UCLA

Department of Statistics Papers

Title

A Graphical Case Statistic for Assessing Posterior Influence

Permalink

<https://escholarship.org/uc/item/8sf3w05c>

Authors

Robert E. Weiss

R. Dennis Cook

Publication Date

2011-10-24

A Graphical Case Statistic for Assessing Posterior Influence

Robert E. Weiss; R. Dennis Cook

Biometrika, Volume 79, Issue 1 (Mar., 1992), 51-55.

Stable URL:

<http://links.jstor.org/sici?sici=0006-3444%28199203%2979%3A1%3C51%3AAGCSFA%3E2.0.CO%3B2-R>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Biometrika is published by Biometrika Trust. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/bio.html>.

Biometrika

©1992 Biometrika Trust

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

A graphical case statistic for assessing posterior influence

BY ROBERT E. WEISS

*Department of Biostatistics, UCLA School of Public Health, Los Angeles,
California 90024-1772, U.S.A.*

AND R. DENNIS COOK

Department of Applied Statistics, University of Minnesota, St. Paul, Minnesota 55108, U.S.A.

SUMMARY

A single number summary of the influence of an observation on the posterior of the parameter of interest is useful, but does not convey all aspects of the effect of deleting the observation. We propose a plot that does convey all the effects of an observation on the posterior.

Some key words: Bayesian data analysis; Bayes's theorem; Dynamic graphics; Influential observations.

1. INTRODUCTION

We consider how to assess the influence of an observation on the posterior of the parameter of interest. Most previous work in a Bayesian context is based on a Kullback–Leibler divergence between the posteriors $p(\beta|Y)$ and $p(\beta|Y_{(i)})$ of the parameters β given the complete data Y and the reduced data $Y_{(i)}$ omitting the i th case (Hartigan, 1968; Pettit & Smith, 1985; Johnson & Geisser, 1985; Pettit, 1986) or between predictive densities (Johnson & Geisser, 1982, 1983; Johnson, 1985). The usual paradigm for the development of Bayesian and classical influence statistics is based on mapping relevant aspects of the full and reduced data to the real line. This basic paradigm has led to a wide variety of influence statistics since it is difficult to portray accurately all aspects of a high dimensional problem with a scalar.

Appropriate graphical displays may remove the practical necessity of restricting individual influence measures to one-dimensional summaries. In this paper we develop a plot that allows assessment of the influence of a single case on the posterior. It does not require specification of a particular influence statistic. Results are given that justify the claim that the suggested plot captures all the influence of a single observation.

The next section motivates the graphical influence statistic for exponential family models with a single parameter. Section 3 extends the graphical influence statistic to the regression framework and provides two optimality results. An example in § 4 illustrates the methodology and the paper closes with discussion.

2. EXPONENTIAL FAMILY INFLUENCE ASSESSMENT WITH A SINGLE PARAMETER

We assume that the observations $y_1, \dots, y_n$ are independent and y_i has sampling density

$$f(y_i|\nu_i(\tau_i)) = \exp \{ \nu_i(\tau_i)y_i - b(\tau_i) + d(y_i) \}, \quad (1)$$

where ν_i , the canonical parameter, is a known function of τ_i , a scalar parameter of interest. For example, τ_i could be the sampling mean of a normal or Poisson observation,

or the probability of success of a binomial. In this section we assume that $\tau_i \equiv \tau$ for all i . The next section relaxes this assumption. We wish to assess the influence of each observation in turn on the posterior distribution of τ .

The usual Bayesian method for assessing the influence of y_i compares $p(\tau|Y)$ and $p(\tau|Y_{(i)})$ using a divergence or distance measure $D\{p(\tau|Y), p(\tau|Y_{(i)})\}$ (Pettit, 1986), to map the influence of y_i into the nonnegative real numbers. Changes in the posterior mean, median, variance and skewness are not furnished by these measures, which can tell us that differences exist, but hide detail.

We are trying to compare the posteriors $p(\tau|Y)$ and $p(\tau|Y_{(i)})$. Since τ is scalar, a simple method is available to assess the effect of y_i : plot the two densities on a graph and compare them. All the effects of deleting a case will be clearly visible, any increase or decrease in median, increase in height at the mode, and changes in shape.

We propose these plots as influence statistics, calling them posterior influence plots. Using dynamic graphics one can comfortably scan many plots, more than enough for most data sets of small to moderate size. If necessary one can precompute the posteriors to facilitate rapid display. In large data sets, scanning one posterior influence plot for each observation may not be a practically useful procedure. In such situations, influence can be assessed by first using possibly several appropriate one-dimensional influence statistics to identify observations that are clearly uninfluential on all summaries. These observations are removed from consideration, leaving the posterior influence plots of the remaining observations for visual inspection. Such screening procedures are usually straightforward to implement.

As long as a scalar parameter or prediction is of interest, then this methodology extends immediately to problems with vector valued parameter. However, when the parameter of interest is of dimension greater than one, modification is needed. This methodology is developed in the next section.

3. INFLUENCE ASSESSMENT IN A REGRESSION FRAMEWORK

We assume a generalized regression model (McCullagh & Nelder, 1989), using model (1), but now $\tau_i = m(x_i, \beta)$ is a function of predictors x_i and regression coefficients β . In this framework we can again construct posterior influence plots of $p(\tau_i|Y)$ and $p(\tau_i|Y_{(i)})$. By construction, these plots display the influence of y_i on τ_i and are interpreted just as the plots of the previous section.

To assess the influence of y_i on the regression coefficients β we could follow the same procedure and construct a posterior influence plot of $p(\beta|Y)$ and $p(\beta|Y_{(i)})$. If the dimension of β is more than one or two, such plots are not possible. However, it turns out that, as demonstrated below, the posterior influence plot of $p(\tau_i|Y)$ and $p(\tau_i|Y_{(i)})$ is a summary of the posterior influence plot of $p(\beta|Y)$ and $p(\beta|Y_{(i)})$. We derive two versions of Bayes's theorem which show that, except for the marginal of τ_i , the posteriors $p(\beta|Y)$ and $p(\beta|Y_{(i)})$ contain no relevant information about the influence of y_i on β .

Bayes's theorem governing the effect of deleting the i th case on $p(\beta|Y)$ is well known,

$$p(\beta|Y) = \frac{p(\beta|Y_{(i)})f(y_i|\tau_i, x_i)}{\tilde{f}(y_i|Y_{(i)})}, \quad (2)$$

where $\tilde{f}(y_i|Y_{(i)})$ is the predictive density for the i th case given the remaining observations. Changing variables in (2) from β to (τ_i, ρ_i) , where ρ_i is chosen to make the transformation

one to one and measurable, and then integrating with respect to ρ_i on both sides of the equation yields

$$p(\tau_i | Y) = \frac{p(\tau_i | Y_{(i)})f(y_i | \tau_i, x_i)}{\int f(y_i | Y_{(i)})}. \quad (3)$$

The posterior $p(\tau_i | Y)$ is affected by the deletion of y_i in exactly the same fashion as the full parameter posterior $p(\beta | Y)$. Dividing (2) by (3) after the change of variables gives

$$p(\rho_i | \tau_i, Y) = p(\rho_i | \tau_i, Y_{(i)}). \quad (4)$$

The results in (3) and (4) show that the full influence of y_i is on the posterior of τ_i , and given τ_i , there is no more influence of y_i . In general, for any function $\theta(\beta)$,

$$p(\theta(\beta) | \tau_i, Y) = p(\theta(\beta) | \tau_i, Y_{(i)}).$$

Thus, conditional on τ_i , deleting the i th case has no effect on the posterior of any function of β .

It is of interest to compare the Kullback divergence between $p(\tau_i | Y)$ and $p(\tau_i | Y_{(i)})$ of the proposed plot to the divergence between $p(\beta | Y)$ and $p(\beta | Y_{(i)})$ and this will lead to a second optimality result. First we introduce a wide class of influence statistics. The divergences of Csiszár (1967) provide a class of potentially useful influence statistics that include the Kullback–Leibler divergences, functions of the Hellinger distances (Devroye & Györfi, 1985, p. 225) and the L_p norms. Define the g -influence of the i th case on $\theta(\beta)$ by

$$D_{\theta(\beta)}(g, i) = \int g \left\{ \frac{p(\theta | Y)}{p(\theta | Y_{(i)})} \right\} p(\theta | Y_{(i)}) d\theta. \quad (5)$$

Assume that g is convex and that $g(1) = 0$, which appear to be minimal assumptions for a sensible influence statistic. See Csiszár (1967) for properties of (5). When $\theta(\beta) = \tau_i$, $D_{\theta(\beta)}(g, i)$ will be written $D(g, i)$.

Consider $D_\beta(g, i)$, the divergence between $p(\beta | Y)$ and $p(\beta | Y_{(i)})$. Changing variables from β to (τ_i, ρ_i) , causes the posterior ratio $p(\tau_i, \rho_i | Y) \{p(\tau_i, \rho_i | Y_{(i)})\}^{-1}$ by (4) to not depend on ρ_i and gives

$$D_\beta(g, i) = D(g, i). \quad (6)$$

If an influence statistic of interest is a member of the class of g -influence measures, then comparing the posteriors $p(\tau_i | Y)$ and $p(\tau_i | Y_{(i)})$ is just as useful as comparing $p(\beta | Y)$ and $p(\beta | Y_{(i)})$. The g -influence is preserved in the change from β to τ_i .

If influence on a subset or function $\theta(\beta)$ is of interest then by convexity of g ,

$$0 \leq D_\theta(g, i) \leq D(g, i). \quad (7)$$

Looking at $p(\tau_i | Y)$ and $p(\tau_i | Y_{(i)})$ will lead one to believe that y_i is more influential than it really is, and the posterior influence plot can be used to determine that particular observations are uninfluential. When interest lies in a possibly multivariate prediction Z_f at predictors X_f , inequality (7) still holds with Z_f in place of $\theta(\beta)$.

4. EXAMPLE

This example uses the vasoconstriction data of Pregibon (1981), who identifies two influential observations, cases 4 and 18, as very influential. Figures 1(a) and 1(b) show posterior influence plots for case 4 and case 31, a case not identified as influential. The complete data marginals were computed using LISP-STAT (Tierney, 1990) with software

Fig. 1. Posterior influence plots for (a) case $i=4$ and (b) case $i=31$ of the vasoconstriction data. Solid line, full data posteriors, $p(\tau_i|Y)$; dashed line, reduced data posteriors, $p(\tau_i|Y_{(i)})$.

based on Tierney & Kadane (1986). The reduced data marginal was computed using (3). The appropriate marginal parameter is $\tau_i = \{1 + \exp(-x_i^T \beta)\}^{-1}$, the probability that an observation at covariate x_i is a 1.

In Fig. 1(a), the reduced data marginal has a mode essentially at zero whereas the complete data marginal, while skewed, peaks near $\tau_4 = 0.006$, and decreases towards zero near zero. Deleting case 4 moves approximately 50% of the posterior mass of τ_4 from between (0.1, 0.3) to less than 0.01. In contrast, the posteriors with and without case 31 have similar shapes but probability moves much farther, from between (0.7, 0.95) to (0, 0.5). Much more probability shifts when case 4 is deleted, but for case 31, the probability shifts across the more important boundary point of $\tau_{31} = 0.5$. In terms of the posterior means of τ_4 and τ_{31} , deleting case 31 shifts the posterior mean of τ_{31} more than deleting case 4 does to the posterior mean of τ_4 .

5. DISCUSSION

Posterior influence plots are not intended to replace one-dimensional summaries. Indeed we have found that it is often useful to display one or two numerical summaries along with $p(\tau_i|Y)$ and $p(\tau_i|Y_{(i)})$. In addition, one-dimensional summaries are needed to select plots for viewing when there are many cases.

The theory also applies if the model (1) is expanded to have an unknown scale parameter ϕ_i , provided only that τ_i is replaced everywhere by the pair (τ_i, ϕ_i) in (3), (4), (6) and (7). The posterior influence plot becomes a comparison of two bivariate distributions, $p(\tau_i, \phi_i|Y)$ and $p(\tau_i, \phi_i|Y_{(i)})$. Work is needed on how best to compare these two bivariate posteriors.

ACKNOWLEDGEMENTS

Much of this paper is taken from the first author's Ph.D. thesis under the guidance of the second author. The first author thanks Rod Little for comments.

REFERENCES

- CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica* **2**, 299–318.
- DEVROYE, L. & GYÖRFI, (1985). *Nonparametric Density Estimation, the L_1 View*. New York: Wiley.
- HARTIGAN, J. A. (1968). Note on discordant observations. *J. R. Statist. Soc. B* **30**, 545–50.
- JOHNSON, W. (1985). Influence measures for logistic regression: Another point of view. *Biometrika* **72**, 59–65.
- JOHNSON, W. & GEISSER, S. (1982). Assessing the predictive influence of observations. In *Statistics and Probability: Essays in Honor of C. R. Rao*, Ed. G. Kallianpur, P. R. Krishnaiah and J. K. Ghosh, pp. 343–58. Amsterdam: North-Holland.
- JOHNSON, W. & GEISSER, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis. *J. Am. Statist. Assoc.* **78**, 137–44.
- JOHNSON, W. & GEISSER, S. (1985). Estimative influence measures for the multivariate general linear model. *J. Statist. Plan. Inf.* **11**, 33–56.
- MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.
- PETTIT, L. I. (1986). Diagnostics in Bayesian model choice. *Statistician* **35**, 183–90.
- PETTIT, L. I. & SMITH, A. F. M. (1985). Outliers and influential observations in linear models. In *Bayesian Statistics 2*, Ed. J. M. Bernardo, M. DeGroot, D. Lindley and A. F. M. Smith, pp. 473–94. Amsterdam: North-Holland.
- PREGIBON, D. (1981). Logistic regression diagnostics. *Ann. Statist.* **9**, 705–24.
- TIERNEY, L. (1990). *LISP-STAT*. New York: Wiley.
- TIERNEY, L. & KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82–6.

[Received January 1991. Revised April 1991]