Influential Observations, Diagnostics and Discordancy Tests

by

Seymour Geisser[1]
University of Minnesota

Technical Report No. 463
January 1986

----

1. Research supported by National Institute of Health Grant No. GM25271.

Influential Observations, Diagnostics and Discordancy Tests

by

Seymour Geisser[1]
University of Minnesota

1. Introduction.

Since the advent of high speed computing it has become much easier and fashionable to quickly and efficiently calculate how individual observations or sets of them influence a statistical analysis. Much attention has been devoted to determining the effect on the estimation of parameters, especially with regard to regression analysis, Cook (1977, 1979), Cook and Weisberg (1980, 1982), Andrews and Pregibon (1978), Hoaglin and Welsch (1978), Belsley, Kuh and Welsch (1980). Johnson and Geisser (1985), Geisser (1985) and Smith and Pettit (1985) discuss parametric inference in the regression problem from a Bayesian viewpoint. The problem of assessing influence with regard to prediction from a Bayesian viewpoint was addressed by Johnson and Geisser (1982) and specifically with regard to regression by Johnson and Geisser (1983). A formal Bayesian decision framework for assessing the influence was briefly introduced by Geisser (1985). In this paper this Bayesian framework for determining the relative influence of an observation or a set of them on decision making or inference will be expanded and reviewed.

One formulation involves decisions or actions that depend on the "true" value of a set of parameters (or a subset of them) of the sampling distribution. A second involves the values of observables, as yet unavailable but that may be

----

1. Research supported by National Institute of Health Grant No. GM25271.

or have been generated from the process under study. In the particular case where inference is restricted to the distribution of a parameter or an observable, use is made of the Kullback-Liebler (1951) divergence as a measure of relative influence. Other diagnostics such as the conditional predictive ordinate Geisser (1980, 1985), Smith and Pettit (1985) will also be discussed.

When highly influential observations are identified, predictive discordancy tests, appropriately conditioned, are defined and used to assess the compatability of such observations with an assumed model. These conditional predictive discordancy tests are of a different nature than the usual frequentist discordancy tests. Some interesting special cases will be featured.

2. Estimative Influence.

From the Bayesian point of view a "formalization" for determining the influence of an observation (or set of them) on decision making or inference about a parameter Θ was briefly set fourth by Geisser (1985). In regard to decision making let $L(d(y),\theta)$ be the loss incurred by making decision d based on observables $y = (y_1,\ldots, y_N)$ at known covariates $x = (x_1,\ldots, x_N)$ when θ is the true value. To ascertain the influence of a particular observable y_i (y_i can also be a subset of observations) we first calculate d_i^* such that

$$\min_d E_\Theta\left[L(d(y_{(i)},\Theta)\right] = \bar{L}(d_i^*)$$

where the expectation is taken over the predictive distribution of Z whose density is

$$p_{(i)}(\theta) = p(\theta|y_{(i)},x_{(i)}) \propto \ell(\theta|y_{(i)},x_{(i)})g(\theta)$$

for $(y_{(i)},x_{(i)})$ being (y,x) with (y_i,x_i) deleted and $\ell(\theta|\cdot)$ and $g(\theta)$ being the likelihood and the prior density respectively. This is then compared to d^* obtained from

$$\min_d E_\Theta[L(d(y),\Theta)] = \bar{L}(d^*)$$

where the expectation is calculated over

$$p(\theta) = p(\theta|y,x) \propto \ell(\theta|y,x)g(\theta),$$

where $\ell(\theta|y,x)$ is the likelihood including y_i. The observables are then ranked according to the defined loss scale, say, $S(d_i^*,d^*)$ which measures how distant the decision d_i^*, made if observation y_i were excluded, is from d^* wherein y_i were included.

When one is primarily interested in reporting the posterior distribution of Θ, one may ask how the distribution of Θ is influenced by the inclusion of y_i. A useful all purpose scalar measure (loss function) is the Kullback-Liebler (1951) directed divergence between the two distributions, Johnson and Geisser (1985)

$$I_i(\theta) = E_\Theta[\ln p_{(i)}(\theta) - \ln p(\theta)]$$

where the expection is taken over $p_{(i)}(\theta)$ and the larger $I_i(\theta)$ the more influential is y_i.

3. Predictive Influence.

To determine the influence of an observation on making decision d(y) we let

$$L(d,z)$$

represent the predictive loss incurred in making decision d(y) when Y = y is observed and Z = z is a set of future realized values. Here we need to calculate the predictive distribution for z,

$$F(z|y,x) = E[F(z|y,\theta)]$$

where the expectation is over the posterior distribution of θ and d^* is such that

$$\min_d E[L(d,Z)] = \bar{L}(d^*)$$

where the expectation is over the predictive distribution of Z. A similar calculation is made for $d_i^*(y_{(i)})$ such that

$$\min_d E[L(d_i,Z)] = \bar{L}(d_i^*)$$

where the expectation is over the predictive distribution of Z calculated only from $y_{(i)}$. Again some measure $S(d_i^*, d_i)$ of the change in the loss function, if any, in decision d_i^* that would have been made if only $y_{(i)}$ were observed to decision d^* if $y = (y_i, y_{(i)})$ were observed, needs to be defined. The observations then could be ranked in order of influence according to S.

4. Influence Measures for Distributions.

When one is primarily interested in reporting the posterior distribution of θ or the predictive distribution of Z there are a number of potentially useful measures of the change induced by adding observation y_i to $y_{(i)}$. For example in the estimative and predictive modes respectively one can calculate

$$\sup_{\theta} |P_{(i)}(\theta|y_{(i)}) - P(\theta|y)| = H_i(\Theta)$$

$$\sup_{z} |F_{(i)}(z|y_{(i)}) - F(z|y)| = H_i(Z)$$

and rank the influence of the observations based on H_i, with the largest being the most influential. Another very useful measure is the Kullback-Liebler directed divergence

$$I_i(\Theta) = E[\ln p_{(i)}(\theta|y_{(i)}) - \ln p_{(i)}(\theta|y)]$$

Johnson and Geisser (1985), Geisser (1985) in the estimative mode and

$$I_i(Z) = E[\ln f_{(i)}(z|y_{(i)}) - \ln f(z|y)],$$

Johnson and Geisser (1982, 1983), Geisser (1985) in the predictive mode where the expectation is over the $p_i(\theta|y_{(i)})$ and $F_{(i)}(z|y_{(i)})$ respectively.

Attempts have been made to calibrate the Kullback distance, McCulloch (1985), using a binary variate or normal variate. For example for $n(x|\mu,\sigma^2)$, a normal density with mean μ and variance σ^2, set

$$I(\mu) = E\left[\ln \frac{n(x|0,1)}{n(x|\mu,1)}\right] = k.$$

then $\mu = \sqrt{2k}$ and one can calculate $1 - \Phi(\sqrt{2k})$ as a function of k which indicates how the probability change from .5 to 0 as a function of k. This may have some value in assessing the magnitude of the influence. One way to think about this is that a given value of k is calibrated with a change from even odds to odds of $(1 - \Phi\sqrt{2k})/\Phi(\sqrt{2k})$.

5. Predictive Discordancy Measures.

Another diagnostic, Geisser (1980),

$$d_i = f_{(i)}(y_i|y_{(i)})$$

called the Conditional Predictive Ordinate (CPO) ranks the discordancy of observations--the smaller the value of d_i the more influential it is, see also Smith and Pettit (1985). A companion diagnostic is essentially the tail area in many applications, although not always necessarily so,

$$P_i = \Pr[f_{(i)}(Z|y_{(i)}) \leq f_{(i)}(y_i|y_{(i)})].$$

Here the smaller P_i the more discordant y_i is from $y_{(i)}$. But observations can be very discordant and yet exert little influence. Conversely observations can be highly influential and be quite compatible. Other diagnostics that perhaps are chosen to reflect a discrepancy from what is expected under the model such as

$$y_i - E(y_i|y_{(i)}),\ y_i - \text{Mode}(y_i|y_{(i)}) \text{ or } y_i - \text{Median}(y_i|y_{(i)})$$

may also be used.

6. Conditional Predictive Discordancy (CPD) Tests.

Presumably if an observation y_i has been made under suspicious circumstances which are identifiable, P_i as given above can be considered a significance test for the discordancy of the observation. However when observations are being ransacked to find either those that are highly influential or have high discordancy values it would be more appropriate to take this into account in constructing a CPD test. Hence if observation y_i were identified on the basis of being the most influential or the most discordant, it would be wise for a test of the discordancy of y_i with $y_{(i)}$ on the basis of model M, to condition on how y_i was chosen. Hence a suggested test is to calculate, conditional on how y_C was selected,

$$P_C = \Pr[Z \in R_C|y_{(C)},M,C]$$

where R_C is a region dictated by the method of choice C which selected y_C. One implementation of the above is calculating

$$P_C = \Pr[f(Z|y_{(C)},M) \geq f(y_C|y_{(C)},M)|C].$$

We present a rather simple illustration. Let $Y_1,\ldots, Y_N$ be $N(\theta,1)$ and assume θ is $N(\beta,\tau^2)$ where β and τ^2 are presumed known. Then the predictive distribution of Z is $N(a,b^2)$ where

$$a = \frac{\tau^2 \bar{y} + \frac{1}{N}\beta}{\tau^2+N} \quad \text{and} \quad b^2 = 1 + \tau^2(N\tau^2+1)^{-1}.$$

Hence for Z based on $y_{(i)}$ rather than y we have $Z \sim N(a_i,b^2)$ where

$$a_i = \frac{\tau^2\bar{y}_{(i)} + \frac{1}{N-1}\beta}{\tau^2 + \frac{1}{N-1}}, \quad b^2 = 1 + \tau^2[1 + (N-1)\tau^2]^{-1} .$$

In this instance all the methods of choice C previously discussed select that y_i which maximizes

$$|y_i - a_i| \quad \text{or} \quad \left(\frac{y_i-a_i}{b}\right)^2 = v_i^2.$$

Since $v_C^2 \geq v_{C-1}^2$ where v_{C-1}^2 is the second largest and conditional on $y_{(i)}$

$$\left(\frac{Z - a_i}{b}\right)^2 \sim \chi_1^2$$

we further condition this on the fact that $V_C^2 \geq v_{C-1}^2$ and compute the discordancy significance test level as

$$\Pr[V_C^2 \geq v_C^2 | V_C^2 \geq v_{C-1}^2] = P_C.$$

or

$$P_C = \frac{1 - F(v_C^2)}{1 - F(v_{C-1}^2)}$$

where $F(\cdot)$ is the distribution function of a χ_1^2 random variable.

7. Multiple Regression.

Consider a normal linear regression situation where

$$\begin{aligned} Y &= X\beta + e, & e &\sim N(0,\sigma^2 I) \\ Y' &= (Y_1,\dots, Y_N), & e' &= (e_1,\dots, e_N) \\ x_i' &= (x_{i1},\dots, x_{ip}), & \beta' &= (\beta_1,\dots, \beta_p) \end{aligned}$$

and

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{pmatrix} = \begin{pmatrix} x_1' \\ \vdots \\ x_N' \end{pmatrix}$$

with assumed prior density for β and σ^2, say

$$g(\beta,\sigma^2).$$

The first step in assessing the influence of individual observations with regard to the estimation of β alone, say, is the computation of the posterior densities $p(\beta) = p(\beta|y,X)$ and $p_{(i)}(\beta) = p(\beta|y_{(i)},X_{(i)})$ where $X_{(i)}$ is X with the ith row deleted. Next we compute

$$I_i(\beta) = I(p_{(i)},p) = E[\ln p_{(i)}(\beta) - \ln p(\beta)].$$

Also it is easy to show that

$$I_i(\beta,\sigma^2) = I_i(\sigma^2) + E[I_i(\beta|\sigma^2)]$$

where

$$I_i(\sigma^2) = E[\ln p_{(i)}(\sigma^2) - \ln p(\sigma^2)],$$

and

$p_{(i)}(\sigma^2)$ and $p(\sigma^2)$ refer to $p_{(i)}(\sigma^2|y_{(i)},X_{(i)})$ and $p(\sigma^2|y,X)$ respectively; similarly

$$I_i(\beta|\sigma^2) = E[\ln p_{(i)}(\beta|\sigma^2) - \ln p(\beta|\sigma^2)]$$

and $I_i(\beta|\sigma^2)$ above is averaged over the density $p_{(i)}(\sigma^2)$. This partition often helps to pinpoint the sources of influence. Details of this approach with examples have been worked out for the multivariate general linear model by Johnson and Geisser (1985).

For prediction it is necessary to calculate the predictive distribution of Z, the m×1 future vector to be observed for a given W, an m×p matrix, i.e.

$$Z = W\beta + e^* \qquad\qquad e^* \sim N(0,\sigma^2 I)$$

with and without y_i. Consequently,

$$f_{(i)}(z) = f_{(i)}(z|W,y_{(i)},X_{(i)}) = \int f(z|W,\beta,\sigma^2)p_{(i)}(\beta,\sigma^2)d\beta d\sigma^2$$

$$f(z) = f(z|W,y,X) = \int f(z|W,\beta,\sigma^2)p(\beta,\sigma^2)d\beta d\sigma^2.$$

One then calculates

$$I_i(Z) = E(\ln f_{(i)}(Z) - \ln f(Z)].$$

If W is unknown but can be assigned probabilities then this can be incorporated into the assessment. If this is not the case, it has been found useful to set W = X, i.e. to essentially ascertain the effect of predicting back on the original set of independent variables as indicative of an overall assessment. The details of this procedure are given by Johnson and Geisser (1982, 1983).

For the purpose of demonstration we use the "non-informative" prior

$$g(\beta,\sigma^2) \propto \frac{1}{\sigma^2} .$$

Let x_i' be the ith row of X then define

$$v_i = x_i'(X'X)^{-1}x_i, \qquad (N-p)s^2 = (y-\hat{y})'(y-\hat{y}),$$
$$\hat{\beta} = (X'X)^{-1}X'y, \qquad \hat{y} = X\hat{\beta}, \qquad \hat{y}_i = x_i'\hat{\beta},$$
$$t_i^2 = \frac{(\hat{y}_i - y_i)^2}{(N-p)s^2(1-v_i)} .$$

Using these results we can calculate the various measures of influence previously defined. First we obtain $2I_i(\beta,\sigma^2)$ which is the sum of the following two expressions,

$$2I_i(\sigma^2) = C + (N-1-p)t_i^2(1-t_i^2)^{-1} + (N-p)\ln(1-t_i^2)$$

$$2E[I_i(\beta|\sigma^2)] = K + (N-1-p)\frac{v_i t_i^2}{(1-v_i)(1-t_i^2)} + \frac{v_i}{1-v_i} + \ln(1-v_i)$$

where C and K are constants independent of the deleted observation. Although an explicit expression for $I_i(\beta)$ is not obtainable, the following approximation, based on a "best" scaled multivariate normal approximation to a multivariate student distribution which minimizes the Kullback-Liebler divergence, Johnson and Geisser (1983), should be adequate

$$2\hat{I}_i(\beta) = \frac{(N-p-2)v_i}{1-v_i}t_i^2 + \ln(1-v_i) + p\left[\frac{N-p-2}{N-p-3} + \ln\frac{N-p-3}{N-p-2} - 1 - \ln(1-t_i^2) - \frac{t_i^2(N-p-2)}{N-p-3}\right]$$

$$- \frac{v_i}{1-v_i}\left[\frac{(N-p-2)(t_i^2-1)}{(N-p-3)}\right] .$$

For the predictive influence function a similar "best" multivariate normal approximation to a multivariate student distribution is utilized. This results in

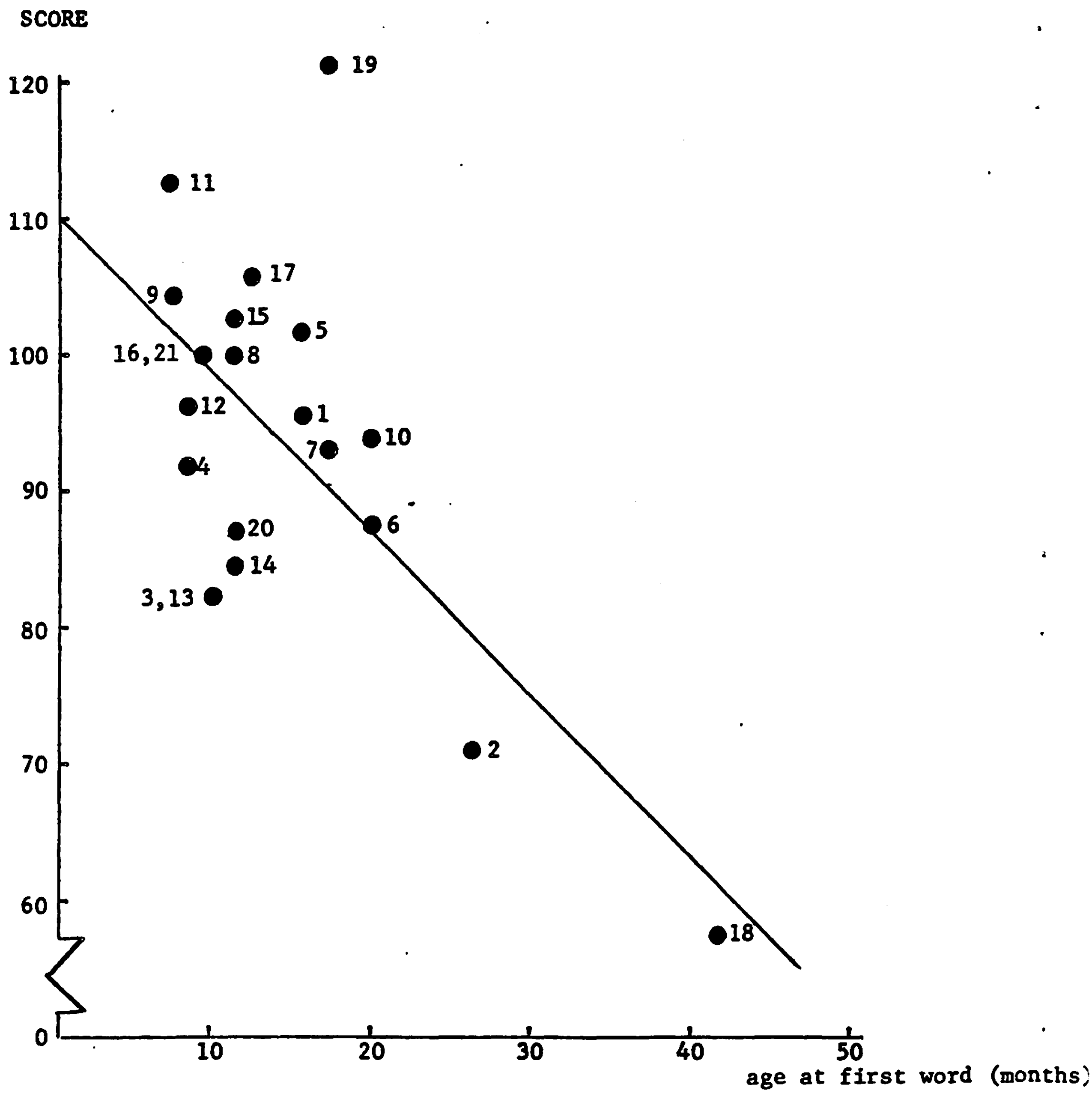
$$2\hat{I}_i(Z) = \frac{(N-p-2)v_i t_i^2(N-p-4)}{2(1-v_i)(N-p-3)} + \left[\frac{v_i(N-p-2)}{2(1-v_i)(N-p-3)} - \ln\left(1 + \frac{v_i}{2(1-v_i)}\right)\right]$$

$$+ N\left[\frac{N-p-2}{N-p-3}(1-t_i^2) - \ln\frac{N-p-2}{N-p-3}(1-t_i^2) - 1\right].$$

In previous papers, Johnson and Geisser (1983, 1985), the formulas are given for a set of k deleted observations. A set of data, Fig. 1, of 21 observations from a study of cyanotic heart disease in children taken from Mickey, Dunn and Clark (1967) was analyzed previously, Johnson and Geisser (1985), for influential observations. Some of the estimative influence measures and the critical component of other diagnostics such as the CPO are presented in Table 1 in compressed form for this data set.

Figure I: Scatter Plot of Gesell Adaptive Score Data

(Source: Mickey, Dunn, and Clark (1967))



Fitted regression line uses all 21 observations:

estimated slope = 1.127
estimated intercept = 109.87

Table 1

Influential cases for the Gesell adaptive score data

Case	$I_i(\beta)$	$I_i(\sigma^2)$	$I_i(\sigma^2,\beta)$	t_i^2
19	1.43	4.75	5.85	.42
18	2.90	.25	3.17	.04
13	.25	.23	.51	.11

Note that $\hat{I}_i(\beta)$, $I_i(\sigma_2)$ and $I_i(\sigma^2,\beta)$ are correct only up to an additive constant. As was indicated previously case 18 is most influential for the estimation of β but this is entirely due to its distance from the rest of the observations and in no way induces suspicion for that case. Case 19 is by far the most influential with respect to the estimation of σ^2 and jointly with respect to (β,σ^2). A number of measures including inspection of Figure 1 indicate the potential discordancy of case 19. We note that the predictive distribution of

$$(N-p-1)t_{19}^2(1-t_{19}^2)^{-1} \text{ is an } F(1,N-p-1) \text{ variate}$$

if we were suspicious of the case prior to its being observed. We assume that we were not and that our suspicion was raised only after diagnostics whose values indicated highest influence or maximum possible discordancy. In this situation, since t_{19}^2 and t_{13}^2 yield the first and second most discrepant diagnostic values we compute the P-value for discordancy of case 19 to be

$$P_{19} = P[F \geq 13.03 | F \geq 2.224] = .013.$$

This is small enough to cause concern about whether case 19 is compatible with the rest of the observations assuming the adequacy of the model for them.

8. Translated Exponential Distribution.

Let $y_1, \ldots, y_N$ be a random sample from

$$f(y|\alpha,\gamma) = \alpha e^{-\alpha(x-\gamma)} \qquad y > \gamma,\ \alpha > 0.$$

Let $y_1, \ldots, y_d$ be fully observed and $Y_{d+1}, \ldots, Y_N$ censored at $y_{d+1}, \ldots, y_N$ respectively.

Let

$$m = \min(y_1, \ldots, y_d)$$

and assume that

$$m < \min(y_{d+1}, \ldots, y_N).$$

Let the prior density be

$$g(\gamma,\alpha) = g(\gamma|\alpha)g(\alpha)$$

where

$$g(\gamma|\alpha) = N_0 \alpha e^{\alpha N_0(\gamma - m_0)}, \qquad \gamma < m_0$$

and

$$g(\alpha) \propto \alpha^{d_0-2} e^{-\alpha N_0(\bar{y}_0-m_0)} \qquad \alpha > 0,\ \bar{y}_0 > m_0$$

where $1 < d_0 \le N_0$. Then Geisser (1984a) obtains for the posterior densities

$$p(\gamma|\alpha) \propto e^{\alpha N^*(\gamma-m^*)} \qquad \gamma < m^*$$

$$p(\alpha) \propto \alpha^{d^*-2} e^{-\alpha N^*(\bar{x}^*-m^*)} \qquad \bar{x}^* > m^*,\ \alpha > 0$$

for $1 < d^* \le N^*$, d^* $d_0 + d$, $N^* = N_0 + N$, $m^* = \min(m_0,m)$, $\bar{y}^* = (N_0+N)^{-1}(N_0\bar{y}_0+N\bar{y})$ and $N\bar{y} = \sum_{i=1}^{N} y_i$.

For the noninformative prior

$$g(\gamma,\alpha) \propto \alpha^{-1},$$

$m^* \longrightarrow m$, $\bar{y}^* \longrightarrow \bar{y}$, $d^* \longrightarrow d$, $N^* \longrightarrow N$. We can calculate

$$I_i(\alpha,\gamma) = I_i(\alpha) + EI_i(\gamma|\alpha)$$

where the expectation sign is over the posterior marginal distribution of α. This can be done in terms of the proper prior using the starred values. We shall, for the sake of comparison, give the results for the noninformative prior. For a censored y_i, $i = d+1,\ldots, N$

$$I_i(\alpha) = (d-1)[r_i - 1 - \log r_i]$$

where

$$r_i = \frac{N(\bar{y}-m)}{N(\bar{y}-m) - (y_i - m)} > 1.$$

Hence for α alone, the largest censored observation is the most influential among the censored ones. Further

$$EI_i(\gamma|\alpha) = k$$

independent of y_i since

$$p_i(\gamma|\alpha) = \alpha(N-1)e^{\alpha(N-1)(\gamma-m)}.$$

Hence the largest censored value is the most influential among censored values for the estimation α alone or for α and γ jointly.

The marginal posterior density of γ is easily shown to be

$$p(\gamma) = (d-1)(\bar{y}-m)^{d-1}/(\bar{y}-\gamma)^d$$

and hence

$$I_i(\gamma) = \ln(\bar{y}-m) - \ln(\bar{y}_{(i)}-m)$$

where $\bar{y}_{(i)}$ is the mean of all the observations with y_i deleted. Hence the maximum of all censored observations is the most influential amongst them for the estimation of γ alone. For y_i such that $i = 1,\ldots, d$, $y_i \neq m$ and $\psi(\cdot)$ is the digamma function

$$I_i(\alpha) = \ln(d-2) - \psi(d-2) + (d-2)(r_i-1-\ln r_i)$$

which is an increasing function of r_i for $r_i > (d-1)/(d-2)$. If $1 < r_i \leq (d-1)/(d-2)$ then the influence is insignificant and it is not worth considering y_i as a particularly influential observation. Hence the largest uncensored y is the most influential among uncensored observations for the estimation of α. Therefore one need only compare the maximum influence among uncensored with that among censored and noting that if the observables are equal the censored one must be more influential.

However for (γ,α) or γ alone, we note that $I_i(\gamma|\alpha)$ is unbounded for $y_i = m$, because of the change in support when m is deleted. Hence the smallest observation is the most influential amongst all observations for the estimation of γ or (γ,α) overall. This fact indicates that the result obtained from the K-L divergence must be treated with caution in such cases.

If we consider prediction this problem of a change in support will not affect the influence measure. The predictive distribution of future observable Z is easily calculated to be, Geisser (1984),

$$F(z) = \begin{cases} \frac{1}{N+1}\left(\frac{\bar{y}-m}{\bar{y}-z}\right)^{d-1} & z \leq m \\ 1 - \frac{N^d(\bar{y}-m)^{d-1}}{(N+1)[z-m+N(\bar{y}-m)]^{d-1}} & z > m. \end{cases}$$

Since Z is supported over the whole real line even if m is deleted we do not have the difficulty of an unbounded I(Z). This also indicates that assessing the influence regarding the totality of parameters $\theta = (\alpha,\gamma)$ of a distribution can differ from assessing the influence of predicting a future observation. At any rate the actual calculation is quite tedious for the predictive influence function. For example when $y_i \neq m$ for $i = 1,\ldots,$ d and $d > 3$

$$\begin{aligned} I_i(Z) &= K - N^{-1}[1+(d-1)(N-1)]\ln(\bar{y}_{(i)}-m) \\ &- \frac{d(N-1)^{d-1}(\bar{y}_{(i)}-m)^{d-2}}{(m-x_i)^{d-2}}\left[\ln\frac{(N-1)(\bar{y}_{(i)}-m)}{N(\bar{y}-m)} + \sum_{j=1}^{d-3}\frac{1}{j}\left(\frac{m-y_i}{(N-1)(\bar{y}_{(i)}-m)}\right)^j\right. \\ &- \frac{d}{N}\left(\frac{\bar{y}_{(i)}-m}{\bar{y}_{(i)}-\bar{y}}\right)^{d-2}\left[\ln\frac{\bar{y}_{(i)}-m}{\bar{y}-m} - \sum_{j=1}^{d-3}\frac{1}{j}\left(\frac{\bar{y}_{(i)}-\bar{y}}{\bar{y}_{(i)}-m}\right)^j\right]. \end{aligned}$$

where K is a constant independent of i. For $y_i = m$, substitute $\bar{y}_{(m)}$ for $\bar{y}_{(i)}$ and m_2, the second smallest uncensored value, for m above with the added assumption that m_2 is also smaller than any uncensored value. In general this is a difficult calculation to make although it appears that it will be a maximum either for m or the largest uncensored value. A similar calculation for the censored values indicates that the largest censored value will be the most influential among censored values. Hence we may say that the single most influential value for the prediction of a future value will either be m or M, the largest among all values.

9. Conditional Predictive Ordinate (CPO).

For an uncensored value $y_i \neq m$ the CPO

$$d_i = f_{(i)}(y_i | y_{(i)}) = \frac{(d-2)(N-1)^{d-1}(\bar{y}_{(i)}-m)^{d-2}}{N[N(\bar{y}-m)]^{d-1}}$$

which clearly shows that the largest $y_i \neq m$, $i = 1,\dots,d$ has the smallest CPO. For $y_i = m$ and m_2 smaller than any uncensored value we obtain

$$d_m = \frac{d-2}{N} \frac{(\bar{y}_{(m)}-m_2)^{d-2}}{(\bar{y}_{(m)}-m)^{d-1}}$$

and

$$\min(d_i) = \min(d_m, d_{M_u}),$$

where M_u is the largest uncensored observation. For the censored observation $i = d+1,\dots,N$

$$d_i = \frac{(d-1)(N-1)^d}{N} \frac{(\bar{y}_{(i)}-m)^{d-1}}{(N(\bar{y}-m))^d}$$

and if the largest uncensored value is about the same as the largest censored value then its CPO will be smaller. Basically the diagnostic will choose either the largest value or the smallest value.

10. Discordancy Tests for Translated Exponential Variates.

For a conditional predictive test for the discordancy of the smallest observation we obtain significance level

$$P_m = \Pr\,[Z \leq m \mid Z \leq m_2] = \left(\frac{\bar{y}_{(m)} - m_2}{\bar{y}_{(m)} - m}\right)^{d-2} .$$

We illustrate this with some data from Kabe (1970) on lifetimes in hours of 5 pieces of a metal material. The values are 525, 603, 621, 648, 663. In this case we calculate

$$P_m = .023$$

Kabe, using the frequentist approach and Dixon's (1950, 1951) test statistic

$$t = \frac{m_2 - m}{M - m},$$

calculates an exact significance level to be $\alpha = 0.027$ (this appears to be erroneous with the correct result being .0164).

For a CPD test for the largest observation

$$P_M = \Pr[Z \geq M | Z > M_2] = \left[\frac{M_2 - m + (N-1)(\bar{y}_{(M)} - m)}{M - m + (N-1)(\bar{y}_{(M)} - m)}\right]^c$$

where M_2 is the second largest observation and $c = d-2$ or $d-1$ depending on whether M was an uncensored or censored observation. As an example we present some data representing an analysis of the chemical phosphorous as a component of carbon steel. The data, given by Likes (1966), in 10^6 multiples, are 4, 6.33, 7, 7, 9, 9.33, 25.

We calculate

$$P_M = .048$$

and compare this with the α level of the usual frequentist test statistic

$$t = \frac{M - M_2}{M - m} = .746$$

as given by Likes, where $\alpha = (N-1)(N-2)\, B\left(\frac{2-t}{1-t}, N-2\right)$

for B(.,.) the beta function and α=.050 for this particular example.

If γ is known one can calculate from the predictive distribution of Z

$$P_M = \left[\frac{M_2 - \gamma + (N-1)(\bar{y}_{(M)} - \gamma)}{M - \gamma + (N-1)(\bar{y}_{(M)} - \gamma)}\right]^{c+1}$$

where c is defined as for the case when γ is unknown.

11. CPD Tests For Combinations of Largest and Smallest.

In order to derive discordancy tests for 2 observations at a time, i.e. combinations of the smallest and largest, we need the joint predictive distribution of two future observations. The preliminary relevant calculations are:

$$\Pr[Z_1 \le z_1, Z_2 \le z_2] = \left(\frac{\bar{y}-m}{\bar{y}-v}\right)^{d-1} + \frac{N}{N+2}\left[\frac{N(\bar{y}-m)}{N\bar{y}+z_1+z_2-(N+2)v}\right]^{d-1}$$

$$- \frac{N}{N+1}\left[\frac{N(\bar{y}-m)}{N(\bar{y}-v)+z_2-v}\right]^{d-1} - \frac{N}{N+1}\left[\frac{N(\bar{y}-m)}{N(\bar{y}-v)+z_1-v}\right]^{d-1}$$

for $\max(z_1, z_2) = v \le m$,

$$\Pr[Z_1 \le z_1, Z_2 > z_2] = \frac{N}{(N+1)(N+2)}\left[\frac{N(\bar{y}-m)}{N(\bar{y}-z_1)+z_2-z_1}\right]^{d-1}$$

for $z_1 \le m \le z_2$,

$$\Pr[Z_1 > z_1, Z_2 > z_2] = \frac{N}{N+2}\left[\frac{N(\bar{y}-m)}{N(\bar{y}-m)+z_1+z_2-2m}\right]^{d-1}$$

for $\min(z_1, z_2) \ge m$.

For a joint discordancy test of the smallest and largest (m,M) we calculate from above using the fact that Z_1 and Z_2 are exchangeable,

$$P_{m,M} = \Pr[Z_1 \le m,\ Z_2 > M \mid Z_1 \le m_2,\ Z_2 > M_2]$$

$$= \left[\frac{(N-2)(\bar{y}_{(M,m)}-m_2) + M_2 - m_2}{(N-2)(\bar{y}_{(M,m)}-m) + M - m}\right]^c$$

where $c = d-2$ if M is censored, and $d-3$ if M is uncensored and $\bar{y}_{(M,m)}$ is the mean of all the observations excluding m and M.

To illustrate this we use the same data on the 5 test pieces of metal that we used for testing the discordancy of the minimum. We obtain

$$P_{m,M} = .062.$$

The usual frequentist test depends on the statistic

$$T = \frac{M_2-m_2}{M-m}$$

and $\alpha = P(T \le t)$ where

$$\alpha = 1 - (N-1)!(1-t)^2 \sum_{j=1}^{N-3} \frac{(-1)^{j+1} j[N-1-(N-j-2)t]^{-1}}{(j+1)!(N-3-j)!(1+jt)!},$$

a result incorrectly given by Kabe (1970) and corrected by Barnett and Lewis (1978). For this case $\alpha = .071$.

For the two smallest (m, m_2) where m_3 is the third smallest and assuming $m_3 \le \min(y_{d+1}, \ldots, y_N)$ it seems plausible to calculate

$$P_{m,m_2} = \Pr[Z_1 \le m, m \le Z_2 \le m_2 | Z_1 \le Z_2 \le m_3]$$

$$= (N-2)\left[\left(\frac{\bar{y}_{(m,m_2)}-m_3}{\bar{y}_{(m,m_2)}-m}\right)^{d-3} - \left(\frac{N(\bar{y}_{(m,m_2)}-m_3)}{N(\bar{y}_{(m,m_2)}-m)+m_2-m}\right)^{d-3}\right]$$

for $N \geq d > 3$.

For the two largest (M,M_2), we calculate, for M_3 the third largest,

$$P_{M,M_2} = \Pr[Z_1 > M, M_2 < Z_2 \le M | Z_1 > Z_2 \ge M_3]$$

$$= 2\frac{\{\Pr[Z_1 > M, Z_2 > M_2] - \Pr[Z_1 > M, Z_2 > M\ ]\}}{\Pr[Z_1 > M_3, Z_2 > M_3]} \ .$$

Then for

$$c = \begin{cases} d-1 & \text{if } M \text{ and } M_2 \text{ are censored} \\ d-2 & \text{if one of } M \text{ or } M_2 \text{ is censored} \\ d-3 & \text{if } M \text{ and } M_2 \text{ are uncensored,} \end{cases}$$

$$P_{M,M_2} = 2[(N-2)(\bar{y}_{(M,M_2)}-m) + 2(M_3-m)]^c$$

$$\times\ [[(N-2)(\bar{y}_{(M,M_2)}-m)+M\ +M_2-2m]^{-c} - [(N-2)(\bar{y}_{(M,M_2}-m)+2(M-m)]^{-c}.$$

For the case where γ is known P_{M,M_2} is calculated as above but with γ substituted for m and $c+1$ for c.

It is to be noted that all these CPD tests can be given in terms of the proper prior by merely substituting m^*, $\bar{y}^*$, d^*, N^* for m, $\bar{y}$, d and N respectively. It is to be remarked that other regions may also be plausible for the calculation of significance.

REFERENCES

Andrews, D. and Pregibon, D. (1978). Finding outliers that matter. J. Roy. Statist. Soc. B, **40**, 85-93.

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). Regression Diagnostics. New York: Wiley.

Cook, R.D. (1977). Detection of influential observations in linear regression. Technometrics, **19**, 15-18.

_______ (1979). Influential observations in linear regression. J. Amer. Statist. Assoc. **74**, 169-174.

Cook, R.D., and Weisberg, S. (1980). Finding influential cases in regression: a review. Technometrics, **22**, 495-508.

_______ (1982). Residuals and Influence in Regression. New York: Chapman and Hall.

Dixon, W.J. (1950). Analysis of extreme values, Annals of Math. Statistics **21**, 488-506.

_______ (1951). Ratios involving extreme values, Annals of Math. Statistics **22**, 68-78.

Geisser, S. (1980). Discussion Sampling and Bayes' inference in scientific modelling and robustness, by G.E.P. Box. J. Roy. Statist. Soc, A, **143**, 416-417.

_______ (1984). Predicting Pareto and exponential observables, Canadian Journal of Statistics, **12**, 143-152.

_______ (1985). On the predicting of observables: A selective update, in: Bayesian Statistics 2, with discussion. ed. J.M. Bernardo et al. North-Holland, 203-230.

Hoaglin, D.C. and Welsch, R.E. (1978). The hat matrix in regression and ANOVA. Amer. Statist. **32**, 17-22: Corrigenda, **145**.

Johnson, W. and Geisser, S. (1982). Assessing the predictive influence of observations. Statistics and Probability Essays in Honor of C.R. Rao. (Kallianpur, Krishnaiah, and Ghosh, eds.) Amsterdam: North Holland, 353-358.

_______ (1983). A predictive view of the detection and characterization of influential observations in regression analysis. J. Amer. Statist. Assoc., **78**, 137-144.

______ (1985). Estimative influence measures for the multivariate general linear model. J. Stat. Plan. and Inf. 11, 33-56.

Kabe, D.G. (1970). Testing outliers from an exponential population. Metrika, 15, 15-18. (77, 81, 82, 85, 86, 87)

Kullback, S. and Liebler, R.A. (1951). On information and sufficiency. Ann Math. Statist. 22, 79-86.

Likes, J. (1966). Distribution of Dixon's statistics in the case of an exponential population, Metrika, 11, 46-54.

McCulloch, R. (1985). Model influence in Bayesian statistics, unpublished doctoral dissertation, University of Minnesota.

Mickey, M.R., O.J. Dunn and V. Clark (1967). Note on the use of stepwise regression in detecting outliers. Comput. Biomedical Res. 1, 105-109.

Smith, A.F.M. and Pettit, L.I. (1985). Outliers and influential observations in linear models, Bayesian Statistics 2, ed. J.M. Bernardo et al. North-Holland, 473-494.