

Objective Bayesian Estimation for the Number of Species

Kathryn Barger* and John Bunge†

Abstract. Objective priors have been used in Bayesian models for estimating the number of species in a population, but they have not been examined in depth. Here we derive the form of two objective priors, using Bernardo’s reference method and Jeffreys’ rule, based on the mixed-Poisson likelihood used in the single-abundance-sample species problem. These derivations are based on asymptotic results for estimates of integer-valued parameters. The factored form of these priors justifies the use of independent prior distributions for the parameter of interest (the number of species) and the nuisance parameters (of the stochastic abundance distribution). We find that the reference prior is preferable overall to the prior resulting from Jeffreys’ rule. Although a comprehensive objective Bayesian approach can become analytically intractable for more complicated models, the essence of the approach can be upheld in practice. We analyze several datasets to show that the method can be implemented in practice and that it yields good results, comparable with current competing methods.

Keywords: Jeffreys’ prior, mixed-Poisson, noninformative prior, reference prior, species richness estimation

1 Introduction

Estimating the number of classes in a population arises in numerous applications. One of the most important examples is the “species problem” in biology, where we estimate the number of species or other taxa in a population. Other examples can be found in areas such as linguistics, to estimate the number of words in an author’s vocabulary (Efron and Thisted 1976); numismatics, to estimate the number of dies in an ancient coinage (Esty 1986); software development, to estimate the number of faults in computer software (Lloyd et al. 1999); and many more. In addition, methods used in the species problem overlap to a large extent with methods used in capture-recapture estimation, although we do not enter into that application here. In this paper members of the population of interest will be referred to as individuals and the types of individuals in the population will be called species. Methodology used in the species problem can be applied to any population (not necessarily composed of animals), as long as it is partitioned into a finite number of types. In this paper we consider estimation of the total number of species based on a single sample of abundances from a closed population.

Generally speaking we prefer parametric to nonparametric approaches to this problem, since parametric models provide strong smoothing of the data that tends to stabilize

*Department of Statistical Science, Cornell University, Ithaca, NY, <mailto:kjb34@cornell.edu>

†Department of Statistical Science, Cornell University, Ithaca, NY, <mailto:jab18@cornell.edu>

the estimate (prediction) of the zero count or number of unobserved species. Model selection is usually empirical, though, since the underlying biological models are debatable, and we return to this issue in Section 5. In the parametric context, we prefer objective Bayesian methods. Apart from general principles, such procedures in this problem offer greater numerical (computational) stability, and more intuitive results than their frequentist counterparts – e.g., posterior estimates are always integers, and intervals for the number of species are asymmetric and guaranteed to be above the observed number of species. We now consider distinctions between the available approaches in a little more detail.

1. Nonparametric, frequentist estimation. The most-used statistic in this area is the Abundance-based Coverage Estimator (ACE) and its variants, due to [Chao and Lee \(1992\)](#). This method has appeared in countless articles. The estimators are intuitive and easy to compute, but are sensitive to the inclusion of outliers (high frequency or abundance counts), and little diagnostic information is available. More recently, nonparametric maximum likelihood (NPMLE) methods have been developed by [Mao and Lindsay \(2007\)](#), [Böhning and Schön \(2005\)](#) and [Wang and Lindsay \(2005\)](#). These methods do not specify a parametric form for the distribution of species' abundances, but estimate the underlying abundance distribution as a finite mixture of point masses. The number of point masses can be determined by various methods, including the Bayesian Information Criterion or through the use of a penalty parameter. Apart from the question of the number of point masses, error estimation is done by resampling, which typically produces a certain proportion of huge outliers which must be trimmed in some way; we discuss this further in Section 5.
2. Parametric, frequentist estimation. This is the oldest and most broadly studied approach ([Bunge and Barger 2008](#)). The method requires specifying a model for the species' abundance distribution and estimating the model parameters, usually by maximum likelihood, and hence model selection and numerical optimization (for parameter estimation) are two challenges of this approach. [Fisher et al. \(1943\)](#) used a gamma mixed-Poisson model for the distribution of species obtained in a sample, and work on this model has continued to the present day ([Wang 2010](#)). Various other mixing distributions have been proposed, including the generalized inverse Gaussian distribution ([Sichel 1997](#)), as well as the inverse Gaussian, Pareto, lognormal, and finite mixtures of exponential distributions ([Bunge and Barger 2008](#)). Some recent examples of the use of these methods for microbial communities appear in [Hong et al. \(2006\)](#) and [Behnke et al. \(2006\)](#).
3. Nonparametric, Bayesian estimation. Outside of the species problem, a standard nonparametric Bayesian solution requires specification of a Dirichlet process prior ([Ferguson 1973](#)) assigning probability distributions to function spaces. In a problem closely related to species richness estimation, namely capture-recapture, [Tardella](#) and co-authors (see [Farcomeni and Tardella \(2010\)](#) and references therein) have studied nonparametric (and parametric) Bayesian methods for estimating the population size. Their methods appear to make structural use of the known, finite

number of “trapping occasions” in the capture-recapture framework. In one setup, for example, they use a binomial model, with the binomial probabilities following some distribution F ; the likelihood is then reparameterised using canonical moments up to the number of trapping occasions, yielding a finite dimensional parameter space. The (mixed) Poisson model considered here can be regarded as a limiting form of the binomial model, and hence it may be possible to extend their approach to our setting, but this is a topic for future research. From another perspective Lijoi et al. (2007) gave a nonparametric Bayesian solution for estimating the probability of discovering a new species under further sampling, and subsequently Favaro et al. (2009) provided a nonparametric Bayesian method for estimating the number of new species to be observed in a new sample of size m , conditional on an initial sample of size n .

4. Parametric, Bayesian estimation. Bayesian estimation for the species problem involves placing a prior on the number of species and on the parameters of the abundance distribution. Let the marginal priors for the number of species C and the nuisance parameter vector η be denoted $\pi(C)$ and $\pi(\eta)$, respectively. The joint prior will be denoted $\pi(C, \eta)$. The Bayesian approach to this problem has several advantages. Posterior interval estimates are restricted to be within the parameter space, i.e., the estimate of C is guaranteed to be at least as large as the observed number of species, and credible intervals are asymmetric, as intuition suggests. As opposed to a maximum likelihood estimate of C , the Bayesian method naturally produces a posterior distribution for C which is discrete, lending itself easily to integer valued estimates. (It is possible to redefine the parameter space so as to require integer valued maximum likelihood estimates, but the Bayesian method of assigning all mass to the integers seems more natural.) Furthermore, computation of Bayesian estimates is more straightforward than the often chancy numerical optimization required for maximum likelihood, especially in higher-dimensional parameter spaces. Still, model selection remains a hurdle and perhaps more importantly, the use of specific priors needs to be justified. All of the priors used to date in the species problem assume C and η are independent; i.e. $\pi(C, \eta) = \pi(C)\pi(\eta)$. A fundamental difference in this paper is that the priors are derived jointly instead of being assumed *a priori* independent. The Jeffreys’ and reference rules allow for this derivation. Nonetheless, due to the form of the likelihood, the joint prior factors into a product of two independent priors for C and η .

In the literature there have been several suggestions and implementations of priors on C . Priors that have been used include the negative binomial (Hill (1979), Lewins and Joanes (1984), Raftery (1987), Rodrigues et al. (2001), Wang et al. (2007)), Poisson (Madigan and York (1997), Raftery (1988)), geometric (Zhang and Stern 2009), bounded uniform (Zhang and Stern 2005), and improper uniform (Boender and Rinnooy Kan (1987), Tardella (2002), Wang et al. (2007)). Jeffreys’ proposed prior (Jeffreys (1939/1961), pp. 238-239), which takes the form $\pi(C) \propto 1/C$, has also been widely used (George and Robert (1992), Madigan and York (1997), Smith (1991), Tardella (2002), Wang et al. (2007)). Priors of the form $\pi(C) \propto 1/C^r$, where r is a nonnegative constant, were discussed in Wang

et al. (2007). This includes the improper uniform prior when $r = 0$ and Jeffreys' proposed prior when $r = 1$. Rissanen (1983) additionally proposed an objective prior based on coding theory and this has been used by Tardella (2002) and Madigan and York (1997). Recently Quince et al. (2008) carried out an in-depth study using the uniform improper prior for C , in conjunction with several different parametric species abundance distributions (see Section 2.1 for definitions), notably the $\log t$ and generalized inverse Gaussian distributions (also using flat priors for the nuisance parameters). Quince et al. (2008) fit their models to the entire observed frequency dataset without truncating large outlying frequencies, and then applied model-selection criteria to obtain a final analysis. We return to the issue of truncation in Section 5 below.

Finally, Berger et al. (2008) study the general problem of objective Bayesian inference for discrete parameters. They analyze the problem “by embedding the original model into a model with a continuous parameter. [...] [W]e can then apply the ordinary reference prior theory [...] and appropriately discretize the resulting continuous reference prior (if necessary).” They describe four approaches for this. Referring to our previous work (Barger and Bunge 2008), they find that our approach (described and extended herein) will yield the same objective prior as their Approach 3, which “proceeds by choosing a consistent linear (or some other simple) estimate[...]; finding the asymptotic sampling distribution of [the estimator]; pretending that [the parameter] is continuous in this asymptotic distribution and finding the corresponding reference prior.” The interesting questions raised by this connection are topics for future research.

Within the Bayesian framework, we choose to use parametric models with objective, or noninformative priors. Objective priors are useful in many cases when prior information is unavailable or controversial. Also, objective priors can be used in a sensitivity analysis for comparison with other, subjective priors. One disadvantage of objective priors is that they are often improper, and so we are required to show that the posterior integrates to unity in order to allow valid inference. In this paper we present a prior which is jointly derived for all model parameters based on a single notion of objectivity. Methods due to Bernardo (1979) and to Jeffreys (1946) are used to derive objective priors without assuming that the the priors for the number of species and the nuisance parameter are independent. In Barger and Bunge (2008) we derived the Jeffreys' and reference priors for some specific, low-dimensional cases, and applied the results, but here we derive these in full generality (Theorem 2.1 in particular) and study them in more depth. We find the reference prior to be logical and usable, and we demonstrate this with results from applied data analyses.

2 Reference Prior

The reference prior (Bernardo 1979) is an objective prior which is based on maximizing the expected entropy of an experiment. This can also be interpreted as maximizing the missing information from the experiment, where information is defined in the same

way as Shannon entropy (Lindley 1956). Under regularity conditions the reference prior can be obtained via the Fisher information (Bernardo and Ramón 1998); here we use a surrogate information matrix, due to Lindsay and Roeder (1987), that allows for a combination of integer- and continuous-valued parameters under further regularity conditions. The derivation of the reference prior takes into account the order of interest of the parameters; in our problem C is the parameter of interest and η is a nuisance parameter. In the case where η is a vector, a further ordering of the components in η must be assigned; however, in specific cases it has been shown that this ordering is not important even though there is no guarantee that different orderings will yield the same reference prior (Irony and Singpurwalla 1997).

In general, for a model $p(x|\phi, \lambda)$ with interest parameter ϕ and nuisance parameter λ , the joint prior $\pi(\phi, \lambda)$ is derived in steps. First, a conditional prior $\pi(\lambda|\phi)$ is derived from the model. Then, the marginal model $p(x|\phi)$ is obtained after integrating out the nuisance parameter, and this model is used to find the prior $\pi(\phi)$. This method is outlined in Bernardo and Ramón (1998); we apply it to the species problem in Section 2.2.

2.1 Likelihood and information

In this section we define notation and specify the likelihood function. Let C be the unknown number of species in the population, and assume C is finite. Let X_i be the sample abundance of species i , that is, the number of individuals collected from the i th species, $i = 1, \dots, C$. These abundances are assumed to arise independently from an F -mixed Poisson distribution, where F is a parametric c.d.f. with m -dimensional parameter vector η , so that the marginal distribution of X_i given η is

$$p_\eta(x) = \int \frac{e^{-\lambda} \lambda^x}{x!} dF(\lambda|\eta), x = 0, 1, 2, \dots$$

Consider $x_i, i = 1, \dots, C$, a realization of the X_i s. We can only observe the number of individuals contributed to the sample by a species when the contribution is greater than zero. The observed data are therefore $n_j = \sum_{i=1}^C I(x_i = j)$ for $j \geq 1$, where $I(x_i = j)$ is 1 if $x_i = j$ and 0 otherwise. Thus, n_j represents the number of species that contribute j individuals to the sample. The observed number of species is $w = \sum_{j \geq 1} n_j$, and the observed number of individuals is $n = \sum_{j \geq 1} j n_j$. Both w and n are random in this model and we will retain this assumption throughout. Taking n to be fixed results in a multinomial-based model (Sandland and Cormack 1984).

The F -mixed Poisson model assumes heterogeneity among the mean abundances λ_i by using a distribution dependent on a vector parameter η . The likelihood can be written as

$$L(C, \eta; \text{data}) = \frac{C!}{(C-w)!n_1!n_2!\dots} \prod_{i=1}^C p_\eta(x_i).$$

Sanathanan (1972) demonstrates that this likelihood can be written as

$$\begin{aligned} L(C, \eta; \text{data}) &= \binom{C}{w} (1 - p_\eta(0))^w (p_\eta(0))^{C-w} \frac{w!}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} \left(\frac{p_\eta(j)}{1 - p_\eta(0)} \right)^{n_j} \quad (1) \\ &= A(w|C, \eta) B(n_1, n_2, \dots | \eta). \end{aligned}$$

The derivation of the reference and Jeffreys' priors depends on an asymptotic result that can be satisfied by showing the maximum likelihood estimators have an asymptotically normal distribution. The variance of this limit distribution can be expressed as the inverse of the information matrix for the parameters. The Fisher information is typically defined for likelihoods which are differentiable with respect to the parameters. The species likelihood is not differentiable in C since C is a discrete parameter taking only positive integer values. However, information for discrete parameters can be defined using the linear difference score which "mimics in the integer parameter setting the role of the usual score function in a continuous parameter model" (Lindsay and Roeder (1987), p. 758). The species likelihood can be shown to satisfy the linear difference property.

Using the methods in Lindsay and Roeder (1987) to determine the information matrix for multiparameter models with both integer-valued and continuous parameters, we obtain the information for the parameters C and η (Barger and Bunge 2008),

$$R(C, \eta) = \begin{pmatrix} C^{-1} \frac{1 - p_\eta(0)}{p_\eta(0)} & \left(-\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T \\ -\frac{\partial}{\partial \eta} \log p_\eta(0) & C \varrho(\eta) \end{pmatrix} \quad (2)$$

where $\frac{\partial}{\partial \eta} \log p_\eta(0)$ is the column vector of partial derivatives,

$$\left[\left(\frac{\partial}{\partial \eta_1} \log p_\eta(0), \frac{\partial}{\partial \eta_2} \log p_\eta(0), \dots, \frac{\partial}{\partial \eta_m} \log p_\eta(0) \right) \right]^T,$$

and

$$\varrho(\eta) = E_X \left[\left(\frac{\partial}{\partial \eta} \log p_\eta(X) \right)^2 \right]$$

is $(m \times m)$ where expectation is taken with respect to p_η . Notice that the diagonal elements of this partitioned matrix contain elements which factor into a function of C times a function of η , the nuisance parameter. This factorization in the elements of the information matrix will become important in deriving the prior for C and η . Also, note that the information is written as a function of the abundance distribution p_η . This also simplifies calculations since our derivations will involve p_η instead of the entire likelihood function.

The result in (2) coincides with the asymptotic results in Sanathanan (1972). Assuming Sanathanan's regularity conditions, the distribution of $(C^{-1/2}(\hat{C} - C), C^{1/2}(\hat{\eta} - \eta))$

is asymptotically normal $N(0, \Sigma)$, where \hat{C} and $\hat{\eta}$ are maximum likelihood estimates of C and η , respectively. For η having dimension m , Σ^{-1} is the $(m + 1) \times (m + 1)$ matrix

$$\Sigma^{-1} = \begin{pmatrix} \frac{1 - p_\eta(0)}{p_\eta(0)} & \left(-\frac{\partial}{\partial \eta} \log p_\eta(0)\right)^T \\ -\frac{\partial}{\partial \eta} \log p_\eta(0) & \varrho(\eta) \end{pmatrix} \tag{3}$$

The matrix in (3) coincides with (2) up to a multiplicative constant on the diagonal.

2.2 General reference case

Theorem 3 is the main result of this paper and shows the form of the joint reference prior for a model with the nuisance parameter η having m components. This derivation is based on the general result for deriving a reference prior for continuous-valued parameters (Bernardo and Ramón 1998), the information for integer-valued parameters (Lindsay and Roeder 1987), and the likelihood from the species model described in Section 2.1.

Theorem 3. *Let $L(C, \eta; x)$ be the species likelihood, where C is an integer-valued parameter representing the number of species in the population and $\eta = (\eta_1, \dots, \eta_m)$ is the nuisance parameter vector from the stochastic abundance distribution. The quantity of interest is C . Let $S(C, \eta)$ be the inverse of the information matrix for the parameters. Then*

$$S(C, \eta) = R(C, \eta)^{-1} = \begin{pmatrix} Cs(\eta)_{11} & s(\eta)_{12} \\ s(\eta)_{21} & C^{-1}s(\eta)_{22} \end{pmatrix}$$

where all of the $s(\eta)$ depend on the abundance distribution; $s(\eta)_{11}$ is a scalar, $s(\eta)_{12}$ is $1 \times m$, $s(\eta)_{21}$ is $m \times 1$, and $s(\eta)_{22}$ is $m \times m$. If S_j is the $(j \times j)$ upper matrix of $S(C, \eta)$ and $H_j = S_j^{-1}$, then each H_j is a $(j \times j)$ matrix with the form

$$H_j = \begin{pmatrix} C^{-1}h(\eta)_1 & h(\eta)_2 \\ h(\eta)_3 & CH(\eta) \end{pmatrix}$$

where $h(\eta)_1$ is a scalar, $h(\eta)_2$ is $1 \times (j - 1)$, $h(\eta)_3$ is $(j - 1) \times 1$, and $H(\eta)$ is $(j - 1) \times (j - 1)$.

A) *The conditional reference priors are*

$$\pi(\eta_m | C, \eta_1, \dots, \eta_{m-1}) \propto \varrho(\eta)_{mm}^{1/2}$$

and

$$\pi(\eta_k | C, \eta_1, \dots, \eta_{k-1}) \propto \exp \left[\int \dots \int (\log h_{kk}^{1/2}) \right]$$

$$\times \left\{ \prod_{j=k+1}^m \pi(\eta_j | C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \Big]$$

where $d\eta_{\mathbf{k}+1} = d\eta_{k+1} \times \dots \times d\eta_m$ if all of the $\pi(\eta_k | C, \eta_1, \dots, \eta_{k-1})$, $k = 1, \dots, m$ are proper. If any of the conditional reference priors are not proper, then a compact approximation is required for the corresponding integrals.

B) The marginal reference prior for C is

$$\pi(C) \propto C^{-1/2} \quad (4)$$

where all of the conditional priors are proper or a compact approximation is used for the corresponding integrals.

There are several immediate results from Theorem 3. The form of $\pi(C)$ is the same regardless of what abundance distribution p_η is used. The conditional reference priors are the reference priors for C independent and identically distributed replicates from p_η . Also, the joint prior, $\pi(C, \eta)$ factors into two independent priors for C and η . One drawback to this method is that finding a reference prior for a vector-valued η can be difficult. If the nuisance parameter is a scalar, we can find precise forms for the joint reference prior for some examples (Barger and Bunge 2008).

There is no guarantee that the resulting joint reference prior will be proportional to a proper probability distribution. We can already see that $\pi(C)$ is not integrable. Consider any prior for C which has the form $C^{\frac{m-1}{2}}$. With $m = 0$, this reduces to the marginal reference prior for C . Using this more general form of the prior in (4), we can show that the posterior distribution can be integrated with respect to C .

The posterior for the model is

$$\begin{aligned} \pi(C, \eta | \text{data}) &\propto \pi(C, \eta) L(C, \eta; \text{data}) \\ &= C^{\frac{m-1}{2}} \pi(\eta) \frac{C!}{(C-w)!} (p_\eta(0))^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_\eta(j))^{n_j} \end{aligned}$$

where C takes values $w, w+1, w+2, \dots$ and $\eta \in \mathbb{R}^m$. We need to show $\int d\pi(C, \eta | \text{data}) < \infty$. For simplicity we show the calculation for one-dimensional η but the extension to m dimensions is straightforward. We can choose the order of integration since the

integrand is positive. We begin with the iterated integral

$$\begin{aligned}
 & \int_{\mathbb{R}} \sum_{C \geq w} \pi(C, \eta | x) d\eta \\
 &= \int_{\mathbb{R}} \sum_{C \geq w} \pi(C, \eta) \frac{C!}{(C-w)!} (p_{\eta}(0))^{C-w} \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_{\eta}(j))^{n_j} d\eta \\
 &= \int_{\mathbb{R}} \pi(\eta) \frac{1}{\prod_{j \geq 1} n_j!} \prod_{j \geq 1} (p_{\eta}(j))^{n_j} \\
 &\quad \times \sum_{C \geq w} C^{\frac{m-1}{2}} \frac{C!}{(C-w)!} (p_{\eta}(0))^{C-w} d\eta \tag{5} \\
 &= \int_{\mathbb{R}} \pi(\eta) L(\eta; \text{data}) d\eta \tag{6}
 \end{aligned}$$

where $\sum_{C \geq w} C^{\frac{m-1}{2}} \frac{C!}{(C-w)!} (p_{\eta}(0))^{C-w}$ in (5) are moments of the negative binomial distribution (without the normalizing constant). Since all of the moments of the negative binomial exist, the sum is always finite. This is only a first step in showing the posterior is proper. Additionally, the expression in (6) must be integrable with respect to η . Since the integrand in (6) is reduced to a proper prior distribution $\pi(\eta)$ times a likelihood function $L(\eta; \text{data})$ (the marginal likelihood with respect to η), we know the posterior is proper. Thus, as long as the prior for the nuisance parameter $\pi(\eta)$ is proper, then the posterior will be finite. Every model used in the examples in Section 5 has a proper prior for the nuisance parameter except for the inverse Gaussian model, for which we use the improper prior from Gutiérrez-Peña and Rueda (2003). This is the known reference prior for jointly distributed observations from an inverse Gaussian distribution, and it is known to yield a proper posterior in that setting (Liseo 1993). We have not shown this analytically for the mixed-Poisson likelihood, but our numerical computations, at least, presented no difficulties using this prior.

3 Jeffreys' prior

The Jeffreys' prior (Jeffreys 1946) is considered objective and is based on invariance under one-to-one reparameterization. The principle behind Jeffreys' rule is that any prior for a parameter ϕ should yield an equivalent result if a model with a transformed parameter is used. The Jeffreys' prior for a scalar, continuous-valued parameter is defined to be proportional to the square root of the Fisher information. For a parameter ϕ and model $p(x|\phi)$, the Jeffreys' prior is

$$\pi(\phi) \propto h(\phi)^{1/2} \tag{7}$$

where

$$h(\phi) = \int_X p(x|\phi) \left(\frac{\partial}{\partial \phi} \log p(x|\phi) \right)^2 dx. \tag{8}$$

For multidimensional models, the determinant of the Fisher information matrix can be used in place of $h(\phi)$, which preserves the invariance property. However, the use of Jeffreys' prior in multivariate models is controversial. Use of independent priors, where Jeffreys' rule is applied to each parameter separately, has been recommended (Bernardo and Smith 2000). The Jeffreys' prior is often improper, so integrability of the posterior must be shown when using this prior. Despite the difficulties in multivariate settings and in showing the resulting posterior is proper, the Jeffreys' prior is widely used when implementing a noninformative Bayesian approach. In the following section we adapt the Jeffreys' prior to accommodate an integer parameter model.

3.1 General Jeffreys' case

Using the multivariate Jeffreys' rule we find that the Jeffreys' prior (Barger and Bunge 2008) has a different form than the reference prior for the case of the integer-valued parameter in the species problem. (Note that we distinguish Jeffreys' prior as the prior derived using the multivariate Jeffreys' rule, not the prior that Jeffreys' proposed in his book *Theory of Probability* (1939/1961) discussed in Section 1.) The Jeffreys' prior can be derived by taking the square root of the determinant of the information matrix in (2). This yields

$$\pi(C, \eta) \propto \pi(C)\pi(\eta) = C^{\frac{m-1}{2}} \pi(\eta)$$

where

$$\pi^2(\eta) = |\rho(\eta)| \times \left| \frac{1 - p_\eta(0)}{p_\eta(0)} - \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right)^T (\rho(\eta))^{-1} \left(\frac{\partial}{\partial \eta} \log p_\eta(0) \right) \right|,$$

so that $\pi(\eta)$ is a function of η only and depends on the species' abundance distribution. Hence the Jeffreys' prior can be written as a product of two independent priors. For $m \geq 0$ $\pi(C)$ is improper; in fact it is an increasing function of C for $m > 1$, and we discuss this in the next section. Finally, since the marginal Jeffreys' prior is not integrable in C , we must still show that the posterior is finite in each particular setting (the posterior is at least integrable with respect to C in general, as shown in Section 2.2).

4 Comparison of priors

We look first at commonalities between the two approaches. The reference and Jeffreys' priors are equivalent for one-parameter problems. In the species problem, having only one unknown parameter corresponds to assuming that all nuisance parameters are known, i.e., $m = 0$, and in this case $\pi(C) \propto C^{-1/2}$ for both methods. More generally, for any m , it is remarkable that both the reference and Jeffreys' methods result in factorization of the joint prior into independent priors for the parameter of interest and the nuisance parameter, and that the factorization is not affected by the choice of abundance distribution. Finding the exact form of $\pi(\eta)$ can be a difficult problem, though,

especially when working with a vector η . For the Jeffreys' method, a vector η increases the complexity of the determinant in computing the prior, and for the reference method, each additional component in η requires additional integration. In cases where the prior is difficult to calculate, one can use the marginal prior for the number of species $\pi(C)$ in combination with another, preferably objective, prior for the nuisance parameter. Still, if proper priors are not used, the posterior must be checked for integrability.

From a global perspective we are inclined to prefer the reference approach. According to [Bernardo and Ramón \(1998\)](#), “The declared objective of reference Bayesian analysis is to specify a prior distribution such that [...] the information provided by the data should dominate the prior information, because of the ‘vague’ nature of the prior knowledge”; furthermore reference analysis is “the only available method to derive non-subjective posterior distributions which satisfy all [of the following] *desiderata*”: invariance, consistent marginalization, consistent sampling properties, generality, and admissibility. Such considerations are especially important in the species problem, because the ecological literature provides little substantive theory on which to base the statistical analysis of species richness. The reference prior has been successful over the Jeffreys' prior in other multiparameter problems ([Irony and Singpurwalla 1997](#)). In particular, in this problem the marginal reference prior for C does not depend on the dimension of η , which seems reasonable (or at least simple), while the Jeffreys' version depends on m and (although flat for $m = 1$) is actually increasing in C for $m > 1$. If we consider $\pi(C)$ in this case as the limit of a proper prior on a bounded parameter space, then the interpretation seems to be that larger (in fact arbitrarily large) values of C are always more likely *a priori*, a stance which the investigator may not wish to adopt. The increasing property also makes numerical computation of the posterior more challenging (we return to this in [Section 5](#)).

At the suggestion of an anonymous referee, we carried out a small simulation comparison of the behavior of the two approaches. We used the one-parameter exponential-mixed Poisson (geometric) model with $f(\lambda|\theta) = \exp(-\lambda/\theta)/\theta$ and hence $p_\theta(j) = (1/(1+\theta))(\theta/(1+\theta))^j$, $\theta > 0$, $j = 0, 1, \dots$ (see [Barger and Bunge \(2008\)](#) for a detailed treatment of this model). The Jeffreys' prior is proportional to $\theta^{1/2}/(1+\theta)$, and the reference prior is proportional to $C^{-1/2}(\theta(1+\theta))^{-1/2}$; in both cases the posterior is proper. We simulated datasets from this model at $C = 100$ and $C = 1000$, for $\theta = 0.5, 1, 5, 10$ ($p_\theta(0) = 2/3, 1/2, 1/6, 1/11$). At each design point we generated 200 replicate datasets. We analyzed the full dataset in each case rather than setting a maximum frequency (upper truncation point) τ ; see [Section 5](#) regarding this issue. We considered frequentist coverage rates for nominal 95% highest posterior density regions, along with the width of the regions. Essentially we found that the HPD regions for both the reference and Jeffreys' approaches achieved nearly exact 95% (frequentist) coverage, and that HPD widths in the two cases were virtually identical, at every design point of the simulation. We attribute this similar behavior mainly to the simplicity of the geometric model. To obtain better resolution regarding potential divergence of these procedures (in terms of frequentist criteria) it will be necessary to consider more complex models and a more extensive simulation design, which is a topic for future research. However, we note that certain empirical data analyses (using more complex models) do

Table 1: Priors for nuisance parameters in Bayesian procedures. The reference prior is used for C in all models, $\pi(C) \propto C^{-1/2}$.

Abundance distribution	prior for nuisance parameters
gamma $f(\lambda \theta_1, \theta_2) = \frac{\exp\{-\lambda/\theta_2\}\lambda^{\theta_1-1}}{\Gamma(\theta_1)\theta_2^{\theta_1}}$	$\pi(\theta_1) \propto 1/\pi(1 + \theta_1^2), \theta_1 > 0$ $\pi(\theta_2) \propto 1/\pi(1 + \theta_2^2), \theta_2 > 0$
inverse Gaussian $f(\lambda \theta_1, \theta_2) = \left(\frac{\theta_1}{2\pi\lambda^3}\right)^{1/2} \exp\left\{-\frac{\theta_1(\lambda-\theta_2)^2}{2\theta_2^2\lambda}\right\}$	$\pi(\theta_1) \propto \theta_1^{-1}, \theta_1 > 0$ $\pi(\theta_2) \propto \theta_2^{-3/2}, \theta_2 > 0$
two-mixed exponential $f(\lambda \theta_1, \theta_2, \alpha) = \alpha \frac{\exp\{-\lambda/\theta_1\}}{\theta_1} + (1 - \alpha) \frac{\exp\{-\lambda/\theta_2\}}{\theta_2}$	$\pi(\theta_1) \propto \theta_1^{-1/2}(1 + \theta_1)^{-1}, \theta_1 > 0$ $\pi(\theta_2) \propto \theta_2^{-1/2}(1 + \theta_2)^{-1}, \theta_2 > 0$ $\pi(\alpha) \propto 1, 0 < \alpha < 1$

display such divergence, and we return to this in Section 5 below.

5 Examples

We consider three examples to demonstrate the use of the reference prior, and to compare it with the parametric models from [Bunge and Barger \(2008\)](#), the nonparametric mixture model from [Böhning and Schön \(2005\)](#), and the nonparametric coverage-based estimators from [Chao and Lee \(1992\)](#). This is the first comparison of parametric objective Bayesian, parametric frequentist, and nonparametric frequentist methods, to the best of our knowledge. We give complete results only for the reference prior, for the reasons discussed above, but we also describe the Jeffreys' result in one case.

For the parametric procedures, models are selected using the method described in [Bunge and Barger \(2008\)](#) (see below for details). The same abundance distribution is used in both the parametric frequentist and parametric Bayesian methods. (Alternatively, model selection could have been performed separately for the Bayesian analysis using a criterion such as DIC, but keeping the model consistent within each example allows direct comparison of the results.) The priors used for the nuisance parameters are shown in Table 1. For all models, we assume that the prior distributions for each of the components of the nuisance parameter are independent. Note that for the inverse Gaussian model, the prior used for the nuisance parameter is the multivariate reference prior for the two components ([Gutiérrez-Peña and Rueda 2003](#)), and independence among components of the nuisance parameter prior is implied.

The three data sets we analyze have previously appeared in the literature. The Christmas Bird Count data from [Chao and Bunge \(2002\)](#) reports observed bird sightings in Fort Meyers, Florida in 1989. This data set contains 20,042 observed individuals, 126 observed species, and a maximum frequency of 3,877. The Lepidoptera data were used in

Fisher et al. (1943) to determine the number of species of Lepidoptera in Rothamsted, England from 1922 to 1936. This data set contains 15,609 observed individuals, 240 observed species, and a maximum frequency of 2,349. The final dataset is a microbial sample from the waters near Disco Island (Stoeck et al. 2007) in which interest lies in determining the number of different protistan species. The sample was taken in July 2003. This data set contains 2,294 observed individuals, 149 observed species, and a maximum frequency of 65.

Datasets of this type usually exhibit a large number of rare species and a small number of abundant ones, leading to a long right tail in the data (high frequency counts). All existing estimation methods (including those used here) are affected by this right tail, so typically a right truncation point, called τ , is selected, and only frequency counts up to τ are used in the analysis. (The number of species with sample abundances greater than τ is added to the final estimate.) This is an unresolved issue akin to the treatment of outliers in linear models; we discuss it in detail in Bunge and Barger (2008). Here we used values of τ selected by the criteria described therein (essentially goodness-of-fit); this yielded $\tau = 221, 112,$ and 58 for the Christmas Bird Count, Lepidoptera and Disco Island data respectively. We also used $\tau = 10$ in all cases since this is the value recommended by Chao and used in her SPADE program for nonparametric analysis (Chao and Shen 2003)

The goodness-of-fit procedure described in Bunge and Barger (2008) also yields a best-fitting parametric abundance distribution in each case. For these datasets the selected distributions were as follows: at $\tau = 10$, gamma (Christmas Bird Count and Lepidoptera) or inverse Gaussian (Disco Island); at higher τ s, finite mixture of two exponentials (all three datasets). Finally, we selected between the two versions of the nonparametric Abundance-Based Coverage Estimator, ACE and ACE1, according to criteria from Chao and Lee (1992); essentially ACE1 is selected if the coefficient of variation of the data exceeds 0.8.

For Bayesian estimation we use MCMC to simulate from the posterior distributions. Expressions for the full conditionals for some of the models are given in Barger and Bunge (2008). The posterior samples are taken to have an approximate effective sample size (Kass et al. 1998) of 2,500. Acceptance rates for parameters are kept below 40%. We use Bayesian posterior modes as point estimates and highest posterior density regions as the corresponding interval estimates.

For the parametric maximum likelihood and nonparametric coverage-based methods, the reported interval estimates are log-transformed (Chao 1987). Profile likelihood methods could be used in the former case to obtain intervals that more accurately reflect the skewed distribution of the estimates, but we have not yet implemented these in our software. The reported estimates for the NPMLE are accompanied by trimmed interval estimates (Dankmar Böhning, personal communication, April 2009), where outliers in the bootstrap resampling have been removed. These estimates were computed using software provided by Böhning and Kuhnert (personal communication).

Figure 1 shows the analysis of each of the three data sets. Our first observation is that for higher τ , the Bayesian and parametric maximum-likelihood point estimates and

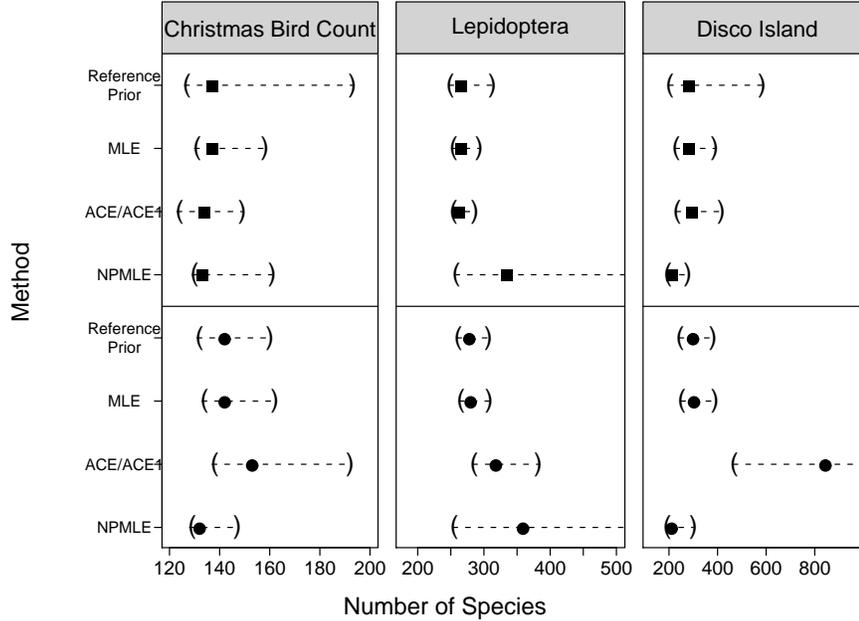


Figure 1: Comparison of methods on three datasets.

Square = estimate at $\tau = 10$; dot = estimate at optimal τ . Dotted line segment = confidence interval (frequentist) or HPD interval (Bayesian).

intervals agree quite well for all three datasets. The Bayesian intervals can be sensitive (Christmas Bird Count, Disco Island) to the loss of information incurred by arbitrarily truncating the data at $\tau = 10$.

The coverage-based nonparametric estimates are reasonable at $\tau = 10$ but tend to be unstable at higher τ (Disco Island). Indeed it is known (Bunge and Barger 2008) that ACE and its variants are sensitive to outliers and diverge to infinity as higher frequencies are included in the data; this is the reason for the default cutoff $\tau = 10$ for these procedures. Finally, the NPML performs reasonably well in some cases but not in others (Lepidoptera); this is mainly due to the heuristic nature of the procedures for selecting the number of components in the final NPML model (Böhning and Schön 2005), and for trimming large outlying values of the estimates produced during bootstrap error estimation.

For comparison, we also carried out the Jeffreys' prior-based analysis of the Lepidoptera data, again using the gamma mixed-Poisson or negative binomial model. The sampler required hand-tuning and did not converge at all at $\tau = 10$, but at $\tau = 112$ we obtained a posterior mode of 342 with SE = 666.9, and 95% HPD interval (272,910). As noted above, we believe that these results reflect the increasing (as well as improper)

nature of the Jeffreys' prior for C : here $m = 2$ so the prior is \sqrt{C} .

6 Conclusion

We have presented a fully Bayesian method using two kinds of objective priors for the problem of estimating the number of classes in a population. The priors are derived based on different notions of objectivity as elaborated by [Jeffreys \(1946\)](#) and [Bernardo \(1979\)](#). Our results are based on asymptotic theory due to [Lindsay and Roeder \(1987\)](#), which provides an information matrix when both integer- and continuous-valued parameters are present. We show that both kinds of objective priors have independent components for parameter of interest and the nuisance parameters.; i.e., $\pi(C, \eta) = \pi(C)\pi(\eta)$, although the method we use to derive the priors does not assume this restriction. Independent priors have been used in all of the previous Bayesian research on the species problem, and we now can provide some justification for this practice.

Although the Bayesian approach is currently one of the less-used methods in applied species estimation, we show that it is practical and yields point estimates that are comparable to frequentist results. Bayesian and frequentist interval estimates are also similar, after using a log transformation (or profile likelihood methods) for the frequentist procedures to account for asymmetric sampling distributions of the estimates.

We recommend the reference prior over the Jeffreys' method for this multiparameter problem. The Jeffreys' estimates tend to be numerically unstable, and the priors for both the parameter of interest and the nuisance parameter are more complicated and difficult to interpret. The reference approach gives a more elegant solution here; in particular the marginal reference prior for C does not depend on the dimension of the nuisance parameter.

We hope to see more use of Bayesian models in the future in this area. Full joint priors can be derived analytically for models with one nuisance parameter, but more complex models are usually needed when dealing with larger data sets. Our suggestion for assigning a prior $\pi(\eta)$, when η is a vector, is at this point somewhat arbitrary and requires more investigation.

Appendix: Proof of Theorem 2.1

We will need the inverse of the information matrix in (2). Denote this inverse by $S(C, \eta) = R(C, \eta)^{-1}$. Taking the inverse of this partitioned matrix, we have

$$S(C, \eta) = \begin{pmatrix} F_{11}^{-1} + F_{11}^{-1}F_{12}E^{-1}F_{21}F_{11}^{-1} & -F_{11}^{-1}F_{12}E^{-1} \\ -E^{-1}F_{21}F_{11}^{-1} & E^{-1} \end{pmatrix}$$

where $E = F_{22} - F_{21}F_{11}^{-1}F_{12}$, and F_{ij} , $i, j = 1, 2$ are the elements of the partitioned information matrix. Now,

$$\begin{aligned} E &= C\varrho(\eta) - \left(-\frac{\partial}{\partial\eta} \log p_\eta(0)\right) \left(\frac{1}{C} \frac{1-p_\eta(0)}{p_\eta(0)}\right)^{-1} \left(-\frac{\partial}{\partial\eta} \log p_\eta(0)\right)^T \\ &= C \left(\varrho(\eta) - \frac{p_\eta(0)}{1-p_\eta(0)} \left(\frac{\partial}{\partial\eta} \log p_\eta(0)\right) \left(\frac{\partial}{\partial\eta} \log p_\eta(0)\right)^T \right). \end{aligned}$$

The elements of this matrix are

$$\begin{aligned} S(C, \eta)_{11} &= C \frac{p_\eta(0)}{1-p_\eta(0)} + \left(C \frac{p_\eta(0)}{1-p_\eta(0)}\right) \left(-\frac{\partial}{\partial\eta} \log p_\eta(0)\right)^T E^{-1} \\ &\quad \times \left(-\frac{\partial}{\partial\eta} \log p_\eta(0)\right) \left(C \frac{p_\eta(0)}{1-p_\eta(0)}\right) \\ &= Cs(\eta)_{11}, \end{aligned}$$

$$\begin{aligned} S(C, \eta)_{12} &= -\left(C \frac{p_\eta(0)}{1-p_\eta(0)}\right) \left(-\frac{\partial}{\partial\eta} \log p_\eta(0)\right)^T E^{-1} \\ &= s(\eta)_{12}, \end{aligned}$$

$$\begin{aligned} S(C, \eta)_{21} &= -E^{-1} \left(-\frac{\partial}{\partial\eta} \log p_\eta(0)\right) \left(C \frac{p_\eta(0)}{1-p_\eta(0)}\right) \\ &= s(\eta)_{21}, \end{aligned}$$

and

$$S(C, \eta)_{22} = E^{-1} = \frac{1}{C} s(\eta)_{22}.$$

Thus, the information matrix has the form

$$S(C, \eta) = \begin{pmatrix} Cs(\eta)_{11} & s(\eta)_{12} \\ s(\eta)_{21} & \frac{1}{C}s(\eta)_{22} \end{pmatrix}.$$

The next result follows from proposition 3 in [Bernardo and Ramón \(1998\)](#). If S_j is the $(j \times j)$ upper matrix of $S(C, \eta)$ and $H_j = S_j^{-1}$, then each H_j is a $(j \times j)$ matrix with the form

$$H_j = \begin{pmatrix} \frac{1}{C}h(\eta)_1 & h(\eta)_2 \\ h(\eta)_3 & CH(\eta) \end{pmatrix}$$

where $h(\eta)_1$ is a scalar, $h(\eta)_2$ is $1 \times (j-1)$, $h(\eta)_3$ is $(j-1) \times 1$, and $H(\eta)$ is $(j-1) \times (j-1)$.

The conditional reference priors are

$$\pi(\eta_m | C, \eta_1, \dots, \eta_{m-1}) \propto \varrho(\eta)_{mm}^{1/2}$$

and

$$\begin{aligned}
 & \pi(\eta_k|C, \eta_1, \dots, \eta_{k-1}) \\
 & \propto \exp \left[\int \cdots \int \log \left(C h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \\
 & = \exp \left[\int \cdots \int \left(\log C + \log h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \\
 & = \exp \left[\log C \int \cdots \int \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \\
 & \quad \times \exp \left[\int \cdots \int \left(\log h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right] \tag{9} \\
 & \propto \exp \left[\int \cdots \int \left(\log h_{kk}^{1/2} \right) \left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_{k+1} \right]
 \end{aligned}$$

where $\varrho(\eta) = E_X \left[\left(\frac{\partial}{\partial \eta} \log p_\eta(X) \right)^2 \right]$ and $\eta_{\mathbf{k}+1} = d\eta_{k+1} \times \dots \times d\eta_m$ if all of the $\pi(\eta_k|C, \eta_1, \dots, \eta_{k-1})$, $k = 1, \dots, m$ are proper. If any of the conditional reference priors are not proper, than a compact approximation is required for the corresponding integrals. In (9) we are able to consider the first exponential as a constant with respect to η_k since $\left\{ \prod_{j=k+1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} = 1$ if all of these conditional priors are proper or if a compact approximation is used. This means all of the conditional priors for η are functions of η only. In fact, the conditional priors are the reference priors for C i.i.d. replicates from p_η .

The marginal reference prior for C is

$$\begin{aligned}
 \pi(C) & \propto \exp \left[\int \cdots \int \log \left(C^{-1/2} s(\eta)_{11}^{-1/2} \right) \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
 & = \exp \left[\int \cdots \int \left(\log C^{-1/2} + \log s(\eta)_{11}^{-1/2} \right) \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
 & = \exp \left[\log C^{-1/2} \int \cdots \int \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
 & \quad \times \exp \left[\int \cdots \int \left(\log s(\eta)_{11}^{-1/2} \right) \left\{ \prod_{j=1}^m \pi(\eta_j|C, \eta_1, \dots, \eta_{j-1}) \right\} d\eta_1 \right] \\
 & \propto C^{-1/2}
 \end{aligned}$$

where all of the conditional priors are proper or a compact approximation is used for the corresponding integrals.

References

- Barger, K. and Bunge, J. (2008). “Bayesian estimation of the number of species using noninformative priors.” *Biometrical Journal*, 50: 1064–1076. 768, 770, 772, 774, 775, 777
- Behnke, A., Bunge, J., Barger, K., Breiner, H.-W., Alla, V., and Stoeck, T. (2006). “Microeukaryote community patterns along an O₂/H₂S gradient in a supersulfidic anoxic fjord (Framvaren, Norway).” *Applied and Environmental Microbiology*, 72: 3626–3636. 766
- Berger, J., Bernardo, J., and Sun, D. (2008). “Reference Priors for Discrete Parameter Spaces.” Technical Report, Duke University. 768
- Bernardo, J. M. (1979). “Reference posterior distributions for Bayesian inference.” *Journal of the Royal Statistical Society Series B*, 41: 113–147. 768, 779
- Bernardo, J. M. and Ramón, J. M. (1998). “An introduction to Bayesian reference analysis: Inference on the ratio of multinomial parameters.” *The Statistician*, 47: 101–135. 769, 771, 775, 780
- Bernardo, J. M. and Smith, A. F. M. (2000). *Bayesian Theory*. New York: Wiley. 774
- Boender, C. G. E. and Rinnooy Kan, A. H. G. (1987). “A multinomial Bayesian approach to the estimation of population and vocabulary size.” *Biometrika*, 74: 849–856. 767
- Böhning, D. and Schön, D. (2005). “Nonparametric maximum likelihood estimation of population size based on the counting distribution.” *Journal of the Royal Statistical Society Series C*, 54: 721–737. 766, 776, 778
- Bunge, J. and Barger, K. (2008). “Parametric models for estimating the number of classes.” *Biometrical Journal*, 50: 971–982. 766, 776, 777, 778
- Chao, A. (1987). “Estimating the population size for capture-recapture data with unequal catchability.” *Biometrics*, 43: 783–791. 777
- Chao, A. and Bunge, J. (2002). “Estimating the number of species in a stochastic abundance model.” *Biometrics*, 58: 531–539. 776
- Chao, A. and Lee, S.-M. (1992). “Estimating the number of classes via sample coverage.” *Journal of the American Statistical Association*, 87: 210–217. 766, 776, 777
- Chao, A. and Shen, T. J. (2003). Program SPADE (Species Prediction And Diversity Estimation), Program and Users Guide published at <http://chao.stat.nthu.edu.tw>. 777

- Efron, B. and Thisted, R. (1976). “Estimating the number of unseen species: How many words did Shakespeare know?” *Biometrika*, 63: 435–447. 765
- Esty, W. W. (1986). “Estimation of the size of a coinage: A survey and comparison of methods.” *Numismatic Chronicle*, 146: 185–215. 765
- Farcomeni, A. and Tardella, L. (2010). “Reference Bayesian methods for recapture models with heterogeneity.” *Test*, 19: 187–208. 766
- Favaro, S., Lijoi, A., Mena, R., and Prünster, I. (2009). “Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior.” *Journal of the Royal Statistical Society Series B*, 71: 993–1008. 767
- Ferguson, T. S. (1973). “A Bayesian analysis of some nonparametric problems.” *Annals of Statistics*, 1: 209–230. 766
- Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943). “The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population.” *Journal of Animal Ecology*, 12: 42–58. 766, 777
- George, E. I. and Robert, C. P. (1992). “Capture-recapture estimation via Gibbs sampling.” *Biometrika*, 79: 677–683. 767
- Gutiérrez-Peña, E. and Rueda, R. (2003). “Reference priors for exponential families.” *Journal of Statistical Planning and Inference*, 110: 35–54. 773, 776
- Hill, B. M. (1979). “Posterior moments of the number of species in a finite population and the posterior probability of finding a new species.” *Journal of the American Statistical Association*, 74: 668–673. 767
- Hong, S.-H., Bunge, J., Jeon, S.-O., and Epstein, S. S. (2006). “Predicting microbial species richness.” *Proceedings of the National Academy of Sciences*, 103: 117–122. 766
- Irony, T. Z. and Singpurwalla, N. D. (1997). “Non-informative priors do not exist: A dialogue with José M. Bernardo.” *Journal of Statistical Planning and Inference*, 65: 159–189. 769, 775
- Jeffreys, H. (1939/1961). *Theory of Probability*. Oxford: University Press, 1961 edition. 767
- (1946). “An invariant form for the prior probability in estimation problems.” *Proceedings of the Royal Society of London Series A*, 186: 453–461. 768, 773, 779
- Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. (1998). “Markov Chain Monte Carlo in practice: A roundtable discussion.” *American Statistician*, 52: 93–100. 777
- Lewins, W. A. and Joanes, D. N. (1984). “Bayesian estimation of the number of species.” *Biometrics*, 40: 323–328. 767

- Lijoi, A., Mena, R., and Prünster, I. (2007). “Bayesian nonparametric estimation of the probability of discovering a new species.” *Biometrika*, 94: 769–786. 767
- Lindley, D. V. (1956). “On a measure of the information provided by an experiment.” *Annals of Mathematical Statistics*, 27: 986–1005. 769
- Lindsay, B. G. and Roeder, K. (1987). “A unified treatment of integer parameter models.” *Journal of the American Statistical Association*, 82: 758–764. 769, 770, 771, 779
- Liseo, B. (1993). “Elimination of nuisance parameters with reference priors.” *Biometrika*, 80: 295–304. 773
- Lloyd, C. J., Yip, P. S. F., and Chan, K. S. (1999). “Estimating the number of faults: Efficiency of removal, recapture, and seeding.” *IEEE Transactions on Reliability*, 48: 369–376. 765
- Madigan, D. and York, J. C. (1997). “Bayesian methods for estimation of the size of a closed population.” *Biometrika*, 84: 19–31. 767, 768
- Mao, C. X. and Lindsay, B. G. (2007). “Estimating the number of classes.” *Annals of Statistics*, 35: 917–930. 766
- Quince, C., Curtis, T. P., and Sloan, W. T. (2008). “The rational exploration of microbial diversity.” *International Society for Microbial Ecology Journal*, 2: 997–1006. 768
- Raftery, A. E. (1987). “Inference and prediction for a general order statistic model with unknown population size.” *Journal of the American Statistical Association*, 82: 1163–1168. 767
- (1988). “Inference for the binomial N parameter: A hierarchical Bayes approach.” *Biometrika*, 75: 223–228. 767
- Rissanen, J. (1983). “A universal prior for integers and estimation by minimum description length.” *Annals of Statistics*, 11: 416–431. 768
- Rodrigues, J., Milan, L. A., and Leite, J. G. (2001). “Hierarchical Bayesian estimation for the number of species.” *Biometrical Journal*, 43: 737–746. 767
- Sanathanan, L. (1972). “Estimating the size of a multinomial population.” *Annals of Mathematical Statistics*, 43: 142–152. 770
- Sandland, R. L. and Cormack, R. M. (1984). “Statistical inference for Poisson and multinomial models for capture-recapture experiments.” *Biometrika*, 71: 27–33. 769
- Sichel, H. S. (1997). “Modelling species-abundance frequencies and species-individual functions with the generalized inverse Gaussian-Poisson distribution.” *South African Statistical Journal*, 31: 13–37. 766

- Smith, P. J. (1991). “Bayesian analyses for a multiple capture-recapture model.” *Biometrika*, 78: 399–407. [767](#)
- Stoeck, T., Kasper, J., Bunge, J., Leslin, C., Ilyin, V., and Epstein, S. (2007). “Protistan diversity in the arctic: A case of paleoclimate shaping modern biodiversity?” *Public Library of Science ONE*, 2: e728. [777](#)
- Tardella, L. (2002). “A new Bayesian method for nonparametric capture-recapture models in presence of heterogeneity.” *Biometrika*, 89: 807–817. [767](#), [768](#)
- Wang, J. and Lindsay, B. G. (2005). “A penalized nonparametric maximum likelihood approach to species richness estimation.” *Journal of the American Statistical Association*, 100: 942–959. [766](#)
- Wang, J. P. (2010). “Estimating species richness by a Poisson-compound gamma model.” *Biometrika*, 97: 727–740. [766](#)
- Wang, X., He, C. Z., and Sun, D. (2007). “Bayesian population estimation for small sample capture-recapture data using noninformative priors.” *Journal of Statistical Planning and Inference*, 137: 1099–1118. [767](#)
- Zhang, H. and Stern, H. (2005). “Investigation of a generalized multinomial model for species data.” *Journal of Statistical Computation and Simulation*, 75: 347–362. [767](#)
- (2009). “Sample size calculation for finding unseen species.” *Bayesian Analysis*, 4: 763–792. [767](#)

Acknowledgments

We are grateful to the Editor, Associate Editor, and two anonymous referees for their careful reading of this paper and many useful suggestions for improvements. We thank Thorsten Stoeck of the University of Kaiserslautern for the Disco Island data, and Dankmar Böhning of the University of Reading for the NPMLE software. This research was conducted using the resources of the Cornell University Center for Advanced Computing, which receives funding from Cornell University, New York State, the National Science Foundation, and other leading public agencies, foundations, and corporations. The work was funded in part by NSF grant DEB-0816638 to JB.

