The Korean Journal of Applied Statistics (2011) DOI: 10.5351/KJAS.2011.24.1.137
24*(1), 137–143*

Order-Restricted Inference with Linear Rank Statistics in Microarray Data

Moonsu Kang[1]

[1]National Cancer Center Research Institute

(Received May 2010; accepted November 2010)

Abstract

The classification of subjects with unknown distribution in a small sample size often involves order-restricted constraints in multivariate parameter setups. Those problems make the optimality of a conventional likelihood ratio based statistical inferences not feasible. Fortunately, Roy (1953) introduced union-intersection principle(UIP) which provides an alternative avenue. Multivariate linear rank statistics along with that principle, yield a considerably appropriate robust testing procedure. Furthermore, conditionally distribution-free test based upon exact permutation theory is used to generate p-values, even in a small sample. Applications of this method are illustrated in a real microarray data example (Lobenhofer *et al.*, 2002).

Keywords: Microarray, union-intersection principle, linear rank statistics, permutation.

1. Introduction

In a small sample like DNA microarray data with unknown distribution, order-restricted inference problems often appear in complex ways. To study gene expression patterns across various treatment groups with order constraints weakens the effectiveness of standard statistical inference and as a result, calls for different perspectives (Ghosh, 2003). Nonstandard methods are proposed to classify genes reflecting the concept of order-restricted inference without any assumptions of specific forms (Sen, 2008; Silvapulle and Sen, 2005; Sen, 2006; Sen *et al.*, 2007; Kang and Sen, 2008). Linear rank statistics based on UIP propose the distribution-insensitive clustering of genes. It is also possible to construct a locally most powerful rank test using a suitable rank scores along with UIP, though it is too difficult to construct an optimal test based on Uniformly Most Powerful(UMP) (Sidak *et al.*, 1999; Krishnaiah and Sen, 1985). Gene expression levels are compared of more than 2 groups using exact tests of homogeneity. By using exact permutation distribution theory, a conditionally distribution-free test based upon proposed test statistics is used to generate p-values and as a result is amenable in the setup of a small sample size. It is also computationally tractable and statistically robust.

2. Preliminary Notation

Consider a DNA microarray experiment having expression data on K genes for n mRNA samples. The gene expression data are in a $K \times n$ matrix $X = (X_{k,i})$, with rows corresponding to genes

[1]Research Institute, National Cancer Center, Madu 1-dong, Ilsandong-gu, Goyang-si, Gyeonggi-do 410-769, Korea. E-mail: moonsukang0223@gmail.com

and columns corresponding to individual microarray experiments, where x_{ki} denotes the expression measure of gene k in sample i, $i = 1, \ldots, n, k = 1, \ldots, K$. The expression measures x_{ki}'s are assumed to be preprocessed. For comparing several groups, a general model consists of G (> 2) groups of subjects, each subject having K genes. For simplicity, we assume that there are no missing values resulting in $n_{gk} = n_g$, $\forall k$. Let $n = \sum_{g=1}^{G} n_g$ be the total number of subjects in the pooled sample. A row vector $\boldsymbol{X}_k = (X_{k,1}, X_{k,2}, \ldots, X_{k,n_1}, \ldots, X_{k,n})$ represents the pooled sample at gene k. In this pooled sample, define $\boldsymbol{R}_k = (R_{k,1}, \ldots, R_{k,n_1}, R_{k,n_1+1}, \ldots, R_{k,n})$, where $R_{k,i}$ is the rank of $X_{k,i}$ in the pooled sample among all the n observations in the k^{th} gene.

3. Linear Rank Statistics with UIP

We want to find out the true profile of a gene to one of a specified set of candidate profiles. Without loss of generality, we focus on monotone increasing pattern among more than 2 groups. Let $\mu_{k,i} = E(X_{k,i})$ denote the mean expression level of the k^{th} gene in the i^{th} observation. Let μ_{gk} is the mean expression level of the k^{th} gene in the g^{th} group. For the k^{th} gene (or position), we can formulate H_{0k} *vs.* H_{1k} as below.
$H_{0k} : \mu_{1k} = \mu_{2k} = \cdots = \mu_{Gk}$ *vs.* $H_{1k} : \mu_{1k} \leq \mu_{2k} \leq \cdots \leq \mu_{Gk}$, where $\boldsymbol{\mu}_k = (\mu_{1k}, \ldots, \mu_{Gk})'$.
The $(G-1) \times G$ matrix is given by

$$\boldsymbol{A} = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$

These hypotheses can be restated as the following two hypotheses.

$$H_{0k} : \boldsymbol{\theta}_k = \boldsymbol{A}\boldsymbol{\mu}_k = \bigcap_{j=1}^{G-1} H_{0jk} = \mathbf{0} \qquad vs. \qquad H_{1k} : \boldsymbol{\theta}_k = \boldsymbol{A}\boldsymbol{\mu}_k = \bigcup_{j=1}^{G-1} H_{1jk} \geq \mathbf{0},$$

where $H_{0jk} : \theta_{jk} = \mu_{j+1,k} - \mu_{j,k} = 0$ *vs.* $H_{1jk} : \theta_{jk} = \mu_{j+1,k} - \mu_{j,k} \geq 0$. These hypotheses are written in terms of finite UIP. However, an infinite UIP will be formulated as well. These hypotheses can be restated as the following two hypotheses. For a given $\boldsymbol{a}$,

$$H_{0k} : \boldsymbol{\theta}_k = \boldsymbol{A}\boldsymbol{\mu}_k = \bigcap_{\boldsymbol{a} \in \Re^{+G}} H_{0\boldsymbol{a}k} = \mathbf{0} \qquad vs. \qquad H_{1k} : \boldsymbol{\theta}_k = \boldsymbol{A}\boldsymbol{\mu}_k = \bigcup_{\boldsymbol{a} \in \Re^{+G}} H_{1\boldsymbol{a}k} \geq \mathbf{0},$$

where $H_{0\boldsymbol{a}k} : \boldsymbol{a}'\boldsymbol{\theta}_k = 0$ *vs.* $H_{1\boldsymbol{a}k} : \boldsymbol{a}'\boldsymbol{\theta}_k \geq 0$. The UIP assumes that for testing $H_{0\boldsymbol{a}k}$ *vs.* $H_{1\boldsymbol{a}k}$, we have an optimal test. However, the underlying density of gene expression levels $X_{k,i}, i = 1, \ldots, n, k = 1, \ldots, K$ are completely unknown with unknown variance. In this framework, it is hard to construct either an optimal test based on UMP or a similar test using Uniformly Most Powerful Invariant(UMPI). In these senses, non-parametrics yield robust statistical inference procedures that are distribution free (Huber and Ronchetti, 1981). Fortunately, the null hypothesis H_{0k} is a hypothesis of invariance (under suitable groups of transformation that map the sample space onto itself). Then it is possible to construct a test for $H_{0\boldsymbol{a}k}$ *vs.* $H_{1\boldsymbol{a}k}$, that is, the locally most powerful rank test(LMPR) test for each $\boldsymbol{a}$. By definition, a test is LMPR if among the class of rank test, it is UMP for H_0 against a class $H_{1\epsilon}$ of alternatives that are indexed by a parameter $\triangle$, such that $0 < \triangle < \epsilon$, $\epsilon > 0$ (Sidak *et al.*, 1999; Silvapulle and Sen, 2005). LMPR properties may not be

available for restricted alternatives (Sidak *et al.*, 1999; Krishnaiah and Sen, 1985). However, UIP-based LMPR test can handle such a problem. Even though each sample size n_g differs by group, all the $n(=\sum_{g=1}^{G} n_g)$ observations $\boldsymbol{X}_k = (X_{k,1}, X_{k,2}, \ldots, X_{k,n_1}, \ldots, X_{k,n})$ for each gene k in the pooled sample are i.i.d r.v's under the null hypothesis. Under the null hypothesis of homogeneity, the joint distribution of n observations for each gene k, remains invariant under any permutation. This permutation distribution can be obtained by considering every possible $n!$ permutations of the pooled sample observations among G groups. Hence, conditionally distribution-free tests can be constructed by an appeal to this permutational invariance. We denote this conditional probability law by P_n. For each gene k, define a multivariate linear rank statistics T_{gk}, $g = 1, \ldots, G$, $k = 1, \ldots, K$ as follow. For a suitable rank scores $a(k)$, assuming $\bar{c}_n = 1/n \sum_{i=1}^{n} c_{ig} = 0$,

$$T_{gk} = \sum_{i=1}^{n} (c_{ig} - \bar{c}_n)\, a(R_{k,i}) = \sum_{i=1}^{n} c_{ig}\, a(R_{k,i}), \tag{3.1}$$

where

$$c_{ig} = \begin{cases} \dfrac{1}{n_g}, & \text{if } \; i = \displaystyle\sum_{l=1}^{g-1} n_l + 1, \ldots, \sum_{l=1}^{g} n_l, \\ 0, & \text{otherwise} \end{cases}$$

and $\boldsymbol{T} = (T_{1k}, \ldots, T_{Gk})'$. The mean of T_{gk} is given by

$$\begin{aligned} E_{P_n}(T_{gk}) &= (E_{P_n}(a(R_{k,i}))) \sum_{i=1}^{n} (c_{ig} - \bar{c}_n) \\ &= \left(\frac{1}{n} \sum_{i=1}^{n} a(R_{k,i}) \right) \left(\sum_{i=1}^{n} c_{ig} \right) \\ &= 0. \end{aligned}$$

The variance of T_{gk} is given by

$$\begin{aligned} V_{P_n}(T_{gk}) &= E_{P_n}(T_{gk})^2 \\ &= V_{P_n}(a(R_{k,i})) \sum_{i=1}^{n} (c_{ig})^2 + \sum_{1 \le i \ne i' \le n} (c_{ig})(c_{i'g}) E_{P_n}(a(R_{k,i}) a(R_{k,i'})) \\ &= \left(\frac{1}{n} \sum_{i=1}^{n} a^2(R_{k,i}) \right) \sum_{i=1}^{n} (c_{ig})^2 + \sum_{1 \le i \ne i' \le n} (c_{ig})(c_{i'g}) \left(-\frac{1}{n(n-1)} \sum_{i=1}^{n} a^2(R_{k,i}) \right) \\ &= \left(\frac{1}{n-1} \sum_{i=1}^{n} a^2(R_{k,i}) \right) \cdot \left(\frac{n-1}{n} \sum_{i=1}^{n} (c_{ig})^2 - \frac{1}{n} \sum_{1 \le i \ne i' \le n} (c_{ig})(c_{i'g}) \right) \\ &= (\boldsymbol{A}_n^2) \left(\frac{n - n_g}{n \cdot n_g} \right), \end{aligned}$$

where

$$\frac{(n-1)}{n} \sum_{i=1}^{n} (c_{ig})^2 - \frac{1}{n} \sum_{1 \le i \ne i' \le n} c_{ig} c_{i'g} = \frac{(n-1)}{n} \sum_{i=1}^{n} (c_{ig})^2 - \frac{1}{n} \left(\left(\sum_{i=1}^{n} c_{ig} \right)^2 - \sum_{i=1}^{n} (c_{ig})^2 \right)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n}(c_{ig})^2 - \frac{1}{n}\sum_{i=1}^{n}(c_{ig})^2 \\
&= \frac{1}{n_g} - \frac{1}{n} \\
&= \frac{(n-n_g)}{n \cdot n_g}
\end{aligned}
$$

and $\boldsymbol{A}_n^2 = 1/(n-1)\sum_{i=1}^{n} a^2(R_{k,i})$.
For $1 \le g \ne g' \le G$, the covariance of T_{gk} and $T_{g'k}$ is

$$
\begin{aligned}
\mathrm{Cov}_{P_n}(T_{gk}, T_{g'k}) &= E_{P_n}(T_{gk}, T_{g'k}) \\
&= E_{P_n}\left(\sum_{i=1}^{n} c_{ig}a(R_{k,i}) \sum_{i=1}^{n} c_{ig'}a(R_{k,i})\right) \\
&= E_{P_n}\left(\sum_{i=\sum_{g=1}^{g-1} n_g - 1}^{\sum_{g=1}^{g} n_g} c_{ig}a(R_{k,i}) \sum_{i'=\sum_{g=1}^{g'-1} n_g - 1}^{\sum_{g=1}^{g'} n_g} c_{i'g'}a(R_{k,i})\right) \\
&= \left(\sum_{i=\sum_{g=1}^{g-1} n_g - 1}^{\sum_{g=1}^{g} n_g} c_{ig}\right)\left(\sum_{i'=\sum_{g=1}^{g'-1} n_g - 1}^{\sum_{g=1}^{g'} n_g} c_{i'g'}\right)(E_{P_n}(a(R_{k,i})a(R_{k,i'}))) \\
&= \boldsymbol{A}_n^2\left(-\frac{1}{n}\right).
\end{aligned}
$$

Hence, the permutation variance of $\boldsymbol{T}_k$ is given by

$$\mathbf{V}_k = \mathrm{Var}(\mathbf{T}_k) = \boldsymbol{A}_n^2 \boldsymbol{C}_n,$$

where

$$
\begin{aligned}
\boldsymbol{C}_n &= \sum_{i=1}^{n}(\boldsymbol{c}_i - \bar{c}_n\mathbf{1}_n)(\boldsymbol{c}_i - \bar{c}_n\mathbf{1}_n)' = \left(\frac{\delta_{gg'}n - n_g}{n \cdot n_g}\right), \\
\delta_{gg'} &= \begin{cases} 1, & \text{if } \ 1 \le g = g' \le G, \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}
$$

$\boldsymbol{c}_i = (c_{i1}, \ldots, c_{iG})'$, a $G\times$ 1 matrix $\mathbf{1}_n = (1, \ldots, 1)'$ and

$$
\boldsymbol{C}_n = \begin{pmatrix}
\frac{n-n_1}{n\cdot n_1} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots \\
-\frac{1}{n} & \frac{n-n_2}{n\cdot n_2} & -\frac{1}{n} & -\frac{1}{n} & \cdots \\
-\frac{1}{n} & -\frac{1}{n} & \frac{n-n_3}{n\cdot n_3} & -\frac{1}{n} & \cdots \\
\vdots & \vdots & \vdots & \ddots & \cdots \\
-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-n_G}{n\cdot n_G}
\end{pmatrix}.
$$

If we define $\boldsymbol{T}_k$ in terms of the vector $\boldsymbol{c}_i$, $\boldsymbol{T}_k$ is $\sum_{i=1}^{n}(\boldsymbol{c}_i - \bar{c}_n\mathbf{1}_n)a(R_{k,i})$. The mean of $\boldsymbol{T}_k$ is

$$
\begin{aligned}
E_{P_n}(\boldsymbol{T}_k) &= (\boldsymbol{c}_i - \bar{c}_n\mathbf{1}_n)E_{P_n}(a(R_{k,i})) \\
&= \mathbf{0}.
\end{aligned}
$$

For $1 \leq k \leq k' \leq K$, the covariance matrix of $\boldsymbol{T}_k$ and $\boldsymbol{T}_{\boldsymbol{k'}}$ is

$$\mathrm{Cov}_{P_n}(\boldsymbol{T}_k, \boldsymbol{T}_{k'}) = \boldsymbol{C}_n \times \sqsubseteq_{k,k'},$$

where $\sqsubseteq_{k,k'} = 1/(n-1)\sum_{i=1}^{n}(a(R_{k,i}) - \bar{a}_n)(a(R_{k',i}) - \bar{a}_n)$ and $\bar{a}_n = \sum_{i=1}^{n} a(i)$. The matrix $\boldsymbol{V}_n (= ((\sqsubseteq_{k,k'})))$ is a P_n-invariant and known matrix. Let $\boldsymbol{T}_n = (\boldsymbol{T}_1, \ldots, \boldsymbol{T}_K)'$. Define the $G \times K$ matrix $\boldsymbol{T}_n^0 = \sum_{i=1}^{n}(\boldsymbol{c}_i - \bar{c}_n \mathbf{1}_n)\boldsymbol{a}_n(\boldsymbol{R}_i)$ as the transpose matrix of $\boldsymbol{T}_n$, where $\boldsymbol{a}_n(\boldsymbol{R}_i) = (a_{n_1}(R_{1,i}), \ldots, a_{n_K}(R_{K,i}))'$. By using the concept of a multivariate linear rank statistics, the mean and the covariance matrix of $\boldsymbol{T}_n^0$ are defined as below.

$$E_{P_n}(\boldsymbol{T}_n^0) = \mathbf{0}_{G \times K},$$
$$\mathrm{Cov}_{P_n}(\boldsymbol{T}_n^0) = \boldsymbol{C}_n \bigotimes \boldsymbol{V}_n.$$

4. Order-Restricted Inference

Given the invariance of $\boldsymbol{V}_n$ under P_n, we adapt the UIP to formulate a rank test for $H_{0k} : \boldsymbol{\theta}_k = \mathbf{0}$ *vs.* $H_{1k} : \boldsymbol{\theta}_k \geq \mathbf{0}$. Let $\boldsymbol{Z}_k = \boldsymbol{A}\boldsymbol{T}_k$ and $\boldsymbol{S}_k = \boldsymbol{A}\boldsymbol{V}_k\boldsymbol{A}'$. Let $\wp = \{1, \ldots, G-1\}$, and for every $a : \emptyset \subseteq a \subseteq \wp$, let a' be its complement and $|a|$ its cardinality. For each a, partition $\boldsymbol{Z}_k$ and $\boldsymbol{S}_k$ as

$$\boldsymbol{Z}_k = \begin{pmatrix} \boldsymbol{Z}_{ka} \\ \boldsymbol{Z}_{ka'} \end{pmatrix}, \qquad \boldsymbol{S}_k = \begin{pmatrix} \boldsymbol{S}_{kaa} & \boldsymbol{S}_{kaa'} \\ \boldsymbol{S}_{ka'a} & \boldsymbol{S}_{ka'a'} \end{pmatrix}$$

and write

$$\boldsymbol{Z}_{ka:a'} = \boldsymbol{Z}_{ka} - \boldsymbol{S}_{kaa'}\boldsymbol{S}_{ka'a'}^{-1}\boldsymbol{Z}_{ka'},$$
$$\boldsymbol{S}_{ka:a'} = \boldsymbol{S}_{kaa} - \boldsymbol{S}_{kaa'}\boldsymbol{S}_{ka'a'}^{-1}\boldsymbol{S}_{ka'a}.$$

The test statistics for the k^{th} gene is

$$L_k = \sum_{\emptyset \subseteq a \subseteq \wp} I(\boldsymbol{Z}_{ka:a'} > \mathbf{0}, \boldsymbol{S}_{ka'a'}^{-1}\boldsymbol{Z}_{ka'} \leq \mathbf{0})(n\boldsymbol{Z}'_{ka:a'}\boldsymbol{S}_{kaa:a'}^{-1}\boldsymbol{Z}'_{ka:a'}), \tag{4.1}$$

rejecting the null hypothesis for large positive values. By reference to the $n!/(n_1! \cdots n_G!)$ conditionally (permutationally) equally likely realizations of $\boldsymbol{R}_k$ for each k, we can enumerate $\boldsymbol{T}_k$ (and hence L_k); this generates the exact conditional (permutational) null distribution P_n of L_k, so that the test based on L_k is conditionally distribution free(CDF). Now p-value can be computed as below.

$$P_k = \Pr(L_k \geq l_k), \tag{4.2}$$

where L_k is a test statistic from the permuted distribution and l_k is an observed test statistic. The behavior of L_k under alternatives depends on the stochastic ordering of $\boldsymbol{\mu}_k$ and these statistics may not be exact distribution-free nor have identical probability laws. However, for every $i < i', X_{k,i'} - X_{k,i}$ has a distribution tilted to the right so that

$$E\{L_k | H_{1k}\} \geq 0, \quad k = 1, \ldots, K.$$

This motivates us to use tests based on L_k using the right hand side critical region (Sen, 2008, 2006; Sen *et al.*, 2007). A proper multiple testing procedure may be applied to the set of dependent p-values. The procedure is used to determine which gene has a monotone increasing pattern among the groups. The choice of rank scores $a(k)$ determine if a test statistic is locally most powerful (Sidak *et al.*, 1999; Krishnaiah and Sen, 1985). For example, the Wilcoxon rank test is LMPR when the density is logistic and the normal score test is LMPR when the density is normal. For the test for the linear trend, the Jonckheere test might be tenable (Odeh, 1972). However, without the linear ordering or the logistic density, the LMPR property might not work for the Jonckheere test.

Table 5.1. Linear rank statistics using different scores

α	π_0	Test statistics	Storey	bh95
0.05	0.3	Linear rank statistics (Uniform score)	0.049	0.089
		... (Normal)	0.030	0.035
		... (Logistic)	0.084	0.003
	0.5	Linear rank statistics (Uniform score)	0.066	0.041
		... (Normal)	0.046	0.040
		... (Logistic)	0.112	0.042
	0.7	Linear rank statistics (Uniform score)	0.098	0.052
		... (Normal)	0.050	0.049
		... (Logistic)	0.170	0.089
0.01	0.3	Linear rank statistics (Uniform score)	0.043	0.010
		... (Normal)	0.007	0.007
		... (Logistic)	0.008	0.008
	0.5	Linear rank statistics (Uniform score)	0.058	0.009
		... (Normal)	0.009	0.008
		... (Logistic)	0.113	0.035
	0.7	Linear rank statistics (Uniform score)	0.088	0.010
		... (Normal)	0.010	0.009
		... (Logistic)	0.174	0.040

5. Numerical Study

Mitogenesis in hormone-responsive 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, show that the steroid hormone estrogen plays a important role of generating Mitogenesis (Lobenhofer *et al.*, 2002). For the purpose of illustration, the data set in Lobenhofer *et al.* (2002) is analyzed. The data consists of 1900 genes measured at 6 time points with 8 observations ($n = 8$) each time point. Gene expression levels are log-transformed. However, the dataset to which we applied the analysis contains 1000 genes and 5 time points (1, 4, 12, 24, 36 hours after estrogen stimulation), at which each group has $4, 3, 2, 2$ and 1 observations, respectively. The pattern of interest is whether or not mean gene expression levels have monotone nondecreasing profile over time. We then express these in term of inequalities among the expected expression levels at 5 time points. Based on $12!/(4!3!2!2!)$ permutationally equally likely realizations for each gene k, we enumerate a test statistics L_k. Linear rank statistics along with corresponding p-values were computed based on different score statistics; uniform(Wilcoxon)(U), Normal(N), and logistic(L), respectively. Storey's FDR (storey) and Benjamini and Hochberg (1995)'s FDR(bh95) were computed for those p-values (Storey, 2002, 2003; Storey *et al.*, 2004; Benjamini and Hochberg, 1995). Table 5.1 displays comparison of FDR procedures with application to breast data, where π_0 was defined as the proportion of true null hypotheses. FDR procedures were computed at a preassigned level $\alpha = 0.05$ and 0.01. The study was performed for $\pi_0 = 0.3$, 0.5 and 0.70. FDR procedures such as storey and bh95 produced relatively the same results at all levels of α and π_0. Linear rank statistics with normal scores control the FDR (less than α) at any preassigned level α under all configurations of π_0. On the other hand, those with a uniform score or logistic score failed to control the FDR. The choice of normal score statistics achieved LMPR property in the data.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society: Series B*, **57**, 289–300.

Ghosh, D. (2003). Penalized discriminant methods for the classification of tumors from microarray experiments, *Bioinformatics*, **59**, 992–1000.

Huber, P. J. and Ronchetti, E. M. (1981). *Robust Statistics*, Wiley Series in Probability and Statistics, New York.

Kang, M. and Sen, P. K. (2008). Kendall tau type rank statistics in genomic data, *Applications of Mathematics*, **3**, 207–221.

Krishnaiah, P. R. and Sen, P. K. (1985). *Handbook of Statistics 4: Nonparametric Methods*, North-Holland, Netherlands.

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 17-estradiol, *Molecular Endocrinology*, **16**, 467–471.

Odeh, R. E. (1972). On the power of Jonckheere's k-sample test against ordered alternatives, *Biometrika*, **59**, 467–471.

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

Sen, P. K. (2006). Robust statistical inference for high dimensional data models with application to genomics, *Austrian Journal of Statistics*, **35**, 197–214.

Sen, P. K. (2008). Kendall's tau in high-dimensional genomic parsimony, *Institute of Mathematical Statistics, Collection Series*, **3**, 251–266.

Sen, P. K., Tsai, M. T. and Jou, Y. S. (2007). High-dimension, low sample size perspectives in constrained statistical inference, *Journal of the American Statistical Association*, **102**, 686–694.

Sidak, Z., Sen, P. K. and Hajek, J. (1999). *Theory of Rank Tests, Second Edition(Probability and Mathematical Statistics)*, Academic Press, San Diego.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference: Inequality, Order, and Shape Restriction*, Wiley-Interscience, New York.

Storey, J. (2002). A direct approach to false discovery rates, *Journal of the Royal Statistical Society: Series B*, **64**, 479–498.

Storey, J. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value, *Annals of Statistics*, **3**, 2013–2035.

Storey, J., Taylor, J. E. and Siegmund, D. (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach, *Journal of the Royal Statistical Society, Series B*, **66**, 187–205.