to measure.
For example in fishery risk assessments prior distributions are commonly specified for the
virgin or pre-exploited biomass of the stock (see for example Punt and Walker, 1997). There is,
however, no direct way of measuring this parameter – inferences can only be made with the
assistance of a fisheries model. Punt and Hilborn (1997) accordingly recommend that
considerable care is taken when specifying the prior. Rarely, however, does one see evidence of
this in the literature.

The analyst may be able to avoid these problems by considering alternative model structures
relative to the parameters that require prior distributions, and casting prior information in a form
which is readily modified by the sample information (Sharefkin, 1983). The latter will usually
involve the use of non-informative priors in the sense of Box and Tiao (1973), in other words a
prior that provides little information relative to what is expected to be provided by the
experiment (or sample).

The likelihood function and falsification principles

In ecological risk assessment, the form of the likelihood function reflects the analyst beliefs'
concerning fundamental ecological processes. The likelihood function is therefore a reflection
of model (epistemic) uncertainty, and as noted in section 2.2, this is the most intractable of all
uncertainty in the risk assessment process.

Gardner and O'Neill (1983) suggest that the likelihood function should be chosen on the basis
of the probabilistic properties of the system in question, the empirical distribution of data, and
any information on the expected distribution of system behaviours. They also note, however,
that this type of information is seldom available. Accordingly the form of the likelihood
function may be the most subjective choice in any ecological risk assessment. Unfortunately
there is no easy solution to this problem. Gelman *et al* (1995) advocate comparing different
likelihood models within an analysis. Again, however, it is rare to see this policy adopted in a
risk assessment[17], presumably because of time and cost constraints.

It is important to note that both the classical and Bayesian schools suffer from the same problem
here: the choice of model form - be it the form of the likelihood function (Bayesian) or the form
of the sampling distribution (classical). Dennis (1996), however, suggests that there are very
real differences because Bayesian models are less vulnerable to falsification (in the sense of
Popper's scientific method). This aspect of the debate is absolutely crucial because it is only
through rigorous testing of risk assessment predictions, that ecological risk assessment can

---

and genetic engineering). By contrast they typically underestimate the risks associated with familiar
activities such as smoking or drinking.
[17] See Crome *et al* (1996) for a notable exception.

defend itself against accusations of being pseudo-quantitative or pseudo-scientific nonsense (Holdway, 1997).

Dennis (1996) asserts that the classical approach to statistical inference (refer to section 2.3) places the emphasis on the model of the random process that is assumed to generate the data. The main purpose of the predictions made by the classical statistician is therefore to challenge this model. If the actual data produce outcomes which are extreme under this model, then the model itself is suspect. By contrast the Bayesian approach places its emphasis on the variability of the parameter. Dennis therefore suggests that it is more difficult for a Bayesian statistician to judge his or her models in light of extreme values – what is at fault, the model or the prior distribution?

Dennis' arguments are well made (particularly in light of the growing tide of anti-classical sentiment in the ecological risk assessment literature) but they are a little simplistic. In the first instance classical inference tests do not always provide unambiguous evidence for or against a particular model. In Appendix B for example a classical hypothesis test is conducted on relationship between Polychlorinated Biphenyl (PCB) concentrations and egg shell thickness. Assuming a normal linear regression model, the slope of the regression line was found to be –0.00028. A test of the null hypothesis that there is no relation between the PCB concentration and egg shell thickness (ie slope of the regression line = 0) resulted in a test statistic value of –2.08 with significance probability 0.042. This is sufficient to reject the null hypothesis in question at the 95% significance level – but only just. There are two potential problems with this result:

1. it is not very conclusive, indeed omitting a single data point results in a test statistic value of –1.256 with a significance probability of 0.214 (refer to Appendix B). This suggests that there is now insufficient evidence to reject the null hypothesis at the 95% significance level – what is at fault the data point or the model?;

2. it is dichotomous (the evidence is either sufficient or insufficient) and therefore open to misinterpretation. An inexperienced analyst might not realise that variability in the data can have an important bearing on the results of analysis, as in the case above. This can lead to erroneous conclusions of zero risk which are not in fact warranted (Hill, 1996).

Dennis (1996) also seems to imply that a residual-based diagnostic analysis (to ascertain whether or not a model makes accurate predictions) is not possible with a Bayesian approach. There is no reason to believe, however, that model predictions based on bayesian inference techniques are any less amenable to a diagnostic analysis than model predictions based on classical inference techniques.

Perhaps what is more important is that a model is compared against new information (as opposed to the information against which it was calibrated), and makes bold predictions capable of refutation (Corkett, 1997). In this context testing the accuracy of models is more concerned with model structure, than with which school of statistical inference the analyst adheres to. Fisheries risk assessments, for example, both classical and Bayesian, suffer the same problems of refutation (Collie, 1988). The credibility of ecological risk assessment lies in continually ground-truthing its predictions. The objective is to ensure that the risk assessment models remain 'well corroborated' until such time as they are falsified. In order to do this the models must be formulated so that they are relevant to the problem, and make predictions that can be readily tested.

5   CONCLUSIONS & SUMMARY

This document has examined Bayesian statistical inference techniques with a view to their use in the ballast water risk assessment currently being developed by CRIMP. This document was motivated by the fact that a quantitative assessment of ballast water risks must use inductive risk assessment techniques (because there is no relevant database on which to base empirical, frequentist techniques) and may have to employ subjective interpretations of probability (which is not possible with classical inference techniques).

The use of subjective probability in ecological risk assessment often raises questions of scientific validity – many statisticians discourage the use of subjective probability on the grounds that it is not 'objective' and therefore unscientific. It is important to recognise, however, that all statistical inference involves a number of important subjective decisions. For example one of the most important judgements that must be made is the functional form of the probability model(s) used – be it the likelihood function (Bayesian) or the sampling distribution of the statistic (classical). In either case this decision is largely subjective.

The use of subjective probability in quantified ecological risk assessment can therefore hardly be dismissed as unscientific – all statistical approaches to uncertainty and risk rely on some subjective input. Consequently appeals for objectivity do not provide a very compelling case against the use of Bayesian methods in ecological risk assessment.

Conversely the use of classical inference techniques in ecological science has been criticised on the grounds that:

1. frequentist and equi-likelihood interpretations of probability are hardly ever applicable to ecological systems; and,

2. classical estimation techniques assume that a single true parameter value exists, when in reality most physical and biological parameters are inherently variable.

In practise, however, probability need only be conceptualised in frequentist or equi-likelihood terms. The analogy need not be direct for the probability model to be useful one – indeed in most classical risk assessment simulations it hardly ever is. Furthermore Bayesian analysts make very similar assumptions regarding single parameter values (albeit implicitly) whenever they assume independent and identically distributed likelihood functions. Again there is very little to distinguish between the two statistical paradigms on these grounds.

Bayesian techniques are often cited as being well suited to complex, multi-parameter models, which are sometimes needed in ecological science, because of the ease with which nuisance parameters can be integrated out of the joint posterior distribution function. This may be true with a limited set of parameters (say 2, 3 or 4) but in order to derive the joint distribution function of a large number of parameters, the analyst must first assign prior probability distributions to each parameter and then account for any correlation between them. This may prove no simpler than performing a classical simulation of the same multi-parameter model.

Bayesian techniques have also been quoted as producing 'common sense' interpretations of uncertainty, and as being superior to classical estimation methods in data sparse situations. The first of these arguments, although important to ecological risk assessment, is far from proven. The literature does not appear to provide any conclusive evidence to support the contention that lay-persons are able to interpret the results of a Bayesian analysis any better than the results of a classical analysis (for example in the form of a controlled experiment).

Bayesian techniques can provide useful results in data sparse situations (that often characterise ecological risk problems), but this is very much dependant on the 'accuracy' of the prior distribution function. If data are truly sparse the likelihood function may be very diffuse, and in these situations the prior distribution function (even of a non-informative prior) exerts considerable influence on the shape of the posterior distribution. Since the functional form assumed for the likelihood function ultimately determines the shape of a non-informative prior, the resulting posterior is likely to prove a very subjective interpretation of the information contained in the data. Data poor situations may therefore pose considerable problems to Bayesian and classical ecologists alike.

Bayesian posterior distributions can, however, be updated very quickly as and when more information is made available to the analyst. As such Bayesian techniques are well suited to the iterative nature of quantitative risk assessment whereby risk estimates are made but then continually adjusted with time. Furthermore the Bayesian posterior distribution immediately directs the analyst to the full distributional qualities of any parameter estimate, as opposed to point estimates, confidence intervals or dichotomous hypothesis-test results. Equivalent distribution qualities can be achieved with classical techniques by simply estimating the moments of the appropriate distribution, but the approach is not as direct as in the Bayesian case.

If the case for or against Bayesian techniques in quantified ecological risk assessment rested solely on the arguments above, one might conclude that Bayesian approaches had a very slight edge on more traditional classical approaches. There remains, however, the issue of the prior distribution. In situations where there is sufficient data to produce a well defined likelihood curve, the potential problems associated with a prior distribution can usually be avoided by using a non-informative prior. Similarly if the data allow a frequency interpretation for the prior, whose parameters represent the synthesis and analysis of existing data, then the analyst is unlikely to be criticised for adopting a Bayesian approach.

In data scare situations, however, including those situations in which there is very little evidence to support the analyst's prior beliefs, Bayesian risk assessments are probably less repeatable than classical approaches. It is difficult therefore to define when and where Bayesian approaches are better suited to quantitative ecological risk assessment than more traditional approaches. On balance each case is probably best approached on merits bearing in mind that:

4. both classical and Bayesian risk assessments require important subjective decisions of the analyst – the extent to which these decision dominate the analysis, however, is dependant on the availability and quality of data;

5. classical inference techniques cannot be used with subjective interpretations of probability, and since this is a valid component of the risk analyst's tool box, Bayesian techniques form an important alternative approach to quantified ecological risk assessment; and,

6. Bayesian statistical inference is well suited to the iterative development of quantitative risk assessment and quickly emphasises the full distribution qualities of uncertain parameters. The results of a Bayesian analysis, however should be judged in light of the data that was available to the analyst and the extent to which an independent analyst might arrive at the same or similar conclusions.

REFERENCES

Anderson J. L. (1998), Embracing Uncertainty: The Interface of Bayesian Statistics and Cognitive Psychology, *Conservation Ecology [Online], 2* (1):2.

Berger J. O. and Berry D. A. (1988), Statistical Analysis and the Illusion of Objectivity, *American Scientist,* **76**:159-165.

Box G. E. P. and Tiao G. C. (1973), *Bayesian Inference in Statistical Analysis*. John Wiley and Sons Inc., New York.

Carlton J. T., Reid D. M. and van Leeuwen H. (1995), *The Role of Shipping in the Introduction of Nonindigenous Aquatic Organisms to the Coastal Waters of the United States (Other Than the Great Lakes) and an Analysis of Control Options*, Report No. CG-D-11-95, National Technical Information Service, Springfield, Virginia 22161.

Collie J. (1988), Book Review: Dynamics of Marine Fish Populations, *Natural Resource Modelling, 2* (4):673-677.

Cordue P. L. and Francis R. I. C. C. (1994), Accuracy and Choice in Risk Estimation for Fisheries Assessment, *Can. J. Fish. Aquat. Sci.,* **51**:817-829.

Corkett C. J. (1997), Managing the Fisheries by Social Engineering: a Re-Evaluation of the Methods of Stock Assessment, *Journal of Applied Ichthyology,* **13**:159-170.

Cox D. R. and Hinkley D. V. (1974), *Theoretical Statistics*. Chapman and Hall, London.

Crome E. H. J., Thomas M. R. and Moore L. A. (1996), A Novel Bayesian Approach to Assessing Impacts of Rain Forest Logging, *Ecological Applications, 6* (4):1104-1123.

Dennis B. (1996), Should Ecologists Become Bayesians, *Ecological Applications, 6* (4):1095-1103.

Dilks D. W., Canale R. P. and Meier P. G. (1992), Development of Bayesian Monte Carlo Techniques for Water Quality Model Uncertainty, *Ecological Modelling,* **62**:149-162.

Edwards A. W. F. (1992), *Likelihood*. John Hopkins Press Ltd., London.

Edwards D. (1996), The First Data Analysis Should Be Journalistic, *Ecological Applications, 6* (4):1090-1094.

Edwards W., Lindman H. and Savage L. J. (1963), Bayesian Statistical Inference for Psychological Research, *Psychological Review,* **70**:193-242.

Ellison A. M. (1996), An Introduction to Bayesian Inference for Ecological Research and Environmental Decision Making, *Ecological Applications, 6* (4):1036-1046.

Evans M., Hastings N. and Peacock B. (1993), *Statistical Distributions*. John Wiley & Sons, New York.

Faber M., Manstetten R. and Proops J. (1992), 'Toward an Open Future: Ignorance, Novelty and Evolution' in Costanza R., Norton B. G. and Haskell B. D. (eds), *Ecosystem Health: New Goals for Environmental Management*, pp. 73-96. Island Press, Washington D. C.

Freund J. E. (1992), *Mathematical Statistics*. Prentice Hall, New Jersey.

Gardner R. H. and O'Neill R. V. (1983), Parameter Uncertainty and Model Predictions: A Review of Monte Carlo Results, in Beck M. B. and van Straten G. (eds), *Uncertainty and Forecasting of Water Quality,* pp. 245-257. Springer-Verlag, Berlin.

Gazey W. J. and Staley M. J. (1986), Population Estimation From Mark-Recapture Experiments Using a Sequential Bayes Algorithm, *Ecology,* **67** (4):941-951.

Gelman A., Carlin J. B., Stern H. S. and Rubin D. B. (1995), *Bayesian Data Analysis*. Chapman and Hall, London.

Good I. J. (1959), Kinds of Probability, *Science,* **129** (3347):443-447.

Hampton J. M., Moore P. G. and Thomas H. (1973), Subjective Probability and Its Measurement, *Journal of the Royal Statistical Society A,* **136** (Part 1):21-42.

Hayes K. R. (1997), *A Review of Ecological Risk Assessment Methodologies*, CRIMP Technical Report No. 13, CSIRO Division of Marine Research, Hobart.

Hayes K. R. and Hewitt C. L. (1998), *Risk Assessment Framework for Ballast Water Introductions,* CRIMP Technical Report Number 14, CSIRO Division of Marine Research, Hobart.

Hill R. A. (1996), From Science to Decision Making: The Applicability of Bayesian Methods to Risk Assessment, *Human and Ecological Risk Assessment,* **2** (4):636-642.

Holdway D. A. (1997), Truth and Validation in Ecological Risk Assessment, *Environmental Management,* **21** (6):816-819.

Iongh H. H. and Van Zon J. C. J. (1993), Assessment of Impact of the Introduction of Exotic Fish Species in North-East Thailand, *Aquaculture and Fisheries Management,* **24**:279-289.

Jaykus L. (1996), The Application of Quantitative Risk Assessment to Microbial Food Safety Risks, *Critical Reviews in Microbiology,* **22** (4):279-293.

Jeffreys H. (1961), *Theory of Probability*. Clarendon Press, Oxford.

Johnson D. H. (1977), *Some Bayesian Statistical Techniques Useful in Estimating Frequency and Density,* Special Scientific Report - Wildlife No. 203, United States Department of Interior, Fish and Wildlife Service, Washington D. C.

Johnson D. H. (1989), An Empirical Bayes Approach to Analysing Recurring Animal Surveys, *Ecology,* **70** (4):945-952.

Kaplan S. (1997), The Words of Risk Analysis, *Risk Analysis,* **17** (4):407-417.

Kot M., Lewis M. A. and Driessche P. van den (1996), Dispersal Data and the Spread of Invading Organisms, *Ecology,* **77** (7):2027-2042.

Law A. M. and Kelton D. W. (1991), *Simulation Modelling and Analysis,* McGraw-Hill, New York.

Lindley D. V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint: Part I. Probability*. Cambridge University Press, Cambridge.

Ludwig D. (1996), Uncertainty and the Assessment of Extinction Probabilities, *Ecological Applications,* **6** (4):1067-1076.

McKaye K. R., Makwinja D. R., Menyani W. W. and Mhone O. K. (1985), On the Possible Introduction of Non-Indigenous Zooplankton-Feeding Fishes into Lake Malawi, Africa, *Biological Conservation,* **33**:289-307.

McPherson G. (1990), *Statistics in Scientific Investigation: Its Basis, Application and Interpretation*. Springer-Verlag, New York.

Moore D. R. J. (1996), Using Monte Carlo Analysis to Quantify Uncertainty in Ecological Risk Assessment: Are We Gilding the Lily or Bronzing the Dandelion?, *Human and Ecological Risk Assessment,* **2** (4):628-633.

Morgan M. G. and Henrion M. (1990), *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk Assessment and Policy Analysis*. Cambridge University Press, Cambridge.

Murphy K. (1997), *The Survival and Sampling of Zooplankton in Ballast Water*, BSc thesis submitted to the Department of Zoology at the University of Tasmania.,

Open University (1995), *Elements of Statistics*. The Bath Press, Avon.

Pascual M. A. and Kareiva P. (1996), Predicting the Outcome of Competition Using Experimental Data: Maximum Likelihood and Bayesian Approaches, *Ecology,* **77** (2):337-349.

Pate-Cornell M. E. (1996), Uncertainties in Risk Analysis: Six Levels of Treatment, *Reliability Engineering and System Safety,* **54**:95-111.

Power M. (1996), Probability Concepts in Ecological Risk Assessment, *Human and Ecological Risk Assessment,* **2** (4):650-654.

Power M., Dixon G. D. and Power G. (1994), Perspectives on Environmental Risk Assessment, *Journal of Aquatic Ecosystem Health ,* **3**:69-79.

Press S. J. (1989), *Bayesian Statistics: Principles, Models and Applications*. John Wiley & Sons, New York.

Punt A. E. and Hilborn R. (1997), Fisheries Stock Assessment and Decision Analysis: the Bayesian Approach. *Reviews in Fish Biology and Fisheries*, **7**(1): 35-63.

Punt A. E. and Walker T. I (1997), Stock Assessment and Risk Analysis for the School Shark Off Southern Australia, CSIRO Division of Marine Research, Hobart.

Rasmussen N. C. (1981), The Application of Probabilistic Risk Assessment Techniques to Energy Technologies, *Ann. Rev. Energy,* **6**:122-38.

Reckhow K. H. (1996), Improved Estimation of Ecological Effects Using an Empirical Bayes Method, *Water Resources Bulletin,* **32** (5):929-935.

Reckhow K. H. and Chapra S. C. (1983), Confirmation of Water Quality Models, *Ecological Modelling,* **20**:113-133.

Riseborough R. W. (1972), 'Effects of Environmental Pollutants upon Animals Other than Man.' *Proceedings of the 6th Berkeley Symposium on Mathematics and Statistics, VI.* pp. 443-463. University of California Press, California.

Rodier D. J. and Zeeman M. G. (1994), Ecological Risk Assessment, in Cockerham L. G. and Shane B. S. (eds), *Basic Environmental Toxicology,* 581-604. CRC Press, Boca Raton.

Schnute J. T. and Hilborn R. (1993), Analysis of Contradictory Data Sources in Fish Stock Assessment, *Can. J. Fish. Aquat. Sci.,* **50**:1916-1923.

Schmitt S. A. (1969), *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics*. Addison Wesley, Reading, Massachusetts.

Sharefkin M. (1983), Uncertainty and Forecasting of Water Quality: Reflections of an Ignorant Bayesian, in Beck M. B. & Straten G. V. (eds), *Uncertainty and Forecasting of Water Quality,* pp.373-379. Springer-Verlag, New York.

Slovic P. (1987), Perception of Risk, *Science,* **236**:280-285.

Smith A. F. M (1984), Present Position and Potential Developments: Some Personal Views Bayesian Statistics, *Journal of the Royal Statistical Society A ,* **147** (Part 2):245-259.

Solow A. R. and Gaines A. G. (1995), An Empirical Bayes Approach to Monitoring Water Quality, *Environmetrics,* **6**:1-5.

Stow C. A., Carpenter S. R. and Lathrop R. C. (1997), A Bayesian Observation Error Model to Predict Cyanobacterial Biovolume From Spring Total Phosphorous in Lake Mendota, Wisconsin, *Can. J. Fish. Aquat. Sci.,* **54**:464-473.

Suter G. (1993), *Ecological Risk Assessment*, Lewis Publishers, Michigan.

Sutton C. A., Murphy K., Martin R. B. and Hewitt C. L. (1998), *A Review and Evaluation of Ballast Water Sampling Protocols,* CRIMP Technical Report No. 18, Centre for Research on Introduction Marine Pests, Hobart, Tasmania.

Vose D. (1996), *Quantitative Risk Assessment: A Guide to Monte Carlo simulation Modelling*. John Wiley & Sons, New York.

Walters C. and Ludwig D. (1994), Calculation of Bayes Posterior Probability Distributions for Key Population Parameters, *Can. J. Fish. Aquat. Sci.,* **51**:713-722.

Weathers K. and Reeves E. (1996), The Defence of the Great Lakes Against the Invasion of
    Nonindigenous Species in Ballast Water, *Marine Technology,* **33** (2):92-100.

Welcomme R. L. (1992), A History of International Introductions of Inland Aquatic Species,
    ICES Mar. Sci. Symp., **194**:3-14.

Wilson R. and Crouch E. A. C. (1987), Risk Assessment and Comparisons: An Introduction,
    Science, **236**:267-270.

Wonham J., Walton W. C., Frese A. M. and Ruiz G. M. (1996), *Transoceanic Transport of
    Ballast Water: Biological and Physical Dynamics of Ballasted Communities and the
    Effectiveness of Mid-Ocean Exchange,* Final report submitted to US Fish & Wildlife
    Service and the Compton Foundation, Smithsonian Environmental Research Centre,
    Edgewater, MD 21037.

Young P. (1983), The Validity and Credibility of Models for Badly Defined Systems, in Beck
    M. B. & van Straten G. (eds), *Uncertainty and Forecasting of Water Quality,* pp. 69-98.
    Springer-Verlag, Berlin.

APPENDIX A   THE AXIOMS OF PROBABILITY

The purpose of any axiom system is to allow fruitful mathematical representations of the real world. Classical examples include the axioms of Euclidean geometry regarding lines and points and the subsequent deductions regarding geometrical figures (Lindley, 1965).

In what follows the probability of an event A (eg an observation or outcome of an experiment) is written as P(A), the probability of an event B as P(B) and so forth. The letter S represents the set of all possible outcomes of a system or experiment (the sample space).

Axiom 1

The probability of any event is a non-negative real number. This is $P(A) \geq 0$ for any subset A of S.

Axiom 2

$P(S) = 1$

Axiom 3

If $A_1, A_2, A_3, \ldots\ldots$ is a finite or infinite sequence of mutually exclusive subsets of S then:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots\ldots$$

APPENDIX B   CLASSICAL INFERENCE EXAMPLE

The data displayed Figure B1 are taken from the Open University (1995) but were originally published in Risebrough (1972). The data show the thickness of 65 Anacapa pelican eggs (in millimetres) plotted against the concentration (in parts per million) of Polychlorinated Biphenyl (PCB), an industrial pollutant.

Figure B1   Shell thickness and PCB concentrations measured in 65 Anacapa pelican eggs
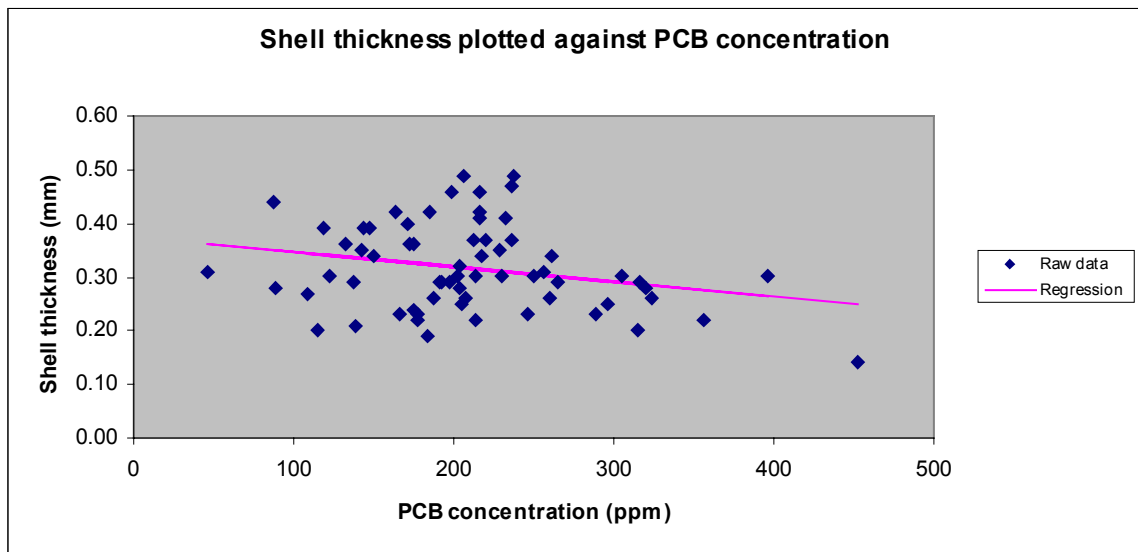
This example will demonstrate some classical inference techniques based on the data above. The objective is to investigate the relationship between shell thickness and PCB concentration. We are assuming that shell thickness (the response variable) can be predicted from PCB concentration (explanatory constants) using a normal linear regression model. Classical inference techniques will be used to estimate the parameters of this model (the slope of the regression line β, the intersect α and the underlying variance σ^2), investigate the sampling distribution of β, and test the hypothesis that shell thickness and PCB concentration are not in fact related (ie $\beta = 0$).

Assuming a normal linear regression model, the random variables Y_i for each fixed x_i are given by

$$Y_i = \alpha + \beta x_i + W_i \quad ,$$

where; Y = response variable (shell thickness)
        x = explanatory constants (PCB concentration)
        β = the slope of the regression line
        α = the intersect of the regression line
        W = normal random error term with mean 0 and variance σ^2.

This relationship may be equivalently expressed as

$$Y_i \approx N(\alpha + \beta x_i, \sigma^2) \quad ,$$

such that

$$f(y_i / x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left[\frac{y_i - (\alpha + \beta x_i)}{\sigma} \right]^2 \right\} \quad ,$$

where α, β and σ are the same for each i.

There are several different ways to estimate the parameters of the this model. The maximum likelihood method can be used for all three parameters, least squares estimation can be used for α and β (giving the same result as the maximum likelihood method), whilst the method of moments can be used for σ. To obtain maximum likelihood estimates of α and β requires partially differentiating the likelihood function

$$L(y_i / x_i) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left[\frac{y_i - (\alpha + \beta x_i)}{\sigma} \right]^2 \right\} \quad ,$$

or its logarithm (which is easier), with respect to α and β, equating the expressions to zero, and then solving them. The resulting estimates are

$$\hat{\beta} = \frac{n\left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)}$$

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} y_i - \hat{\beta} \sum_{i=1}^{n} x_i}{n} \quad .$$

By introducing the notation

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 \quad ,$$

and

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) \quad ,$$

these estimates can be equivalently expressed as

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \quad ,$$

and

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x} \ .$$

By writing

$$\hat{B} = \frac{S_{xY}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{y})}{S_{xx}} = \sum_{i-1}^{n}\left(\frac{x_i - \overline{x}}{S_{xx}}\right)Y_i$$

it can be seen that the estimator[18] $\hat{B}$ is a linear combination of n independent normal random variables Y_i. It follows that $\hat{B}$ has a normal distribution with expected value

$$E\left(\hat{B}\right) = \sum_{i=1}^{n}\left[\frac{x_i - \overline{x}}{S_{xx}}\right]E(Y_i / x_i) = \beta \quad ,$$

and variance

$$\text{var}\left(\hat{B}\right) = \sum_{i=1}^{n}\left[\frac{x_i - \overline{x}}{S_{xx}}\right]^2 \text{var}(Y_i / x_i) = \frac{\sigma^2}{S_{xx}} \quad .$$

The quantity

$$Z = \frac{\hat{\beta} - \beta}{\sigma / \sqrt{S_{xx}}} \quad ,$$

is therefore a standard normal random variate. Under the assumptions of the normal linear regression model it can also be shown that

$$\frac{n\hat{\sigma}^2}{\sigma^2} \quad ,$$

is a value of a chi-squared random variable with two degrees of freedom. This random variable is independent of $\hat{B}$ allowing substitution into the definition of student's t distribution such that

$$t = \frac{\dfrac{\hat{\beta} - \beta}{\sigma / \sqrt{S_{xx}}}}{\sqrt{\dfrac{n\hat{\sigma}^2}{\sigma^2} / (n - 2)}} = \frac{\hat{\beta} - \beta}{\hat{\sigma}}\sqrt{\frac{(n - 2)S_{xx}}{n}} \quad ,$$

is a value of a random variable having the t distribution with n-2 degrees of freedom. Replacing $\hat{\sigma}^2$ with its method of moments estimate

---

[18] The term estimate is used to denote a number obtained from data. An estimator refers to a random variable expressing an estimating formula.

$$s^2 = \frac{\sum\limits_{i=1}^{n} (y_i - \hat{y})^2}{n - 2} \quad , \qquad\qquad\qquad [B1]$$

provides an important result for the sampling distribution of β without the nuisance parameter σ

$$\frac{\hat{\beta} - \beta}{S / \sqrt{S_{xx}}} \approx t(n - 2) \quad . \qquad\qquad\qquad [B2]$$

This distribution can now be used to derive confidence intervals for the slope parameter β, and to test the null hypothesis that the response variable and explanatory constants are not in fact related, ie that β = 0.

Using the data illustrated in Figure B1, the maximum likelihood estimates of the slope and intersect parameters are $\hat{\beta} = -0.00028$ and $\hat{\alpha} = 0.3749$, such that the normal linear regression model is

$$Y_i = 0.3749 - 0.00028 x_i + W_i \quad .$$

Having derived the sampling distribution of β the analyst is able to test the null hypothesis that shell thickness and PCB concentration are not in fact related, ie β = 0. The standard error of the estimate given by equation B1 is

$$s = \sqrt{\frac{\sum\limits_{i=1}^{n} (y_i - \hat{y})^2}{n - 2}} = \sqrt{\frac{0.388}{63}} = 0.0785 \quad .$$

Under the null hypothesis β = 0 the value of the test statistic is calculated from equation B2

$$t = \frac{(-0.00028) - 0}{0.0785 / \sqrt{340341.75}} = -2.08$$

Comparing the value of the test statistic against student's t distribution with 63 degrees of freedom yields a total significance probability[19] of 0.042. This result suggests that at the 95% significance level there is sufficient evidence to reject the null hypothesis. The results of the test, however, are not very compelling – the value of the significance probability is only just significant, suggesting that the evidence portrayed in Figure B1 is not conclusive. Indeed re-running the analysis but omitting the data point in the bottom right corner of Figure B1, resulted in slope and intersect estimates of $\hat{\beta} = -0.000184$ and $\hat{\alpha} = 0.357$, such that the normal linear regression model is

$$Y_i = 0.357 - 0.000184 x_i + W_i \quad .$$

---

[19] If the statistical experiment were to be repeated numerous times, and if the null hypothesis were true, the Significance Probability (SP) represents the proportion of future experiments that would offer less support for the null hypothesis than the experiment that was in fact performed.

Under the same null hypothesis, however, the value of the test statistic becomes -1.256, which yields a total significance probability of 0.214. Thus by omitting one data point from the analysis there is no longer sufficient evidence to reject the null hypothesis.

APPENDIX C   BALLAST SAMPLING PROBLEM

Table C1   Conditional probabilities for ballast sampling problem

Ballast tank status	Sample result = Positive (Pos)	Sample result = Negative(Neg)
Infected (I)	P(Pos/I) = 0.70	P(Neg/I) = 0.30
Uninfected (U)	P(Pos/U) = 0.05	P(Neg/U) = 0.95

Table C2   Calculating the posterior probability of ballast tank infection (one sample)

Alternative	Prior	Pr(Neg/alt)	Joint	Posterior
Tank infected (I)	0.5	0.30	0.150	0.240
Tank not infected (U)	0.5	0.95	0.475	0.760
			Σ 0.625	Σ 1.00

Table C3   The effect of collecting more samples on the posterior probability

Alternative	Prior	Pr(Neg/alt)	Joint	Posterior
Tank infected (I)	0.240	0.30	0.072	0.091
Tank not infected (U)	0.760	0.95	0.722	0.909
			Σ 0.794	Σ 1.00
Tank infected (I)	0.091	0.30	0.027	0.031
Tank not infected (U)	0.909	0.95	0.864	0.969
			Σ 0.891	Σ 1.00
Tank infected (I)	0.031	0.30	0.009	0.010
Tank not infected (U)	0.969	0.95	0.921	0.990
			Σ 0.930	Σ 1.00

APPENDIX D   JOURNEY SURVIVAL PROBLEM

The length of time (T) that a population is expected to survive in the ballast tank can be viewed as a random variable. For a some species, the number of individuals that survive to time t has been observed to decline exponentially as t increases, such that

$$P(T \geq t) = C \exp\left(-\frac{t}{\mu}\right) \quad , \qquad\qquad t \geq 0$$

where $1/\mu = \lambda$ = the death rate of the population. Let p(t) represent the probability distribution function of the random variable T, such that

$$\int_{t}^{\infty} p(x).dx = P(T \geq t) = C \exp\left(-\frac{t}{\mu}\right) \quad .$$

Differentiating with respect to t (using Part 1 of the fundamental theorem of calculus) yields

$$p(t) = \frac{1}{\mu} C \exp\left(-\frac{t}{\mu}\right) \quad .$$

C is a normalising constant which ensures that the second axiom of probability is satisfied such that

$$1 = \int_{0}^{\infty} \frac{1}{\mu} C \exp\left(-\frac{t}{\mu}\right).dt$$

$$= \frac{1}{\mu} C \int_{0}^{\infty} \exp\left(-\frac{t}{\mu}\right).dt$$

$$= \lim_{R \to \infty} \frac{1}{\mu} C \left[-\mu \exp\left(-\frac{t}{\mu}\right)\right]_{0}^{\infty}$$

$$= -C \lim_{R \to \infty} \left[\exp\left(-\frac{R}{\mu}\right) - 1\right]$$

$$= C \qquad .$$

According to this model the random variable T is said to follow an exponential distribution with expected value μ such that the probability of observing a survival time t given by

$$p(t) = \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right) \quad .$$

If individuals are still alive at the end of the vessel's journey, the appropriate probability is provided by the right tail of the exponential distribution; the probability that the survival time is at least some value x

$$P(t \rangle x) = 1 - F(x)$$

$$= 1 - \left[1 - \exp\left(-\frac{x}{\mu} \right) \right]$$

$$= \exp\left(-\frac{x}{\mu} \right) \quad .$$

For n independent observations of extinction time in the ballast tank, $\mathbf{t} = (t_1, t_2, ...t_n)$, and m independent observations of journeys which the species survived, $\mathbf{x} = (x_1, x_2, ... x_m)$, the likelihood function for μ is given by

$$l(\mu / \mathbf{t}, \mathbf{x}) = \prod_{i=1}^{n} \left[\frac{1}{\mu} \exp\left(-\frac{t_i}{\mu} \right) \right] \times \prod_{j=1}^{m} \left[\exp\left(-\frac{x_j}{\mu} \right) \right]$$

$$= \left(\frac{1}{\mu} \right)^n \exp\left[-\sum_{i=1}^{n} \frac{t_i}{\mu} \right] \times \exp\left[-\sum_{j=1}^{m} \frac{x_j}{\mu} \right]$$

$$= \left(\frac{1}{\mu} \right)^n \exp\left[-\frac{1}{\mu} (\sum_{i=1}^{n} t_i + \sum_{j=1}^{m} x_j) \right] \quad .$$

Journey survival model #1

Assuming a uniform prior distribution for μ, $p_0(\mu) = k$, on the support $0 \leq \mu \leq \infty$ and applying Bayes theorem, we obtain

$$p(\mu / \mathbf{t}, \mathbf{x}) = c \left(\frac{1}{\mu} \right)^n \exp\left(-\frac{F}{\mu} \right) \quad ,$$

where $F = (\Sigma t_i + \Sigma x_j)$, and k has been subsumed into the normalising constant c which is given by

$$1 = \int_0^\infty c \left(\frac{1}{\mu} \right)^n \exp\left(-\frac{F}{\mu} \right) d\mu \quad .$$

The integral is evaluated by making the change of variable $F/\mu = z$ as follows

$$1 = \int_\infty^0 -c \left(\frac{z}{F} \right)^n \exp(-z) \frac{F}{z^2} dz$$

$$= \frac{c}{F^{n-1}} \int_0^{\infty} z^{n-2} \exp(-z)dz \quad .$$

Using the Euler definition of the gamma function

$$\Gamma(\alpha) = \int_0^{\infty} z^{\alpha-1} \exp(-z)dz \quad ,$$

where α is a positive integer and $\Gamma(\alpha) = (\alpha - 1)!$, this reduces to

$$1 = \frac{c}{F^{n-1}}(n-2)! \quad ,$$

such that

$$c = \frac{F^{n-1}}{(n-2)!} \quad .$$

The posterior probability distribution of μ given the data vectors $\mathbf{t}$ and $\mathbf{x}$ is therefore given by

$$p(\mu / \mathbf{t}, \mathbf{x}) = \frac{F^{n-1}}{(n-2)!}\left(\frac{1}{\mu}\right)^n \exp\left(-\frac{F}{\mu}\right) \quad ,$$

where $F = (\Sigma t_i + \Sigma x_j)$, $t \geq 0$, $x \geq 0$ and n is a positive integer.

Journey survival model #2

Assuming a uniform prior distribution for μ, $p_0(\mu) = k$, on the support $0 \leq \mu \leq M$ and applying Bayes theorem we obtain

$$p(\mu / \mathbf{t}, \mathbf{x}) = c\left(\frac{1}{\mu}\right)^n \exp\left(-\frac{F}{\mu}\right) \quad ,$$

where $F = (\Sigma t_i + \Sigma x_j)$, and k has been subsumed into the normalising constant c which is given by

$$1 = \int_0^M c\left(\frac{1}{\mu}\right)^n \exp\left(-\frac{F}{\mu}\right)d\mu \quad .$$

This integral has no closed form solution and must be evaluated numerically. For the distribution illustrated in Figure 3.3 this was achieved using Simpson's Rule with 300 unit increments.

Journey survival model #3

Assuming a non-informative prior distribution for μ, $p_0(\mu) \propto 1/\mu$, on the support $0 < \mu < \infty$ and applying Bayes theorem, we obtain

$$p(\mu/\mathbf{t},\mathbf{x}) = c\left(\frac{1}{\mu}\right)^{n+1} \exp\left(-\frac{F}{\mu}\right) \quad,$$

where $F = (\Sigma t_i + \Sigma x_j)$, and c is the normalising constant such that

$$1 = \int_0^\infty c\left(\frac{1}{\mu}\right)^{n+1} \exp\left(-\frac{F}{\mu}\right) d\mu \quad.$$

The integral is evaluated by making the change of variable $F/\mu = z$ as follows

$$1 = \int_\infty^0 -c\left(\frac{z}{F}\right)^{n+1} \exp(-z)\frac{F}{z^2} dz$$

$$= \frac{c}{F^n} \int_0^\infty z^{n-1} \exp(-z) dz \quad.$$

Using the Euler definition of the gamma function

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} \exp(-z) dz \quad,$$

where α is a positive integer and $\Gamma(\alpha) = (\alpha - 1)!$, this reduces to

$$1 = \frac{c}{F^n}(n-1)! \quad,$$

such that

$$c = \frac{F^n}{(n-1)!} \quad.$$

The posterior probability distribution of μ given the data vectors $\mathbf{t}$ and $\mathbf{x}$ is therefore given by

$$p(\mu/\mathbf{t},\mathbf{x}) = \frac{F^n}{(n-1)!}\left(\frac{1}{\mu}\right)^{n+1} \exp\left(-\frac{F}{\mu}\right) \quad,$$

where $F = (\Sigma t_i + \Sigma x_j)$, $t \geq 0$, $x \geq 0$ and n is a positive integer.

Journey survival model #4

Assuming a non-informative prior distribution for μ, $p_0(\mu) \propto 1/\mu$, on the support $0 < \mu < M$ and applying Bayes theorem, we obtain

$$p(\mu / \mathbf{t}, \mathbf{x}) = c\left(\frac{1}{\mu}\right)^{n+1} \exp\left(-\frac{F}{\mu}\right) \quad,$$

where $F = (\Sigma t_i + \Sigma x_j)$, and c is the normalising constant such that

$$1 = \int_0^M c\left(\frac{1}{\mu}\right)^{n+1} \exp\left(-\frac{F}{\mu}\right) d\mu$$

This integral has no closed form solution and must be evaluated numerically. For the distribution illustrated in Figure 3.5 this was achieved using Simpson's Rule with 300 unit increments.

APPENDIX E1   BINOMIAL MODEL

Variate	Y is a binomial random variable (the number of success in n Bernoulli trials), θ is the probability of success at each trial.

Likelihood function $l(\theta/y)$

Binomial model:

$$\binom{n}{y}\theta^y(1-\theta)^{n-y}$$

Conjugate prior probability $p_0(\theta)$

Beta distribution:

$$\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

Non-informative prior probability $p_0(\theta)$

$$\propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$$

Posterior probability $p(\theta/y)$

Beta distribution:

$$\frac{1}{B(y+\alpha,n-y+\beta)}\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

Derivation of posterior distribution

Bayes theorem states:

$$p(\theta/y) \propto L(\theta/y)p_0(\theta)$$

$$\propto \theta^y(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}.$$

The normalising constant is therefore given by

$$\frac{1}{\int_0^1 \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}d\theta} \quad .$$

Using the Beta function identity

$$B(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1}dx ,$$

and its relation to the Gamma function

$$B(m,n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)},$$

the posterior distribution becomes

$$p(\theta / y) = \frac{1}{B(y+\alpha, n-y+\beta)} \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)} \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta+1}.$$

Derivation of uninformative prior distribution

$L = \ln l(\theta/y) = \text{constant} + y\ln\theta + (n-y)\ln(1-\theta)$ such that

$$\frac{\partial L}{\partial \theta} = \frac{y}{\theta} - \frac{(n-y)}{(1-\theta)}$$

$$\frac{\partial^2 L}{\partial \theta^2} = -\frac{y}{\theta^2} - \frac{(n-y)}{(1-\theta)^2} \quad .$$

Since $E(Y) = n\theta$ it follows that

$$J(\theta) = \mathop{E}_{y/\theta}\left(-\frac{\partial^2 L}{\partial \theta^2}\right) = \frac{-n\theta}{\theta} - \frac{(n-n\theta)}{(1-\theta)^2}$$

$$\propto \theta^{-1}(1-\theta)^{-1},$$

where $J(\theta)$ is Fisher's information measure. Using Jeffrey's rule

$$p_0(\theta) \propto J^{\frac{1}{2}}(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}},$$

which is Beta (½, ½) distribution with normalising constant $B(½, ½)^{-1}$.

Posterior distribution with non-informative prior $p(\theta/y)$

Beta (y+½, n-y+½) distribution

$$p(\theta / y) = \frac{\Gamma(n+1)}{\Gamma\left(y+\frac{1}{2}\right)\Gamma\left(n-y+\frac{1}{2}\right)} \theta^{y-\frac{1}{2}}(1-\theta)^{n-y-\frac{1}{2}} \quad 0 < \theta < 1$$

APPENDIX E2   POISSON MODEL

Variate	Y is a Poisson random variable (the number of events within a specified time interval t), λ is the average rate of these events
Likelihood function $l(\theta/y)$	Poisson model: $$\frac{\lambda^y}{y!}\exp(-\lambda)$$
Conjugate prior probability $p_0(\theta)$	Gamma distribution: $$\frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\frac{\lambda}{\beta})$$
Non-informative prior probability $p_0(\theta)$	$$\propto \lambda^{-\frac{1}{2}}$$
Posterior probability $p(\theta/y)$	Gamma $(n\bar{y}+\alpha, \beta+n)$ distribution: $$\frac{(\beta+n)^{\alpha+n\bar{y}}}{\Gamma(\alpha+n\bar{y})}\lambda^{(\alpha+n\bar{y}-1)}\exp\left[-(\beta+n)\lambda\right]$$
Derivation of posterior distribution	Suppose $\mathbf{y}=(y_1...y_n)$ is a set of n independent frequencies each distributed as a Poisson distribution with mean λ. Then given $\mathbf{y}$, the likelihood is $$l(\lambda/\mathbf{y})=\prod_{i=1}^{n}\frac{\lambda^{y_i}}{y_i!}\exp(-\lambda)$$ $$=\frac{\lambda^{\Sigma y_i}}{y_1!.y_2!...y_n!}\exp(-n\lambda)$$ $$\propto \lambda^{n\bar{y}}\exp(-n\lambda),$$ where $\bar{y}=\dfrac{\Sigma y_i}{n}$. Using Bayes theorem $$p(\lambda/\mathbf{y})\propto p_0(\lambda)l(\lambda/\mathbf{y}) ,$$

it follows that

$$p(\lambda / \mathbf{y}) \propto \lambda^{\alpha-1} \exp(-\beta\lambda) \lambda^{n\bar{y}} \exp(-n\lambda)$$

$$\propto \lambda^{(n\bar{y}+\alpha-1)} \exp\left[-(\beta+n)\lambda\right].$$

The normalising constant is therefore given by

$$c = \left[\int_0^\infty \lambda^{(n\bar{y}+\alpha-1)} \exp\left(\left[-(\beta+n)\lambda\right]d\lambda\right)\right]^{-1}.$$

Making the change of variable $x = (\beta + n)\lambda$, the integral becomes

$$I = \int_0^\infty \left[\frac{x}{(\beta+n)}\right]^{(n\bar{y}+\alpha-1)} \exp(-x)(\beta+n)dx$$

$$= \frac{(\beta+n)}{(\beta+n)^{(n\bar{y}+\alpha-1)}} \int_0^\infty x^{(n\bar{y}+\alpha-1)} \exp(-x)dx$$

$$= \frac{\Gamma(n\bar{y}+\alpha)}{(\beta+n)^{(n\bar{y}+\alpha)}},$$

such that

$$c = \frac{(\beta+n)^{(n\bar{y}+\alpha)}}{\Gamma(n\bar{y}+\alpha)},$$

and

$$p(\lambda / \mathbf{y}) = \frac{(\beta+n)^{(n\bar{y}+\alpha)}}{\Gamma(n\bar{y}+\alpha)} \lambda^{(n\bar{y}+\alpha-1)} \exp\left[-(\beta+n)\lambda\right],$$

which is a Gamma $(n\bar{y} + \alpha, \beta + n)$ distribution.

Derivation of uninformative prior distribution

$L(\lambda/y) = \ln l(\lambda/y) = \text{constant} + ny\ln\lambda - n\lambda$ such that

$$\frac{\partial^2 L}{\partial \lambda^2} = -\frac{n\bar{y}}{\lambda^2}.$$

Since $E(Y) = \lambda$ it follows that

$$J(\lambda) = E_{y/\lambda}\left(-\frac{\partial^2 L}{\partial \lambda^2}\right) = \frac{n\lambda}{\lambda^2}$$

$$\propto \lambda^{-1},$$

where $J(\lambda)$ is Fisher's information measure. Using Jeffrey's rule

$$p_0(\lambda) \propto J^{\frac{1}{2}}(\lambda) \propto \lambda^{-\frac{1}{2}}.$$

Posterior distribution
with non-informative
prior $p(\theta/y)$

Gamma $(n\bar{y} + \frac{1}{2}, n)$ distribution

$$p(\lambda/\mathbf{y}) = \frac{(n)^{(n\bar{y}+\frac{1}{2})}}{\Gamma(n\bar{y}+\frac{1}{2})} \lambda^{(n\bar{y}-\frac{1}{2})} \exp[-(n\lambda)]$$

APPENDIX E3   EXPONENTIAL MODEL

Variate	**Y is a exponential random variable (the waiting time between consecutive events occurring in a Poisson process) μ is the mean waiting time between events.**

Likelihood function $l(\theta/y)$

Exponential model:

$$\frac{1}{\mu}\exp(-\frac{y}{\mu})$$

Conjugate prior probability $p_0(\theta)$

Inverted gamma distribution:

$$\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\mu}\right)^{\alpha+1}\exp\left(-\frac{\beta}{\mu}\right)$$

Non-informative prior probability $p_0(\theta)$

$$\propto \frac{1}{\mu}$$

Posterior probability $p(\theta/y)$

Inverted gamma $(n+\alpha, n\bar{y}+\beta)$ distribution :

$$\frac{(n\bar{y}+\beta)^{n+\alpha}}{\Gamma(n+\alpha)}\left(\frac{1}{\mu}\right)^{n+\alpha+1}\exp\left[-\frac{1}{\mu}(n\bar{y}+\beta)\right]$$

Derivation of posterior distribution

Suppose $\mathbf{y} = (y_1....y_n)$ is a set of independent and identically distributed observations on the waiting time T between consecutive events in a Poisson process, where $E(T) = \mu$ = the mean waiting time, then

$$l(\mu/\mathbf{y}) = \prod_{i=1}^{n}\frac{1}{\mu}\exp\left(-\frac{y_i}{\mu}\right)$$

$$= \left(\frac{1}{\mu}\right)^n\exp\left(-\frac{\sum_{i=1}^{n}y_i}{\mu}\right)$$

$$= \left(\frac{1}{\mu}\right)^n\exp\left(\frac{n\bar{y}}{\mu}\right) \qquad .$$

Using Bayes theorem

$$p(\mu/\mathbf{y}) \propto p_0(\mu)l(\mu/\mathbf{y}) \quad,$$

it follows that

$$p(\mu/\mathbf{y}) \propto \left(\frac{1}{\mu}\right)^{\alpha+1} \exp\left(-\frac{\beta}{\mu}\right)\left(\frac{1}{\mu}\right)^{n} \exp\left(\frac{n\bar{y}}{\mu}\right)$$

$$\propto \left(\frac{1}{\mu}\right)^{n+\alpha+1} \exp\left[-\frac{1}{\mu}(n\bar{y}+\beta)\right] \quad.$$

The normalising constant is given by

$$c = \left\{\int_0^{\infty}\left(\frac{1}{\mu}\right)^{n+\alpha+1} \exp\left[-\frac{1}{\mu}(n\bar{y}+\beta)\right]\right\}^{-1} \quad.$$

Making the change of variable $z = (n\bar{y}+\beta)/\mu$ the integral becomes

$$\mathbf{I} = \int_{\infty}^{0} -\left(\frac{z}{n\bar{y}+\beta}\right)^{n+\alpha+1} \exp(-z)\frac{(n\bar{y}+\beta)}{z^2}dz$$

$$= \frac{1}{(n\bar{y}+\beta)^{n+\alpha}} \int_0^{\infty} z^{n+\alpha-1} \exp(-z)dz$$

$$= \frac{\Gamma(n+\alpha)}{(n\bar{y}+\beta)^{n+\alpha}} \quad.$$

The normalising constant is therefore given by

$$c = \frac{(n\bar{y}+\beta)}{\Gamma(n+\alpha)} \quad,$$

such that

$$p(\mu/y) = \frac{(n\bar{y}+\beta)^{n+\alpha}}{\Gamma(n+\alpha)}\left(\frac{1}{\mu}\right)^{n+\alpha+1} \exp\left[-\frac{1}{\mu}(n\bar{y}+\beta)\right] \quad,$$

which is an Inverted gamma $(n+\alpha, n\bar{y}+\beta)$ distribution.

| Derivation of uninformative prior distribution | $L(\mu/\mathbf{y}) = \ln l(\mu/\mathbf{y}) = n\ln\left(\frac{1}{\mu}\right) - \frac{n\bar{y}}{\mu}$ such that |

$$\frac{\partial^2 L}{\partial\mu^2} = \frac{n}{\mu^2} - \frac{2n\bar{y}}{\mu^3} \quad.$$

Since $E(Y) = \mu$ it follows that

$$J(\mu) = \underset{\mu/y}{E}\left(-\frac{\partial^2 L}{\partial \mu^2}\right) = \frac{n}{\mu^2}$$

$$\propto \mu^{-2} \quad ,$$

where $J(\mu)$ is Fisher's information measure. Using Jeffrey's rule

$$p_0(\mu) \propto J^{\frac{1}{2}}(\mu) \propto \mu^{-1} \quad .$$

Posterior distribution   Inverted gamma $(n, n\bar{y})$ distribution
with non-informative
prior p(θ/y)
$$p(\mu / \mathbf{y}) = \frac{(n\bar{y})^n}{\Gamma(n)}\left(\frac{1}{\mu}\right)^{n+1} \exp\left(-\frac{n\bar{y}}{\mu}\right)$$

APPENDIX E4   NORMAL MODEL (σ^2 KNOWN, μ UNKNOWN)

Variate	**Y is a normally distributed random variable whose mean μ is unknown but whose variance σ^2 is known**

Likelihood function $l(\theta/y)$ — Normal model:

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]$$

Conjugate prior probability $p_0(\theta)$ — Normal model

$$\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left[-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\right]$$

Non-informative prior probability $p_0(\theta)$ — Constant

$$p_0(\mu/\sigma)=c$$

Posterior probability $p(\theta/y)$ — Normal distribution

$$\frac{\sigma_r}{\sqrt{2\pi}}\exp\left[-\frac{\sigma_r^2}{2}(\mu-\overline{\mu})^2\right]$$

where $\quad \overline{\mu}=\left(\dfrac{\overline{y}}{\sigma_1^2}+\dfrac{\mu_0}{\sigma_0^2}\right)\sigma_r^2$, $\ \sigma_1=\dfrac{\sigma}{\sqrt{n}}\ $ and $\ \sigma_r^2=\left(\dfrac{1}{\sigma_1^2}+\dfrac{1}{\sigma_0^2}\right)$.

Derivation of posterior distribution

Suppose $\mathbf{y}=(y_1...y_n)$ is a set of n independent observations of a normal random variate Y with unknown mean μ, but known variance σ^2. Then given $\mathbf{y}$, the likelihood is

$$l(\mu/y)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]$$

$$=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right].$$

But $\displaystyle\sum(y_i-\mu)^2=\sum(y_i-\overline{y})^2+n(\mu-\overline{y})^2$ such that

$$l(\mu / \mathbf{y}) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\mu - \bar{y})^2 \right] \right\}.$$

Given the data $\mathbf{y}$ and the variance σ then

$$\left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n,$$

and

$$\exp\left[-\frac{1}{2\sigma^2} \sum (y_i - \bar{y})^2 \right] \quad,$$

are fixed constants independent of μ such that

$$l(\mu / \mathbf{y}) \propto \exp\left[-\frac{n}{2\sigma^2} (\mu - \bar{y})^2 \right]$$

$$\propto \exp\left[-\frac{1}{2\sigma_1^2} (\mu - \bar{y})^2 \right],$$

where $\quad \sigma_1 = \dfrac{\sigma}{\sqrt{n}}$.

The conjugate prior probability is $N(\mu_0, \sigma_0^2)$ such that

$$p_0(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left[-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2 \right]$$

$$\propto \exp\left[-\frac{1}{2\sigma_0^2} (\mu - \mu_0)^2 \right].$$

Using Bayes theorem

$$p(\mu / \mathbf{y}) \propto p_0(\mu) l(\mu / \mathbf{y}),$$

it follows that

$$p(\mu / \mathbf{y}) \propto \exp\left\{ -\frac{1}{2} \left[\left(\frac{\mu - \bar{y}}{\sigma_1} \right)^2 + \left(\frac{\mu - \mu_0}{\sigma_0} \right)^2 \right] \right\}.$$

Using the following identity

$$A(z - a)^2 + B(z - b)^2 = (A + B)(z - c)^2 + \frac{AB}{(A + B)} (a - b)^2,$$

where $c = \dfrac{(Aa + Bb)}{(A + B)}$, it can be shown that

$$\sigma_1^{-2}(\mu - \bar{y})^2 + \sigma_0^{-2}(\mu - \mu_0)^2 = (\sigma_1^{-2} + \sigma_0^{-2})(\mu - \bar{\mu})^2 + d,$$

where $\bar{\mu} = \dfrac{\bar{y}\sigma_1^{-2} + \mu_0\sigma_0^{-2}}{(\sigma_1^{-2} + \sigma_0^{-2})}$

$$d = \frac{(\sigma_1^{-2}\sigma_0^{-2})(\bar{y} - \mu_0)^2}{(\sigma_1^{-2} + \sigma_0^{-2})}.$$

Given the data **y**, the variance σ and the parameters of the prior distribution, μ_0 σ_0, d and $\bar{\mu}$ are constants independent of μ. It follows that

$$p(\mu / \mathbf{y}) \propto \exp\left[-\frac{1}{2}(\sigma_1^{-2} + \sigma_0^{-2})(\mu - \bar{\mu})^2\right].$$

The normalising constant c is given by

$$c = \left\{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\sigma_1^{-2} + \sigma_0^{-2})(\mu - \bar{\mu})^2\right] d\mu\right\}^{-1}.$$

The integral is evaluated by making the change of variable

$$z = (\sigma_1^{-2} + \sigma_0^{-2})^{\frac{1}{2}}(\mu - \bar{\mu}),$$

such that

$$\frac{dz}{d\mu} = (\sigma_1^{-2} + \sigma_0^{-2})^{\frac{1}{2}}.$$

Thus

$$I = \int_{-\infty}^{\infty} \exp(-\frac{1}{2}z^2)(\sigma_1^{-2} + \sigma_0^{-2})^{-\frac{1}{2}} dz$$

$$= 2(\sigma_1^{-2} + \sigma_0^{-2})^{-\frac{1}{2}}\int_{0}^{\infty} \exp(-\frac{1}{2}z^2) dz.$$

Using the Gamma function identity

$$\Gamma(\frac{1}{2}) = 2^{\frac{1}{2}}\int_{0}^{\infty} \exp(-\frac{1}{2}r^2) dr,$$

and the fact that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ the integral becomes

$$I = \frac{\sqrt{2\pi}}{(\sigma_1^{-2} + \sigma^{-2})^{\frac{1}{2}}} .$$

Thus

$$c = \frac{(\sigma_1^{-2} + \sigma_0^{-2})}{\sqrt{2\pi}} ,$$

such that

$$p(\mu / \mathbf{y}) = \frac{(\sigma_1^{-2} + \sigma_0^{-2})^{\frac{1}{2}}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\sigma_1^{-2} + \sigma_0^{-2})(\mu - \overline{\mu})^2 \right] ,$$

which is a Normal distribution with mean $\overline{\mu}$ and variance $(\sigma_1^{-2} + \sigma_0^{-2})$.

Derivation of uninformative prior distribution

Given a set of n independent observations of a normal random variate Y with unknown mean μ, then

$$l(\mu / \mathbf{y}) \propto \exp\left[-\frac{n}{2\sigma^2}(\mu - \overline{y})^2 \right] .$$

Therefore

$$L = \ln l(\mu / \mathbf{y}) \propto \frac{-n}{2\sigma^2}(\mu - \overline{y})^2 ,$$

and

$$\frac{\partial^2 L}{\partial \mu^2} = -\frac{n}{\sigma^2} = \text{constant} .$$

Since E[constant] = constant it follows that the non-informative prior for the mean of a normal distribution (with known variance) is simply a constant.

Posterior distribution with non-informative prior p(θ/y)

Normal distribution $(\mu = \overline{y}, \sigma_1 = \frac{\sigma}{\sqrt{n}})$

$$p(\mu / \mathbf{y}) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[\frac{1}{2\sigma_1^2}(\mu - \overline{y})^2 \right]$$

APPENDIX E5   NORMAL MODEL (μ KNOWN, σ^2 UNKNOWN)

Variate	Y is a normally distributed random variable whose mean μ is known but whose variance σ^2 is unknown
Likelihood function $l(\theta/y)$	Normal model $$\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]$$
Conjugate prior probability $p_0(\theta)$	Inverted gamma distribution $$\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp\left(-\frac{\beta}{\sigma^2}\right)$$
Non-informative prior probability $p_0(\theta)$	$$\propto \frac{1}{\sigma}$$
Posterior probability $p(\theta/y)$	Inverted gamma $(\frac{n}{2}+\alpha,\beta+\frac{ns^2}{2})$ distribution $$\frac{\left(\beta+\frac{ns^2}{2}\right)^{\frac{n}{2}+\alpha}}{\Gamma\left(\frac{n}{2}+\alpha\right)}\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\alpha+1}\exp\left[-\left(\beta+\frac{ns^2}{2}\right)\frac{1}{\sigma^2}\right]$$
Derivation of posterior distribution	Suppose $\mathbf{y} = (y_1...y_n)$ is a set of n independent observations of a normal random variable Y with known mean μ, but unknown variance σ^2, then given $\mathbf{y}$ the likelihood function is $$l(\sigma^2/\mathrm{y},\mu)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(y_i-\mu)^2\right]$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{ns^2}{2\sigma^2}\right) \quad ,$$

where $s^2 = \dfrac{\sum(y_i - \mu)^2}{n}$ is the variance.

The conjugate prior probability is inverted Gamma such that

$$p_0(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp\left(-\frac{\beta}{\sigma^2}\right)$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp\left(-\frac{\beta}{\sigma^2}\right) \quad .$$

Using Bayes theorem

$$p(\sigma^2/\mathrm{y},\mu) \propto p_0(\sigma^2)l(\sigma^2/\mathrm{y},\mu) \quad ,$$

it follows that

$$p(\sigma^2/\mathrm{y},\mu) \propto \left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp\left(-\frac{\beta}{\sigma^2}\right)\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}}\exp\left(-\frac{ns^2}{2\sigma^2}\right)$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\alpha+1}\exp\left[-\left(\beta+\frac{ns^2}{2}\right)\frac{1}{\sigma^2}\right] \quad .$$

The normalising constant is given by

$$c = \left\{\int_0^\infty \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\alpha+1}\exp\left[-\left(\beta+\frac{ns^2}{2}\right)\frac{1}{\sigma^2}\right]d\sigma^2\right\}^{-1} \quad .$$

The integral is evaluated by making the change of variable

$$z = \frac{1}{\sigma^2} \quad ,$$

such that

$$I = -\int_\infty^0 z^{\frac{n}{2}+\alpha+1}\exp\left[-\left(\beta+\frac{ns^2}{2}\right)z\right]\frac{dz}{z^2}$$

$$= \int_0^\infty z^{\frac{n}{2}+\alpha-1} \exp\left[-\left(\beta + \frac{ns^2}{2}\right)z\right]dz \quad .$$

Using the Gamma function identity

$$b^{-a}\Gamma(a) = \int_0^\infty x^{a-1}\exp(-bx)dx \quad ,$$

it follows that

$$\int_0^\infty z^{\frac{n}{2}+\alpha-1}\exp\left[-\left(\beta + \frac{ns^2}{2}\right)z\right]dz = \left(\beta + \frac{ns^2}{2}\right)^{-\frac{n}{2}+\alpha}\Gamma\left(\frac{n}{2}+\alpha\right) \quad ,$$

such that

$$c = \frac{\left(\beta + \dfrac{ns^2}{2}\right)^{\frac{n}{2}+\alpha}}{\Gamma\left(\dfrac{n}{2}+\alpha\right)} \quad ,$$

and

$$p(\sigma^2/\mathrm{y},\mu) = \frac{\left(\beta + \dfrac{ns^2}{2}\right)^{\frac{n}{2}+\alpha}}{\Gamma\left(\dfrac{n}{2}+\alpha\right)}\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\alpha+1}\exp\left[-\left(\beta + \frac{ns^2}{2}\right)\frac{1}{\sigma^2}\right] \quad .$$

Derivation of uninformative prior distribution

Given a set of n independent observations of a normal random variate Y with unknown variance σ^2 then

$$l(\sigma/\mathrm{y},\mu) \propto \left(\frac{1}{\sigma}\right)^n \exp\left(\frac{-ns^2}{2\sigma^2}\right) \quad ,$$

where $s^2 = \sum_{i=1}^n (y_i - \mu)^2 / n$ is the variance. It follows that

$$L = \ln l(\sigma/\mathrm{y},\mu) \propto n\ln\left(\frac{1}{\sigma}\right) - \frac{ns^2}{2\sigma^2} \quad ,$$

such that

$$\frac{\partial^2 L}{\partial\sigma^2} \propto \frac{n}{\sigma^2}\left[1 - \frac{3s^2}{\sigma^2}\right] \quad .$$

Since $E(s^2) = \sigma^2$ for large n, it follows (approximately) that

$$J(\sigma) = E_{\sigma}\left(-\frac{\partial^2 L}{\partial \sigma^2}\right) \propto \frac{1}{\sigma^2} \quad,$$

where $J(\sigma)$ is Fisher's information measure. Using Jeffrey's rule

$$p_0(\sigma) \propto J^{\frac{1}{2}}(\sigma) \propto \frac{1}{\sigma} \quad.$$

Posterior distribution with non-informative prior $p(\theta/y)$

Inverted gamma $(\dfrac{n}{2}, \dfrac{ns^2}{2})$ distribution

$$p(\sigma^2/\mathrm{y}, \mu) = \frac{\left(\dfrac{ns^2}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\dfrac{n}{2}\right)} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left[-\left(\frac{ns^2}{2}\right)\frac{1}{\sigma^2}\right]$$

APPENDIX E6   NORMAL MODEL (BOTH μ AND σ^2 UNKNOWN)

Variate	**Y is a normally distributed random variable whose mean μ is unknown and whose variance σ^2 is unknown**

Likelihood function $l(\theta, \sigma/y)$

Normal model

$$l(\mu, \sigma / \mathrm{y}) \propto \left(\frac{1}{\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

Non-informative prior probabilities $p_0(\theta)$

Assuming non-informative prior probabilities are independent

$$p_0(\mu) \propto \text{constant}$$

$$p_0(\sigma) \propto \frac{1}{\sigma}$$

Posterior probability (with non-informative prior) $p(\theta, \sigma/y)$

$$p(\mu, \sigma / \mathbf{y}) = \frac{\sqrt{\dfrac{n}{2\pi}}\left(\dfrac{\upsilon s^2}{2}\right)^{\frac{\upsilon}{2}}}{\left[\dfrac{1}{2}\Gamma\left(\dfrac{\upsilon}{2}\right)\right]}\sigma^{-(N+1)}\exp\left\{-\frac{1}{2\sigma^2}\left[\upsilon s^2 + n(\mu - \bar{y})^2\right]\right\}$$

Derivation of posterior distribution (with non-informative prior)

Suppose $\mathbf{y} = (y_1 ... y_n)$ is a set of n independent observations of a normal random variate Y with unknown mean μ and unknown variance σ^2. Then given $\mathbf{y}$ the likelihood is

$$l(\mu, \sigma / \mathbf{y}) \propto \left(\frac{1}{\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right]$$

$$\propto \left(\frac{1}{\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\mu - \bar{y})^2\right]\right\}$$

$$\propto \left(\frac{1}{\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu - \bar{y})^2\right]\right\} \qquad ,$$

where $s^2 = \displaystyle\sum_{i=1}^{n}(y_i - \bar{y})/(n-1)$ is the sample variance.

Assuming non-informative prior for μ and σ (and that their prior distributions are independent) the posterior distribution becomes

$$p(\mu, \sigma / \mathbf{y}) \propto \left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\mu - \bar{y})^2 + (n-1)s^2\right]\right\}$$

$$\propto \left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\mu - \bar{y})^2 + \upsilon s^2\right]\right\} \quad,$$

where $\upsilon = (n-1)$.

The normalising constant is therefore given by

$$c = \left\{\int\limits_{-\infty}^{\infty}\int\limits_{0}^{\infty}\left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\mu - \bar{y})^2 + \upsilon s^2\right]\right\}d\sigma d\mu\right\}^{-1} \quad.$$

The integral is evaluated in the following manner

$$\int\limits_{-\infty}^{\infty}\int\limits_{0}^{\infty} f(\sigma)g(\mu)d\sigma d\mu = \int\limits_{0}^{\infty} f(\sigma)d\sigma \int\limits_{-\infty}^{\infty} g(\mu)d\mu \quad,$$

such that

$$I = \int\limits_{0}^{\infty}\left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left(-\frac{\upsilon s^2}{2\sigma^2}\right)d\sigma \int\limits_{-\infty}^{\infty} \exp\left[-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right]d\mu$$

$$= \int\limits_{0}^{\infty}\left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left(-\frac{\upsilon s^2}{2\sigma^2}\right)d\sigma \int\limits_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{\mu - \bar{y}}{\sigma/\sqrt{n}}\right)^2\right]d\mu \quad.$$

Using the normal integral identity

$$\int\limits_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x - \eta}{c}\right)^2\right]dx = \sqrt{2\pi}c \quad,$$

this becomes

$$I = \int\limits_{0}^{\infty}\left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left(-\frac{\upsilon s^2}{2\sigma^2}\right)d\sigma.\frac{\sqrt{2\pi}\sigma}{\sqrt{n}}$$

$$= \sqrt{\frac{2\pi}{n}}\int\limits_{0}^{\infty}\left(\frac{1}{\sigma}\right)^{n} \exp\left(-\frac{\upsilon s^2}{2\sigma^2}\right)d\sigma \quad.$$

To evaluate the remaining integral make the change of variable $a = \upsilon s^2 / 2$ such that

$$I = \sqrt{\frac{2\pi}{n}} \int_0^\infty \left(\frac{1}{\sigma}\right)^n \exp\left(-\frac{a}{\sigma^2}\right) d\sigma \qquad .$$

Using the following gamma integral identity

$$\int_0^\infty x^{-(p+1)} \exp\left(-ax^{-2}\right) dx = \frac{1}{2} a^{-\frac{p}{2}} \Gamma\left(\frac{p}{2}\right) \qquad ,$$

the integral becomes

$$I = \sqrt{\frac{2\pi}{n}} \frac{1}{2} (a)^{-\frac{(n-1)}{2}} \Gamma\left(\frac{(n-1)}{2}\right)$$

$$= \sqrt{\frac{2\pi}{n}} \frac{1}{2} \left(\frac{\upsilon s^2}{2}\right)^{-\frac{\upsilon}{2}} \Gamma\left(\frac{\upsilon}{2}\right) \qquad ,$$

such that

$$c = \frac{\sqrt{\frac{n}{2\pi}} \left(\frac{\upsilon s^2}{2}\right)^{\frac{\upsilon}{2}}}{\frac{1}{2} \Gamma\left(\frac{\upsilon}{2}\right)} \qquad .$$

The posterior distribution function is therefore given by

$$p(\mu, \sigma / \mathbf{y}) = \frac{\sqrt{\frac{n}{2\pi}} \left(\frac{\upsilon s^2}{2}\right)^{\frac{\upsilon}{2}}}{\frac{1}{2} \Gamma\left(\frac{\upsilon}{2}\right)} \left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\mu - \bar{y})^2 + \upsilon s^2\right]\right\} .$$

APPENDIX F   JOURNEY SURVIVAL SIMULATION

The journey survival model introduced in section 3.1 is

$$P_t = P_0 \exp\left(-\frac{t}{\mu}\right) \quad,$$

where P_t is the ballast tank population after t days, P_0 is the initial inoculum and μ is the species life expectancy. The objective is to introduce the uncertainty regarding the parameter μ into the model output. The uncertainty regarding this parameter is described by the posterior distribution function

$$p(\mu / \mathbf{t}, \mathbf{x}, \mathbf{y}) \propto \left(\frac{1}{\mu}\right)^{n+1} \exp\left(-\frac{F}{\mu}\right) \quad t > 0, 0 \leq \mu \leq M \quad.$$

which has the kernal of an inverted gamma distribution.

The simulation proceeds in the following manner:

1. generate a random variate ϕ from a Gamma distribution with shape parameter b and scale parameter c;

2. the variate $\mu = 1/\phi$ is an inverted gamma variate;

3. if μ is greater than M (eg 100) discard it and return to step 1, otherwise substitute μ into the journey survival model;

4. calculate P_t given the inoculum P_0 and the journey duration t;

5. repeats steps 1 to 4 one thousand times and collate the resultant vector P_{t1}....P_{t1000} in a size frequency distribution.

Gamma variates (γ: b, c) for the case where c is an integer can be computed from

$$\gamma : b, c \approx -b \ln\left(\prod_{i=1}^{c} R_i\right) = \sum_{i=1}^{c} - b \ln R_i \quad,$$

where R_i are independent unit rectangular variates (Evans *et al*, 1993). Independent unit rectangular variates can be easily generated in MicroSoft Excel using the RND( ) function.

Random inverted gamma variates can be generated by using the change of variable $\mu = 1/\phi$. The proof is as follows: Suppose μ has an inverted gamma distribution with shape parameter b and scale parameter c such that

$$\Pr(\mu < a) = k \int_0^a \left(\frac{1}{\mu}\right)^{c+1} \exp\left(-\frac{b}{\mu}\right) d\mu \quad,$$

where k is a normalising constant.

Make the change of variable $\phi = 1/\mu$ such that

$$d\phi = \frac{-1}{\mu^2}d\mu \quad .$$

Now

$$\Pr(\mu < a) = \Pr\left(\phi > \frac{1}{a}\right)$$

$$= k\int_{\frac{1}{a}}^{\infty} -\phi^{c+1}\exp(-b\phi)\frac{1}{\phi^2}d\phi$$

$$= k\int_{-\infty}^{1/a}\phi^{c-1}\exp(-b\phi)d\phi$$

$$= k\left[\int_{-\infty}^{0}\phi^{c-1}\exp(-b\phi)d\phi + \int_{0}^{\frac{1}{a}}\phi^{c-1}\exp(-b\phi)d\phi\right]$$

$$= 0 + k\int_{0}^{\frac{1}{a}}\phi^{c-1}\exp(-b\phi)d\phi \qquad ,$$

thereby establishing that ϕ has a gamma distribution with shape parameter b and scale parameter c, as required.

APPENDIX G   INTEGRATING OUT NUISANCE PARAMETERS

Consider inferences regarding a normal random variate Y in which both μ and σ are unknown. Interest commonly centres on μ ($=\theta_1$) whilst σ ($=\theta_2$) is simply a nuisance parameter. The joint posterior distribution function $p(\theta, \sigma / \mathbf{y})$ assuming non-informative priors for θ and σ is given by (Appendix E6)

$$p(\mu, \sigma / \mathbf{y}) \propto \left(\frac{1}{\sigma}\right)^{(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\mu - \bar{y})^2 + (n-1)s^2\right]\right\} \quad .$$

Integrating over σ yields the marginal posterior distribution function for μ

$$p(\mu / \mathbf{y}) = \int_0^\infty \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[n(\mu - \bar{y})^2 + (n-1)s^2\right]\right\} d\sigma \quad .$$

The integral is evaluated by making the change of variable $z = A/2\sigma^2$ where $A = n(\mu - \bar{y})^2 + (n-1)s^2$ such that

$$I = \int_\infty^0 \left(\frac{2z}{A}\right)^{-\frac{1}{2}(n-1)} \exp(-z) - \frac{A}{2z^2} dz$$

$$= \left(\frac{2}{A}\right)^{\frac{n}{2}} \int_0^\infty z^{-\frac{1}{2}(n-1)} \exp(-z) dz$$

$$= \left(\frac{2}{A}\right)^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)$$

$$\propto A^{-\frac{n}{2}}$$

$$\propto \left[n(\mu - \bar{y})^2 + (n-1)s^2\right]^{-\frac{n}{2}}$$

$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{\upsilon s^2}\right]^{-\frac{1}{2}(\upsilon+1)} \quad ,$$

where $\upsilon = (n-1)$.

By making the change of variable

$$t = \frac{(\mu - \bar{y})}{s/\sqrt{n}} \quad ,$$

this is recognisable as the kernel of students t distribution with υ degrees of freedom

$$p(\mu/\mathbf{y}) \propto \left[1 + \frac{t^2}{\upsilon}\right]^{-\frac{\upsilon+1}{2}} \quad .$$

It is interesting to note that the distribution of this quantity t is identical to that derived by classical sampling theory. Both the Bayesian posterior distribution for t and its classical counterpart are independent of the nuisance parameter σ. It is important to note, however, that in the Bayesian approach σ is eliminated through integration, which can be used for any set of nuisance parameters[20]. In classical sampling theory independence occurs purely through good fortune; in general terms quantities which are functions of the data and the parameters of interest, but whose sampling distributions do not involve nuisance parameters, do not usually exist ((Box and Tiao, 1973).

---

[20] Box and Tiao (1973) provide a word of caution in this respects: if $p(\theta_1/\theta_2, \mathbf{y})$ is very sensitive to changes in θ_2 it is important to carefully examine the marginal posterior distribution $p(\theta_2/\mathbf{y})$ prior to eliminating θ_2 to make inferences about θ_1.