The Korean Communications
in Statistics Vol. 7, No. 2, 2000
pp. 575-583

Robust Nonparametric Regression Method using Rank Transformation[1]

Dongryeon Park[2]

Abstract

Consider the problem of estimating regression function from a set of data which is contaminated by a long-tailed error distribution. The linear smoother is a kind of a local weighted average of response, so it is not robust against outliers. The kernel M-smoother and the lowess attain robustness against outliers by down-weighting outliers. However, the kernel M-smoother and the lowess requires the iteration for computing the robustness weights, and as Wang and Scott(1994) pointed out, the requirement of iteration is not a desirable property. In this article, we propose the robust nonparametric regression method which does not require the iteration. Robustness can be achieved not only by down-weighting outliers but also by transforming outliers. The rank transformation is a simple procedure where the data are replaced by their corresponding ranks. Iman and Conover(1979) showed the fact that the rank transformation is a robust and powerful procedure in the linear regression. In this paper, we show that we can also use the rank transformation to nonparametric regression to achieve the robustness.

Keywords : Rank transformation, Robust regression, Nonparametric regression, Local polynomial smoothing

1. Introduction

Consider the problem of estimating regression function from a set of data which is contaminated by a long-tailed error distribution. Let $Y_i \in \boldsymbol{R}$ be the dependent variable and $X_i \in \boldsymbol{R}$ be the independent variable, $i = 1, \cdots, n$. Suppose that there exists a smooth function m such that

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \cdots, n. \tag{1}$$

1) This research was supported by the Hanshin University Research Grants in 2000.

2) Associate Professor, Department of Statistics, Hanshin University, Osan 447-791, KOREA.
E-mail : drpark@hucc.hanshin.ac.kr

In recent years, the size of data sets are growing fast, so the number of the outliers included in the data sets are also increasing. Therefore the problem of handling bad or influential data points is growing even faster than the size of data sets. Thus the practical importance of robust procedures that automatically handle outliers is ever growing (Rousseeuw and Leroy, 1987).

The linear smoother is a kind of a local weighted average of response, so it is not robust against outliers. Robustness against outliers can be attained by down-weighting outliers. Hardle(1984) and Hardle and Gasser(1984) extend the basic idea of M estimation. They reduce the influence of outlying observations by the use of a non-quadratic bounded loss function. A robust kernel M-smoother, $\widehat{m}^M(x)$ is defined as

$$\widehat{m}^M(x) = \arg\min_\theta \sum_{i=1}^{n} K_h(x - x_i)\, \rho(Y_i - \theta). \tag{2}$$

Here $K_h(\cdot) = (1/h)K(\cdot/h)$ where K is a kernel function, and $\rho(\cdot)$ is the loss function such as

$$\rho(x) = \begin{cases} (1/2)\,x^2, & \text{if } |x| \le c \\ c|x| - (1/2)\,c^2, & \text{if } |x| > c \end{cases} \tag{3}$$

Another method that down-weights the influence of large residuals is the locally weighted regression, or lowess, which was proposed by Cleveland(1979). The basic idea of lowess is down-weighting the influence of large residuals by iterative fitting the local polynomial regression. When we compute the lowess fitting, $\widehat{m}^L(x)$, each observations (x_i, y_i) are assigned neighborhood weights $w_i(x)$. The robustness weights, r_i depend on the residuals $\hat{\varepsilon}_i = y_i - \hat{y}_i$ from the current fitting and they are given by

$$r_i = B\left(\frac{\hat{\varepsilon}_i}{6s}\right) \tag{4}$$

where $s =$ median$|\hat{\varepsilon}_i|$ and B is the bisquare weight function,

$$B(x) = \begin{cases} (1-x^2)^2, & \text{if } |x| < 1 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

In the next fitting, weighted local fitting is carried out using the weight $r_i w_i(x)$ for (x_i, y_i), and this process is repeated until the fitted curve converges. The kernel M-smoother is implicitly defined, so it requires iterative numerical methods, and the lowess also requires the iteration for computing the robustness weights. However, as Wang and Scott(1994) pointed out, the requirement of iteration is not a desirable property. In this article, we propose the robust nonparametric regression method which does not require the iteration. Robustness can be achieved not only by down-weighting outliers but also by transforming outliers. The rank transformation is a simple procedure where the data are replaced by their corresponding ranks. Iman and Conover(1979) showed the fact that the rank transformation is a robust and powerful procedure in the linear regression. In this paper, we show that we can also use the rank transformation to nonparametric regression to achieve the robustness.

2. Procedure

2.1 The rank transformation approach to the parametric regression

The rank transformation procedure in Iman and Conover(1979) is following. Suppose we have data (x_i, y_i), $i = 1, \ldots, n$. The dependent variable y_i are replaced by their corresponding ranks. $R(y_i)$ is the rank assigned to ith value of Y. Likewise, independent variable x_i are also replaced by their ranks. $R(x_i)$ is the rank assigned to ith value of X. Tied observations are assigned average ranks. Then the usual least squares regression analysis is performed entirely on these ranks. The regression equation based on ranks is given by

$$\hat{R}(Y) = (n+1)/2 + \hat{\beta}_1(R(X) - (n+1)/2) \qquad (6)$$

In the case of no ties in the assignment of ranks, $\hat{\beta}_1$ in (6) is simply Spearman's rho. To get the predicted value $\hat{y}_*$ for the arbitrary value x_*, x_* should be replaced by its rank, $R(x_*)$ by the following interpolation rule.

1. If $x_* < x_{(1)}$, then $R(x_*) = R(x_{(1)})$.
2. If $x_* > x_{(n)}$, then $R(x_*) = R(x_{(n)})$.
3. If $x_* = x_{(i)}$ for some i, then $R(x_*) = R(x_{(i)})$. (7)
4. If $x_{(i)} < x_* < x_{(i+1)}$, $i = 1, \ldots, n-1$, then

$$R(x_*) = R(x_{(i)}) + [R(x_{(i+1)}) - R(x_{(i)})] \times [(x_* - x_{(i)})/(x_{(i+1)} - x_{(i)})]$$

Plug in $R(x_*)$ into the regression equation (6) yields $\hat{R}(y_*)$. Now we transform the predicted rank $\hat{R}(y_*)$ into a predicted value $\hat{y}_*$ as follows.

1. If $\hat{R}(y_*) < R(y_{(1)})$, then $\hat{y}_* = y_{(1)}$.
2. If $\hat{R}(y_*) > R(y_{(n)})$, then $\hat{y}_* = y_{(n)}$.
3. If $\hat{R}(y_*) = R(y_{(i)})$, then $\hat{y}_* = y_{(i)}$ for some i. (8)
4. If $R(y_{(i)}) < \hat{R}(y_*) < R(y_{(i+1)})$, $i = 1, \ldots, n-1$, then

$$\hat{y}_* = y_{(i)} + (y_{(i+1)} - y_{(i)}) \times [(\hat{R}(y_*) - R(y_{(i)}))/(R(y_{(i+1)}) - R(y_{(i)}))]$$

Iman and Conover(1979) showed that the rank transform approach has an obvious advantage when the dependent variable is a monotonic function of the independent variable and this monotonic relationship is nonlinear in nature. However, they emphasized that true outliers which make linear data highly non-monotonic could cause problems for rank regression and in such cases it would seem reasonable to use other robust regression. Therefore, for the case of monotone nonlinear output where the observations cannot be dismissed as outliers, the rank transform approach to regression can produce good results.

2.2 The rank transformation approach to the nonparametric regression

In the ordinary least square regression, we put data together and use all of them in estimating parameters. $R(x_i)$ and $R(y_i)$ in (7) and (8) are ranks based on the whole set of data. However, in the nonparametric regression the observations beyond the small neighborhood of x do not play any role in estimating $\widehat{m}(x)$. The nonparametric regression $\widehat{m}(x)$ can be defined as

$$\widehat{m}(x) = n^{-1} \sum_{j \in N_x} W_j(x) Y_j \tag{9}$$

where $W_j(x)$ denotes a sequence of weights and N_x is a neighborhood of x. If the global bandwidth is used, then $N_x = \{i : x - h \le x_i \le x + h\}$, and if $k-NN$ smoother is used, then $N_x = \{i : x_i$ is one of the k nearest observation to $x\}$. We use only the local information around x in the estimation procedure. Therefore the rank based on the whole set of data is not appropriate for the rank transform approach to the nonparametric regression. The rank only using the data inside of the neighborhood of x is more suitable for estimating $\widehat{m}(x)$ by the nonparametric regression. The rank transformation procedure for estimating $\widehat{m}(x)$ is following.

1. Construct the neighborhood of x, N_x.
2. Assign the rank to x_j and y_j as $R(x_j)$ and $R(y_j)$, respectively. Here $j \in N_x$.
3. Apply the nonparametric regression procedure to rank transformed data $(R(x_j), R(y_j))$, $j \in N_x$.
4. Convert the estimate results of Step 3 to the original scale using (7) and (8).

3. Application to Data

As the robust nonparametric regression, the most commonly used method is probably the lowess, because it is built into the S statistical language. The performance of the lowess is really excellent and actually it is one of the best robust procedure. In this section, we compare the performance of the rank transformation approach and the lowess using the simulated data and the real data. The local quadratic fitting was used for the lowess.

Example 1. A random sample of size n is simulated from the model

$$Y = \exp(X) + \varepsilon$$

with $X \sim \text{Unif}(0,5)$ and ε has the Cauchy distribution with location parameter 0 and scale parameter 5. In this simulation model, the dependent variable is a nonlinear monotonic function of the independent variable, so the rank transformation approach should have a obvious advantage under this circumstance (Iman and Conover, 1979). The noise was simulated from the Cauchy distribution, so some extreme values make the data highly non-monotonic and, according to Iman and Conover(1979), this can cause a bad influence to rank transformation. As the nonparametric procedure for rank transformation, the local quadratic regression (Fan, 1992) was used. The $k-NN$ rule was used for both the rank transform method and the lowess, so the fraction f of the total number of data points that used to estimate $\widehat{m}(x)$ is constant for all x. The smoothing parameter $f = .5$ was used for both methods. During the simulation example, we applied the rank transformation to Y variable only, because we found that this produces better results. $N = 200$ Monte Carlo runs were made and the average value of $\frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{m}(x_i)|$ over the simulation runs are on the Table 1.

	$n = 20$	$n = 50$	$n = 100$
rank transform	7.4914	4.7399	3.5796
lowess	14.9426	2.6018	2.0301

Table 1 Monte Carlo Mean Absolute Deviation for two methods

A set of simulated data for $n = 100$ case are shown in Figure 1, along with two estimated curves. The rank transformation method shows a conspicuous performance at the small sample size. As the sample size increases, the range of ranks also increases. Therefore the rank transformation method can not hold a superior position any more for the large sample case.

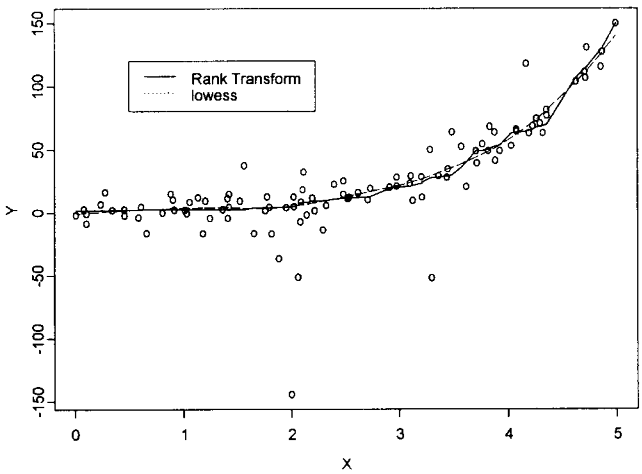

Figure 1 A simulated data set using Cauchy error distribution and two estimates.

Example 2. A random sample of size n is simulated from the model

$$Y = X^3(1-X)^3 + \varepsilon$$

with $X \sim$ Unif$(0,1)$ and ε has the Cauchy distribution with location parameter 0 and scale parameter 0.001. In this simulation model, the dependent variable is not monotonic function of the independent variable. As in Example 1, the local quadratic regression was used for the nonparametric procedure of rank transform approach. The smoothing parameter is chosen by $f = .4$ for both methods. $N = 200$ Monte Carlo runs were made and the average value of $\frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{m}(x_i)|$ over the simulation runs for each methods are on the Table 2.

	$n = 20$	$n = 50$	$n = 100$
rank transform	0.0017	0.0009	0.0007
lowess	0.0031	0.0009	0.0003

Table 2 Monte Carlo Mean Absolute Deviation for two methods

A set of simulated data for $n = 100$ case are shown in Figure 2, along with two estimated curves. Even though the non-monotonic relationship between X and Y, the rank transform approach shows a good performance.

Example 3. We considered the motorcycle impact data which is one of the S-Plus data set. The data are obtained from a study of the effectiveness of helmets in collisions. The X values are time measurements in milliseconds after a simulated impact, and the Y values are measurements of head acceleration in units of g (9.8 meters/ sec ^{2}). The sample size is $n = 133$. The data are shown in Figure 3 along with two estimated curves. The data are clearly heteroscedastic. Although the outliers are not as large as in the simulated data sets, the data nonetheless present a difficult challenge for robust nonparametric regression (Wang and Scott, 1994). The local quadratic regression was used for the rank transform approach, and the smoothing parameter f is chosen by $f = .2$. It seems that two curves track the data quite well. Since the true curve is unknown, it is difficult to judge which one is better. However, the rank transform approach seems to underestimate the true curve at the peak.

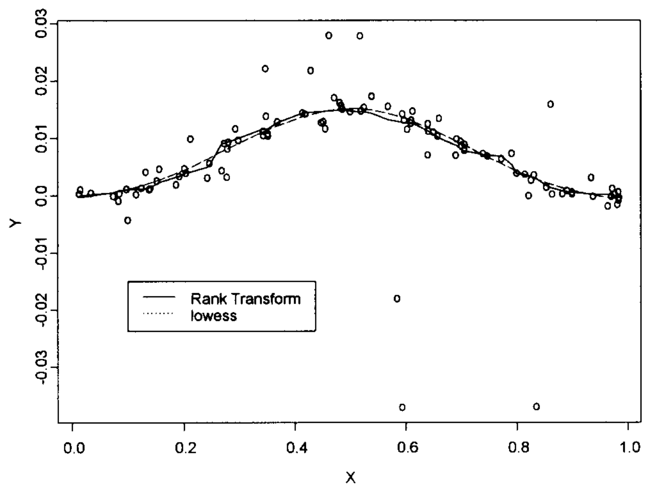

Figure 2 A simulated data set using Cauchy error distribution and two estimates.

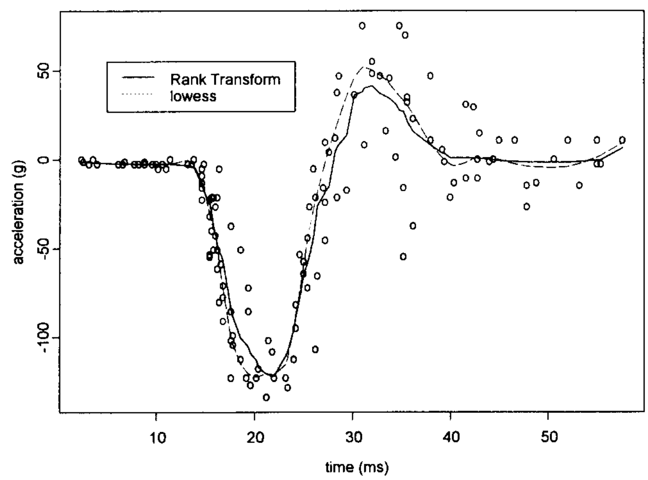

Figure 3 Motorcycle impact data with two estimated curves

4. Conclusion

The rank transform approach to parametric regression has a great advantage only on monotone data, so if data have long tailed error distribution, the rank transform approach to parametric regression does not quite work well. However, the performance of the rank transform approach to nonparametric regression does not depend on the monotonicity of data. The simulated data sets of Example 1 and Example 2 have the Cauchy error distribution and the true curve of Example 2 is even non-monotone. The real data of Example 3 is also non-monotone and the variance of the error is not constant. However, the rank transform approach showed excellent performance for all Example cases. Moreover, the rank transform approach has a great advantage at the small sample size. The lowess achieves robustness by iteratively reweighted least square fitting, but the rank transform approach does not require the iteration and its procedure is very simple. The extension to the multivariate case looks easy and straightforward.

References

[1] Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots", *Journal of the American Statistical Association*, 74, 829-836.

[2] Cleveland, W. S. and Devlin, S. J. (1988), "Locally Weighted Regression: An -Approach to Regression Analysis by Local Fitting", *Journal of the American Statistical Association*, 83, 596-610.

[3] Fan, J. Q. (1992), "Design-adaptive Nonparametric Regression", *Journal of the American Statistical Association*, 87, 998-1004.

[4] Hardle, W. (1984), "Robust Regression Function Estimation", *Journal of Multivariate Analysis*, 14, 169-180

[5] Hardle, W. (1990), *Applied Nonparametric Regression*, Cambridge, U.K.: Cambridge University Press.

[6] Hardle, W. and Gasser, T. (1984), "Robust Non-Parametric Function Fitting", *Journal of the Royal Statistical Society*, Ser. B, 46, 42-51.

[7] Iman, R. L. and Conover, W. J. (1979), "The Use Of the Rank Transform in Regression", *Technometrics*, 21, 499-509.

[8] Rousseeuw, P. J. and Lervoy, A. M. (1987), *Robust Regression and Outlier Detection*, John Wiley, New York

[9] Scott, D. W. (1992), *Multivariate Density Estimation*, John Wiley, New York

[10] Wang, F. T. and Scott, D. W. (1994), "The L_1 Method for Robust Nonparametric Regression", *Journal of the American Statistical Association*, 89, 65-76.