Statistical Inference for Peakedness Ordering Between Two Distributions

Myongsik Oh [1]

Abstract

The concept of dispersion is intrinsic to the theory and practice of statistics. A formulation of the concept of dispersion can be obtained by comparing the probability of intervals centered about a location parameter, which is peakedness ordering introduced first by Birnbaum (1948). We consider statistical inference concerning peakedness ordering between two arbitrary distributions. We propose nonparametric maximum likelihood estimator of two distributions under peakedness ordering and a likelihood ratio test for equality of dispersion in the sense of peakedness ordering.

Keywords: Isotonic regression, order restricted inference, peakedness, stochastic ordering.

1. Introduction

The concept of dispersion is intrinsic to the theory and practice of statistics. A formulation of the concept of dispersion can be obtained by comparing the probability of intervals centered about a location parameter, typically mean or median of the distribution. It seems natural to interpret dispersion in terms of the distance of a random variable X from a location parameter μ, that is, the magnitude of $|X - \mu|$. One might be interest in comparing such dispersions of two or more distributions.

Following Birnbaum (1948), a random variable X is said to be more peaked about $a \in \mathbf{R}^1$ than Y about $b \in \mathbf{R}^1$ if, for all $x \geq 0$,

$$F(x+a) - F(-x+a) \geq G(x+b) - G(-x+b). \tag{1.1}$$

where F and G are the distribution functions of X and Y, respectively.

Proschan (1965), Karlin (1968), Bickel and Lehmann (1979), Shaked (1982), and Schweder (1982), among others, have considered properties and connections with other orderings. The statistical inference concerning peakedness ordering, however, has received little attention. We note that (1.1) is equivalent to

$$P(|X-a| \leq t) \geq P(|Y-b| \leq t) \text{ for every } t \in (0, \infty).$$

We see so-called stochastic ordering between two random variables, $|X-a|$ and $|Y-b|$. We say a random variable Y is more dispersed about ν than X about μ if $|Y-\nu|$ is stochastically larger than $|X-\mu|$. Based on this fact, El Barmi and Rojo (1996) considered the likelihood ratio tests for peakedness in multinomial populations. The usage of their result is, however, restricted because of the following several reasons. First they provided the test for equality

[1] Associate Professor, Department of Statistics, Pusan University of Foreign Studies, Pusan 608-738

of two multinomial parameters against peakedness ordering rather than the equality in the sense of peakedness ordering. This will be discussed fully later in Section 3. Second, they considered only the case that the location parameter is assumed to be at center of the distribution, although peakedness ordering can be defined about at any location parameter. This seems to be based upon the fact that dispersion ordering can be explained by peakedness ordering about median. Finally they considered only multinomial populations. Even though their result can be extended to the arbitrary distribution functions, the extension to both estimation and testing problem requires some modifications, which also discussed later.

In this article we consider estimation of general distribution functions under peakedness ordering. The estimation procedure does not require that the location parameters should be at the center of the distributions. As widely known some estimator for one-sample problem may not satisfy consistency. See Rojo and Samaniego (1991). Here we only consider two-sample problem here, which is of particular interest in practice. We assume the location parameters which we will use in comparing peakedness are assumed to be known. In Section 2, maximum likelihood estimation of two general distribution functions under peakedness ordering is discussed. Though its estimation procedure is basically the same as given by El Barmi and Rojo (1996), it requires some modifications. In section 3, the likelihood ratio test for equality in peakedness against peakedness ordering in discrete setting is discussed. The difference between test of El Barmi and Rojo (1996) is fully discussed. In section 4, real data is analyzed for illustrative purpose.

2. Estimation of Distribution Functions

Let F and G be distribution functions with known location parameter μ_x and μ_y. Without loss of generality we assume that $\mu_x = \mu_y = \mu$ and both random samples are observed at $-\infty < t_1 < t_2 < \cdots < t_k < +\infty$. Let δ_{1i} (δ_{2i}) be the number of observations from F (G) distribution at t_i. The ordinary nonparametric maximum likelihood estimator can be obtained by finding F and G which maximize

$$\prod_{i=1}^{k}\{F(t_i) - F(t_i-)\}^{\delta_{1i}}\{G(t_i) - G(t_i-)\}^{\delta_{2i}}. \tag{2.1}$$

Now our problem is to find F and G which maximize (2.1) subject to (1.1). This can be achieved easily by a reparametrization.

Consider imaginary data points $t_{i+k} = 2\mu - t_i$ with $\delta_{1,i+k} = \delta_{2,i+k} = 0$ for $i = 1, \ldots, k$. Let $-\infty < s_1 < s_2 < \cdots < s_l \le s_{l+1} < \ldots < s_{2l} < \infty$ be ordered distinct values of $t_i, i = 1, \ldots, 2k$, except the case that μ is equal to one of t_i's so that $\mu = s_l = s_{l+1}$. We observe that $l \le k$, $s_l < \mu < s_{l+1}$ (or possibly $s_l = \mu = s_{l+1}$) and $s_i = 2\mu - s_{2l-i+1}$, for $i = 1, \ldots, l$.

For $j = 1, \ldots, 2l$, let

$$d_{1j} = \sum_{i \in \{1,2,\ldots,2k\}: t_i = s_j} \delta_{1i} \quad \text{and} \quad d_{2j} = \sum_{i \in \{1,2,\ldots,2k\}: t_i = s_j} \delta_{2i}.$$

Then (2.1) can be rewritten as

$$\prod_{i=1}^{2l}\{F(s_i) - F(s_i-)\}^{d_{1i}}\{G(s_i) - G(s_i-)\}^{d_{2i}}. \tag{2.2}$$

The peakedness ordering (1.1) can be expressed as, for $j = 0, \ldots, l-1$,

$$F(s_{l+1+j}) - F(s_{l-j}-) \ge G(s_{l+1+j}) - G(s_{l-j}-). \tag{2.3}$$

Now we are going to find F and G which maximize (2.2) subject to (2.3).

Let $\theta_{11} = F(s_{l+1}) - F(s_l-)$, $\theta_{21} = G(s_{l+1}) - G(s_l-)$, $\phi_{11} = (F(s_l) - F(s_l-))/(F(s_{l+1}) - F(s_l-))$ and $\phi_{21} = (G(s_l) - G(s_l-))/(G(s_{l+1}) - G(s_l-))$, and for $j = 1, \ldots, l-1$,

$$\begin{aligned} \theta_{1,j+1} &= F(s_{l-j}) - F(s_{l-j}-) + F(s_{l+j+1}) - F(s_{l+j+1}-), \\ \theta_{2,j+1} &= G(s_{l-j}) - G(s_{l-j}-) + G(s_{l+j+1}) - G(s_{l+j+1}-), \\ \phi_{1,j+1} &= \frac{F(s_{l-j}) - F(s_{l-j}-)}{\theta_{1,j+1}}, \quad \phi_{2,j+1} = \frac{G(s_{l-j}) - G(s_{l-j}-)}{\theta_{2,j+1}}. \end{aligned}$$

Then (2.2) becomes

$$\prod_{i=1}^{2}\prod_{j=0}^{l-1}\left[\theta_{i,j+1}^{d_{i,l-j}+d_{i,l+j+1}} \cdot \phi_{i,j+1}^{d_{i,l-j}}(1-\phi_{i,j+1})^{d_{i,l+j+1}}\right] \tag{2.4}$$

with convention $0^0 = 1$, and the restriction (2.3) is equivalent to

$$\sum_{j=0}^{\ell}\theta_{1,j+1} \geq \sum_{j=0}^{\ell}\theta_{2,j+1} \text{ for } \ell = 0, \ldots, l-1, \tag{2.5}$$

$$0 \leq \phi_{i,j+1} \leq 1 \text{ for } i = 1, 2, j = 0, \ldots, l-1. \tag{2.6}$$

Noting that $\sum_{j=0}^{l-1}\theta_{1,j+1} = \sum_{j=0}^{l-1}\theta_{2,j+1} = 1$, we see a stochastic ordering between θ_{1j}'s and θ_{2j}'s. Moreover, restrictions (2.5) and (2.6) do not relate θ's and ϕ's. This means that the maximization of (2.4) can be achieved by maximizing two parts (one involves θ's only and the other ϕ's only) separately.

First consider the estimation of ϕ's, which is just a binomial problem. The ML estimate, $\phi^*_{i,j+1}$, for $i = 1, 2$, $j = 0, \ldots, l-1$, of ϕ_{ij}, is given by $d_{i,l-j}/(d_{i,l-j} + d_{i,l+j+1})$ provided that $d_{i,l-j} + d_{i,l+j+1} > 0$. Next, let $E_{\mathbf{w}}(\mathbf{x}|A)$ denote the projection of $\mathbf{x}$ onto A provided it exists and is unique. See Robertson, Wright and Dykstra (1988) for details of projection theory. Let $\mathbf{d}_1 = (d_{1l} + d_{1,l+1}, d_{1,l-1} + d_{1,l+2}, \ldots, d_{11} + d_{1,2l})$, $\mathbf{d}_2 = (d_{2l} + d_{2,l+1}, d_{2,l-1} + d_{2,l+2}, \ldots, d_{21} + d_{2,2l})$, $m = \sum_{j=0}^{l-1}(d_{1,l-j} + d_{1,l+j+1})$, and $n = \sum_{j=0}^{l-1}(d_{2,l-j} + d_{2,l+j+1})$. Robertson and Wright (1981) gave the ML estimate of two-sample nultinomial parameters under stochastic ordering using Fenchel duality. See also Barlow and Brunk (1972). Using their notation the ML estimates of θ's under (2.5) are given by for $j = 0, \ldots, l-1$,

$$\begin{aligned} \theta^*_{1,j+1} &= (d_{1,l-j} + d_{1,l+j+1}) \cdot E_{\mathbf{d}_1}\left(\frac{m\mathbf{d}_1 + n\mathbf{d}_2}{(m+n)\mathbf{d}_1}|\mathcal{A}\right)_j, \\ \theta^*_{2,j+1} &= (d_{2,l-j} + d_{2,l+j+1}) \cdot E_{\mathbf{d}_2}\left(\frac{m\mathbf{d}_1 + n\mathbf{d}_2}{(m+n)\mathbf{d}_2}|\mathcal{I}\right)_j, \end{aligned}$$

where $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \ldots, x_k + y_k)$, $\mathbf{x}/\mathbf{y} = (x_1/y_1, \ldots, x_k/y_k)$, $\mathcal{I} = \{\mathbf{x} \in \mathbf{R}^k : x_1 \leq x_2 \leq \cdots \leq x_k\}$, $\mathcal{A} = \{\mathbf{x} \in \mathbf{R}^k : -\mathbf{x} \in \mathcal{I}\}$, and $E(\cdot)_j$ denotes the jth component of $E(\cdot)$. Since random samples are folded about a given location it is likely to have missing components in $\mathbf{d}_1$ and $\mathbf{d}_2$. For the estimation procedure when missing values are present, see Lee (1987).

Now for $j = 0, \ldots, l-1$

$$\begin{aligned} F^*(s_{l-j}) - F^*(s_{l-j}-) &= \theta^*_{1,j+1}\phi^*_{1,j+1}, \quad & F^*(s_{l+j+1}) - F^*(s_{l+j+1}-) &= \theta^*_{1,j+1}(1-\phi^*_{1,j+1}), \\ G^*(s_{l-j}) - G^*(s_{l-j}-) &= \theta^*_{2,j+1}\phi^*_{2,j+1}, \quad & G^*(s_{l+j+1}) - G^*(s_{l+j+1}-) &= \theta^*_{2,j+1}(1-\phi^*_{2,j+1}) \end{aligned}$$

It follows immediately from Robertson and Wright (1981) that under the hypothesis of stochastic ordering the ML estimators of θ's are strongly consistent. Since ϕ's are also strongly consistent, F^* and G^* are strongly consistent too.

Next we consider the estimation of distribution when two distributions are equal in the sense of peakedness ordering, which means that, for $j = 0, \ldots, l-1$,

$$\begin{aligned} & F(s_{l-j}) - F(s_{l-j}-) + F(s_{l+j+1}) - F(s_{l+j+1}-) \\ & = G(s_{l-j}) - G(s_{l-j}-) + G(s_{l+j+1}) - G(s_{l+j+1}-) \end{aligned} \tag{2.7}$$

Using the same reparametrization scheme we can show that (2.7) is equivalent to

$$\theta_{1,j+1} = \theta_{2,j+1} \text{ for } j = 0, \ldots, l-1, \tag{2.8}$$

with (2.6). Then the likelihood is consist of two part; one is multinomial likelihood and the others are product binomial. The ML estimate, $\theta^\circ_{1,j+1}$, for $j = 0, \ldots, l-1$, of $\theta_{1,j+1}$ is given by

$$\theta^\circ_{1,j+1} = \theta^\circ_{2,j+1} = \frac{d_{1,l-j} + d_{1,l+j+1} + d_{2,l-j} + d_{2,l+j+1}}{m+n}.$$

Note that $\phi^\circ_{i,j+1} = \phi^*_{i,j+1}$, for $j = 0, \ldots, l-1$. Hence we have, for $j = 0, \ldots, l-1$,

$$\begin{aligned} F^\circ(s_{l-j}) - F^\circ(s_{l-j}-) &= \theta^\circ_{1,j+1}\phi^\circ_{1,j+1}, & F^\circ(s_{l+j+1}) - F^\circ(s_{l+j+1}-) &= \theta^\circ_{1,j+1}(1-\phi^\circ_{1,j+1}), \\ G^\circ(s_{l-j}) - G^\circ(s_{l-j}-) &= \theta^\circ_{1,j+1}\phi^\circ_{2,j+1}, & G^\circ(s_{l+j+1}) - G^\circ(s_{l+j+1}-) &= \theta^\circ_{1,j+1}(1-\phi^\circ_{2,j+1}). \end{aligned}$$

The ML estimator under the equality assumption in the sense of peakedness ordering is also strongly consistent.

3. Hypothesis Testing

Assume that F and G have support on the fixed set $(t_1, \ldots, t_k)$ and that each point has positive probability, so that we are concerned with discrete distributions with common support. In this section we consider the likelihood ratio test for equality in a sense of peakedness ordering of two distributions against peakedness ordering. The hypotheses are, for all $x > 0$ and given μ,

$$\begin{aligned} H_0 &: F(x+\mu) - F(-x+\mu) = G(x+\mu) - G(-x+\mu) \\ H_1 &: F(x+\mu) - F(-x+\mu) \geq G(x+\mu) - G(-x+\mu) \\ &\quad \text{with strict inequality holds for at least one } x. \end{aligned}$$

After reparametrization given random samples, H_0 is related to (2.8) and H_1 to (2.5). The test rejects H_0 for large value of

$$T = 2m\sum_{j=0}^{l-1}\hat{\theta}_{1,j+1}(\log\theta^*_{1,j+1} - \log\theta^\circ_{1,j+1}) + 2n\sum_{j=0}^{l-1}\hat{\theta}_{2,j+1}(\log\theta^*_{2,j+1} - \log\theta^\circ_{2,j+1}),$$

where $m = \sum_{i=1}^{2l} d_{1i}$, $n = \sum_{i=1}^{2l} d_{2i}$,

$$\begin{aligned} \hat{\theta}_{1,j+1} &= F_m(s_{l-j}) - F_m(s_{l-j}-) + F_m(s_{l+j+1}) - F_m(s_{l+j+1}-), \\ \hat{\theta}_{2,j+1} &= G_n(s_{l-j}) - G_n(s_{l-j}-) + G_n(s_{l+j+1}) - G_n(s_{l+j+1}-), \end{aligned}$$

F_m and G_n are empirical distributions of F and G, respectively.

Now we need to find the asymptotic null distribution of test statistic T to find a critical value. It follows immediately from Robertson and Wright (1981) that

$$\lim_{m,n\to\infty} P[T \geq t] = \sum_{i=1}^{l} P_s(i,l;\boldsymbol{\theta})P[\chi^2_{l-i} \geq t] \qquad (3.1)$$

provided that $m/(m+n) \to a \in (0,1)$. We note that $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_l)$, and that $P(i,l;\boldsymbol{\theta})$ is the probability that $E_{\boldsymbol{\theta}}(W|\mathcal{A})$ has exactly i distinct levels, where $\mathbf{W} = (W_1, \ldots, W_l)$ and $W_1, \ldots, W_l$ are independent normal variables with zero means and variances $\theta_1^{-1}, \ldots, \theta_l^{-1}$, respectively. This is so-called level probability. To find a critical value we need to know the value of $\boldsymbol{\theta}$, which is unknown because the null distribution does not specify the common distributions.

Robertson and Wright (1983) showed that the equal-weights null distribution of chi-bar-square test statistics provide reasonable approximation for the case of unequal sample sizes if the sample size are not too different for the simple ordering. So we recommend to use equal-weight level probabilities for finding a critical value or p-value. The equal weight level probability can be obtained by the following recursive relationship.

$$\begin{aligned} P(1,l) &= \frac{1}{l}, \quad p(l,l) = \frac{1}{l!}, \\ P(i,l) &= \frac{1}{l}P(i-1,l-1) + \frac{l-1}{l}P(i,l-1) \text{ for } i = 2, \ldots, l-1. \end{aligned}$$

Finally one might use the least favorable distribution for a conservative test. The least favorable test is given by

$$\sup_{H_0} \lim_{m,n\to\infty} P[T \geq t] = \frac{1}{2}(P[\chi^2_l \geq t] + P[\chi^2_{l-1} \geq t]).$$

4. Example

In this section, a data of lung cancer mortality in South Australia is analyzed to illustrate the proposed method. This data has been used in El Barmi and Rojo (1996). Apparently, a peak is seen at age interval 65-69 for male and at age 70-74 for female. We are going to compare peakedness about at age interval 65-69. Under H_0 the MLE for parameters are

$$\begin{aligned} \mathbf{p}^\circ &= (0.0141, 0.0282, 0.0762, 0.1205, 0.1750, 0.2024, 0.1873, 0.1289, 0.0674) \\ \mathbf{q}^\circ &= (0.0141, 0.0282, 0.0957, 0.1174, 0.1510, 0.2024, 0.2114, 0.1320, 0.0478) \end{aligned}$$

and under H_1

$$\begin{aligned} \mathbf{p}^* &= (0.0086, 0.0288, 0.0750, 0.1240, 0.1721, 0.2084, 0.1842, 0.1326, 0.0663) \\ \mathbf{q}^* &= (0.0384, 0.0256, 0.1024, 0.1024, 0.1598, 0.1811, 0.2238, 0.1152, 0.0512). \end{aligned}$$

We have $T = 3.979238$ with $l = 6$. The p-value is 0.3471 for equal-weights approximation. The result shows no solid evident that male is more peaked about age 65-69 than female.

5. Concluding remarks

The likelihood ratio tests concerning stochastic ordering for general distributions have been studied by several authors. Dykstra, Madsen and Fairbanks (1982) and Franck (1984) are among others. Their results can be applied with some modifications if needed. We will not pursuit this subject in this paper.

In this paper we did not consider the case of unknown location parameter, such as median, which is more frequently encountered situation in practice. We may also impose symmetry assumption, which is quite common in nonparametric setting. These problems are, however, nontrivial extension of the procedure given in this article.

REFERENCES

Barlow, R. E. and Brunk, H. D. (1972). The isotonic regression problem and its dual. *Journal of the American Statistical Association*, **67**, 140-147.

Bickel, P. J. and Lehmann, .E. L. (1979). Descriptive statistics for nonparametric models IV. Spread, in *Contributions to Statistics*, Jaroslaw Hajek Memorial Volume, J. Jureckova, ed. Reidel, Dordrecht, 33-40.

Birnbaum, Z.W. (1948). On random variables with comparable peakedness. *Annals of Mathematical Statistics*, **19**, 76-81.

Dykstra, R. L., Madsen, R. W. and Fairbanks, K. (1982). A nonparametric likelihood ratio test. *Journal of Statistical Computation and Simulation*, **18**, 247-64.

El Barmi, H. and Rojo. J. (1996) Likelihood ratio tests for peakedness in multinomial populations *Journal of Nonparametric Statistics* **7**, 221-237.

Franck, W. E. (1984). A likelihood ratio test for stochastic ordering. *Journal of the American Statistical Association*, **79**, 686-691.

Karlin, S. (1968). *Total Positivity*, Stanford University Press, CA.

Lee, C. I. C. (1987). Maximum Likelihood Estimates for Stochastically Ordered Multinomial Populations with Fixed and Random Zeros. *Foundation of Statistical Inference*, (I. B. MacNeill and G. J. Umphrey eds.), 189-197.

Proschan, F. (1965). Peakedness of distributions of convex combinations, *Annals of Mathematical Statistics*, **36**, 1703-1706.

Robertson, T., and F. T. Wright (1981). Likelihood ratio test for and against stochastic ordering between multinomial populations, *The Annals of Statistics*, **9**, 1248-1257.

Robertson, T. and Wright, F. T. (1983). On approximation of the level probabilities and associated distributions in order restricted inference, *Biometrika*, **70**, 597-606.

Robertson, T.; Wright, F. T.; Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, Chichester.

Rojo, J. and Samaniego, F. (1991). On nonparametric maximum likelihood estimation of a distribution uniformly stochastically smaller than a standard. *Statistics and Probability Letters* , **11**, 267-271.

Schweder, T. (1982). On the dispersion of mixtures, *Scandinavian Journal of Statistics*, **9**, 165-169.

Shaked, M. (1982). Dispersive ordering of distributions, *Journal of Applied Probability*, **19**, 310-320.