A CASE STUDY IN BAYESIAN SENSITIVITY:
FISH RESPONSE TO LAKE ACIDIFICATION

MICHAEL LAVINE

Institute of Statistics and Decision Sciences
Duke University

DP #89-15

Key Words and Phrases: Bayesian robustness, Bayesian sensitivity, binomial regression, logistic regression.
Abstract.

The purpose of this paper is to compare the importance of two sensitivity issues in a Bayesian analysis. The first issue is the choice of a probability distribution to represent prior uncertainty. Even assuming that there is a "true" distribution on the parameter space that represents our prior uncertainty about the "true" regression function, we are surely unable to describe that distribution with complete accuracy. Instead, for convenience, we often choose a distribution that is close to the true one in the hope that this approximation does not lead us too far astray.

The second issue is the use of a parametric family of regression functions to represent the set of all regression functions that we deem plausible. When we believe the true regression function to be close, in some appropriate sense, to a parametric family of regression functions, then, for convenience, we often restrict attention to the parametric family and ignore the other plausible functions in the hope that this simplification does not lead us too far astray.

Both issues have been treated separately elsewhere, the first in the Bayesian robustness literature, the second in the Bayesian nonparametric literature. Lavine (1988) provides a unified framework for dealing with them both. The current paper shows how to implement Lavine's ideas in a problem previously studied by Reckhow (1987, 1988), of a Bayesian analysis of fish response to lake acidification. For Adirondack lakes that were known previously to have supported brook trout, the probability \( p \) that a lake continues to support brook trout was modelled as a logistic function of the lake's pH and calcium. Prior opinions were elicited from an expert. Data were collected on a large number of lakes and a Bayesian analysis performed. For the purpose of this paper we simplify the original problem somewhat by ignoring calcium and modelling the probability that a lake still supports brook trout as
a function only of the lake's pH.

Our results in this example indicate that the second sensitivity issue can be much more important than the first, and that the usual Bayesian sensitivity analyses, by concentrating only on the first, are overlooking a potentially important source of posterior uncertainty.
1 Mathematical Framework.

This section describes how Lavine's (1988) general paradigm for studying Bayesian robustness applies to our particular problem. For lakes $i = 1, \ldots, n$ let $x_i$ and $y_i$ be respectively the pH of the $i$-th lake and a Bernoulli random variable indicating whether the $i$-th lake supports brook trout. Let $R$ be the set of all monotonically increasing regression functions:

$$R = \{ r : \mathbb{R} \to [0, 1] \}$$

where $r(x)$ is the probability that a lake of pH $x$ has brook trout, and $x_1 \geq x_2 \Rightarrow r(x_1) \geq r(x_2)$. Clearly, this model is not appropriate for all values of $x$. We apply it only to $x \in [3.5, 8]$, where it is reasonable. Let $R_B \subset R$ be the set of logistic regression functions,

$$R_B = \{ r_\beta \in R : \beta \in B \equiv \mathbb{R}^2; \ r_\beta(x) = 1/(1 + \exp(\beta_1(\beta_0 - x))) \}$$

and let $\pi_0^m$ be a probability measure on $B$.

We constructed $\pi_0^m$ from a fisheries expert's answers to questions like

"Given 100 lakes in the Adirondacks that have supported brook trout populations in the past, and if all 100 lakes have pH = 5.6 and calcium concentration = 130 $\mu$eq/L, what number do you now expect continue to support the brook trout population?"

For reasons of simplicity, and because both the expert and our preliminary analysis indicated that it was much less important than pH, we chose to ignore information about calcium.

Let $\alpha_0$ and $\alpha_1$ be i.i.d. on $[-1, 1]$ with density $(10 - |9\alpha|)/11$ and let $\beta_0 = \alpha_0 + k_0$ and $\beta_1 = k_1 \exp(\alpha_1)$, where $k_0 = 4.91413559$ and $k_1 = 4.653863$. The joint density of the $\alpha$'s is transformed into $\pi_0^m$. The constants $k_0$ and $k_1$ determine the mode of $\pi_0^m$ and come from a
linear regression of the noninfinite logits of the expert's responses as a function of pH. The rest of $\pi_0^m$ is chosen for convenience and to make Figure 1.1 look plausible.

Figure 1.1 shows the nine logistic curves corresponding to $\alpha_i \in \{-1, 0, 1\}$, $i = 1, 2$. The points on the plots are the expert's prior assessments. The interpretation of $\pi_0^m$ is that the curve in the center of the figure is given 10 times the prior weight of the ones directly above, below, to the right and to the left. And the center curve is given 100 times the weight of the curves in the corners.

A prior $\pi$ is a probability measure on $R$. Any such probability measure can be thought of as the $R$ marginal of a probability measure $\pi^* \times R$; which in turn can be thought of as the product of a marginal and a conditional, $\pi^* = \pi^m \times \rho^c$, where $\pi^m$ is a probability measure on $B$ and $\rho^c(\cdot | \beta)$ is a probability measure on $R$. Let $\pi_0^m(\cdot | \beta)$ be the degenerate measure that assigns mass 1 to $\rho_0$. Then the usual Bayesian analysis, treating $\pi_0^m$ as a "prior" on $B$, is equivalent to one that uses prior $\pi_0$, defined to be the $R$ marginal of $\pi_0^m$. $\pi_0 \equiv \pi_0^m \times \rho_0$.

We address the first sensitivity issue, the choice of $\pi_0^m$, by constructing a class $\Gamma^m$ of probability measures $\pi^m$ on $B$, where each $\pi^m \in \Gamma^m$ is close to $\pi_0^m$ in some appropriate sense. We address the second sensitivity issue with a class $\Gamma^c$ of probability measures $\rho^c$ on $R$, where each $\rho^c \in \Gamma^c$ is close to $\rho_0$. Then we measure sensitivity by computing upper and lower bounds on posterior predictive probabilities over all priors that are $R$ marginals of the measures in

$$\Gamma_1 = \{\pi^m \times \pi_0^c : \pi^m \in \Gamma^m\}$$

and again over priors that are $R$ marginals of

$$\Gamma_2 = \{\pi_0^m \times \rho^c : \rho^c \in \Gamma^c\}.$$ 

A more complete analysis might also consider $\Gamma_3 = \{\pi^m \times \rho^c : \pi^m \in \Gamma^m, \rho^c \in \Gamma^c\}$ but
our purpose here is to compare the two sensitivity issues rather than compute an overall sensitivity index.

Choosing $\Gamma^m$ and $\Gamma^e$ can be a difficult task. It has been suggested that the classes should be chosen to yield a reasonable range of prior predictions. Accordingly, we have chosen $\Gamma^m$ and $\Gamma^e$ to yield reasonable, and roughly equivalent, bounds on prior predictions.

For each $\beta$ let $\ell_\beta$ and $u_\beta$ be functions defined by

$$\ln(u_\beta(x)/(1-u_\beta(x))) = 1 + \ln(r_\beta(x)/(1-r_\beta(x)))$$

and

$$\ln(\ell_\beta(x)/(1-\ell_\beta(x))) = -1 + \ln(r_\beta(x)/(1-r_\beta(x))),$$

which imply

$$u_\beta(x) = c r_\beta(x)/(1 - r_\beta(x) + c r_\beta(x)) \quad \text{and} \quad \ell_\beta(x) = r_\beta(x)/(c - c r_\beta(x) + r_\beta(x)).$$

Define $N_\beta$, a neighborhood of $r_\beta$, by

$$N_\beta = \{r \in \mathbb{R} : \forall x \ell_\beta(x) \leq r(x) \leq u_\beta(x)\}$$

That is, $N_\beta$ is the set of all regression functions within plus or minus 1 of $r_\beta$ in log odds. Figure 1.2 shows $\ell_\beta$ and $u_\beta$ for two values of $\beta$.

Now define

$$\Gamma^m = \{\pi^m : (1/2)\pi^m_0 \leq \pi^m \leq 2\pi^m_0\}$$

and

$$\Gamma^e = \{\pi^e : \pi^e(N_\beta|\beta) = 1\}.$$ 

Figure 1.3 shows the expert's prior assessments, and the bounds on prior predictions implied by $\Gamma_1$ and $\Gamma_2$. The two classes of priors give very similar bounds on prior predictions. By
looking at the bounds on posterior predictions we can determine the relative importance of the two sensitivity issues, and whether either of them is important enough to be worrisome.

Note that under $\Gamma_2$, upper and lower bounds on posterior predictive probabilities are attained when $\pi^c$ is degenerate. In effect, we associate $\beta$ not with $r_\beta$, but with some other $r \in N_\beta$. Therefore, the integrated likelihood function

$$\text{lik}(\beta) = \int_{R} \text{Pr(Data|r)} \pi^c(dr|\beta)$$

becomes the conditional probability of the Data given the regression function associated with $\beta$.

2 Posterior Predictions.

The data are observations of pH and presence of brook trout on 73 Adirondack lakes. Figure 2.1 shows the data and the bounds on posterior predictions from the two types of prior classes. Obviously, posterior predictions are highly sensitive to slight changes in the regression function. To understand more clearly what is happening, look at Figures 2.2, 2.3, and 2.4. The first shows the usual likelihood function, using $\pi_0$ as the prior. The next two figures show the likelihood function when computing the lower and upper bounds on the predictive probability of success, using $\Gamma_2$, at pH = 5. They are quite different. To see why, we look at the two places where the likelihood function is maximized, at $\alpha' = (1.0, -0.2)$ and $\alpha'' = (-0.2, -1.0)$. Figure 2.5a shows the regression function used at $\alpha'$ for minimizing the predictive probability at $x = 5$. Figure 2.5b shows the regression function used at $\alpha'$ for maximizing that probability. Figure 2.6 shows the corresponding curves for $\alpha''$. The curve in Figure 2.5a has a much lower $r(5)$ and a much higher likelihood than the curve in Figure 2.6a. Similarly, the curve in Figure 2.5b has a much lower $r(5)$, and lower likelihood, than the one in Figure 2.6b.
We do not suggest that the priors that achieve the bounds are sensible priors, or that the regression functions in the support of those priors are sensible regression functions. However, the fact that the bounds are far apart suggests that inferences are very sensitive to the exact specification of the family of regression functions, and that the usual parametric analysis, based solely on $R_B$ may be missing a significant amount of posterior uncertainty. Figures like 2.5 and 2.6 can help to discover which features of the regression function are most important.

One possibility suggested by the figures is that the bounds may be extremely sensitive to the specification of the tails of the regression functions, where a small change in a probability can make an enormous change in the likelihood. We explore this possibility in the next section, first by trying a three parameter model which allows for the possibility of misclassification, and then by trying a new prior, one that is centered close to the maximum likelihood $\beta$.

3 Two Alternative Models.

The sensitivity that we see may arise because the prior and data set are in conflict with each other: the mle is far from the mode of $\pi_0^\infty$. To explore that possibility we tried two alternative models. First we tried extending the model with a parameter $\epsilon$, representing the probability that a lake with brook trout is correctly identified as such. Thus, if $p$ is the probability that a lake supports brook trout, then $\epsilon p$ is the probability that it is reported as supporting brook trout. We elicited a prior for $\epsilon$ from a fisheries expert. It is uniform on [.75,1]. Values of $\epsilon p$ are determined by the single prior for $\epsilon$ and by the classes of priors, $\Gamma_1$ and $\Gamma_2$, for $p$. Upper and lower bounds on both prior and posterior predictive probabilities $\epsilon p$ are shown in Figures 3.1 and 3.2. These figures indicate that the analysis is still much more sensitive to the regression function than to the choice of $\pi_0^\infty$. 

7
As a second alternative, we used the original two parameter model, but with a prior centered close to the mle. This prior does not represent the expert's opinion but does allow us to see what would have happened had the expert's opinion not been in conflict with the data. The new prior is obtained by taking the distribution for $\alpha$ as in Section 1 but redefining $\beta_0 = \alpha_0 + 6$ and $\beta_1 = \exp(\alpha_1)$. The new prior is centered at $\beta = (6, 1)$, which is close to the mle under $\pi_0$.

Ranges of prior and posterior predictive probabilities are shown in Figure 3.3 and Figure 3.4. Again, the regression function is more important than $\pi_0^\alpha$. The reason is that the likelihood function can vary enormously, even for functions within $N(6,1)$. Figure 3.5 shows two regression functions in that neighborhood. One regression function is chosen to maximize the likelihood function in that neighborhood, and yields a value of $\Pr[\text{Data} | r] = 2.5 \times 10^{-16}$. The other is chosen to give a small value of the likelihood, and has $\Pr[\text{Data} | r] = 5.9 \times 10^{-20}$. Obviously, any Bayesian analysis, and any analysis based on the likelihood function, will be very sensitive to the choice of $r$ from within each neighborhood.

Discussion

We have seen that posterior predictions can be very sensitive to the specification of the regression function, much more so than to the specification of a distribution on the parameter space. Therefore, we need to think more carefully about the simplification we make when using a parametric family of regression functions to represent all plausible regression functions. And we need to think about the evidence the data provide, focussing on issues like the evidence that an observation at $x_1$ can provide about the probability of success at $x_2$.

The same sort of problem arises in many statistical analyses. Consider, for example, a normal location problem. We initially think of the model $X|\theta \sim N(\theta, 1)$ but are not sure
that the data are truly normal. Around each normal density, indexed by $\theta$, we imagine a neighborhood, $N_\theta$, of "nearby" densities, and consider the set of inferences as we substitute other densities in $N_\theta$ for the normal.

Now suppose there has been one observation $x$, and consider how much information $x$ provides for distinguishing between two specific values $\theta_1$ and $\theta_2$. According to the normal model the information is given by the ratio $L = \exp(-(1/2)(x - \theta_1)^2) / \exp(-(1/2)(x - \theta_2)^2)$. If $x$ is in the tail of either or both distributions then $L$ is far from 1, and is very sensitive to the specification of the tail density. A small change in the tail density can make a big change in $L$, and in how we perceive the evidence distinguishing between $\theta_1$ and $\theta_2$.

But we do not really believe our tail specification to be an accurate representation of prior belief. Therefore, we must consider how much information $x$ really provides for distinguishing $\theta_1$ from $\theta_2$. The question can only be answered in the context of the specific inference problem at hand. In general, however, the answer may not correspond to the ratio of any pair of sampling densities. One possibility, that consistently underestimates the discriminatory power of the data, is to replace $f(x|\theta)$ with something that is like the normal density, or some other density, near $\theta$, but is constant, or falling very slowly, far from $\theta$. The function need not be a proper density.
6 Bibliography.


Michael Lavine
ISDS
322 Old Chemistry
Duke University
Durham, NC 27706
Figure 1.2

alpha = (1.0, -0.2)

two neighborhoods

alpha = (-0.2, -1.0)
Figure 2.2

likelihood function for the logistic curves
Figure 2.4

likelihood function for the maximizing curves
Figure 2.5b

minimizing regression function

Figure 2.5a

maximizing regression function
\[ \alpha = (1.0, -0.2) \]
Figure 2.6

Figure 2.6a

maximizing regression function

Figure 2.6b

minimizing regression function
alpha = (-0.2, -1.0)
Figure 3.1

Prior bounds on predictive probabilities: 3 parameters

Prob. of brook trout

pH

conditional class
marginal class
Figure 3.2

class

posterior bounds on predictive probabilities: 3 parameters
Figure 3.3

prior bounds on predictive probabilities: new prior mean

conditional class
marginal class

prob. of brook trout

pH