Statistics & Probability Letters 13 (1992) 203–208
North-Holland

19 February 1992

Rank order probabilities for the dispersion problem

Subhash C. Kochar
Indian Statistical Institute, New Delhi 110 016, India

George Woodworth
University of Iowa, Iowa City, IA 52242, USA

Received December 1990
Revised February 1991

Abstract: First we consider the two-sample dispersion problem. Let X and Y be independent random variables of continuous type with distribution functions F and G, respectively, satisfying $G(x)=F(\sigma x)$ for all x where $\sigma>0$ and $F(x)=1-F(-x)$, for all x. Let X_1, X_2 (Y_1, Y_2) be two independent copies of X (Y) and define $p(\sigma)=P[X_1 \leqslant Y_1 \leqslant Y_2 \leqslant X_2]$. It is shown that $p(\sigma)$ is nondecreasing in σ implying $P[X_1 \leqslant Y_1 \leqslant Y_2 \leqslant X_2] \geqslant P[Y_1 \leqslant X_1 \leqslant X_2 \leqslant Y_2]$ for $\sigma \geqslant 1$. Next, we consider the bivariate scale problem where (X, Y) has joint c.d.f. $F(x, y)$ satisfying $F(x, y)=H(x, \sigma y)$, where $H(x, y)$ is bivariate symmetric and also $H(x, y)=\bar{H}(-x, -y)$ for every (x, y). Sufficient conditions are given under which $p(\sigma)$ is nondecreasing in σ. These results have been applied to study the properties of some nonparametric tests for the two-sample dispersion and paired-sample problems.

Keywords: Mood's test for dispersion, U-statistic, bivariate symmetry.

1. Introduction

Let $X=(X_1,\dots,X_m)$ and $Y=(Y_1,\dots,Y_n)$ be two independent random samples drawn from absolutely continuous populations with densities $f(\cdot)$ and $g(\cdot)$, respectively. $F(\cdot)$ and $G(\cdot)$ denote the corresponding distribution functions. We use the notation $\bar{F}$ for $1-F$, $W=(W_1,\dots,W_{m+n})$ denotes the order statistics of the combined sample $(X, Y)=(X_1,\dots,X_m, Y_1,\dots,Y_n)$, and $Z=(Z_1,\dots,Z_{m+n})$ is a random vector of zeros and ones whose ith component, Z_i, is 0 if W_i comes from $f(\cdot)$ and 1 if W_i comes from $g(\cdot)$.

Let $z=(z_1,\dots,z_{m+n})$ be a fixed vector of zeros and ones; we define the transpose of z by $z^t=(z_1^t,\dots,z_{m+n}^t)$, where $z_i^t=z_{m+n+1-i}$. $P(z)=P[Z=z]$ denotes the probability of the rank order z.

In a series of papers Savage (1956, 1957) and Savage, Sobel and Woodworth (1966) gave many interesting results on rank order probabilities for the two-sample problem. They mainly considered cases when the two distributions are ordered by monotone likelihood ratio property or when the distributions are symmetric differing only in their location parameters.

In Section 2, we consider the two-sample scale problem when both the distributions are symmetric about the origin. Some monotonicity results are obtained for the rank order probabilities and applications of these results to some testing problems are considered. In the last section, the bivariate scale problem has been considered.

0167-7152/92/$05.00

2. The two-sample scale problem

Lemma 2.1. *Let f and g be symmetric about the origin, then $P[z] = P[z^t]$.*

Proof. Proof is simple and hence omitted. □

In particular, if $G(x) = F(\sigma x)$, $\sigma > 0$, and f is symmetric about zero, then $P[0011] = P[1100]$ and $P[0101] = P[1010]$, when $m = n = 2$. The only information about σ is contained in the arrangements [0110] and [1001].

Kochar and Gupta (1986) have shown that when $m = n$, the Mood's statistic for the two-sample-scale problem is a linear function of the U-statistic

$$S_{22} = \left[\binom{m}{2}\binom{n}{2}\right]^{-1} \sum \phi(X_i, X_j; Y_k, Y_l), \tag{2.1}$$

where the summation is over $1 \leqslant i < j \leqslant m$, $1 \leqslant k < l \leqslant n$ and where

$$\phi(x_i, x_j; y_k, y_l) = \begin{cases} 1 & \text{if } z = [0110], \\ -1 & \text{if } z = [1001], \\ 0 & \text{otherwise}. \end{cases} \tag{2.2}$$

As a result, $E[S_{2,2}] = P[0110] - P[1001]$. To study the power properties of the Mood's test, we discuss the monotonicity properties of the functional $p(\sigma) = P[0110]$ in the following theorem.

Theorem 2.2. *Let $G(x) = F(\sigma x)$, $\sigma > 0$, where f is symmetric about the origin. Then $p(\sigma) = P[0110]$ is nondecreasing in σ.*

Proof.

$$p(\sigma) = 2P[X_1 \leqslant (Y_1, Y_2) \leqslant X_2]$$

$$= 2\int\int_{x \leqslant y} [G(y) - G(x)]^2 \, \mathrm{d}F(x) \, \mathrm{d}F(y) \tag{2.3}$$

$$= 2\int\int_{y \leqslant x} [G(x) - G(y)]^2 \, \mathrm{d}F(x) \, \mathrm{d}F(y). \tag{2.4}$$

From (2.3) and (2.4), we get

$$p(\sigma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [G(x) - G(y)]^2 \, \mathrm{d}F(x) \, \mathrm{d}F(y)$$

$$= 2\int_{-\infty}^{\infty} G^2(x) \, \mathrm{d}F(x) - 2\left[\int_{-\infty}^{\infty} G(x) \, \mathrm{d}F(x)\right]^2$$

$$= 2P[YYX] - 2P^2[YX]$$

$$= 2P[YYX] - \tfrac{1}{2} \quad \text{(from Lemma 2.1)}.$$

Now

$$P[YYX] = \int_{-\infty}^{\infty} F^2(\sigma x)\,\mathrm{d}F(x)$$
$$= \int_{-\infty}^{0} F^2(\sigma y)\,\mathrm{d}F(y) + \int_{0}^{\infty} F^2(\sigma x)\,\mathrm{d}F(x)$$

(making the transformation $x = -y$, in the first integral and using symmetry)

$$= \int_{0}^{\infty} \left[\bar{F}^2(\sigma x) + F^2(\sigma x)\right] \mathrm{d}F(x).$$

Since for $x \geqslant 0$, the integrand $[F^2(\sigma x) + \bar{F}^2(\sigma x)]$ is nondecreasing in σ, the required result follows. □

Corollary 2.3. *Under the conditions of the above theorem,* $P[0110] - P[1001]$ *is nondecreasing in* σ. *This implies*:

(i) $P[0110] > P[1001]$ *for* $\sigma > 1$. (2.5)

(ii) $\phi(X_1, X_2; Y_1, Y_2)$ *is stochastically increasing in* σ.

Proof. As $G(x) = F(\sigma x)$, $X \overset{\text{dist}}{=} \sigma Y$. Hence

$$p(\sigma) = P[0110] = 2P[\sigma Y_1 \leqslant (Y_2, Y_3) \leqslant \sigma Y_4]$$

and

$$p[1001] = 2P[Y_1 \leqslant (\sigma Y_2, \sigma Y_3) \leqslant Y_4] = 2P\left[\frac{Y_1}{\sigma} \leqslant (Y_2, Y_3) \leqslant \frac{Y_4}{\sigma}\right] = p\left(\frac{1}{\sigma}\right).$$

Since $p(\sigma)$ is nondecreasing in σ, it follows that $p(\sigma) - p(1/\sigma) = P[0110] - P[1001]$ is nondecreasing in σ. (i) follows from this since for $\sigma > 1$, $p(\sigma) > p(1/\sigma)$, (ii) follows from the fact that statistic ϕ takes only three values $+1$,0 and -1 with the property that for $\sigma_1 < \sigma_2$, $P_{\sigma_1}[\phi = 1] \leqslant P_{\sigma_2}[\phi = 1]$ and $P_{\sigma_1}[\phi = -1] \geqslant P_{\sigma_2}[\phi = -1]$. □

It follows from the properties of U-statistics and this corollary that the test based on large values of $S_{2,2}$ (or equivalently on the Mood's statistic in case $m = n$) is consistent for testing H_0: $\sigma = 1$ against the alternative H_1: $\sigma > 1$ when both the distributions are symmetric about the same point. It will be a virtue if one can extend the result (ii) of this corollary to prove the monotonicity property of the power function of the Mood's test for $m, n > 2$. However, we are unable to establish this result.

Remark. It can be shown that a sufficient condition for (2.5) to hold for non-symmetric distributions on $[0, \infty)$ is that $\log f$ is convex.

Kochar and Gupta (1986) generalized the Mood's test as follows. Let c and d be fixed integers such that $2 \leqslant c \leqslant m$, $2 \leqslant d \leqslant n$ and define

$$Q_{c,d}(x_1, \ldots, x_c, y_i, \ldots, y_d) = \begin{cases} 1 & \text{if } z_1 = z_{c+d} = 0, \\ -1 & \text{if } z_1 = z_{c+d} = 1, \\ 0 & \text{otherwise}. \end{cases}$$

Let $S_{c,d}$ be the corresponding U-statistic. We have the following desirable property for the measure of dispersion $q_{c,d}(\sigma) = E[S_{c,d}]$.

Theorem 2.4. *Let $G(x)=F(\sigma x)$, where f is symmetric about the origin. Then*

$$q_{c,d}(\sigma)=E[S_{c,d}]=P[z_1=z_{c+d}=0]-P[z_1=z_{c+d}=1]$$

is nondecreasing in $\sigma>0$. In particular, $q_{c,c}(\sigma)>0$ for $\sigma>1$.

Proof. It is easy to see that

$$\begin{aligned} q_{c,d} &= P[\max(Y_1,\dots,Y_d)\leqslant \max(X_1,\dots,X_c)]-P[\min(Y_1,\dots,Y_d)\leqslant \min(X_1,\dots,X_c)] \\ &= c\left[\int_{-\infty}^{\infty}\left[G^d(x)F^{c-1}(x)+\bar{G}^d(x)\bar{F}^{c-1}(x)\right]\mathrm{d}F(x)-1\right. \qquad (2.6) \\ &= 2c\int_0^{\infty}\left[F^d(\sigma x)F^{c-1}(x)+\bar{F}^d(\sigma x)\bar{F}^{c-1}(x)\right]\mathrm{d}F(x)-1 \qquad (2.7) \end{aligned}$$

as the integrand in (2.6) is symmetric about zero. The required result follows since for $x\geqslant 0$, the integrand in (2.7) is a nondecreasing function of σ.

For $\sigma>1$, $q_{c,c}(\sigma)>q_{c,c}(1/\sigma)=-q_{c,c}(\sigma)$. Hence the result. □

This result establishes the consistency of the test $S_{c,c}$ for the one-sided two-sample scale problem in case of symmetric distributions.

3. The bivariate scale model

We extend the two-sample scale model to the bivariate situation, allowing for dependence between X and Y. The joint c.d.f. $F(x, y)$ of X and Y is of the form

$$F(x,y)=H(x,\sigma y) \qquad (3.1)$$

where $H(x, y)$ is a bivariate c.d.f. satisfying

$$H(x,y)=H(y,x)\quad \forall(x,y) \qquad (3.2)$$

and $\sigma>0$, is a constant.

In case of independence, this model reduces to the two-sample scale problem as considered earlier. We give some sufficient conditions under which $p(\sigma)=P_\sigma[X_1<(Y_1, Y_2)<X_2]$ is nondecreasing in σ, where (X_i, Y_i), $i=1, 2$, are independent observations from F. The following lemmas are easy to prove, hence the proofs are ommitted.

Lemma 3.1. *Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed. Then*

$$P[0110]=2P[(Y_1,Y_2)\leqslant X_1]-P[Y_2\leqslant X_1, Y_1\leqslant X_2]-P^2[Y_1\leqslant X_1]. \quad \square$$

Lemma 3.2. *Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed according to a distribution which is symmetric about* (0, 0), *that is,*

$$P[X\leqslant x, Y\leqslant y]=P[X\geqslant -x, Y\geqslant -y]\quad \forall(x,y). \qquad (3.3)$$

Then the following are true:

(a) $P[Y_2\leqslant X_1, Y_1\leqslant X_2]=P[X_2\leqslant Y_1, X_1\leqslant Y_2]$.

(b) $P[Y\leqslant X]=P[X\leqslant Y]=\frac{1}{2}$.

(c) $P[0110]-P[1001]=2\left[P[(Y_1,Y_2)\leqslant X_1]-P[(X_1,X_2)\leqslant Y_1]\right]$. □

Lemma 3.3. *Let* (X_1, Y_1) *and* (X_2, Y_2) *be i.i.d. observations from a distribution F satisfying* (3.1) *and* (3.2). *Then*

$$P_\sigma[(X_1, X_2) \leqslant Y_1] = \eta(1/\sigma)$$

where

$$\eta(\sigma) = P_\sigma[(Y_1, Y_2) \leqslant X_1]. \quad \square \tag{3.4}$$

Theorem 3.4. *Let* (X_1, Y_1) *and* (X_2, Y_2) *be i.i.d. observations from a distribution F satisfying* (3.1) *and* (3.3). *Then a sufficient condition for* $\eta(\sigma)$ *to be nondecreasing in* σ *is*

$$P[Y \leqslant 0 \mid X = x] \geqslant \tfrac{1}{2} \quad \textit{for } x \geqslant 0. \tag{3.5}$$

Proof. Let $H_1(x) = H(x, \infty)$, $H_2(x) = H(\infty, x)$ and $h_i = H_i'$, $i = 1, 2$.

$$\begin{aligned}\eta(\sigma) &= P_\sigma[Y_1 \leqslant X_2, Y_2 \leqslant X_2]\\ &= \sigma\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x_2} h(x_2, \sigma y)\,\mathrm{d}y\right]H_2(\sigma x_2)\,\mathrm{d}x_2\\ &= \sigma\left[\int_{-\infty}^{0}\left\{\int_{-\infty}^{x_2} h(x_2, \sigma y)\,\mathrm{d}y\right\}H_2(\sigma x_2)\,\mathrm{d}x_2 + \int_{0}^{\infty}\left\{\int_{-\infty}^{x_2} h(x_2, \sigma y)\,\mathrm{d}y\right\}H_2(\sigma x_2)\,\mathrm{d}x_2\right].\end{aligned}$$

Making the substitution $x = -x_2^*$ and $y = -y^*$ in the first integral; and using (3.3), we get

$$\eta(\sigma) = \int_0^\infty\left[\int_{-\infty}^{\sigma x} h(x, y)\,\mathrm{d}y\right]H_2(\sigma x) + \left[\int_{\sigma x}^{\infty} h(x, y)\,\mathrm{d}y\right]\bar{H}_2(\sigma x)\,\mathrm{d}x. \tag{3.6}$$

Differentiating (3.6) with respect to σ and assuming that differentiation under the integral sign is permissible, we get

$$\begin{aligned}\eta'(\sigma) &= \int_0^\infty x\left[H_2(\sigma x) - \bar{H}_2(\sigma x)\right]h(x, \sigma x)\,\mathrm{d}x\\ &\quad + \int_0^\infty x h_2(\sigma x)\left\{\int_{-\infty}^{\sigma x} h(x, y)\,\mathrm{d}y - \int_{\sigma x}^{\infty} h(x, y)\,\mathrm{d}y)\right\}\,\mathrm{d}x.\end{aligned}$$

For $x \geqslant 0$, the integrand in the first integral is nonnegative.

A sufficient condition for the second integral to be nonnegative for all $x \geqslant 0$, $\sigma > 0$ is

$$\int_{-\infty}^{\sigma x} h(x, y)\,\mathrm{d}y - \int_{\sigma x}^{\infty} h(x, y)\,\mathrm{d}y \geqslant 0 \quad \text{for } x \geqslant 0,$$

That is,

$$\int_{-\infty}^{\sigma x}\frac{h(x, y)}{h_1(x)}\,\mathrm{d}y \geqslant \tfrac{1}{2} \quad \text{for } x \geqslant 0. \tag{3.7}$$

This will hold for all $x \geqslant 0$ and $\sigma > 0$ if (3.5) holds. $\square$

Corollary 3.5. *If H satisfies* (3.2), *then under the conditions of Theorem* 3.4,

$$P[0110] - P[1001] = 2[\eta(\sigma) - \eta(1/\sigma)]$$

is nondecreasing in σ. *In particular*, $P[0110] \geqslant P[1001]$. $\square$

Proof follows from Lemmas 3.2, 3.3 and Theorem 3.4.

Remark. If F is such that $P[Y \leqslant y \mid X = x]$ is nondecreasing in x for each fixed y, we say Y is stochastically decreasing in X (see Barlow and Proschan, 1981). Since $P[Y \leqslant 0 \mid X = 0] = \frac{1}{2}$ if (X, Y) is symmetric about (0, 0), (3.5) will hold if Y is stochastically decreasing in X.

Example. Bivariate normal distribution with $E(X) = E(Y)$, $\text{Var}(X) = \sigma^2\ \text{Var}(Y)$, and $\text{Cov}(X, Y) \leqslant 0$ satisfies the conditions of Theorem 3.4 and Corollary 3.5.

References

Barlow, R.E. and F. Proschan (1981), *Statistical Theory of Reliability and Life Testing: Probability Models* (To Begin With, Spring Field, MD).

Kochar, S.C. and R.P. Gupta (1986), Some competitors of the Mood test for the two-sample scale problem, *Comm. Statist.—Theory Methods* **15**, 231–239.

Savage, I.R. (1956), Contributions to the theory of rank order statistics: the two sample case, *Ann. Math. Statist.* **27**, 590–615.

Savage, I.R. (1957), Contributions to the theory of rank order statistics; the trend case, *Ann. Math. Statist.* **28**, 968–977.

Savage, I.R., M. Sobel and G. Woodworth (1966), Fine structure of the orderings of probabilities of rank orders in the two-sample case, *Ann. Math. Statist.* **37**, 98–112.