ELSEVIER

Statistics & Probability Letters 46 (2000) 257–261

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

# On dispersive ordering between order statistics in one-sample and two-sample problems

Baha-Eldin Khaledi, Subhash Kochar *

*Stath-Math Unit, Indian Statistical Institute, 7, SJS Sansanwal Marg, New Delhi-110016, India*

Received February 1999; received in revised form March 1999

**Abstract**

Let $X_{i:n}$ denote the $i$th-order statistic of a random sample of size $n$ from a continuous distribution with distribution function $F$. It is shown that if $F$ is a decreasing failure rate (DFR) distribution, then $X_{i:n}$ is *less dispersed* than $X_{j:m}$ for $i \leqslant j$ and $n-i \geqslant m-j$. Let $Y_{j:m}$ denote the $j$th-order statistic of a random sample of size $m$ from a continuous distribution $G$. We prove that if $F$ is less dispersed than $G$ and either $F$ or $G$ is DFR, then $X_{i:n}$ is less dispersed than $Y_{j:m}$ for $i \leqslant j$ and $n-i \geqslant m-j$. © 2000 Elsevier Science B.V. All rights reserved

*Keywords:* Hazard rate ordering; DFR distribution; Exponential distribution

## 1. Introduction

Order statistics play a central role in statistics and a lot of work has been done in the literature on different aspects of this problem. For a glimpse of this, see the two volumes of papers on this topic by Balakrishnan and Rao (1998a,b).

Throughout this paper we shall be assuming that all random variables under consideration are nonnegative and their distribution functions are strictly increasing on $(0,\infty)$ or on some interval of $(0,\infty)$. We shall use "increasing" ("decreasing") to mean "nondecreasing" ("nonincreasing").

One of the basic criteria for comparing variability in probability distributions is that of dispersive ordering. Let $X$ and $Y$ be two random variables with distribution functions $F$ and $G$, respectively. Let $F^{-1}$ and $G^{-1}$ be their right continuous inverses (quantile functions). We say that $X$ is less *dispersed* than $Y$ ($X \overset{\text{disp}}{\leqslant} Y$) if $F^{-1}(\beta) - F^{-1}(\alpha) \leqslant G^{-1}(\beta) - G^{-1}(\alpha)$, for all $0 \leqslant \alpha \leqslant \beta \leqslant 1$. This means that the difference between any two quantiles of $F$ is smaller than the difference between the corresponding quantiles of $G$. A consequence of $X \overset{\text{disp}}{\leqslant} Y$ is that $|X_1 - X_2|$ is stochastically smaller than $|Y_1 - Y_2|$ and which in turn implies $\mathrm{var}(X) \leqslant \mathrm{var}(Y)$ as well as $E[|X_1 - X_2|] \leqslant E[|Y_1 - Y_2|]$, where $X_1, X_2 (Y_1, Y_2)$ are two independent copies of $X$ $(Y)$. For details, see Section 2.B of Shaked and Shanthikumar (1994).

* Corresponding author.

0167-7152/00/$ - see front matter 
PII: S0167-7152(99)00110-8

Let $\overline{F}$ and $\overline{G}$ denote the survival functions and $r_F$ and $r_G$ denote the hazard rate functions of random variables $X$ and $Y$, respectively. We say that $X$ is smaller than $Y$ in the hazard rate ordering (denoted by $X \leqslant_{\text{hr}} Y$) if $\overline{G}(x)/\overline{F}(x)$ is nondecreasing in $x$, which is equivalent to $r_F(x) \geqslant r_G(x)$ for all $x$, if $X$ and $Y$ are continuous random variables. Bagai and Kochar (1986) noted a connection between hazard rate ordering and dispersive ordering. They observed that if $X \leqslant_{\text{hr}} Y$ and either $F$ or $G$ is DFR (decreasing failure rate), then $X \overset{\text{disp}}{\leqslant} Y$.

Let $X_1, \ldots, X_n$ be a random sample of size $n$ from a continuous distribution with distribution function $F$ and let $X_{i:n}$ denote the $i$th-order statistic of this random sample. David and Groeneveld (1982) proved that if $F$ is a DFR distribution, then $\text{var}(X_{i:n}) \leqslant \text{var}(X_{j:n})$ for $i \leqslant j$. Kochar (1996) strengthened this result to prove that under the same condition, $X_{i:n} \overset{\text{disp}}{\leqslant} X_{j:n}$ for $i \leqslant j$.

In this paper we further extend these results to compare the variabilities of order statistics based on samples of possibly different sizes. We consider both, the one-sample as well as the two-sample problems. It is proved in the next section that if $F$ is DFR, then $X_{i:n} \overset{\text{disp}}{\leqslant} X_{j:m}$ for $i \leqslant j$ and $n-i \geqslant m-j$. Let $Y_{j:m}$ denote the $j$th-order statistic of a random sample of size $m$ taken from a probability distribution with continuous distribution function $G$. It is proved in the next section that if $X \overset{\text{disp}}{\leqslant} Y$ and if either $F$ or $G$ is DFR, then $X_{i:n} \overset{\text{disp}}{\leqslant} Y_{j:m}$ for $i \leqslant j$ and $n-i \geqslant m-j$. This result also holds if, instead, we assume that $X \leqslant_{\text{hr}} Y$ and either $F$ or $G$ is DFR.

We shall be using the following results to prove the main results in the next section.

**Theorem 1.1** (Saunders, 1984). *The random variable $X$ satisfies $X \overset{\text{disp}}{\leqslant} X + Y$ for any random variable $Y$ independent of $X$ if and only if $X$ has a log-concave density.*

**Theorem 1.2** (Hickey, 1986). *Let $Z$ be a random variable independent of random variables $X$ and $Y$. If $X \overset{\text{disp}}{\leqslant} Y$ and $Z$ has a log-concave density, then*

$$X + Z \overset{\text{disp}}{\leqslant} Y + Z$$

This result leads to the following corollary.

**Corollary 1.1.** *Let $X_1, X_2; Y_1, Y_2$ be independent random variables with log-concave densities. Then $X_i \overset{\text{disp}}{\leqslant} Y_i$ for $i = 1, 2$ implies*

$$X_1 + X_2 \overset{\text{disp}}{\leqslant} Y_1 + Y_2. \tag{1.1}$$

**Proof.** Since $X_2$ is independent of $X_1$ and $Y_1$ and it has a log-concave density, it follows from Theorem 1.2 that $X_1 \overset{\text{disp}}{\leqslant} Y_1$ implies

$$X_1 + X_2 \overset{\text{disp}}{\leqslant} Y_1 + X_2. \tag{1.2}$$

Using the same argument it follows that $X_2 \overset{\text{disp}}{\leqslant} Y_2$ implies

$$Y_1 + X_2 \overset{\text{disp}}{\leqslant} Y_1 + Y_2. \tag{1.3}$$

Combining (1.2) and (1.3), we get the required result. □

## 2. Main results

Boland et al. (1998) proved that if $X_1,\dots,X_n$ is a random sample of size $n$ from an exponential distribution, then $X_{i:n} \overset{\text{disp}}{\leqslant} X_{j:n}$ for $i \leqslant j$. In the next lemma we extend this result to the case when the order statistics are based on samples of possibly different sizes.

**Lemma 2.1.** *Let $X_{i:n}$ be the ith-order statistic of a random sample of size n from an exponential distribution. Then*

$$X_{i:n} \overset{\text{disp}}{\leqslant} X_{j:m} \quad \textit{for } i \leqslant j \quad \textit{and} \quad n-i \geqslant m-j. \tag{2.1}$$

**Proof.** Suppose we have two independent random samples, $X_1,\dots,X_n$ and $X'_1,\dots,X'_m$ of sizes $n$ and $m$ from an exponential distribution with failure rate $\lambda$. The $i$th-order statistic, $X_{i:n}$ can be written as a convolution of the sample spacings as

$$\begin{aligned} X_{i:n} &= (X_{i:n} - X_{i-1:n}) + \cdots + (X_{2:n} - X_{1:n}) + X_{1:n} \\ &\overset{\text{dist}}{=} \sum_{k=1}^{i} E_{n-i+k}, \end{aligned} \tag{2.2}$$

where for $k=1,\dots,i$, $E_{n-i+k}$ is an exponential random variable with failure rate $(n-i+k)\lambda$. It is a well-known fact that $E_{n-i+k}$'s are independent. Similarly, we can express $X'_{j:m}$ as

$$X'_{j:m} \overset{\text{dist}}{=} \sum_{k=1}^{j} E'_{m-j+k}, \tag{2.3}$$

where again for $k=1,\dots,j$, $E'_{m-j+k}$ is an exponential random variable with failure rate $(m-j+k)\lambda$ and $E'_{m-j+k}$'s are independent. It is easy to verify that $E_{n-i+1} \overset{\text{disp}}{\leqslant} E'_{m-j+1}$ for $n-i \geqslant m-j$.

Since the class of distributions with log-concave densities is closed under convolutions (cf. Dharmadhikari and Joeg-dev, 1988, p. 17), it follows from the repeated applications of Corollary 1.1 that

$$\sum_{k=1}^{i} E_{n-i+k} \overset{\text{disp}}{\leqslant} \sum_{k=1}^{i} E'_{m-j+k}. \tag{2.4}$$

Again since $\sum_{k=i+1}^{j} E'_{m-j+k}$, being the sum of independent exponential random variables has a log-concave density and since it is independent of $\sum_{k=1}^{i} E'_{n-i+k}$, it follows from Theorem 1.1 that the RHS of (2.4) is less dispersed than $\sum_{k=1}^{j} E'_{m-j+k}$ for $i \leqslant j$.

That is,

$$X_{i:n} \overset{\text{dist}}{=} \sum_{k=1}^{i} E_{n-i+k} \overset{\text{disp}}{\leqslant} \sum_{k=1}^{j} E'_{m-j+k} \overset{\text{dist}}{=} X'_{j:m}.$$

Since $X_{j:m}$ and $X'_{j:m}$ are stochastically equivalent, (2.1) follows from this. □

The proof of the next lemma can be found in Bartoszewicz (1987).

**Lemma 2.2.** *Let $\phi: \mathbb{R}_+ \to \mathbb{R}_+$ be a function such that $\phi(0)=0$ and $\phi(x)-x$ is increasing. Then for every convex and strictly increasing function $\psi: \mathbb{R}_+ \to \mathbb{R}_+$ the function $\psi\phi\psi^{-1}(x)-x$ is increasing.*

In the next theorem we extend Lemma 2.1 to the case when F is a DFR distribution.

**Theorem 2.1.** *Let $X_{i:n}$ be the ith-order statistic of a random sample of size n from a DFR distribution F. Then*

$$X_{i:n} \overset{\text{disp}}{\leqslant} X_{j:m} \quad \textit{for } i \leqslant j \quad \textit{and} \quad n-i \geqslant m-j.$$

**Proof.** The distribution function of $X_{j:m}$ is $F_{j:m}(x) = B_{j:m}F(x)$, where $B_{j:m}$ is the distribution function of the beta distribution with parameters $(j, m-j+1)$.

Let $G$ denote the distribution function of a unit mean exponential random variable. Then $H_{j:m}(x)=B_{j:m}G(x)$ is the distribution function of the $j$th-order statistic in a random sample of size $m$ from a unit mean exponential distribution. We can express $F_{j:m}$ as

$$F_{j:m}(x) = B_{j:m}GG^{-1}F(x) = H_{j:m}G^{-1}F(x). \tag{2.5}$$

To prove the required result, we have to show that for $i \leqslant j$ and $n-i \geqslant m-j$,

$$F^{-1}_{j:m}F_{i:n}(x) - x \quad \text{is increasing in } x$$

$$\Leftrightarrow F^{-1}GH^{-1}_{j:m}H_{i:n}G^{-1}F(x) - x \quad \text{is increasing in } x. \tag{2.6}$$

By Lemma 2.1, $H^{-1}_{j:m}H_{i:n}(x)-x$ is increasing in $x$ for $i \leqslant j$ and $n-i \geqslant m-j$. Also the function $\psi(x)=F^{-1}G(x)$ is strictly increasing and it is convex if $F$ is DFR. The required result now follows from Lemma 2.2. □

**Remark.** A consequence of Theorem 2.1 is that if we have random samples from a DFR distribution, then

$$X_{i:n+1} \overset{\text{disp}}{\leqslant} X_{i:n} \overset{\text{disp}}{\leqslant} X_{i+1:n+1} \quad \text{for } i = 1, \ldots, n.$$

In the next theorem we establish dispersive ordering between order statistics when the random samples are drawn from different distributions.

**Theorem 2.2.** *Let $X_1, \ldots, X_n$ be a random sample of size n from a continuous distribution F and let $Y_1, \ldots, Y_m$ be a random sample of size m from another continuous distribution G. If either F or G is DFR, then*

$$X \overset{\text{disp}}{\leqslant} Y \Rightarrow X_{i:n} \overset{\text{disp}}{\leqslant} Y_{j:m} \quad \textit{for } i \leqslant j \quad \textit{and} \quad n-i \geqslant m-j. \tag{2.7}$$

**Proof.** Let $F$ be a DFR distribution. The proof for the case when $G$ is DFR is similar. By Theorem 2.1, $X_{i:n} \overset{\text{disp}}{\leqslant} X_{j:m}$ for $i \leqslant j$ and $n-i \geqslant m-j$. Bartoszewicz (1986) proved that if $X \overset{\text{disp}}{\leqslant} Y$ then $X_{j:m} \overset{\text{disp}}{\leqslant} Y_{j:m}$. Combining these we get the required result. □

Since the property $X \leqslant_{\text{hr}} Y$ together with the condition that either $F$ or $G$ is DFR implies that $X \overset{\text{disp}}{\leqslant} Y$, we get the following result from the above theorem.

**Corollary 2.1.** *Let $X_1, \ldots, X_n$ be a random sample of size n from a continuous distribution F and $Y_1, \ldots, Y_m$ be a random sample of size m from another continuous distribution G. If either F or G is DFR, then*

$$X \leqslant_{\text{hr}} Y \Rightarrow X_{i:n} \overset{\text{disp}}{\leqslant} Y_{j:m}.$$

## Acknowledgements

Baha-Eldin Khaledi would like to thank the Indian Council for Cultural Relations, New Delhi, India, Razi University, Kermanshah, Iran and the Ministry of Culture and Higher Education of the Islamic Republic of Iran, Tehran, Iran for arranging scholarships which enabled him to study at the Indian Statistical Institute, New Delhi, India.

## References

Bagai, I., Kochar, S.C., 1986. On tail ordering and comparison of failure rates. Commun. Statist. Theor. Meth. 15, 1377–1388.

Balakrishnan, N., Rao, C.R., 1998a. Handbook of Statistics 16 – Order Statistics: Theory and Methods. Elsevier, New York.

Balakrishnan, N., Rao, C.R., 1998b. Handbook of Statistics 17 – Order Statistics: Applications. Elsevier, New York.

Bartoszewicz, J., 1986. Dispersive ordering and the total time on test transformation. Statist. Probab. Lett. 4, 285–288.

Bartoszewicz, J., 1987. A note on dispersive ordering defined by hazard functions. Statist. Probab. Lett. 6, 13–17.

Boland, P.J., Shaked, M., Shanthikumar, J.G., 1998. Stochastic ordering of order statistics. In: Balakrishnan, N., Rao, C.R. (Eds.), Handbook of Statistics 16 – Order Statistics: Theory and Methods. Elsevier, New York, pp. 89–103.

David, H.A., Groeneveld, R.A., 1982. Measures of local variation in a distribution: Expected length of spacings and variances of order statistics. Biometrika 69 (1), 227–232.

Dharmadhikari, S., Joeg-dev, K., 1988. Unimodality, Convexity and Applications. Academic Press, New York.

Hickey, R.J., 1986. Concepts of dispersion in distributions; A comparative note. J. Appl. Probab. 23, 914–921.

Kochar, S.C., 1996. Dispersive ordering of order statistics. Statist. Probab. Lett. 27, 271–274.

Saunders, D.J., 1984. Dispersive ordering of distributions. Adv. Appl. Probab. 16, 693–694.

Shaked, M., Shanthikumar, J.G., 1994. Stochastic ordering and its applications. Academic Press, San Diego, CA.