The Korean Journal of Applied Statistics (2009)
22(5), 1097–1102

Robust Inference for Testing Order-Restricted Inference

Moonsu Kang[1]

[1]Korean food and drug administration

(Received June 2009; accepted June 2009)

Abstract

Classification of subjects with unknown distribution in small sample size setup may involve order-restricted constraints in multivariate parameter setups. Those problems makes optimality of conventional likelihood ratio based statistical inferences not feasible. Fortunately, Roy (1953) introduced union-intersection principle(UIP) which provides an alternative avenue. Redescending M-estimator along with that principle yields a considerably appropriate robust testing procedure. Furthermore, conditionally distribution-free test based upon exact permutation theory is used to generate p-values, even in small sample. Applications of this method are illustrated in simulated data and read data example (Lobenhofer *et al.*, 2002)

Keywords: Redescending M-estimator, union-intersection principle, distribution-free.

1. Introduction

To classify subjects across distinct groups(or experimental conditions) with unknown distribution in small sample has received much attention in the literature. On top of that, it may have complex structures which are marred by inequality, order, stochastic ordering, functional and shape constraints or others. Such a model is abound in interdisciplinary fields, in particular in the field of genomics and the contemplated bioinformatics area. For example, microarray data has a lot of standardization and normalization with small sample size so that conventional simple models, such as ANOVA and MANOVA models, may hardly be tenable. The likelihood, sufficiency, and invariance principles are very crucial in finite sample methodology. In complex statistical models involving some such constraints, those properties are not often satisfied. Furthermore, even in asymptotics, lack of some regularity conditions often encounter roadblocks for optimal inferences. As such, it should be noted that the existing literature did not seem to work well. One possible remedy is to utilize the union-intersection principle, which was developed by Roy (1953) as a heuristic method of test construction. It was shown that using the principle accompanies computational advantages and increased adaptability, and greater adaptability to complex nonstandard models. On the other hand, we attempt to handle situations in which a distribution is unknown(or nonnormal) and sample size is very small. In respect to unknown distribution, a robust procedure should be used against

[1]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University Richmond VA 23298. E-mail: moonsukang0223@gmail.com

departures from underlying assumptions usually caused by outliers. It should have good performance under the underlying assumptions and the performance gets worse as the situation departs from the distributional assumptions. In order to test order-restricted inference, a formulation of a linear regression model is appraised by utilizing robust regression estimators. in the next section. Transition from normal theory to nonparametric statistics is shown there. There are various types of robust estimators. M-estimator(maximum likelihood type estimator), L-estimator(linear combinations of order statistics) and R-estimator(estimator based on rank transformation) (Huber, 1981); RM estimator(repeated median) (Siegel, 1982) and LMS estimator(estimator using the least median of squares) (Rousseeuw and Yohai, 1984) have been paid attentions to. It was known that M-estimators are most robust, even in finite sample. Interestingly, among all, redescending M-estimators acheive both high efficiency and robustness properties whereas the Huber M estimator entails loss of efficiency at the cost of robustness. However, there is a computational problem in calculating the redescending M-estimators in a wide variety of linear models including regression, since it is not easy to solve an equation for M-estimator. Using S-estimator as an initial value may help to get solutions. For a comprehensive review of them, please see references (Maronna *et al.*, 2006; Huber, 1973; Andrews, 1974) where other pertinent references have been extensively cited.

In relation to small sample size, we may not use asymptotic distributional theory and then find an alternative course for small sample size perspective. We may apply permutation theory, which gives us (conditionally) distribution-free tests without any parametric inferences or specific distributional assumptions.

The paper is organized as follows. Section 2 shows us the representation of model of our interest in terms of a linear regression model. In Section 3, we introduce redescending M-estimator and discuss asymptotic properties. Section 4 shows an explicit representation for proposed test statistics. Performance of the proposed statistics was compared with that of others in simulation studies and real data study (Lobenhofer *et al.*, 2002) in Section 5. The last section summarizes our result.

2. A Robust Linear Model

In the context of testing order-restricted inference, we may model the data using a linear regression model. Consider the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ is the vector of gene expression levels across G groups and The (known)Design matrix of the $n \times G$ matrix

$$\mathbf{X} = \begin{pmatrix} 1\,0\,0\,0 & 0 & \ldots \\ \vdots\,\vdots\,\vdots\,\vdots & \vdots & \ldots \\ 1\,1\quad 0 & 0 & \ldots \\ \vdots\,\vdots\,\vdots\,\vdots & \vdots & \ldots \\ 1\,0\,1\,0 & 0 & \ldots \\ \vdots\,\vdots\,\vdots\,\vdots & \vdots & \ldots \\ 1\,0\,0\,0 & \ldots & 1 \end{pmatrix}$$

The unknown parameter of the $G \times 1$ matrix

$$\boldsymbol{\beta} = (\mu_1, \delta_2, \delta_3, \ldots, \delta_G),$$

where μ_1 denotes the average response in the first group and δ_j refers to the difference between μ_1 and the average response in the j-th group, $j = 2, \ldots, G$. The vector of independent and identically distributed(i.i.d.) errors with unknown distribution F of the $n \times 1$ matrix

$$\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3, \ldots, \epsilon_n).$$

Without loss of generality, we focus on the increasing pattern of means across groups in the paper. Note that the error distribution is unknown. Thus, robust method is more appropriate than other test statistics relying on classical normal theory. Our hypotheses are represented as

$$H_0 : \delta_2 = \delta_3 = \cdots = \delta_G \quad vs. \quad H_1 : \delta_2 \leq \delta_3 \cdots \delta_G (1).$$

3. Robust Estimators and Related Asymptotic Properties

This section presents proposed redescending M-estimator and the limiting distribution of the estimator. Redescending M-estimators are ψ-type M-estimators, in which ψ functions are non-decreasing near the origin, but decreasing toward 0 (redescend smoothly to zero), so that they usually satisfy $\Psi(\text{x}) = 0$ for all x with $|x| > r$. Due to those properties of ψ functions, they are very efficient, have a high breakdown point, and do not suffer from a masking effect, unlike other robust statistics. The fact that they reject gross outliers, and do not completely ignore moderately large outliers leads to high efficiency, whereas the Huber estimator for several symmetric, wider tailed distributions considers them as still moderate. In fact, the redescending M estimators are more efficient in contrast with the Huber estimator for the Cauchy distribution. We skip theoretical review of the estimator in the section. Let Ψ be a bi-square function, which satisfies crucial assumptions necessary for asymptotic normality: (1) Ψ is bounded and increases from -1 to 1 and (2) Ψ'' is continuous. Let $\boldsymbol{\theta}$ be $(\boldsymbol{\beta}, \sigma)$. Instead of estimating $\boldsymbol{\beta}$ with scale parameter σ fixed, Huber and Dutter (1974) and Huber (1977) introduced more elegant methods so that $\boldsymbol{\beta}$ and a scale σ are estimated at the same time by minimizing

$$\sum \rho \left(\frac{y_i - x_i \boldsymbol{\beta}}{\sigma} \right) \sigma + B_n \sigma, \tag{3.1}$$

where B_n is a suitably chosen sequence of constants with $B = \lim B_n$. An estimating equation for redescending M-estimator is given by

$$\sum \Psi \left(\frac{y_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) \mathbf{x}_i = 0, \qquad \sum \chi \left(\frac{y_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) = B_n, \tag{3.2}$$

where $\Psi = \rho'$ and $\chi(t) = \int_0^t x \partial \Psi(x)$. Deriving asymptotic normality relies heavily on Silvapullé (1985). For more in depth, please refer to Silvapullé (1985). Now, the matrix representation of the regression equation is given by $\mathbf{Y} = \alpha \mathbf{1} + \mathbf{X} \boldsymbol{\delta} + \boldsymbol{\epsilon}$. Under this setup, we may derive asymptotic normality of test statistics.

$$n^{\frac{1}{2}} \left((\hat{\alpha}, \hat{\boldsymbol{\delta}}, \hat{\sigma}) - (\alpha, \boldsymbol{\delta}, \sigma) \right) \to \mathcal{N}(\mathbf{0}, \mathbf{C}), \tag{3.3}$$

where

$$\mathbf{C} = \sigma^{*2} \begin{pmatrix} h_1 & 0 & h_4 \\ 0 & n h_2 (\mathbf{X}'\mathbf{X})^{-1} & 0 \\ h_4 & 0 & h_3 \end{pmatrix},$$

$h_1 = (ab - c^2)^{-2}(b^2 u - 2bcw + c^2 v)$, $h_2 = (u/a^2)$, $h_3 = (ab - c^2)^{-2}(a^2 v - 2acw + c^2 u)$, $h_4 = (ab - c^2)^{-2}(abw - bcu - acv + c^2 w)$, $a = \text{E}(\Psi'(\eta))$, $b = \text{E}(\eta^2 \Psi'(\eta))$, $c = \text{E}(\eta \Psi'(\eta))$, $u = \text{Var}(\Psi(\eta))$, $v = \text{Var}(\chi(\eta))$ and $w = \text{Cov}(\Psi(\eta), \chi(\eta))$ with $\eta = (\varepsilon/\sigma^*)$.

4. Order-Restricted Inference

For testing the null hypothesis, it may be intended to consider alternatives that the vector β belongs to the nonnegative orthant space $\Re^{+(G-1)}$. In the univariate case, an optimal UMP test exists for such one-sided alternative. However, in such a multivariate case, UMP tests do not exist. For example, the Hotelling T^2 will result in a larger set of confidence interval and will entail some loss of efficiency. It's therefore interesting to appraise statistical inference under such restricted setups. UIP (Union-Intersection Principle) formulation of Roy (1953) could be well tailored for statistical inference under the one-sided multivariate alternative hypothesis. Let $\hat{\beta} = (\hat{\alpha}, \hat{\delta})$ the M-estimator. In conjuction to (1) in Section 2, we use UIP to formulate a robust M-test for

$$H_0 : \mathbf{A}\beta = \mathbf{0} \qquad vs. \qquad H_1 : \mathbf{A}\beta \geq \mathbf{0},$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \end{pmatrix}.$$

Let $\mathcal{G} = \{1, \dots, G-1\}$, and for every a: $\emptyset \subseteq a \subseteq \mathcal{G}$, let a' be its complement and $|a|$ its cardinality. For each a: $\emptyset \subseteq a \subseteq \mathcal{G}$, partition $\mathbf{Z}$ and $\mathbf{V}$ as

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_a \\ \mathbf{Z}_{a'} \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} \mathbf{V}_{aa} & \mathbf{V}_{aa'} \\ \mathbf{V}_{a'a} & \mathbf{V}_{a'a'} \end{pmatrix}$$

and write

$$\mathbf{Z}_{a:a'} = \mathbf{Z}_a - \mathbf{V}_{aa'}\mathbf{V}_{a'a'}^{-1}\mathbf{Z}_{a'},$$
$$\mathbf{V}_{aa:a'} = \mathbf{V}_{aa} - \mathbf{V}_{aa'}\mathbf{V}_{a'a'}^{-1}\mathbf{V}_{a'a}.$$

We have the result of the limiting distribution derived in the previous section. In passing, since $n^{1/2}(\mathbf{M}_n - \beta)$ to a G-variate normal law, for n very large, we get

$$\left(n\mathbf{V}^{-1}\right)^{\frac{1}{2}} (\mathbf{Z} - \beta) \xrightarrow{\mathcal{D}} \mathcal{N}_{G-1}(\mathbf{0}, \mathbf{I}). \tag{4.1}$$

Under (4.2), we formulate test statistics T given by

$$\Sigma_{\emptyset \subseteq a \subseteq \mathcal{G}} I\left(\mathbf{Z}_{a:a'} > \mathbf{0}, \mathbf{V}_{a'a'}^{-1}\mathbf{Z}_{a'} \leq \mathbf{0}\right)\left(n\mathbf{Z}'_{a:a'}\mathbf{V}_{aa:a'}^{-1}\mathbf{Z}'_{a:a'}\right). \tag{4.2}$$

Remember that we handle situations when sample size n is too small. As such, asymptotic normality does not hold. For this reason, the permutation distribution theory is fully exploited for small sample size setup. Under the null hypothesis of homogeneity, the joint distribution of all n observations remains invariant under any permutation, leading to manageable testing procedures. It should be noted, however, that the behavior of T under alternatives depends on the stochastic ordering of β. This statistics should be not exact-distribution free. In a very nonparametric set up, conditionally distribution-free tests can be constructed in the following manner. We reject the null hypothesis for large positive values, where T is a test statistic and t is an observed test statistics. Under the alternative, $Y_{i'} - Y_i$, $1 \leq i < i' \leq n$ has a distribution tilted to the right so that $E(T|H_1) \geq 0$. This motivates us to use tests based on T using the right hand side critical region. The distribution

Table 5.1. Simulation study

means of 6 groups	proposed	ANOVA	JT
(1, 2, 3, 4, 5, 6)	2.5e−11	4.1e−10	1.6e−07
(1, 3, 5, 7, 9, 11)	3.1e−15	2.5e−14	4.9e−09
(1, 4, 7, 11, 14, 17)	1.4e−20	7.1e−19	1.4e−09
(1, 5, 9, 13, 17, 21)	1.3e−24	1.4e−22	1.4e−09
(1, 6, 11, 16, 21, 26)	6.1e−25	7.3e−23	1.4e−09

of T under H_0 is generated by the $n!/(n_1!n_2!\cdots n_G!)$ equally likely permutations of observations $Y_i, i = 1, \ldots, n$. Henceforth, proposed method of p-value can be computed based on T. Our testing procedure is given by

$$P = \Pr(T \geq t|H_0), \tag{4.3}$$

where t is an observed test statistics.

5. Simulation Studies

We now carry out simulation studies to illustrate the performance of proposed method and compare with two existing procedures: ANOVA and Jonckheere-Terpstra trend test(JT) which is one of the most common nonparametric test for ordered differences among classes. JT tests the null hypothesis that the distribution of the response variable does not differ among classes. One of the drawbacks in that is that this test works well at large in large sample. The original data consists of 6 groups having 24 observations. Each group has 4 random normal variables with different mean from other groups. We assign population mean to each group in increasing order. As the difference in the means between a pair of groups increases by 1, we computed p-values based on each method and illustrate them in the following table. We may simulate the permuation distribution of the test statistics. When computing p-value for proposed method, the data was permuted with about $24!/(4!)^6$ iterations.

Table 5.1 reports that three procedures all have relatively small p-values. Proposed method is more sensitive to increase in the difference of means between groups. On the other hand, ANOVA method has less conservative test statistics (or smaller p-value) than what we expected, which does not mean that the test is justified theoretically, for example, in terms of likelihood based inferences. ANOVA method includes more broad definition of alternative hypothesis than other procedures. As difference increases, p-value for proposed method decreases accordingly, but Jonckheere-Terpstra trend test(JT) does not reflect this aspect. These results are expected in our theory.

6. A Real Data Example

We consider a genomic model in microarray data analysis as an illustration. The crux of data analysis is to identify differentially expressed genes among a huge number of of genes, tested simultaneously, across experimental conditions. Let us introduce data structure in Lobehofer *et al.* (2002). Mitogenesis in breast cancer cells may be stimulated by the steroid hormone estrogen. The cDNA microarray gene expression levels of a hormone-responsive breast cancer epithelial cell line with a mitogenic dose of estrogen without other confounding growth factors in serum were examined. Gene expression changes were measured at 6 time points 1, 4, 12, 24, 36 and 48 hours after estrogen stimulation. The expression levels of DNA replication fork genes stimulated by estrogen, without growth factors in serum, shows that the steroid hormone estrogen plays a important role

Table 6.1. Real data

proposed	ANOVA	JT
0.001	0.002	0.005

of generating Mitogenesis. The data consists of 6 groups with 8 observations per group. Gene expression levels are log-transformed. For our purpose, our dataset contains 4 observation per each time point. Table 6.1 presents p-values for three methods used in previous section.

It was found that we may reject the null hypothesis for that gene, based on three procedures. Note that the p-value for proposed method seems to be smaller than others.

7. Concluding Remarks

In this paper, we propose robust procedure for testing order-restricted inference, which works well even for finite sample. Though asymptotic normality did not hold for small sample size, proposed method with use of permutation theory has better performance than other trend test procedure like Jonckheere-Terpstra test. In other words, proposed method has smaller p-values which demonstrates that the method is more sensitive to order-restricted inference.

References

Andrews, D. F. (1974). A Robust method for multiple linear regression, *Technometrics*, **16**, 523–531.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*. 1, 799–821.

Huber, P. J. (1977). Robust methods of estimation of regression coefficients, *Statistics*, **8**, 41–53.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Huber, P. J. and Dutter, R. (1974). Numerical solution of robust regression problem, In *Proceedings of the Symposium on Computational Statistics*, Physika Verlag, Vienna, 165–172.

Lobenhofer, E. K., Bennett, L., Cable, P. L., Li, L., Bushel, P. R. and Afshari, C. A. (2002). Regulation of DNA replication fork genes by 17Beta-estradiol, *Molecular Endocrinology*, **16**, 1215–1229.

Maronna, R. A., Martin, D. R. and Yohai, J. V. (2006). *Robust Statistics: Theory and Methods*, Wiley Series in Probability and Statistics, New York.

Rousseeuw, P. J. and Yohai, V. J. (1984). Robust regression by means of S-estimators, In *Robust and Nonlinear Time Series Analysis, Lecture Notes in Statistics*, **26**, Springer-Verlag, (Berlin/New York) 256–272.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis, *The Annals of Mathematical Statistics*, **24**, 220–238.

Silvapullé, M. J. (1985). Asymptotic behavior of robust estimators of regression and scale parameters with fixed carriers, *The Annals of Statistics*, **13**, 1490–1497.

Siegel, A. F. (1982). Robust regression using repeated medians, *Biometrika*, **69**, 242–244.