ENVIRONMETRICS
Environmetrics 2006; **17**: 683–704
Published online 17 February 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/env.786

On using expert opinion in ecological analyses: a frequentist approach

Subhash R. Lele[1*,†] and Kristie L. Allen[2]

[1]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta AB T6G 2G1, Canada*
[2]*Wells Fargo Merchant Payment Solutions, MAC A0347-010, 1200 Montego Way, Bldg C. Walnut Creek, California 94598, USA*

SUMMARY

Many ecological studies are characterized by paucity of hard data. Statistical analysis in such situations leads to flat-likelihood functions and wide confidence intervals. Although, there is paucity of hard data, expert knowledge about the phenomenon under study is many times available. Such expert opinion may be used to strengthen statistical inference in these situations. Subjective Bayesian is one approach to incorporate expert opinion in statistical studies. This approach, aside from the subjectivity, also faces operational problems. Elicitation of the prior is the most difficult step. Another is the lack of a precise quantitative definition of what characterizes an expert. In this paper, we discuss a different approach to incorporating subjective expert opinion in statistical analyses. We argue that it is easier to elicit data than to elicit *a prior*. Such elicited data can then be used to supplement the hard, observed data to possibly improve precision of statistical analyses. The approach suggested here also leads to a natural definition of what constitutes a useful expert. We define a useful expert as one whose opinion adds information over and above what is provided by the observed data. This can be quantified in terms of the change in the Fisher information before and after using the expert opinion. One can, thus, avoid the real possibility of using an expert opinion that adds noise, instead of information, to the hard data. We illustrate this approach using an ecological problem of modeling and predicting occurrence of species. An interesting outcome of this analysis is that statistical thinking helps discriminate between a useful expert and a not so useful expert; expertness need not be decided purely on the basis of experience, fame, or such qualitative characteristics.

KEY WORDS: calibration of an expert; GAP analysis; hierarchical model; logistic regression; presence-absence data

1. INTRODUCTION

Ecologists are increasingly interested in models of ecosystems that relate the probability of occurrence of species to habitat characteristics (Scott *et al.*, 1993; Guisan and Zimmerman, 2000; Fleishman *et al.*, 2001; Scott *et al.*, 2002). Ideally, in these studies managers have access to hard data such as field studies where presence or absence of species along with the associated habitat characteristics is measured. If

*Correspondence to: S. R. Lele, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alta. AB T6G 2G1, Canada.
†E-mail: slele@ualberta.ca

Contract/grant sponsor: National Scientific and Engineering Research Council of Canada.

Received 20 October 2005
Accepted 16 December 2005

such data are available, logistic regression is the natural approach to modeling the probability of presence of a species as a function of habitat covariates. The logistic model has proven its effectiveness in multitude of situations that require modeling of binary data with covariates (Collett, 1991; Hosmer and Lemeshow, 2000) and ecological problems are no exception. However, in practice, estimation of the parameters in Logistic Regression model can be quite difficult if the number of important covariates is large compared to the number of observations (Fleishman *et al.*, 2001). Analysis of rare species also poses a problem. If the number of locations where the species is present is small compared to the number of surveyed locations, then the parameter estimates are quite inaccurate with large standard errors.

Given the enormous costs in terms of time, human effort, and money involved in conducting ecological surveys, observed data are almost always in short supply but usually some untapped expert information is available. Such information is acquired through years of observations and accumulation of general biological knowledge. However, this information is not in the form of hard data but is in the form of opinions and thoughts. The goal of this paper is to suggest a way to exploit this knowledge, provided it is truly useful, to improve the estimation accuracy of logistic regression. These estimates are useful for making effective management decisions about which resources are important for the species under consideration. In addition improved estimates lead to increased accuracy of the predictions as well.

The outline of the paper is as follows. First, we describe the biological situation that motivated this paper. Then issues surrounding elicitation of expert opinion and its incorporation in statistical analysis are discussed. A new approach to incorporate expert opinion, based on elicited data, is proposed. Some theoretical results related to the statistical model, estimation procedure, quantification of the informativeness of the expert, and combining expert opinion are discussed. This is followed by the analysis of the occurrence data on four small mammals (shrews and voles) in Montana. Three different experts provided their best guesses on the occurrence of four different species at a given set of locations based on their knowledge. Two of these experts were ecologists familiar with the species and the study area. The third expert was the GAP analysis program (Scott *et al.*, 1993, 2000, http://www.gap.uidaho.edu) that provided the set of predictions. The human experts as well as the GAP analysis program had access to the habitat characteristics at these locations, but they did not know the true state of the occurrence of the species at these locations. Using the method described in the paper, we compare and contrast these experts in terms of their informativeness. It is shown that a useful expert improves the estimation accuracy as quantified in terms of the standard errors of the estimators and the width of the confidence intervals. Improved estimation leads to improved predictive accuracy of the model. We also illustrate how combining two experts can be either beneficial or detrimental depending on the relationship between the experts. We discuss some further practical issues such as handling of measurement error in the species occurrence data. The paper concludes with summary comments. Although the example discussed in this paper is for binary data, the methodology can be extended to other types of data. The methodology discussed in this paper is also useful in other fields such as satellite image analysis and classification, medical diagnostics, etc.

2. A MOTIVATING EXAMPLE

The scientific problem of interest we consider is quite simple: the prediction of the presence or the absence of the shrews and voles across the state of Montana. Shrews are the smallest of the mammalian predators, some as tiny as a human thumb. They hunt insects and other invertebrates such as earthworms, ants, spiders, and centipedes. The combination of their small size (masked shrews weigh between 2.5 and 5.5 g), high metabolism, and carnivorous habits make them potentially good indicator species. Toxins in the environment may accumulate quite rapidly in shrews because of their small size, large appetites, and high position in the food chain as compared to most small mammals. Seasonal and environmental changes can affect their survival significantly by altering availability of both prey and preferred habitats. High rates of reproduction coupled with vulnerability to environmental change results in constantly expanding and contracting local populations (Nagorsen, 1996; Wilson and Ruff, 1999). Masked shrews in particular occur throughout Alaska, Canada, and the northern regions of the United States. They tend to proliferate most successfully in damp habitats and are found throughout the state of Montana. Their wide range coupled with certain habitat requirements makes them an ideal species for the purpose of this paper.

Habitat covariates were collected for 158 locations spread throughout the state of Montana. Actual presence-absence measurements were collected at 56 sites (Allen *et al.*, 1997). The 56 sites sampled included a wide range of physical and spatial diversity. The variety of habitats ranged from flat arid grasslands to steep lush forests and from alpine meadows to windswept shrub lands. A total of four habitat covariates (elevation, precipitation, latitude, and longitude) are used in this paper. Elevation measures the height above sea level in feet. Elevations in Montana ranged from 2000 to 10 900 feet, with the median being 3600 feet. Precipitation is annual precipitation measured in inches. It ranged from 6 to 80 inches, with a median of 12 inches. Latitude and longitude are proposed as proxy for 'all other' habitat characteristics. Latitude ranges from 44 to 49 and longitude from -104 to -115°.

The goal of this study was to estimate the logistic regression model that relates the habitat covariates to the probability of presence of a species at a given location. This may then be used to predict the presence or absence of the masked shrew at the locations where they were not trapped or, in general, it may be used to predict the effect of habitat changes on the probability of the presence of a species. For further examples of such studies, see Fleishman *et al.*, (2001) and references therein. One can, of course, simply use the logistic regression or an objective Bayesian version of the logistic regression to solve this problem (Hepinstal and Sader, 1997; Fleishman *et al.*, 2001). In this paper we show how expert opinion, if obtained properly, can be used to supplement the logistic-regression approach in a frequentist manner.

3. ON USING EXPERT OPINION

One of the commonly suggested approaches for incorporating expert opinion is the (subjective) Bayesian approach. Hence, we begin by reviewing this approach to incorporating expert opinion into statistical models by eliciting prior distributions of the parameters in the statistical model.

The Bayesian paradigm is the natural and often the standard method for incorporating expert opinion into statistical models (Kadane *et al.*, 1980; Genest and Zidek, 1986; DeGroot, 1988; Steffey, 1992). In this approach, the basic idea is to obtain from experts their prior beliefs regarding the possible values of the *parameters* in the statistical model. These prior beliefs are quantified in terms of a probability distribution, called the 'prior distribution.' In the modeling process, this distribution is then modified in the light of the hard data to obtain the 'posterior distribution' using the Bayes theorem. The posterior distribution signifies the changed beliefs about the parameter values in the light of the data. The posterior distribution is then used for inference about parameters and for prediction.

Many scientists, statisticians, and philosophers feel uncomfortable with the Bayesian approach, especially the subjective Bayesian approach. For a summary of these arguments see, for example, Efron (1986), Dennis (1996), Mayo (1996), Royall (1997), Lele and Das (2000), Lele (2004). For a balanced and unbiased version of these arguments, along with the discussion of the issues associated with the objective Bayesian approach, see Chapter 6 of Barnett (1999) or Chapter 5 of Press (2003). Of course, frequentist approach as advocated by the Neyman–Pearson–Fisher school of thought is not without its distractors. For example, see Berger (1985), Royall (1997), and various papers in Taper and Lele (2004). Aside from the foundational arguments, from a purely operational point of view, there are two further major issues in the implementation of the subjective Bayesian approach.

(a) *Quantification of the expert opinion in terms of prior distributions:* It is well accepted by the practitioners that quantification of *a prior* belief in the form of a probability distribution is a complex problem (Garthwaite and Dickey, 1992). Paraphrasing Yang and Berger (1997) one can summarize the difficulties succinctly as follows.
   (i) Frequently, elicitation of subjective prior distributions is impossible, because of time or cost limitations, or resistance or lack of training of clients.
   (ii) Subjective elicitation can easily result in poor prior distributions, because of systematic elicitation bias and the fact that elicitation typically yields only a few features of the prior, with the rest of the prior (e.g., its functional form) being chosen in a convenient, but possibly, inappropriate, way.
   (iii) In high dimensional problems, the best one can typically hope for is to develop subjective priors for the 'important' parameters, with the unimportant or nuisance parameters being given non-informative priors.

   We provide further discussion of the problem of elicitation of the expert opinion in Section 8. Nevertheless, West (1988) observes that '...it is often (or rather, always) difficult to elicit a full distribution with which an expert is totally comfortable.'

(b) *Quantification of the informativeness of an expert:* From a purely operational point of view, it is hard to imagine that all experts are equally informative about the phenomenon under consideration. Thus it is important to quantify the informativeness of an expert. It is quite possible that an expert, albeit unintentionally, might end up adding noise instead of real information to the hard data. One needs a mechanism to discriminate between experts and then weigh the expert opinion according to its usefulness in the statistical analysis.

The methodology discussed in this paper attempts to address these two issues specifically.

4. ELICIT DATA, NOT PRIOR

Why is it so difficult to elicit *a prior* distribution? It is important to recognize that the concept of *a prior* probability distribution on the parameters of a statistical model is a *statistical* construct that is hard for most scientists to visualize. This makes it difficult for scientists to express their opinions in a precise, clear, and consistent fashion. Thus, it is important that a mechanism be devised that will facilitate scientists to express their opinions in a practical fashion.

If the elicitation of priors is so difficult for the experts in the field, what is more practical? What kind of information can we realistically obtain from an expert with comparative ease? It is often natural for an expert to think in terms of the process or the observations under study. A sensible approach is to ask the expert to provide guess values for the quantities that they are most comfortable with, *the observable data itself*. For a similar sentiment, see Kadane *et al.* (1980) or Kadane and Wolfson (1998). What do we mean by this? Consider some practical situations. Suppose we want to predict and map the presence-absence of a particular species, say the masked shrew (*Sorex Cinereus*). Unfortunately, we do not have enough money to survey every location. However, we may have substantial information on the habitat characteristics at all locations of interest. For example, we may have satellite images for the area depicting the habitat characteristics. Given these images, it is not an impossible task for an expert to provide a guess about the presence or absence of the species at all given locations. In essence, the expert is providing his/her map of the presence-absence of the masked shrew (*sorex cinereus*) at the locations of interest. Similarly for pollutant contamination surveys, experts can use geo-morphological features, soil type, etc. to provide their guess as to whether contamination will be above or below a given threshold. This is apparently quite operational. For example, the possibility of using such expert guesses to improve kriging predictions in geological studies was earlier suggested by two geologists (and, not theoretical statisticians) Kulkarni (1984) and Journel (1986). Also a recent paper by McCoy *et al.* (1999) discusses the application of 'educated guesses' in predicting optimal habitats for sand skink, *Neoseps reynoldsi*. We can then conclude that it is operationally easier to obtain the guess values of the observable data than *a prior* distribution on the parameters of a statistical model. The ease with which we obtained the expert opinion for this paper also attests to the fact that eliciting data from experts is not a difficult task.

Suppose we obtain these guess values from several experts. How should we quantify the informativeness of each expert? To address this issue, we need to develop a statistical model. We will undertake that in the next section.

5. STATISTICAL MODEL FOR COMBINING EXPERT GUESSES WITH OBSERVED DATA

Suppose there are N locations at which we have obtained the expert guesses. Let a subset of these locations of size n be such that we also have the knowledge of the actual observations. Let us denote the expert guesses by $E_1, E_2, \ldots, E_N$. Let the actual observations be denoted by $O_1, O_2, \ldots, O_n$. Let us denote the covariate vectors at the N locations by $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathbf{X}_N$. Let us assume that $O_i \mid X_i \sim f(., X_i, \beta)$ be independent random variables. Let us assume that the expert guesses are related to the true state of nature at those locations by the model $E_\mathrm{i} \mid O_\mathrm{i} \sim g(., O_\mathrm{i}, \eta)$ and are (conditionally on O's and X's) independent of each other. Hence the marginal distribution of the expert guesses can be written as $h(E;\beta,\eta) = \int g(E \mid O;\eta) f(O;X,\beta)\mathrm{d}O$.

The setup described above is a hierarchical model setup used commonly in the measurement error models (Carlin and Louis, 1996). Essentially one can look upon the expert guesses as measurement error introduced observations. Using the first n observations, where we know the truth, one can estimate the measurement error model parameters η. Once the measurement error model parameters are estimated, one can then use the additional $N-n$ observations to supplement the original n observations to possibly improve the statistical inference. The extent of improvement depends on how accurately the measurement error model parameters are estimated, the extent of the measurement error and the number $N-n$ of additional observations. All of this can be made mathematically precise as follows.

6. ESTIMATION OF THE PARAMETERS AND QUANTIFYING INFORMATIVENESS OF THE EXPERT

Given the model description above, we can write down the likelihood function as:

$$L(\beta, \eta; O, E, X) = \prod_{i=1}^{n} g(E_i \mid O_i; \eta) \prod_{i=1}^{n} f(O_i; X_i, \beta) \prod_{i=n+1}^{N} h(E_i; X_i, \beta, \eta) \tag{6.1}$$

One can obtain the full-likelihood estimators of the parameters by maximizing this function. Or, alternatively, one can maximize the first component with respect to the parameter η and obtain $\hat{\eta}$. The second component could then be maximized with respect to the parameter β, holding $\eta = \hat{\eta}$. This second approach corresponds to the pseudo maximum likelihood approach studied by Gong and Samaniego (1981). Both approaches lead to consistent and asymptotically normal estimators under standard regularity conditions and the condition that $n \to \infty$ and $N = Kn$ with fixed K. Appendix 1 provides the exact conditions for consistency and asymptotic normality under the pseudo maximum-likelihood setup. The proof essentially follows the argument given by Bradley and Hart (1962) or Fahrmeir and Kaufmann (1985).

We can also write down the likelihood function for β using only the observed data as $L_n(\beta; O, X) = \prod_{i=1}^{n} f(O_i; X_i, \beta)$. This can be maximized to obtain the maximum-likelihood estimator of β based only on the observed data. The consistency and asymptotic normality of the maximum-likelihood estimator can be proved using standard regularity conditions and the argument as outlined, for example, in Bradley and Hart (1962) or Fahrmeir and Kaufmann (1985). Given the consistency and asymptotic normality of these estimators, we are now in a position to decide if the expert guesses are informative or not.

Let us denote the Fisher information about β, when only the observed data are used, by: $I_n(\beta) = E(-\frac{\partial^2}{\partial \beta^2} \log \prod_{i=1}^{n} f(O_i; X_i, \beta))$. The asymptotic variance of the estimator of β using only the observed data is obtained by the inverse of the Fisher information matrix. We can also compute the Fisher information of (β,η) based on the full-likelihood function (Equation (6.1)) that incorporates the expert-guesses. This can be inverted and the asymptotic variance of the estimator of β based on the combination of observed data and expert guesses can be computed. Due to the costs involved in the estimation of the nuisance parameters η, this asymptotic variance is, in general, not guaranteed to be smaller than the one based only on the observed data. However, by using the expert guesses, we are also increasing the sample size from n to N. Thus, the introduction of the nuisance parameters η does not simply increase the cost of estimation but also comes with the benefit of a larger sample size. It is the balance between these two that decides if the expert guesses are useful or not. Thus, we define an expert to be useful if his/her guesses reduce the variance of the estimator of the parameters of interest β. If one uses the pseudo-likelihood estimator, one can compare the corresponding variance of β with that of the variance of β based only on the observed data. If the variance is reduced, the expert is useful; if not, the expert is not useful and may not be used in the statistical analysis.

6.1. Combining expert opinion

Suppose we have two or more experts who have provided their guesses. It is easy to combine these opinions in the framework discussed earlier. We can simply write down the likelihood using the data

(O_n, E_{1N}, E_{2N}) as follows.

$$L_1 = \prod_{i=1}^{n} g_1(E_{1i} \mid O_i; \eta_1) \prod_{i=1}^{n} g_2(E_{2i} \mid O_i; \eta_2)$$

$$L_2 = \prod_{i=1}^{n} f(O_i; X_i, \beta) \prod_{i=n+1}^{N} h(E_{1i}, E_{2i}; X_i, \beta, \eta_1, \eta_2)$$

where $h(E_1, E_2; \beta, \eta_1, \eta_2) = \int g(E_1|O; \eta_1) g(E_2|O; \eta_2) f(O; X, \beta) \mathrm{d}O$

The likelihood function is given by $L(\beta, \eta_1, \eta_2; O, E_1, E_2, X) = L_1 \times L_2$.

The main assumption here is that the expert opinions, conditional on the true state of nature, are independent. Of course, marginally they will not be independent. We can assess the usefulness of the expert using the Fisher information as discussed previously. When combining experts, we pay the price of estimating more nuisance parameters and so unless there is a substantial and supplementary effect of the two experts, the combined expert need not always be a better expert than a single expert. Just as (weighted) averaging of two different estimators of the same parameter leads to a better estimator only under certain conditions; combining two experts leads to a better estimator only under certain conditions. These conditions are hard to specify in precise form except for simple cases. Essentially they relate to the amount of information about the parameter of interest in the conditional distribution of $E_2 \mid E_1$. We do not attempt to do it in this paper.

In the following, we will analyze one simple but instructive particular case. This will illustrate the conditions under which using the expert guesses may or may not reduce the variance of the estimator. It also illustrates an interesting phenomenon that even an expert whose opinions are negatively correlated with the true observations is useful. Thus, any association between the expert opinion and the true state of nature is information. Aside from this instructive case, we also introduce the model that will be used to analyze the biological dataset described in Section 2.

6.2. *Some particular cases*

6.2.1. Normal–Normal case. Let us consider a simple model where $O_i \sim N(\mu, \sigma^2)$ for $i = 1, 2, \ldots, n$ and $E_i \mid O_i \sim N(\eta O_i, 1)$ for $i = 1, 2, \ldots, N$. It then follows that the marginal distribution of the expert guesses is given by $E_i \sim N(\eta\mu, 1 + \eta^2\sigma^2)$.

Let us start with assuming known σ^2 and η. The score functions for μ with and without the expert opinion are respectively given by

$$\text{(a)} \sum_{i=1}^{n} (O_i - \mu) = 0 \quad \text{and} \quad \text{(b)} \sum_{i=1}^{n} (O_i - \mu) + \frac{\eta}{1 + \eta^2\sigma^2} \sum_{i=n+1}^{N} (E_i - \eta\mu) = 0$$

The corresponding Fisher information is given by

$$\text{(a)}\ I_n(\mu) = \frac{n}{\sigma^2} \quad \text{and} \quad \text{(b)}\ I_N(\mu) = \frac{\left(n + \frac{(N-n)\eta^2}{(1+\eta^2\sigma^2)}\right)^2}{\left(n\sigma^2 + \frac{(N-n)\eta^2}{(1+\eta^2\sigma^2)}\right)}$$

We note some interesting features below.

(i) If $N = n$, then $I_N(\mu) = I_n(\mu)$ as one would expect.
(ii) If the parameter $\eta = 0$, $I_N(\mu) = I_n(\mu)$. That is, one neither gains nor loses by using the expert guesses that are uncorrelated with the actual phenomenon. Of course, this is the case only if we know the value of η. In practice, we need to estimate this parameter and hence if the expert guess is, in fact, uncorrelated with the observations, we lose by using such an expert. An expert who adds noise to the hard data can deteriorate the inferences.
(iii) The Fisher information based on the expert guesses and observed data is a symmetric function of the parameter η. Thus any association between the expert's guesses and observed data is useful. Even a negatively correlated expert guess is useful.

Now let us study the exact conditions under which the expert guesses increase the Fisher information about the parameter μ.

Let $N = Kn$ and assume that $K > 1$. Consider the ratio of the two Fisher informations:

$$\frac{I_N(\mu)}{I_n(\mu)} = \frac{\left(n + \frac{n(K-1)\eta^2}{(1+\eta^2\sigma^2)}\right)^2}{\left(n\sigma^2 + \frac{n(K-1)\eta^2}{(1+\eta^2\sigma^2)}\right)} \cdot \frac{\sigma^2}{n}$$

For further algebraic simplicity, let us assume that $\eta = 1$. Then this ratio reduces to

$$\frac{I_N(\mu)}{I_n(\mu)} = \frac{\left(1 + \frac{(K-1)}{(1+\sigma^2)}\right)^2}{\left(1 + \frac{(K-1)}{(1+\sigma^2)\sigma^2}\right)}$$

This ratio is not always larger than one. For example, suppose $\sigma^2 = 0$, that is, the original observations O_i's provide perfect knowledge about the value of μ. Clearly, if we 'contaminate' these perfect observations with imperfect observations, the expert guesses E_i's that have variance one, we cannot expect the resultant estimator of μ to be better. Of course, this is an extreme example. On the other hand, suppose $\sigma^2 = 1$, then any $K > 1$ provides improvement; larger the K, larger is the improvement. For a given value of σ^2, we can determine the number of expert guesses, the value of N, which will ensure that the Fisher information is larger when we use them than when we use only the observed data. In Figure 1, we plot the values of K, as a function of σ^2, that ensure increase in the Fisher information. For example, suppose $\sigma^2 = 0.2$, this plot suggests that we need at least $N = 5.8n$ expert guesses in order to gain information above and beyond what is provided by the hard data. That is, if we can get expert guesses for $N > 5.8n$ values (and the assumed model is correct), we can improve the estimator of μ, otherwise we may, in fact, deteriorate the original estimator. Expert opinion does not (neither can we expect to) always improve the situation. If the expert ends up adding more noise than signal to the information at hand, then he/she can be harmful instead of being helpful.

Now instead of known η, we consider estimating this parameter using the first n observations. We assume that the parameter $\sigma^2 = 1$ for the sake of exposition. In practice, we will need to estimate that parameter as well.

Step 1: Estimate the parameter η using the first component of the full likelihood, namely,

$$\prod_{i=1}^{n} g(E_i \mid O_i, \eta)$$

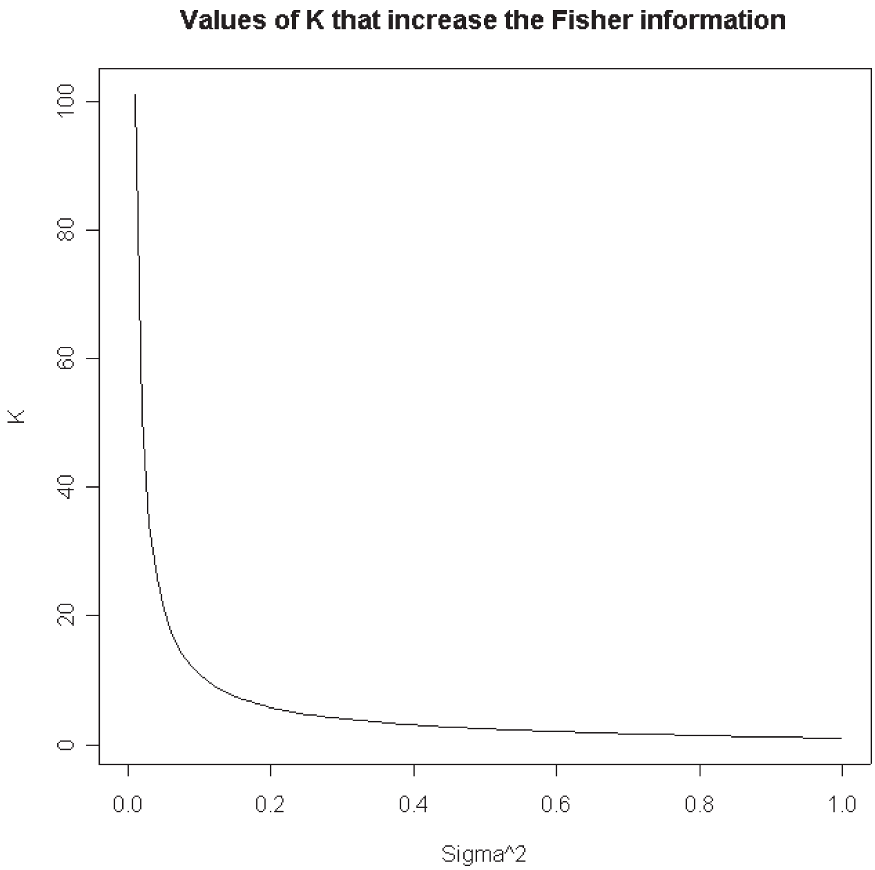

Figure 1. Sample size required for guaranteeing increase in the Fisher information. For example, when $\sigma^2 = 0.2$, we need at least $N = 5.8n$ number of expert guesses to guarantee an increase. Thus, if $n = 10$, we need $N \geq 59$, to assure that expert guesses will increase the accuracy of the estimator. If $N \leq 58$ then, in fact, using of expert guesses will deteriorate the original estimator based on the observed data of size $n = 10$

In the Normal–Normal case we are studying, this estimator can be written explicitly as

$$\hat{\eta} = \frac{\sum_{i=1}^{n} E_i O_i}{\sum_{i=1}^{n} O_i^2}$$

Step 2: Fix the value of $\eta = \hat{\eta}$ and maximize the last two components of the likelihood, namely,

$$\prod_{i=1}^{n} f(O_i; X_i, \beta) \prod_{i=n+1}^{N} h(E_i; X_i, \beta, \hat{\eta})$$

with respect to β.

This corresponds to the pseudo-likelihood estimator as described in Gong and Samaniego (1981). In the Normal–Normal case, it is easy to see that the pseudo-likelihood estimator of μ is given by solving the equation

$$\sum_{i=1}^{n}(O_i - \mu) + \frac{\hat{\eta}}{1+\hat{\eta}^2}\sum_{i=n+1}^{N}(E_i - \hat{\eta}\mu) = 0.$$

This can be explicitly written as

$$\hat{\mu}_N(\hat{\eta}) = \frac{\sum_{i=1}^{n} O_i + \frac{\hat{\eta}}{1+\hat{\eta}^2}\sum_{i=n+1}^{N} E_i}{n + \frac{(N-n)\hat{\eta}^2}{(1+\hat{\eta}^2)}}.$$

Application of the Slutsky's theorem (Serfling, 1980) along with the consistency of $\hat{\eta}$, shows that this estimator is consistent for μ as $n \to \infty$ and $N = Kn$ for some *fixed*K. The finite sample properties such as unbiasedness or mean squared error seem difficult to compute analytically. Whether and by how much this estimator is better than the estimator based only on the observed data depends on the variance of the estimator of η and the value of K. Smaller the variance and larger the value of K, larger are the gains. If the variance is large and K is small, we may, in fact, do worse by using this estimator. In that case, the expert opinion is adding noise to the hard data. In such a case, we are better off not using the expert opinion.

This model also illustrates the fact that one cannot collect only the expert opinion and obtain lots of information about the parameter of interest without any real data collection. For the model considered above notice that the marginal distribution $E_i \sim N(\eta\mu, 1+\eta^2\sigma^2)$ does not allow identification of the parameters (η, μ, σ^2). Thus, given only the expert guesses, one can obtain no information on the parameters of interest, namely (μ, σ^2), let aside estimating them well.

6.2.2. Bernoulli–Bernoulli case. We now describe the model that will be used to conduct the real data analysis. In this case, both the observations and the expert guesses are binary. We also have covariates corresponding to each, sampled as well as not sampled, location.

$$O_i \sim \text{Bernoulli}(p_i) \text{ where } \log\frac{p_i}{1-p_i} = X_i^{\mathrm{T}}\beta$$

$$E_i \mid O_i \sim \text{Bernoulli}(\pi_i) \text{ where } \log\frac{\pi_i}{1-\pi_i} = \eta_0 + \eta_1 O_i$$

The marginal distribution of E_i is given by:

$$P(E_i = 1) = P(E_i = 1 \mid O_i = 0)P(O_i = 0) + P(E_i = 1 \mid O_i = 1)P(O_i = 1)$$

The parameters η_0 and η_1 inform us about the characteristics of the expert. Clearly if $\eta_1 = 0$, expert guesses are unrelated to the actual state of the process. Such an expert cannot, in general, be useful. The usefulness of the expert in the binary data case depends on how often the expert guesses 0's as 0's and 1's as 1's. This corresponds to the sensitivity and specificity of the expert guesses. Of course, not only the values of these parameters but also the accuracy of the estimator of these parameters as well as the accuracy of the estimator based only on the observed data play a role in the actual determination of the 'usefulness' of the expert as discussed previously. We will see an example of this in the Section 7.

One can use the maximum-likelihood estimators for obtaining the estimator of the regression parameters β using only the observed data. These can be obtained using any of the standard programs. The pseudo-likelihood estimator of the regression parameter β can be obtained by first estimating the parameter η using the n observations, where we have the information on O_i's as well as E_i's. This is simply the maximum-likelihood estimation using the Logistic Regression model. Then one can maximize the following function with respect to β, with $\eta = \hat{\eta}$.

$$L(\beta; \hat{\eta}, O, E, X) = \prod_{i=1}^{n} f(O_i \mid X_i, \beta) \prod_{i=n+1}^{N} h(E_i \mid X_i, \beta, \hat{\eta})$$

According to our experience, pseudo-likelihood estimator seems numerically easier to obtain than the maximum-likelihood estimator, especially if the number of covariates is substantial. This is also indicated in Gong and Samaniego (1981).

To decide if and by how much the expert opinion is useful, one can compute the Fisher information with and without expert opinion. Alternatively, one can compute the parametric bootstrap-based standard errors or confidence intervals and compare them. If the expert opinion leads to a reduction in the standard error or reduction in the lengths of the confidence intervals, it should be deemed useful. In the following analysis, we will use the length of the bootstrap confidence interval to measure the usefulness of the experts. We use simple parametric bootstrap percentile confidence intervals (Efron and Tibshirani, 1994). These confidence intervals are known to be valid, that is, they have proper coverage properties. Hence we do not compare the coverage but their lengths. We use the parametric bootstrap confidence intervals instead of the Fisher information because they seem more relevant in finite sample situations. In the limit, of course, they reflect the comparison of Fisher information.

7. ANALYSIS OF THE SPECIES OCCURRENCE EXAMPLE

Recall that we have a total of 158 locations out of which 56 were sampled. In the following, we assume that at the sampled locations, the presence of a species is an actual presence (no misclassification) and absence is a true absence. The method to deal with the measurement error in these observations will be discussed in the next section. The covariates, elevation, precipitation, latitude, and longitude are known at all 158 locations.

We recruited three experts to obtain the expert guesses. Of our three experts, two were university faculty members specializing in small mammals in the region. The third 'expert' was a model called GAP that utilizes various layers of biological information for the state (and surrounding watersheds) and an independent database of small mammals. Kristie Allen, a graduate student in the Department of Biology at The Montana State University, conducted the interviews with these experts and also obtained the predictions from the GAP model. With a Master's degree in Statistics, her training in statistics was adequate for the field of biology, however she had had no special training in eliciting information. For each of the 158 locations, she asked the expert to provide a prediction: Is a particular species of shrews present or absent at that location? The question was easy to convey and easily understood by non-statisticians such as these ecologists and a computer program as well. The ease of communication was due largely to the fact that the question is posed on the same scale as the underlying ecological question. The experts were provided with the covariates and asked to predict the presence of the masked shrew on all 158 sites. We used the Binary-Binary model described in the previous section to conduct the analyses.

Initially standard analysis of the data from the 58 sampled locations was conducted. For each of the species, standard logistic regression analysis was conducted using the 'glm' command in R software (Ihaka and Gentleman, 1996). The relevant covariates were chosen using the stepwise Akaike information criterion ('stepAIC' function in R program with MASS library attached). For none of the species, latitude was an important covariate. We used parametric bootstrap (1000 bootstrap samples) with the percentile method to obtain confidence intervals for each of the parameters. To study the usefulness of the experts, we then augmented the sampled data with the expert guesses. Estimation of the parameters was conducted using the pseudo-likelihood method described in the previous section. We used 'optim' function with Nelder–Mead algorithm to maximize the pseudo-likelihood function. In Tables 1–4, we describe the results. In these tables, we report the estimates of the parameters η and β, and the ratio of the length of the confidence intervals for β based on the expert opinion and the one based only on the sampled locations. Smaller the ratio, shorter the interval when expert opinion is used and hence more useful is the expert.

Following major points may be noted:

(1) *Not all experts are equally useful*: It is clear that Expert A is, in general, most useful because it reduces the length of the confidence interval the most. It seems that Expert B adds almost nothing to the information (except Species 4) whereas Expert C is somewhere between the other two experts. This can be determined solely based on the statistical analysis without recourse to subjective opinions based on fame, experience and so on. To complete the story, after the analysis, we wanted to find out who the experts were and to see if the statistical analysis of their expertness made sense in reality or not. As it turns out, Expert A is a well known researcher in the field; Expert B, on the other hand, is more of a Geographic Information Systems expert rather than a field expert.

Table 1. Analysis of *Sorex Cinereus* with and without the expert opinion

Covariate	GLM	Expert A	Expert B	Expert C
Reliability				
$0 \to 0$	—	0.7235	0.90	0.96
$1 \to 1$		0.6667	0	0.22
Intercept	−71.051	−56.5366	−72.7998	−62.009
	(−168.73, 37.74)	**(0.7402)**	**(0.9549)**	**(0.8804)**
Longitude	−0.5970	−0.4693	−0.608	−0.5192
	(−1.43, − 0.30)	**(0.7305)**	**(0.9561)**	**(0.8800)**
Elevation	0.000634	0.0005722	−0.000693	−0.0005077
	(−0.000259, 0.00185)	**(0.7394)**	**(0.9590)**	**(0.7577)**

The number in bold indicate the ratio of the length of the confidence interval with expert opinion to that without expert opinion. Expert A is the most useful, Expert B is almost ineffective, and Expert C is somewhere in between. Parameter estimates do seem to change with or without the expert opinion.

Table 2. Analysis of *Sorex Vagrans* with and without the expert opinion

Covariate	GLM	Expert A	Expert B	Expert C
Reliability				
$0 \rightarrow 0$	—	1	0.865	0.946
$1 \rightarrow 1$		0.3158	0.05	0.3158
Intercept	−49.817	−19.1620	−48.5453	−34.0188
	−104.856	**0.4900**	**1.0136**	**0.7553**
	−26.955			
Longitude	−0.4339	−0.1611	−0.4206	−0.2903
	−0.9237	**0.4915**	**1.0057**	**0.7526**
	−0.2235			
Precipitation	0.07654	0.05662	0.0813	0.06324
	0.01978	**0.6929**	**1.1347**	**0.7595**
	0.1990			

The number in bold indicate the ratio of the length of the confidence interval with expert opinion to that without expert opinion. Smaller the number more is the gain. Expert A is substantially more useful for this species than *Sorex Cinereus*, Expert B is ineffective, almost detrimental, and Expert C is somewhere in between. Parameter estimates do seem to change quite a bit with the expert opinion.

Expert C is the GAP analysis program. It is satisfying to know that this agrees well with what one finds out from the statistical analysis.

(2) *An expert may not be useful in every situation*: An expert who is useful for one species need not always be useful for another species. For example, Expert A is much more useful for *Sorex Cinereus* and *Sorex Vagrans*, but not so useful for *Sorex Montanus*. On the other hand, Expert B is quite a bit more useful for species *Sorex Montanus* as compared to other species.

(3) *A not so useful expert does not affect the analysis adversely*: At least when n is reasonable, that is when we can calibrate the expert well, a not useful expert does not seem to affect the results adversely. This is remarkable and quite useful from the practical point of view. Probably an analogous result in the subjective Bayesian analysis is that the effect of prior goes away as the sample size increases.

(4) *Same value of η may lead to different amount of usefulness*: Notice that for Species 2 and 3, for expert A the values of η are quite similar. However this expert was very useful for Species 2 but not useful for Species 3. This is because the number of covariates for Species 3 is only one,

Table 3. Analysis of *Clethrionomys gapperi* with and without the expert opinion

Covariate	GLM	Expert A	Expert B	Expert C
Reliability η	—	0.8905	0.730	0.8685
		0.3889	0.111	0.2778
Intercept	−1.3930	−1.2608	−1.4804	−1.4650
	(2.417, −0.689)	**0.9988**	**1.1055**	**1.0132**
Precipitation	0.03336	0.03275	0.03332	0.03609
	(0.0020, 0.087)	**0.9263**	**0.9888**	**0.9615**

The number in bold indicate the ratio of the length of the confidence interval with expert opinion to that without expert opinion. None of the experts seem to make much difference in the inference. This is probably because we only have one parameter in this model and the estimates based on the observed data are already quite good.

Table 4. Analysis of *Sorex Montanus* with and without the expert opinion

Covariate	GLM	Expert A	Expert B	Expert C
Reliability η	—	0.86	0.92	0.72
		0.1667	0.1667	0.1667
Intercept	−61.8461	−61.7046	−68.6299	−56.18056
	(−192.43, 22.70)	**0.8232**	**0.7930**	**0.9494**
Longitude	−0.5517	−0.5507	−0.6148	−0.4996
	(−1.748, −0.199)	**0.8154**	**0.7781**	**0.9447**
Precipitation	−0.08173395	−0.0822	−0.0909	−0.0719
	(−0.403, −0.026)	**0.8671**	**0.9224**	**0.9994**

Notice that Expert B is the best of all the experts for this species. This is in contrast to the performance for other species where generally Experts A and C were better.

precipitation. The original estimates based only on the observed data are quite good. Recall from our simple Normal–Normal example in Section 6 that it is not just the value of η that determines the usefulness of an expert but also the accuracy with which observed data based estimators are obtained. It is difficult to improve upon a highly accurate estimator as compared to not so highly accurate estimator. For Species 2, we have two covariates. The estimators based on the observed data were not very accurate and hence the same value of η led to a substantial improvement. Recall also that the improvement is not only a function of η and N, but is also a function of the design matrix $\mathbf{X}$ which determines the accuracy of the estimator obtained using the observed data as well as the expert guesses. See Lele and Das (2000) for an analysis in the context of linear model with Normal errors.

We next consider the possibility of combining expert opinions to see if such combination leads to a 'super-expert'. In Table 5, we present the results of combining two experts for analyzing Species 1 data. It is clear from this table that one can either gain or lose as a consequence of combining experts. Clearly if the experts are highly correlated with each other, they do not add substantial information to what is available from one of them. On the other hand, if they are useful experts, we expect them to be correlated however, if they are useful experts but not perfectly correlated, sometimes they may help each other out. On the other hand, if they cancel each others' strengths, we may, in fact, lead to a worse expert than any one of them individually. This is borne out in the analysis quite clearly. Expert A and C seem to help each other out, whereas Expert B is pulling others down.

A careful look at the tables would reveal that for one of the species (*Sorex Vagrans*), the regression coefficient estimates using the expert opinion A, in particular, are quite different than the ones based

Table 5. Efficiency gain after combining two experts

Covariates	Experts (A + B)	Experts (A + C)	Expert (B + C)
Intercept	0.8291640	0.6974174	0.9524402
Longitude	0.8297462	0.6998052	0.9400040
Elevation	0.7274376	0.5907153	0.7584286

These results are only for *SorexCinerus*. Combining Experts A and C leads to shorter confidence intervals whereas combining Experts A and B or Experts C and B leads to deterioration of Experts A and C, respectively. Combining experts does not always lead to a better expert.

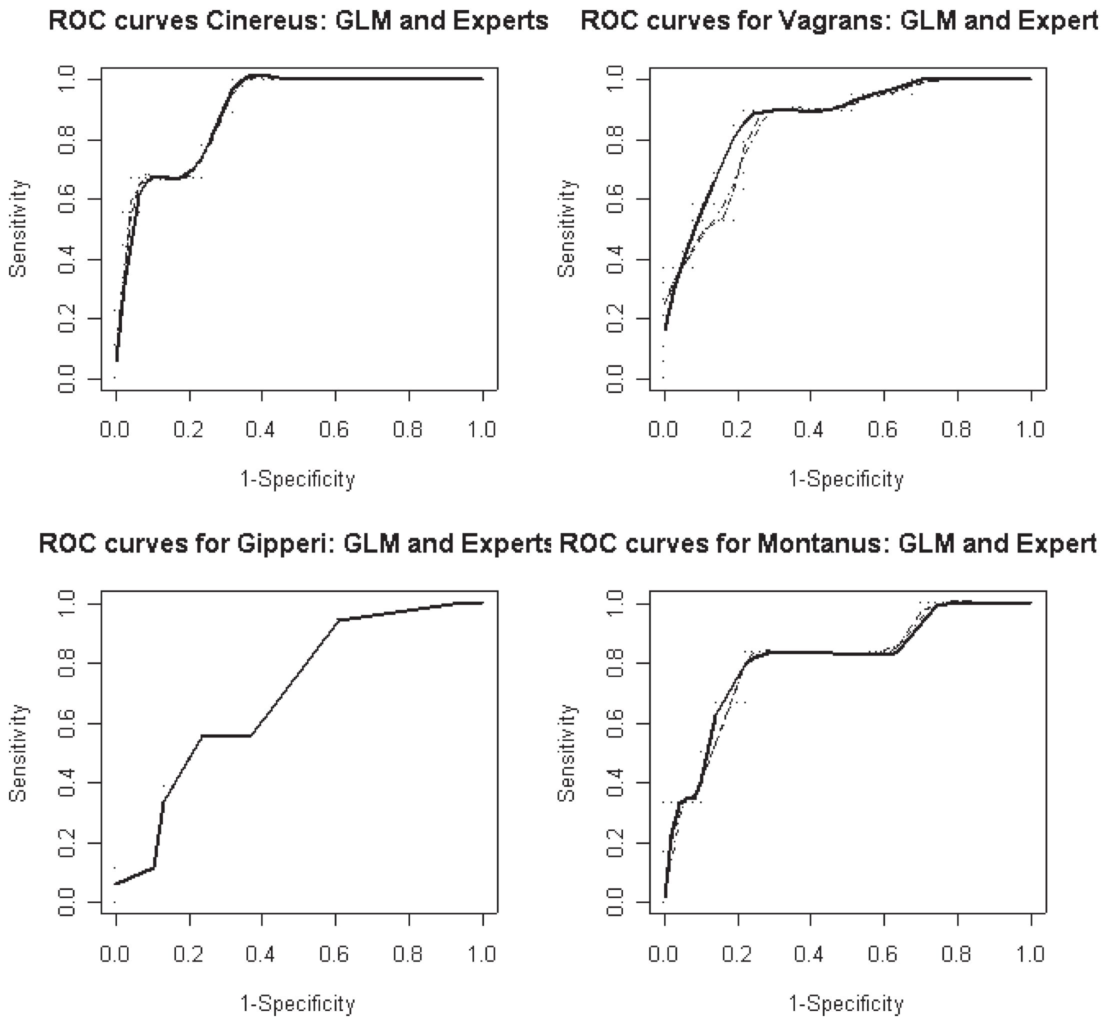

Figure 2. Comparison of ROC curves based on the fitted Logistic Regression models with only observed data and with the expert opinion. Although some of the fitted models have different regression coefficients, the predictive performances are comparable and the most useful expert generally has an ROC curve that dominates other ROC curves

only on the observed data. Although the confidence intervals are substantially smaller when Expert A guesses are used, one may feel uneasy about the substantial change in the estimates. For larger values of n, given the consistency of the estimators, we do not expect this to be a problem but for small to medium sample sizes this could happen. Which model should we choose to use in this case, one with the expert opinion or the one without the expert opinion? Since both are binary data models, one can compare their receiver operating characteristic curves (ROC). If the curves are comparable or if one with the expert opinion A always dominates the one with observed data only, we have additional confidence in the model estimated using the expert opinion. In Figure 2, we show the comparison of ROC curves for all four species, for three experts and the observation only based model. In all four cases, it is clear that the ROC curves are quite comparable and in fact, similar to the conclusion drawn from the confidence interval performance, the ROCs for the Expert A seem to dominate others, even in the case of *Sorex Vagrans*. Although ROC is not the only way to compare performance of two models for binary data, it is certainly a reasonable method. These comparisons indicate that the model based on the expert opinion, although its regression coefficient estimates are quite different, is a reasonable model.

A major concern with the presence-absence data is the possibility of false absences. Just because one did not observe a species does not always mean the species is not present. Similarly there is a real possibility of false identification of the species. The analysis presented thus far assumes that there is no measurement error in the data. A necessary condition for the analysis of data with measurement error is that one has replicate samples at each of the locations. Given these replicate samples, one can estimate the probability of misclassification and hence correct for it. In the dataset that we analyze, such replicate samples are unavailable and hence we could not take into account the measurement error. However, if such replicate samples were available, one can modify the likelihood function accordingly. (The details are algebraically complicated and hence are delegated to the Appendix.) One can then estimate the parameters by maximum-likelihood method. The pseudo-likelihood method does not seem to be applicable in this situation. The rest of the analysis as to the usefulness of an expert etc. can be carried out in a similar fashion.

8. RELATIONSHIP WITH OTHER ELICITATION PROPOSALS

Effective elicitation of expert opinion has been addressed in the Bayesian literature (Winkler, 1967, 1981; Savage, 1971; French, 1980, 1985, 1986; Kadane *et al.*, 1980; Dawid, 1982, 1987; Kahneman *et al.*, 1982; Lindley, 1983, 1985, 1988; Schervish, 1984; West, 1984, 1988, 1992; Genest and Schervish, 1985; West and Harrison, 1989, Chapter 15; West and Crosse, 1992; Chaloner, 1996; Kadane and Wolfson, 1998; Clemen and Reilly, 2001; Winkler, 2003; Tversky and Kahneman, 1974). A special issue of the *Journal of the Royal Statistical Society, Series A* (1998) was dedicated to the subject of elicitation of information from experts. These papers along with the associated discussion provide an excellent overview of the topic from the Bayesian perspective. In the following we summarize the main features of the Bayesian approach to elicitation and compare it with the approach discussed in the present paper.

(1) According to the schemes discussed in these papers, elicitation of the expert opinion is used to construct *priors* that are used in the Bayesian framework of modifying the *prior* to *posterior*. It was explicitly pointed out by Sir David Cox in the discussion of these papers (Cox, 1998) that the major lacuna in these approaches is that the expert opinion is not calibrated. As a consequence such expert opinion may get used even when it is not useful. The approach suggested in the present paper pointedly addresses the issue raised by Sir David Cox by providing precise rule for determining if the expert is useful or not and when to use the expert opinion.
(2) The papers by French (1980), Dawid (1982), DeGroot and Fienberg (1982) among others do explicitly try to study the concept of calibration of an expert. However, these papers and papers by Scheverish (1984), West and Crosse (1992) where an effort has been made to combine the expert opinion with hard data do not provide how this calibration could be used to obtain optimal weights in the combination of the expert opinion and the hard data. The approach suggested in the present paper provides a natural way to combine the expert information with the hard data by using the hierarchical model and the likelihood function.

(3) All of these approaches suggest that the expert opinion be elicited in terms of the probabilities of an event and not the prediction of the actual event. Thus, the hard data and the expert opinion are not on comparable scales. This makes their combination a difficult task. It is also unclear whether the elicited probabilities always satisfy the Kolmogorov axioms or not, and what to do if they are not satisfied. These probabilities are also unlikely to satisfy the transformation rules. For example, if one elicits the probability distribution of the random variable X and also from the same expert elicits probability distribution for a transformation of X, say $g(X)$, that these probability distributions are exact transformations of each other. On the other hand, eliciting data is likely to have less chance of not being transformation invariant because it is easier to think in terms of observables than probabilities.

In the light of the above discussion, elicitation of the expert opinion by eliciting guess values of the realization and its combination with the hard data using hierarchical model and likelihood function is, at the least, a feasible and attractive alternative to the elicitation methods proposed in the Bayesian literature.

9. CONCLUSION

This paper discusses the incorporation of expert opinion in statistical analysis in a frequentist fashion. One of the major difficulties, as acknowledged and explained in detail by various papers referred to in Section 8, with the incorporation of expert opinion lies in its elicitation. We suggest eliciting data rather than eliciting priors, again reflecting sentiments expressed by Kadane *et al.* (1980), Kadane and Wolfson (1998) among others. In the Bayesian approaches to elicitation of data, one either tries to construct priors based on the expert guesses or tries to elicit probability distributions for the observed values. On the other hand, in our proposal, we elicit expert information on the same scale as the observables and hence one can combine them with the observed data using the likelihood function directly. This method also allows for the incorporation of auxiliary information such as environmental habitat characteristics easily. Such use of the auxiliary information is not seen in the Bayesian elicitation approaches. We show that using the likelihood function and hierarchical model, one can then calibrate the expert opinion with the observed data and use even a negatively associated expert. We provide quantification for the ‘usefuleness’ of an expert in terms of gain in the Fisher information. One advantage of this is that one does not have to argue ‘expertness’ on the basis of experience, fame, or some such characteristics that may be open to debate. We illustrate that not all experts are alike; some are more useful than others. Combining experts need not always lead to a super expert. Experts may supplement each other or they may cancel each other out.

Eliciting data from the experts clearly involves additional work by the scientists. One has to obtain information from the expert by conducting interviews. It is not a free lunch. However, if this additional work is taken up and the expert is useful, the benefits can be substantial as shown in the analysis of the example. In our experience, obtaining this additional information is not a difficult task.

It is legitimate, although debatable, to claim that when an expert provides an opinion about presence or absence of a species, he/she has a probability in mind and not a binary outcome. One of the problems in using the probability judgment, as against the binary outcome, is the difficulty in calibration against the observed data. Since replication of the weather events or replication of the species presence-absence events at the same location and at the same time is impossible; it is difficult to calibrate such probability judgment. On the other hand, by eliciting information on the observable scale and using the regression model that relates elicited data to observed data, one can calibrate the expert opinion in the form of elicited data quite easily.

There is a close connection between the methodology suggested in this paper and the method of double sampling (Little and Rubin, 2002, Chapter 2). Suppose the characteristic of the population that is of scientific interest is difficult to measure but there is another characteristic that is easy to measure and is strongly related to the characteristic of interest. In ecological sampling, one can then survey a large area for the simple to measure characteristic and survey a subset of that area for the characteristic of interest as well. Then using the calibration techniques and the larger survey, one obtains a better estimate of the characteristic of interest than one would with only the subsample. The expert opinion, aside from measurement error introduced observation, can also be looked upon as 'an easy to obtain' information that is highly related to the characteristic of interest. This relationship opens up the possibility of using the mathematical and computational machinery available in the missing data literature for utilizing expert opinion.

The use of expert opinion in the form of elicited data also has something to offer back to the missing data literature. For example, missing data, due to censoring, or drop-outs are a common phenomenon in longitudinal data, panel data, or clinical trials. In many of these cases, it might be possible for the physician to provide his/her guesses of the outcome. If these are validated against the observed data, as outlined in this paper, they will help improve inferences over and above what is done using the standard missing data techniques. Essentially, we will be adding information through physician's opinions, albeit calibrated against the complete data.

The methodology discussed in this paper is clearly model dependent. It is imperative that the usual machinery for model selection and model diagnostics should be brought to bear on the problem. Such an analysis will point out an expert who might be selectively misleading; for example, existence of outliers will indicate such an expert. It will be interesting to see if semi-parametric methods such as estimating functions can be used instead of the likelihood function in the context of incorporation of elicited data. It will make this methodology somewhat model robust.

ACKNOWLEDGEMENTS

Subhash R. Lele was partially supported by a grant from the National Scientific and Engineering Research Council of Canada. We would also like to thank the experts Elizabeth Hadly, Rick Douglass, and the Montana GAP Analysis project, Wildlife Spatial Analysis Lab, The University of Montana, Missoula, MT 59812. Dennis Flath, Montana Department of Fish Wildlife and Parks collected the presence-absence data. We are grateful to Mark Taper, Chris Jerde, and Sam Scheiner for important discussions and comments. We are thankful to the editor and the referees for pointing us to the substantial literature on elicitation in the field of management science and statistics.

APPENDIX: CONSISTENCY AND ASYMPTOTIC NORMALITY OF THE PSEUDO-LIKELIHOOD ESTIMATORS

Notice that obtaining the pseudo-likelihood estimator corresponds to solving the following system of estimating function.

$$\frac{\partial}{\partial \eta} l_1(\eta, E_n, O_n) = 0$$

$$\frac{\partial}{\partial \beta} l_2(\beta; O_n, X_n) + \frac{\partial}{\partial \beta} l_3(\beta; E_{N-n}, X_{N-n}) = 0$$

Assuming the interchange of derivative and integration is allowed, the proof of zero unbiasedness is standard (Casella and Berger, 1990)

Proving the existence of a consistent root and asymptotic normality for this set of estimating functions depends on the application of Taylor series expansion, weak law of large numbers for independent but not identically distributed random variables and the Lindberg–Feller central limit theorem. Several different proofs with various conditions have been given in the literature in this context. See, for example, Bradley and Hart (1962) and Fahrmeir and Kaufmann (1985). Here we will use the proof given by Bradley and Hart (1962).

Assumptions

Let $\underline{\theta} = (\underline{\eta}, \underline{\beta})$ denote the full parameter vector of size $P \times 1$. Let us denote the parameter space by Θ. For notational simplicity, we ignore the underscore notation for the vector of a parameter. For all values of the parameter space, assume that the following derivatives exist for all r, s, $t = 1, 2, \dots, P$.

$$\frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_s}\log f(E_i \mid O_i, \eta), \frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_s}\log f(O_i \mid X_i, \beta), \frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_s}\log f(E_i \mid X_i, \beta, \eta)$$

$$\frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_s}\frac{\partial}{\partial\theta_t}\log f(E_i \mid X_i, \beta, \eta), \frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_s}\frac{\partial}{\partial\theta_t}\log f(O_i \mid X_i, \beta)$$

$$\frac{\partial}{\partial\theta_r}\frac{\partial}{\partial\theta_s}\frac{\partial}{\partial\theta_t}\log f(E_i \mid X_i, \beta, \eta)$$

Furthermore for all parameter values assume that

(1)
$$\lim_{n\to\infty}\frac{1}{n}\mathbf{E}\left(g_1 g_1^{\mathrm{T}}\right) \to \mathbf{J}_2(\theta), \lim_{n\to\infty}\frac{1}{n}\mathbf{E}\left(g_2 g_2^{\mathrm{T}}\right) \to \mathbf{J}_2(\theta)$$

where $\mathbf{J}_1(\theta), \mathbf{J}_2(\theta)$ are positive definite matrices. Notice that $\mathbf{E}(g_1 g_2^{\mathrm{T}}) = 0$. Also notice that the sum involved in the matrix $\mathbf{E}(g_2 g_2^{T})$ is of order $O(n)$ because of the assumption $N = Kn$ where K is fixed.

(2)
$$\lim_{n\to\infty}\frac{1}{n}\mathbf{E}(gg^{\mathrm{T}}) \to I(\theta),$$

a positive definite matrix. Notice that, considered as a partitioned matrix, this matrix is a lower triangular matrix with diagonal matrices positive definite under the standard-likelihood theory. The expectations of the third order derivatives are finite.

Given these conditions, one can mimic the proofs as given in Bradley and Hart (1962) to show consistency and asymptotic normality of the pseudo-likelihood estimators. We do not reproduce the steps here.

INCORPORATION OF MEASUREMENT ERROR

In order to deal with the presence of measurement error, one needs replicate observations at each of the locations. Let $Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{im})$ denote the replicate observations at the ith location. The data at hand now consist of $(E_1, E_2, \dots, E_N)$, the expert guesses, and $(Z_1, Z_2, \dots, Z_n)$ where $Z_i = (Z_{i1},$

$Z_{i2}, \ldots, Z_{im}$). In seems natural to assume that given the true state of nature O_i; E_i, and Z_i are independent of each other. Measurement error *process* has nothing to do with the *process* with which an expert guesses the true state of nature. Under this assumption, the following results are straightforward. We use $f(.)$ as a generic notation for the probability function, instead of a clumsy notation with different letters for each density.

$$f(E, Z; X, \beta, \eta, \alpha) = \int f(E \mid O) f(Z \mid O) f(O) \mathrm{d}O$$

$$f(E; X, \beta, \eta) = \int f(E \mid O) f(O) \mathrm{d}O$$

$$f(Z; X, \beta, \alpha) = \int f(Z \mid O) f(O) \mathrm{d}O$$

$$f(E \mid Z) = \frac{f(E, Z)}{f(Z)}$$

Given this, we can then write down the likelihood as follows:

$$L(\beta, \eta, \alpha; E, Z) = \prod_{i=1}^{n} f(E_i \mid Z_i; X_i, \alpha, \beta, \eta) \prod_{i=1}^{n} f(Z_i \mid X_i, \beta, \alpha) \prod_{i=n+1}^{N} f(E_i; X_i, \eta, \beta)$$

Because there is no suitable decomposition of the likelihood as a function of η alone, as it was available when there was no measurement error, one cannot use the pseudo-likelihood estimation method. One has to maximize the full likelihood with respect to all the parameters (β, η, α). One can possibly maximize the full likelihood iteratively as follows:

Step 1: Maximize the second component with respect to (α, β),
Step 2: Substitute these estimators in the likelihood function and maximize with respect to η.
Step 3: Substitute the estimator of η and maximize the full likelihood (not just the second component as in Step 1) with respect to (α, β).
Step 4: Iterate through Steps 2 and 3, until convergence.

Because the estimators in Step 1 and 2 are consistent, this iteration scheme should lead to consistent estimators, although we do not have formal proof for this result.

REFERENCES

Allen K, Weaver TW, Flath DL. 1997. Small mammal capture efficiencies among three trap types. *Intermountain Journal of Sciences* **3**(1): 1–6.
Barnett V. 1999. *Comparative Statistical Inference* (3rd edn). John Wiley and Sons: NY.
Berger JO. 1985. *Statistical Decision Theory*. Springer-Verlag: New York.
Bradley RA, Hart JJ. 1962. The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika* **49**: 205–214.
Carlin BP, Louis TA. 1996. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall/CRC Press: London, UK.
Casella G, Berger R. 1990. *Statistical Inference*. Brooks Cole: USA.

Chaloner K. 1996. The elicitation of prior distributions. In *Case studies in Bayesian Biostatistics*, Berry D, Stangl D (eds). Dekker: New York; 141–156.
Clemen RT, Reilly T. 2001. *Making Hard Decisions with Decision Tools*. Duxbury Thomson Learning: Pacific Grove, CA.
Collett D. 1991. *Modelling Binary Data*. Chapman and Hall/CRC Press: London, UK.
Cox DR. 1998. Discussion on the papers on 'Elicitation'. *The Statistician* **47**: 58–59.
Dawid AP. 1982. The well calibrated Bayesian. *Journal of the American Statistical Association* **77**: 605–610.
Dawid AP. 1987. The difficulty about conjuction. *The Statistician* **36**: 91–98.
DeGroot MH. 1988. A Bayesian view of assessing uncertainty and comparing expert opinion. *Journal of Statistical Planning and Inference* **20**: 295–306.
DeGroot MH, Fienberg SE. 1982. The comparison and evaluation of forecasters. *Journal of the Royal Statistical Society, Series A* **32**: 12–22.
Dennis B. 1996. Discussion: should ecologists become Bayesians? *Ecological Applications* **6**(4): 1095–1103.
Efron B. 1986. Why isn't everyone a Bayesian (with disucssion). *The American Statisticsan*. **40**(1): 1–11.
Efron B, Tibshirani R. 1994. *An introduction to bootstrap*. CRC Press: UK.
Fahrmeir L, Kaufmann H. 1985. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* **13**: 342–368.
Fleishman E, MacNally R, Fay JP, Murphy DD. 2001. Modelling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the great basin. *Conservation Biology* **15**: 1674–1685.
French S. 1980. Updating of belief in the light of someone else's opinion. *Journal of the Royal Statistical Society, Series A* **143**: 43–48.
French S. 1985. Group consensus probability distributions: a critical survey. In *Bayesian Statistics 2*, Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds). Amsterdam: North-Holland.
French S. 1986. Calibration and the expert problem. *Management Science* **32**: 315–321.
Garthwaite PH, Dickey JM. 1992. Elicitation of prior distributions for variable selection problems in regression. *Annals of Statistics* **20**: 1697–1719.
Genest C, Schervish MJ. 1985. Modeling expert judgements for Bayesian updating. *Annals of Statistics* **13**: 1198–1212.
Genest C, Zidek JV. 1986. Combining probability distributions: a critique and an annotated bibliography. *Statistical Science* **1**: 114–135.
Gong G, Samaniego FJ. 1981. Pseudo maximum likelihood estimation: theory and applications. *Annals of Statistics* **9**: 861–869.
Guisan A, Zimmerman NE. 2000. Predictive habitat distribution models in ecology. *Ecological Modeling* **135**: 147–186.
Hosmer DW, Lemeshow S. 2000. *Applied Logistic Regression* (2nd edn). John Wiley & Sons Inc.: New York, NY, USA.
Ihaka R, Gentleman R. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**: 299–314.
Journel AG. 1986. Constrained interpolation and qualitative information—the soft kriging approach. *Mathematical Geology* **18**(3): 269–286.
Kadane JB, Wolfson LJ. 1998. Experiences in elicitation. *The Statistician* **47**: 3–19.
Kadane JB, Dickey JM, Winkler RL, Smith WS, Peters SC. 1980. Interactive elicitation of expert opinion in a normal linear model. *Journal of the American Statistical Association* **75**: 845–854.
Kulkarni RB. 1984. Bayesian Kriging in Geotechnical Problems. In *Geostatistics for Natural Resources Characterization Part 2*, Verly G, David M, Journel AG, Marechal A (eds). D. Reidel Publishing Company: Dordrecht; 261–270.
Lele SR. 2004. Elicit data, not prior: on using expert opinion in ecological studies (with discussion). *The Nature of Scientific Evidence: Statistical, Philosophical and Empirical perspectives*, Taper ML, Lele SR (eds). University of Chicago Press: Chicago, Illinois, USA, 410–435.
Lele SR, Das A. 2000. Elicited data and incorporation of expert opinion for statistical inference in spatial studies. *Mathematical Geology* **32**(4): 465–487.
Lindley DV. 1983. Reconciliation of probability distributions. *Operations Research* **31**: 806–886.
Lindley DV. 1985. Reconcilliation of discrete probability distributions. In *Bayesian Statistics 2*, Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds). North-Holland: Amsterdam.
Lindley DV. 1988. The use of probability statements. In *Accelerated Life Tests and Experts' Opinion in Reliability*, Clarotti CA, Lindley DV (eds). North-Holland: Amsterdam.
Little RJA, Rubin DB. 2002. *Statistical Analysis with Missing Data* (2nd edn). Wiley-Interscience: USA.
Mayo D. 1996. *Error and the Growth of Experimental Knowledge*. University of Chicago Press: USA.
McCoy ED, Sutton PE, Mushinsky HR. 1999. The role of guesswork in conserving the threatened Sand Skink. *Conservation Biology* **13**(1): 190–194.
Nagorsen DW. 1996. *Opossums, Shrews and Moles of British Columbia*. Royal British Columbia Museum Handbook ISSN 118-5114, UBC Press: Vancouver.
Press SJ. 2002. *Subjective and Objective Bayesian Statistics: Principles, Models and Application*. John Wiley and Sons, NY, USA.
Royall RM. 1997. *Statistical Evidence: A Likelihood Primer.* Chapman & Hall. London. UK.
Savage LJ. 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* **66**: 783–801.

Schervish MJ. 1984. Combining expert judgements. Technical report 294, Department of Statistics, Carnegie Mellon University, Pittsburgh.
Scott JM, Davis F, Csuti B, *et al.* 1993. Gap analysis: a geographic approach to protection of biodiversity. *Wildlife Monographs* **123**: 1–41.
Scott JM, Heglund PJ, Morrison ML (eds). 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island press: Washington, DC, USA.
Serfling RJ. 1980. *Approximation Theorems of Mathematical Statistics*. Wiley-Interscience: USA.
Steffey D. 1992. Hierarchical Bayesian modeling with elicited prior information. *Communications in Statistics A* **21**: 799–821.
Taper ML, Lele SR (eds). 2004. *The Nature of Scientific Evidence*. University of Chicago Press: Chicago.
Tversky A, Kahneman D. 1974. Judgement under uncertainty: Heuristics and biases. *Science* **185**: 1124–1121.
West M. 1984. Bayesian aggregation. *Journal of the Royal Statistical Society, Series A* **147**: 600–607.
West M. 1988. Modelling expert opinion. In *Bayesian Statistics 3*, Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds). Oxford University Press: Oxford.
West M. 1992. Modeling agent forecast distributions. *Journal of the Royal Statistical Society, Series B* **54**: 553–567.
West M, Crosse J. 1992. Modelling probabilistic agent opinions. *Journal of the Royal Statistical Society, Series B* **54**: 285–299.
West M, Harrison PJ. 1989. *Bayesian Forecasting and Dynamic Models*. Springer Verlag: New York.
Wilson DE, Ruff S. 1999. *The Smithsonian Book of North American Mammals*. Smithsonian Institution Press: Washington and London.
Winkler RI. 1967. The quantification of judgement: some methodological suggestions. *Journal of the American Statistical Association* **62**: 776–800.
Winkler RL. 1981. Combining probability distributions from dependent information sources. *Management Science* **27**: 479–488.
Winkler RL. 2003. *An Introduction to Bayesian Inference and Decision* (2nd edn). Probabilistic Publishing: Gainesville.
Yang R, Berger J. 1997. A catalogue of noninformative priors. Institute of Statistics and Decision Science. Duke University Discussion papers 97–42.

