Studies in Avian Biology No. 34:136–144

BAYESIAN STATISTICS AND THE ESTIMATION OF NEST-SURVIVAL RATES

Andrew B. Cooper and Timothy J. Miller

Abstract. Bayesian statistical approaches have received little attention in the nest-survival literature despite the growing usage in other fields of ecology. Appealing aspects of Bayesian statistics are that they allow the researcher to quantitatively account for prior knowledge when analyzing data and they calculate the probability of a hypothesis being true or of a parameter taking on a certain range of values given the collected data. While attempting to keep the discussion accessible to non-statisticians, we give an overview of the theory of Bayesian statistics, including discussions of prior distributions, likelihoods, and posterior distributions. We briefly discuss some of the advantages and disadvantages of Bayesian methods relative to alternative approaches. Finally, we describe how Bayesian methods have been applied to estimating age-specific nest survival rates.

Key Words: age-specific, Bayes, hypothesis testing, likelihood, parameter estimation, posterior, prior, survival.

ESTADÍSITICAS BAYESIANAS Y LA ESTIMACIÓN DE TASAS DE SOBREVIVENCIA DE NIDO

Resumen. Enfoques de estadísticas Bayesianas han recibido poca atención en la literatura sobre sobrevivencia de nido, a pesar de su creciente utilización en otros campos de la ecología. Algunos de los motivos por los cuales son atractivas las estadísticas Bayesianas es porque al analizar los datos permiten al investigador a contar cuantitativamente para el conocimiento previo, y también calculan si la probabilidad de que una hipótesis sea verdad o un parámetro, tomando un cierto rango de valores segun los datos colectados. Mientras tratamos de mantener la discusión accesible a no estadistas, proporcionamos un panorama de la teoría de las estadísticas Bayesianas incluyendo discusiones de distribuciones previas, probabilidades, y distribuciones posteriores. Discutimos brevemente algunas de las ventajas y desventajas de los métodos Bayesianos en relación a métodos alternativos. Finalmente, describimos cómo los métodos Bayesianos han sido aplicados en la estimación de tasas de sobrevivencia de nido específicas de edad.

While growing in popularity in many fields of ecology, Bayesian statistics have received only scant attention in the nest-survival literature. Bayesian statistics allow the researcher to formally incorporate prior knowledge into the analysis and then provide results that give the probability of a hypothesis being true or of a parameter taking on a certain range of values. No other statistical approach permits such statements, despite the fact that they are crucial for decision making. This paper gives an overview of the theory and application of Bayesian statistics and then describes one way in which they have been applied to estimating age-specific nest survival rates when the age of the nest is not known.

BAYESIAN STATISTICAL INFERENCE

Based purely on the calculus of probabilities (Casella and Berger 1990), Bayes rule (also known as Bayes theorem) describes the relationship between two conditional probabilities and can be used to calculate the probability of one event occurring given (or conditional on) another event having already occurred. In equation form, we use a vertical line, |, to represent this conditioning. As such, the probability of event A occurring given event B has already occurred would be written as, $P(A|B)$. For two events, A and B, Bayes rule is written as:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Where $P(A)$ and $P(B)$ are the probabilities of event A and event B occurring under all possible conditions, respectively; and $P(A|B)$ and $P(B|A)$ are the conditional probabilities of event A occurring given event B has already occurred and event B occurring given event A has already occurred, respectively. Bayes rule is considered to be a mathematical fact when it refers to generic events. Controversy, however, arises over the application of Bayes rule to statistical inference. For Bayesian statistical inference, the hypothesis in question (i.e., that a parameter Θ, equals a specific value, θ) is treated as one event (A), and observation of data (y) is treated as another event (B). In order to distinguish between a random variable and a specific realization of that random variable, we use Θ to represent the parameter (which is treated as a random variable) but θ to represent the specific value of that parameter, and we use Y to represent data that have not yet been observed (i.e., before the study takes place), and y to represent the actual data in hand. When applying Bayes rule to statistical inference, and using these new symbols to describe events A and B, the equation for Bayes rule appears as:

$$P(\Theta \mid y) = \frac{P(Y \mid \theta) * P(\Theta)}{P(Y)}$$

Bayes rule has four main components when applied to statistical inference, each of which will be explained in greater detail: (1) the prior probability distribution of the parameter values, P(Θ); (2) the probability distribution of the data before it is actually observed given a hypothesized value for the parameter, P(Y | θ); (3) the marginal probability distribution of the data, P(Y); and (4) the posterior probability distribution of the parameter values given the observed data, P(Θ | y).

When the prior probability distribution for the parameter values is assumed to take a parametric form (i.e., normal, lognormal, beta), Bayes rule becomes more complex because we now must condition on the parameter values, known as hyperparameters, which govern the shape of the prior distribution and are symbolized by ω. For example, if one used a normal distribution as a prior, then the mean and variance would be the hyperparameters. If one used a uniform distribution, then the maximum and minimum values would be the hyperparameters. To emphasize the dependency on these hyperparameters, we write Bayes rule as:

$$P(\Theta \mid y,\omega) = \frac{P(Y \mid \theta) * P(\Theta \mid \omega)}{P(Y \mid \omega)}$$

Where P(Θ | ω) represents the prior probability distribution of the parameter values given the values for the hyperparameters, P(Y | θ) represents the probability distribution of the data before they are actually observed given a hypothesized value for the parameter, P(Y | ω) represents the marginal probability distribution of the data given the values for the hyperparameters, and P(Θ | y,ω) represents the posterior probability distribution of the parameter values given the observed data and hyper parameters. Using a parametric prior distribution often decreases the difficulty in computing the posterior distribution and is commonly used for Bayesian mark-recapture and nest-survival studies (Dupis 1995, 2002; He et al. 2001, He 2003). We use this version of Bayes rule throughout the rest of the manuscript.

When the data have not yet been observed (Y) but we have a hypothesis about a specific parameter value, θ, we can describe the probability of the not-yet-observed data occurring given the hypothesized parameter value with the probability distribution, P(Y | θ). However, once we have data in hand (y) and are interested in the potential set of values of the parameter, Θ, we use a different nomenclature and refer to the likelihood, L(Θ | y). The likelihood is commonly used instead of P(Y | θ) in Bayes rule. The likelihood measures how likely different parameter values are given the observed data, and the maximum likelihood estimate (MLE) for a parameter is the parameter value that yields the highest likelihood value. However, it is important to note that the likelihood as a function of the parameter given observed, fixed data is not a probability distribution for the parameter values. In other words, it does not tell us the probability of the parameter taking on specific values. Indeed, this is the reason for using Bayes rule.

The likelihood serves as the basis for many statistical methods used in ecological research today, including testing a null hypothesis of some parameter equaling zero or comparing models that make different assumptions about the parameters (i.e., constant survival rates versus time-varying survival rates). Generalized linear regression models (of which normal and logistic regression are subsets) rely on likelihoods (McCullagh and Nelder 1989). Mark-recapture models are often estimated using likelihoods (Lebreton et al. 1992) as are the nest survival models of Heisey and Nordheim (1995), Dinsmore et al. (2002), and Shaffer (2004). The information-theoretic approaches to model selection (AIC, AIC_c, $QAIC_c$) described in Burnham and Anderson (2002) are based on likelihoods. It is the combining of likelihoods and prior probability distributions which causes much of the controversy between Bayesian and frequentist statisticians.

The prior probability distribution of the parameter values (also less formally called the prior): P(Θ | ω), describes any knowledge or assumptions about the model parameters, and ideally the model structure itself, that exists before the data are observed. The functional form of a prior is usually chosen to match the range of sensible values of the parameters, and the hyperparameters (ω) specify, among other things, the shape, average, and variability of the parameter values before the data are observed.

As such, P(Θ | ω) can be read as the prior probability of the parameter taking on a range of values given the choice of hyperparameters. Bayesian analysis requires that the knowledge or assumptions about model parameters be explicitly and quantitatively stated (Gelman et al. 1995, Ellison 1996) and the hyperparameters are chosen to reflect this. For example, with nest-survival models, priors must be stated for both the probability that a nest is encountered and the probability that a nest survives from one observation point to the next. Models such as those described by Heisey and Nordheim (1995), He et al. (2001), He (2003), and Cao et al. (in press) explicitly model the encounter probabilities in order to account for the fact that some nests do not enter the study because they did not survive (a form of truncation bias) and that the encounter probability may change as a function of time or nest age. A prior on a survival or encounter probability parameter may be that the probability is bounded between zero and one, inclusive, and no value is more probable than another. This could be modeled using the uniform distribution, though beta or Dirichlet distributions, which include the uniform distribution as special cases, are more common for survival and encounter probabilities (Dupis 1995, 2002; He et al. 2001, He 2003). When a prior states only very limited or imprecise knowledge of the potential values of the parameter, they are often described as being diffuse, vague, or flat; when they represent no knowledge, they are called non-informative. A Jeffreys' prior (Jeffreys 1961) is a specific type of non-informative prior and is mentioned here only so that readers may recognize the term if it is encountered in other readings. A subset of Bayesian methods called objective Bayesian methods use only such priors (Link et al. 2002). However, researchers must be careful in that what may at first appear to be a non-informative prior on one parameter may convey a great deal of information about other parameters (Walters and Ludwig 1994). Berger et al. (2001) and Hobert and Casella (1996) discuss some of the difficulties in using diffuse priors.

When warranted, priors may contain more detailed information. For example, one could base the priors on a formal synthesis of previous studies focused on the same or similar species in the same or similar habitats. The priors could be based on a survey of the opinions from a range of experts (Wolfson et al. 1996). In some cases, the priors are based on the subjective belief of the investigator, which in turn, should be based on an understanding of the biological system in question (Cooper et al. 2003). In any case, the specific form of the prior and the justification for this form should be stated in any presentation of the research to ensure the underlying assumptions are transparent to the reader (Link et al. 2002). As will be discussed later, researchers should also assess the sensitivity of their results to the choice of priors.

The marginal probability of the data, P(Y | ω), is obtained by integrating the joint probability distribution of the data and the hypotheses over all possible hypotheses (i.e., that Θ = θ for all possibly values of θ), where the joint probability distribution is the product of the prior and the likelihood. As such, P(Y | ω) does not depend upon the particular hypothesis in question, but is dependent upon the hyperparameters (ω) in the prior probability distribution, (i.e., P(Θ | ω)). In practice, P(Y | ω) is treated as a scaling constant because the data are already observed and the hyperparameters are chosen a priori (Ellison 1996).

The final component of Bayes rule is the posterior probability distribution (less formally called the posterior), P(Θ | y,ω). The posterior can be thought of as a compromise between, or a weighted average of, the prior distribution and the information contained in the data (Gelman et al. 1995). The posterior specifically describes the probabilities associated with possible values (for discrete distributions) or ranges of values (for continuous distributions) for the parameters in question given the data in hand and the prior knowledge of those parameters as defined by the hyperparameters (Link et al. 2002). So, unlike a confidence interval, the posterior distribution permits such concepts as a specific probability that the parameter of interest lies within a specific range, called the Bayesian credibility interval. For example, a 95% credibility interval implies a 95% chance that the true value of the parameter lies within the stated range. It is important to note that the posterior distribution describes only the uncertainty in the parameter estimate, not its variability over spatio-temporal scales (Clark 2005) unless such variability is explicitly incorporated by adding parameters to the model (i.e., time-specific survival rather than constant survival). With non-informative, vague, or flat priors, the mode of the posterior distribution will occur at the same value as the maximum likelihood estimate obtained under the purely likelihood-based methods (Link et al. 2002, Clark 2005).

If the model has more than one parameter, then the posterior distribution actually describes the joint probability of the parameters taking on sets of values, fully accounting for the correlation between the parameter estimates (e.g., the probability of survival falling within some range and encounter probability falling within another range). In many cases, however, the researcher is only interested in one or a few of the many parameters. For example, even though care is taken in modeling the encounter probabilities (e.g., constant over time or ages versus variable over time or ages), the real questions of interest typically center on the survival rate estimates. A similar situation exists with regard to the recapture probabilities in a mark-recapture model designed to estimate survival rates. In these cases, the encounter or recapture probabilities would be called nuisance parameters because they are unknown and must be estimated, but the real interest (and the hypothesis in question) lies elsewhere. The marginal posterior distribution allows one to make statements about the parameter of interest alone and is calculated by integrating over the nuisance parameters (Gelman et al. 1995, Ellison 1996, Hobbs and Hilborn 2006). To calculate the marginal probability of survival falling within a specific range, we would integrate over all possible values for the encounter probabilities, essentially incorporating the uncertainty of the encounter probabilities and their correlation with survival directly into the probability statement for survival alone.

To determine the posterior can be rather challenging. One issue is that few programs are available to perform these analyses. WinBUGS (Spiegelhalter et al. 1995) is one user-friendly program applicable to many Bayesian analyses, but one must still have a familiarity with likelihoods and Bayesian methods in order to use it. The program MARK (White and Burnham 1999) can perform Bayesian estimation of nest survival for the Dinsmore et al. (2002) model, but it only allows for normally distributed priors for the parameters in the logit model for the covariates which define the survival probabilities. Another related issue is that even with such user-friendly packages as WinBUGS, computing the posterior distributions for some models can take on the order of hours for a standard desktop computer (Hobbs and Hilborn 2006). The posteriors are often approximated using an approach called Markov chain Monte Carlo (MCMC) and, in particular, the Metropolis-Hastings algorithm, of which Gibbs sampling is a special case. Readers will come across these terms when reading about Bayesian methods, but the details of these methods are beyond the scope of this manuscript. See Casella and George (1992), Kass et al. (1998), and especially Link et al. (2002) for more complete descriptions.

The computational burden of determining the posterior distribution can be greatly decreased by using priors that are conjugate distributions (or more simply conjugates) for the likelihood. When the prior is a conjugate for the likelihood, the posterior distribution will, by definition, have the same functional form as the prior. For example, a beta-distributed prior is a conjugate for the binomial likelihood. Most mark-recapture and nest-survival models, whether Bayesian or not, use a binomial likelihood. When the prior for survival is defined using a beta distribution, the posterior distribution will always follow a beta distribution because of this conjugacy. This is one reason why the beta distribution or its multivariate relative, the Dirichlet distribution, is often used to define the priors for survival in Bayesian mark-recapture and nest-survival studies (Dupis 1995, 2002; He et al. 2001, He 2003). Another reason is that both these distributions are flexible enough to be used for both informative and non-informative priors (Fig. 1). Although using a beta (or Dirichlet) distribution for the prior with a binomial likelihood will ensure that the posterior is also beta-distributed, the specific shape of beta-distributed posterior will depend on both the prior and the data.

When applying Bayesian methods, it is important to examine the posteriors' sensitivity to the choice of priors. Specifically, it may be useful to apply a range of priors, all of which still conform to the researchers a priori knowledge of the parameters, and examine the changes these different priors cause in the posterior (Link et al. 2002). If the posterior is sensitive to changes in the prior, then the currently available data contain relatively little information about the parameter of interest, possibly due to small sample size (Ellison 1996, Ludwig 1996, Link et al. 2002). In such cases, the posterior is determined mostly by the prior information, and therefore great care must be taken in interpreting the meaningfulness of the posterior. However, it is not uncommon for the data from a well-designed study to overwhelm the information contained in the priors and produce posteriors robust to changes in these priors (Clark 2005).

WHY BOTHER WITH BAYESIAN METHODS?

A long-running discussion concerns the pros and cons of frequentist versus Bayesian methods in both the statistical and ecological journals. The discussion often focuses on the philosophical underpinnings of each (Clark 2005) such as the definition of probability (Ludwig 1996) and whether variables are fixed but unknown as opposed to random (Ellison 1996). Rather than delving into these discussions once again, we refer readers to Dixon and Ellison (1996) and

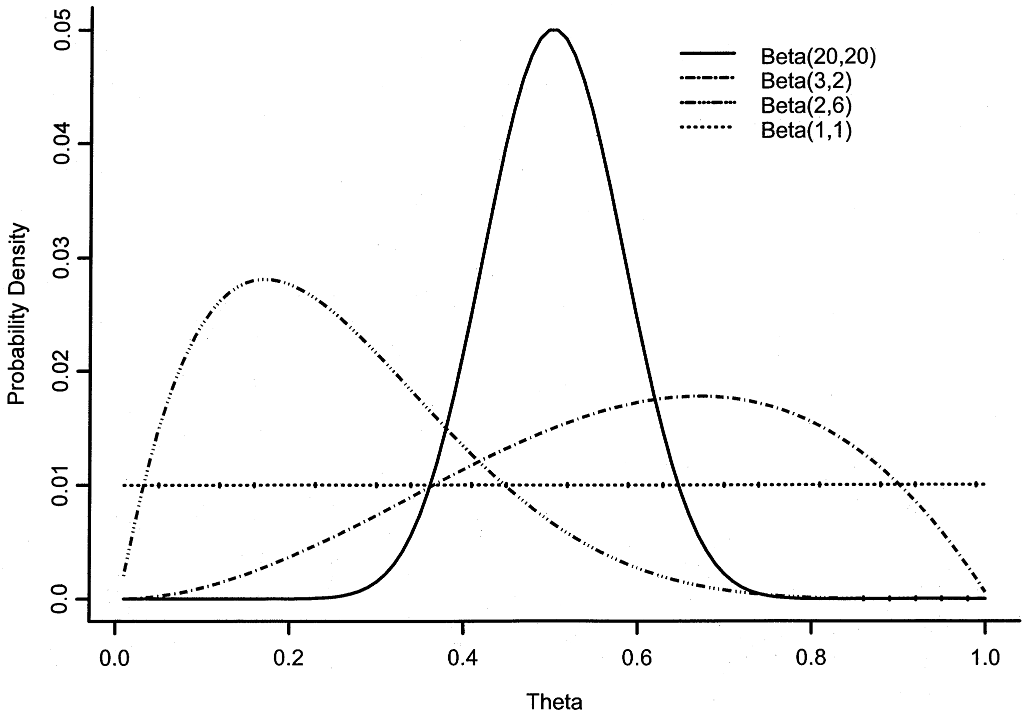

FIGURE 1. The beta-distribution can take on a range of shapes depending on the values of the hyperparameters.

the other papers immediately following it in that issue of *Ecological Applications*. Here, we will focus solely on those points we feel are of practical importance with respect to estimating nest survival rates.

The primary outputs from a frequentist analysis are typically the P-value, the parameter estimate (or effect size), and a confidence interval for the parameter value. The P-value gives the probability of obtaining a value for a test statistic (which is based on both the data and the parameter value) as or more extreme than the one observed, given the null hypothesis is true (i.e., that a given parameter equals zero). If the P-value is small, usually <0.05, then the null hypothesis is rejected because it is unlikely that the value would be observed if the null hypothesis were true. It is important to remember, and often forgotten (as argued by Johnson 2002), that failure to reject the null hypothesis does not translate into support for the null hypothesis (Kass and Raftery 1995, Ellison 1996, Johnson 2002). Similarly, a small P-value does not describe the level of support for the specific estimated parameter value, only that it is unlikely to be the value defined in the null hypothesis (Ellison 1996).

In order to move away from strict hypothesis testing, some (Robinson and Wainer 2002) are advocating more frequent use of confidence intervals for the estimated parameter values. A k% confidence interval (e.g., a 95% confidence interval) implies that if the experiment were repeated ad infinitum, and a k% confidence interval was estimated for each experiment, then k% of those intervals would contain the true value of the parameter. This also implies that 100–k% of those intervals would not contain the true value of the parameter; and for a single experiment, there is no way to determine whether the estimated confidence interval actually contains the true value. Despite continued confusion on this point, as mentioned by Hobbs and Hilborn (2006), the confidence interval does not mean that a k% probability exists that the true value for the parameter lies within the interval, nor does it describe the probability distribution for the parameter (Ellison 1996). A k% Bayesian credibility interval, however, does imply that there is a k% chance that the true value of the parameter lies within the stated range.

Another way in which frequentist methods have moved away from null hypothesis testing is through the use of information theoretic approaches such as AIC and AIC-based model averaging (Burnham and Anderson 2002). These approaches allow researchers to estimate the relative support the data provide for competing models and then make predictions that incorporate the uncertainty as to which is the best model. These approaches, as with P-values and confidence intervals, do not allow researchers to make statements about the relative probability of a parameter (e.g., survival rate) taking on one value or another value or the relative probability of a predicted outcome. Only Bayesian statistics, and the posterior distribution in particular, allow researchers to make probabilistic statements concerning the validity of the null or alternative hypothesis, about specific values of the parameters in question, or about predicted outcomes based on the fitted model (Reckhow 1990, Ellison 1996, Wade 2000, Hobbs and Hilborn 2006). It should also be noted that model averaging is possible in a Bayesian context using what are called Bayes factors (Gelman et al. 1995, Kass and Raftery 1995).

The benefit of being able to make probabilistic statements regarding hypotheses is crucial when it comes to applying the results of research to management decision making. It is not likely to be good enough to simply state that some land-use practice effects nest survival (i.e., the null hypothesis of no effect has been rejected). Managers will wish to know how much survival may be affected. What is the probability that survival will decrease by >5%, by >10%, or by >30%? Rather than knowing that one management option is better than another at increasing nest-survival rates (rejecting the null hypothesis of two management actions producing equal survival rates), managers wish to have an estimate of how much better one management option is over another, and their associated probabilities (e.g., option A has a 75% chance of increasing survival rates by more than 1% compared to option B, but only a 10% chance of increasing survival rates by more than 5% compared to option B). Answering such questions requires a Bayesian framework. If one wishes to use the output of a nest-survival study in a population dynamics model, then one needs to know the relative probability of survival taking on a range of values. Mean effect sizes, standard errors, and the associated confidence intervals from frequentist analyses do not give you this, even when based on AIC model averages. Only the posterior distribution gives you this information.

One of the more subtle differences between the output from a frequentist analysis and a Bayesian analysis centers on the treatment of nuisance parameters, such as the encounter probabilities in the models of Heisey and Nordheim (1995), He et al. (2001), He (2003), and Cao et al. (in press) or the recapture probabilities in mark-recapture models. In a frequentist framework, the maximum likelihood estimate and the standard error of that estimate for the parameter of interest (e.g., the survival rate) is calculated by maximizing the likelihood of all parameters, including the nuisance parameters. Though in practice, nuisance parameters are often removed from the likelihood equation prior to maximization by the use of either sufficient statistics or integration. In the Bayesian framework, however, we can calculate marginal posterior distributions for our parameters of interest as mentioned above. The uncertainty associated with the parameter of interest is assessed by integrating across the posterior distribution of the nuisance parameters (Gelman et al. 1995, Ellison 1996, Hobbs and Hilborn 2006). As such, the uncertainty in the nuisance parameters is propagating directly into the posterior distribution for the parameter of interest. In the case of nest-survival or mark-recapture models, any uncertainty in the encounter or recapture probabilities is propagated directly into the posterior distribution of survival. This may make little difference for simple models (Hobbs and Hilborn 2006), but for more complex models, the differences can be startling (Reckhow 1990:2053) and this difference will be especially noticeable when the parameter of interest is non-linearly correlated with the nuisance parameters (Ludwig 1996).

In addition to the ability to make probability statements about the parameters of interest and propagate the uncertainty in nuisance parameters, Bayesian methods are often touted because of a range of other desirable features. First, unlike frequentist methods, inference from Bayesian methods is not based on asymptotic assumptions. The results from Bayesian methods are valid even for small sample sizes, assuming the models for the data and priors are both correct. Uncertainty will increase with decreasing sample sizes, but the posterior distributions and credibility intervals remain valid. Second, Bayesian methods can be used to estimate a wide class of hierarchical models (e.g., mixed-effects models with random effects that have non-normal distributions) because the conditional structure and computational methods do not require that certain parameters be removed from the likelihood via sufficient statistics or integration, as is often required with non-Bayesian methods. And third, because of the requirement of explicitly stating assumptions in the form of prior distributions and the conditional structure of Bayesian models, increasing the complexity of these models can often be accomplished without decreasing the transparency to the reader

APPLICTION OF BAYESIAN STATISTICS TO NEST-SURVIVAL STUDIES

Despite the fact Bayesian statistics have been applied extensively to mark-recapture models (Dupis 1995, Brooks et al. 2002, Johnson and Hoeting 2003) and band-recovery models (Vounatsou and Smith 1995; Brooks et al. 2000, 2002), Bayesian nest-survival models have been largely unexplored. Any nest-survival model based on a likelihood equation (Bart and Robson 1982, Heisey and Nordheim 1995, Dinsmore et al. 2002, Shaffer 2004, Stanley 2004a) could be analyzed in a Bayesian framework. A review of each of these methods is beyond the scope of this manuscript, but is included elsewhere in this volume (Johnson, *this volume*), but with the careful application of prior knowledge, each of them could be adapted so as to produce formal posterior distributions for survival and the effects of covariates on survival, when applicable. The program MARK (White and Burnham 1999) can, in fact, perform Bayesian estimation of the Dinsmore et al. (2002) model, but only allows prior distributions for the parameters for the covariates in the logit model, and these priors must be normal distributions. Except for the Dinsmore et al. (2002) model in MARK, all applications of Bayesian statistics to nest-survival models have focused on estimating age-specific survival rates, especially when nest age is not known.

Several frequentist techniques exist for estimating age-specific survival. The models described by Dinsmore et al. (2002) and Shaffer (2004) are able to estimate age-specific survival, but only when ages are known, such as through egg floating (Westerskov 1950) or egg candling (Weller 1956, Lokemoen and Koford 1996), and they are unable to accommodate age-specific encounter probabilities (Rotella et al. 2004). The method of Heisey and Nordheim (1995) is able to estimate age-specific nest-survival rates when nest ages are unknown, but the algorithm to solve the likelihood equations often has difficulty converging on an estimate when the incubation period (the number of days between the time when the first egg is laid and the first nestling fledges) is long (He 2003). The solution to the estimation problem for the Heisey and Nordheim (1995) model has been to group ages together and assume constant survival and encounter probabilities for each group. If, however, the interval lengths for the groups are not chosen properly or too many ages are grouped together, this solution can produce biased estimators of survival (Heisey and Nordheim 1990, He et al. 2001, He 2003). The algorithms used for Bayesian models, such as MCMC, do not tend to have such difficulties estimating large numbers of parameters and can therefore successfully estimate age-specific survival rates when nest age is unknown without having to make assumptions about certain ages having equal probabilities.

He et al. (2001) were the first to publish a Bayesian nest-survival model that could estimate age-specific survival rates without knowing nest age. This model makes many of the typical nest-survival model assumptions (i.e., nests are independent, nest fate is independent of nest encounter and visits to the nest, and nest fate is correctly determined). The key differences in the assumptions behind the He et al. (2001) model are that nests of the same age have the same survival and encounter probabilities, nests of different ages may have different survival and encounter probabilities, and nest age need not be known. This model does, however, require that nests be visited daily once they are encountered.

He et al. (2001) were able to estimate age-specific survival and encounter probabilities without knowing age because they assume nests are visited daily and the incubation period is constant and known (i.e., each nest requires the same fixed number of days between the day the first egg is laid and the first nestling fledges, and that number is known a priori). The latter assumption was also used by Heisey and Nordheim (1995). When encountered nests are visited daily and the incubation period is known, the age of a successful nest at first encounter can be deduced. For example, if the incubation period is 26 d, and the nest was determined to be successful on day 10 of observation, then the nest must have been 17-d old when it was discovered (first day of observation). For species with multi-stage nests, age at first encounter might be able to be determined based on the day on which it transitioned from one stage to the next, regardless of whether the nest fails or succeeds.

If the nest is unsuccessful, then nest age at first encounter and at failure can still be placed within a range of values when the nest is not directly aged. For example, if the incubation period is 26 d, and the nest failed on day 22 of observation, then the following scenarios could have occurred: the nest was discovered at age 1, survived from ages 1–21, and failed at age 22; the nest was discovered at age 2 and failed at age 23; the nest was discovered at age 3 and failed at age 24; the nest was discovered at age 4 and failed at age 25. The uncertainty as to which scenario actually occurred can be incorporated directly into the likelihood equation in much the same way as unknown fates are incorporated into mark-recapture models (Lebreton et al. 1992). It is this uncertainty, however, which causes problems for the method of Heisey and Nordheim (1995) when incubation times are long. If a species has clearly defined nesting stages and the researcher knows the stage in which the nest failed but is unable to determine an exact age at first encounter, then this stage information can be included in the model below by limiting the range of possible age at failure to those ages within the observed stage at failure.

The likelihood equation in He et al. (2001) can be constructed by writing the observation history for each nest in probabilistic terms and then multiplying them together, just as one can do with mark-recapture models. We will use the same variables as in He et al. (2001)—δ_i equals the probability that a nest of age i is encountered, q_i equals the probability that a nest of age i fails (note that this is a failure rate rather than a survival rate), and the probability that a nest succeeds equals one minus the sum of all the age-specific failure probabilities. For the following example, assume the incubation period is 26 d as above. For nests that succeeded, we calculate their ages at discovery, so the equation for their contribution to the likelihood, with the actual age substituted for the subscript i would be:

$$\delta_i * (1 - (q_1 + q_2 + q_3 + \ldots + q_{26}))$$

For nests that failed, rather than writing a general equation, we will simply give examples. If a nest failed on the day 22 of observation (as above), that nest's contribution to the likelihood would equal:

$$\delta_1 q_{22} + \delta_2 q_{23} + \delta_3 q_{24} + \delta_4 q_{25}$$

If the nest failed on day five of observation, then that nest's contribution to the likelihood would equal:

$$\delta_1 q_5 + \delta_2 q_6 + \delta_3 q_7 + \ldots + \delta_{19} q_{23} + \delta_{20} q_{24} + \delta_{21} q_{25}$$

The above equation would be read as the probability of the nest being discovered at age 1 (the first observation) and failing at age 5 (the fifth observation) plus the probability of being discovered at age 2 and failing at age 6, all the way up to the probability of being discovered at age 21 and failing at age 25. When nest ages are determined at first encounter but fail, the uncertainty in the age at discovery and failure is removed, and only a single term is required to model that nest's history, $\delta_i q_j$ with the actual ages substituted for the subscripts i and j. Similar equations would be written for every single nest and then multiplied together. The model is flexible enough to be applied to situations when all, some, or no nests are aged at first encounter. The full likelihood equation, as described in He et al. (2001), is the product of all the nests' contributions to the likelihood divided by a scaling variable that equals the sum of all possible combinations of encounter at age and either failure at age or survival to first fledging.

The next component of the Bayesian model in He et al. (2001) is the set of priors. He et al. (2001) use non-informative priors for both the age-specific survival and encounter probabilities. In particular, they use the Dirichlet distribution, which is a multivariate version of the beta distribution with hyperparameters equal to one as depicted in Fig. 1. A problem with this is that the Dirichlet distribution induces a correlation between the age-specific parameters (i.e., survival probabilities are correlated across ages or encounter probabilities are correlated across ages). He (2003), however, uses independent beta distributions for each age-specific survival and encounter probability, thus removing the correlation issue.

With the likelihood and priors now defined as above, He et al. (2001) use the Gibbs sampler to produce the marginal posterior for the age-specific encounter and failure probabilities. The likelihood equation as described above was manipulated by substituting and transforming some variables so that the Gibbs sampler would solve for all the parameters more efficiently, but the details of these substitutions and transformations are beyond the scope of this manuscript. He et al. (2001) demonstrate this method working well with both simulated and real data.

A number of refinements to the He et al. (2001) model have been made. Along with relaxing the assumption of correlated priors on the age-specific encounter and survival probabilities, He (2003) also relaxed the assumption that each nest was visited daily, thus allowing for irregular visits and censoring of failure events (i.e., when the timing of failure events is not known exactly). The model in He (2003), however, can underestimate the age 1 survival probabilities under certain irregular visiting schedules, and Cao and He (2005) present three solutions to this. Cao et al. (in press) extends the He (2003) irregular visit model by incorporating categorical covariates into the survival probabilities. Finally, Cao and He (unpubl. data) expand on Cao et al. (in press) by allowing for both categorical and continuous nest-specific covariates for the survival probabilities. All the above models are currently written as FORTRAN programs, but work is underway to make them more user-friendly (C. Z. He, pers. comm.).

CONCLUSION

Bayesian statistics provide a powerful tool for formally incorporating prior knowledge and allow researchers to make probabilistic statements about the realized outcomes. Being able to calculate the probability of a hypothesis being true or a parameter taking on a range of values is crucial for applying research to management and decision-making.

The algorithms used for Bayesian analysis perform well with even very complex models, which is in large part why the age-specific survival models with unknown age work as well as they do. Computer programs such as WinBUGS and those under development by He and colleagues (C. Z. He, pers. comm.) will make Bayesian methods far more accessible than they have been in the past. Even though writing code in WinBUGS is relatively straight forward for those comfortable with other programming languages (Visual Basic, C++, or even scripts in Splus or R), developing one's own model based on the work of He and colleagues would be no small feat. Developing less complicated models in WinBUGS such as when the age of each nest is known exactly, could be achieved if one has a firm understanding of likelihoods and probability distributions, is comfortable with programming, and understands the wide array of diagnostics (Kass et al. 1998, Link et al. 2002) necessary to examine the adequacy of the posterior distribution. The program MARK has the capability of performing a Bayesian analysis of the Dinsmore et al. (2002) model, however the way in which the priors are required to be defined (i.e., as normal distributions and only on the parameters for the covariates in the logit model, and not the survival rate itself) may limit researchers' ability to adequately incorporate the full range of prior knowledge. If one wishes to use only non-informative priors, the formulation in MARK for the Dinsmore et al. (2002) should be more than adequate. As both science and statistical theory move forward, Bayesian methods hold great promise for helping researchers find solutions to complex problems and provide managers and decision-makers the tools they need to make wise choices.

ACKNOWLEDGEMENTS

We would like the thank Stephanie Jones for giving us the opportunity to participate in this special issue and Chong Z. He for introducing us to this topic. The manuscript was greatly improved by the comments of Stephanie Jones, Brian Dennis, Jay Rotella, Micheal Wunder, and five anonymous reviewers.