TECHNIQUES FOR DEALING WITH DISCORDANT OBSERVATIONS

Richard L. Branham, Jr.
Centro Regional de Investigaciones Científicas y Tecnológicas
Casilla de Correo 131
5500 Mendoza
Argentina

ABSTRACT. To test any theory, such as theories of motion--Newtonian or relativistic--of solar system objects, one must compare the predictions of theory with observation. But discordant observations habitually plague the reducer of astronomical data. To alleviate the baleful effects, particularly harmful when the observations are reduced by the method of least squares, of discordant data investigators almost invariably reject observations whose corresponding (O-C)'s or post-solution residuals exceed a cutoff. But techniques that are insensitive to the assumption that the observational errors are normally distributed, called robust estimation in the literature, have also been developed.

Various techniques for robust estimation exist, but some of them, such as minimization of the 1.5th power of the residuals, are only practical for small data sets because of the amount of computing involved. This study assesses the utility of two techniques for robust estimation that are practical with large data sets: iteratively reweighted least squares and an L_1 solution. The data set used is 21,365 optical observations of the minor planets 6-9 and 15 that Branham (1979) published. The analysis in that publication derived values for a correction to the equinox of the FK4 catalog system at epoch 1950.0, $\Delta E = 0''.283 \pm 0''.342$, and to Newcomb's general precession in longitude, $\Delta p_c = 1''.318 \pm 0''.238$, that are in good agreement with other determinations of these quantities, $\Delta E = 0''.525 \pm 0''.045$ and $\Delta p_c = 1''.10 \pm 0''.15$, that Fricke (1972; 1982) summarized. The distribution of the residuals from the 21,365 observations, which had a mean error of unit weight of $2''.394$, is more peaked, skewed, and heavy-tailed than one would obtain from the normal distribution. The present study uses these same observations and robust estimation to ascertain if the derived ΔE and Δp_c are in reasonable agreement with the other determinations. Although these two yardsticks may not be as precise as one would like, they should nevertheless allow one to weed out techniques or data reduction procedures that result in grossly discordant values.

Iteratively reweighted least squares (Coleman, et al. 1980) is ordinary least squares applied to a data set whose residuals have been

J. Kovalevsky and V. A. Brumberg (eds.), Relativity in Celestial Mechanics and Astrometry, 229–231.

weighted to achieve robustness. If $\underset{\sim}{r}$ represents the vector of the residuals then we minimize $(\underset{\sim}{W}^{1/2}\cdot\underset{\sim}{r})^{T}\cdot(\underset{\sim}{W}^{1/2}\cdot\underset{\sim}{r})$, where $\underset{\sim}{W}$ is a diagonal matrix of the weights, instead of, as with ordinary least squares, $\underset{\sim}{r}^{T}\cdot\underset{\sim}{r}$. To start an iteratively reweighted least squares solution we need a preliminary solution to calculate the residuals because an individual weight depends on the size of the residual. The weights used in this study fall into three categories: the Andrews, Biweight, and Talwar weightings reject large residuals that exceed a cutoff, and the former two assign higher weight to small residuals; the Welsch weighting employs no cutoff, but large residuals receive such small weight that for practical purposes it is zero; the Fair and Huber weightings assign relatively high weight to large residuals. The latter two categories reflect modern ideas that heavy-tailed distributions are common rather than exceptional. For this study a preliminary solution was calculated from all of the data and iterated until a convergence criterion was met. Of the weighting schemes studied thos that apply no cutoff were all unsatisfactory. The others, Andrews, Biweight, and Talwar, all gave good determinations (ΔE = 0".510±0".168, 0".448±0".163, 0".442±0".164; Δp_c = 1".436±0".135, 1".448±0".131, 1".643±0".131 respectively) albeit at the cost of increased computer time. About three times more CPU time was needed by iteratively reweighted least squares than would be by ordinary least squares. The former, therefore, requires a careful study of the distribution of the residuals to determine a proper weighting function.

The other technique studied, L_1 estimation, is a non-least squares procedure based on the principle of minimization of the sum of the absolute values of the residuals. The algorithm for L_1 estimation used in this study is a modification of Dantzig's simplex algorithm for linear programming (Barrodale and Roberts 1974). L_1 estimation may require considerably more computing than ordinary least squares, but the exact amount of CPU time depends critically on the available physical memory. However, the results from the L_1 algorithm were highly satisfactory (ΔE = 0".449 and Δp_c = 1".261), and because L_1 estimation rejects no residuals the investigator is freed from the task of establishing cutoffs or studying the shape of the distribution of the residuals. The technique, therefore, incorporates an invariant algorithm that would be useful for comparison of the results of individual investigations.

REFERENCES.

Barrodale, I., and Roberts, F.D.K. (1974). Commun. ACM 17, 319.

Branham, R.L. (1979). Astron. Pap. Amer. Ephem. 21, Part 3.

Coleman, D., et al. (1980). ACM Trans. Math. Software 6, 327.

Fricke, W. (1972). Annu. Rev. Astron. Astrophys. 10, 101.

Fricke, W. (1982). Astron. Astrophys. 107, L. 13.

DISCUSSION

Kristensen : you recommend weighting the observations by the residuals o-c. In my opinion, weights should be determined by a critical analysis of the series of observations.

Branham : of course, it would be better to analyse the real observation conditions, but we often have no information about these conditions but only pure data.

Kristensen : have you compared the results obtained by least squares with those obtained by L_1 on fictitious examples where errors are constructed by a random numbers generator to be truly Gauss-distributed ?

Branham : we did so, but this comparison was not very convincing. It was shown in 1977 in a paper on various robust methods, that the optimal method for simulated data is often bad for real data. So the only way to select a method is to test it on real data.

Seidelmann : what were your considerations in chosing your method ?

Branham : first because it has been theoretically proven, secondly because it worked on test examples.

Batrakov : would it not be better to use more standard two-step method: first use least squares and eliminate outliers, then, again use least squares ?

Branham : this method works well for eliminating the outliers. But in our case, the distribution is not Gaussian and the two-step method would not work.

Kreinovich : you have used the simplex method for linear programming problem. So your program took 12 hours of computer time. But nowadays, new methods have been invented (ellipse- method, Karkmarkar method). Their use would reduce the computer time.

Branham : I heard about the russian ellipse-method five years ago, but as far as I know, it is not better than the simplex method.

Kreinovich : since then, it has been greatly improved. Another question : you use on L_1 method. Why didn't you choose L_p with p adapted to observations ? One can prove that these methods are optimal.

Branham : we are planning to try $L_{1.5}$ method.

https://doi.org/10.1017/S0074180900148247 Published online by Cambridge University Press