

HAZARD RATE ORDERING OF THE LARGEST ORDER STATISTICS FROM GEOMETRIC RANDOM VARIABLES

BARA KIM,* *Korea University*

JEONGSIM KIM,** *Chungbuk National University*

Abstract

Mao and Hu (2010) left an open problem about the hazard rate order between the largest order statistics from two samples of n geometric random variables. Du *et al.* (2012) solved this open problem when $n = 2$, and Wang (2015) solved for $2 \leq n \leq 9$. In this paper we completely solve this problem for any value of n .

Keywords: Hazard rate order; order statistics; parallel system

2010 Mathematics Subject Classification: Primary 60E15

Secondary 62N05

1. Introduction

Order statistics have been extensively studied when the observations are independent and identically distributed (i.i.d.). However, in some practical situations, the observations are not i.i.d. Due to the complexity of the distribution theory in the non-i.i.d. case, this area of research has received little attention.

The largest order statistics can be used in the survival function (or reliability function) of a parallel system. Stochastic comparisons between the largest order statistic from independent and heterogeneous exponential random variables, and the largest order statistic from i.i.d. exponential random variables, have been investigated by many researchers; see, for example, [2]–[5]. The geometric distribution can be regarded as the discrete counterpart of the exponential distribution, therefore it is natural to investigate whether (and how) stochastic comparison results for exponential random variables will also hold for geometric random variables.

Let X_1, \dots, X_n be independent geometric random variables with parameters p_1, \dots, p_n , respectively, i.e.

$$\mathbb{P}(X_i = k) = p_i(1 - p_i)^{k-1}, \quad k = 1, 2, \dots$$

Let Y_1, \dots, Y_n be i.i.d. geometric random variables with common parameter $\bar{p} = (\sum_{i=1}^n p_i)/n$, the arithmetic mean of the p_i 's. Let $X_{n:n}$ and $Y_{n:n}$ be the largest order statistics of X_1, \dots, X_n and Y_1, \dots, Y_n , respectively, i.e.

$$X_{n:n} = \max\{X_1, \dots, X_n\} \quad \text{and} \quad Y_{n:n} = \max\{Y_1, \dots, Y_n\}.$$

In this paper we are interested in the hazard rate ordering between $X_{n:n}$ and $Y_{n:n}$. For a discrete random variable X that takes value in $\mathbb{N} = \{1, 2, \dots\}$, its hazard rate function is

Received 2 November 2017; revision received 16 March 2018.

* Postal address: Department of Mathematics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea.

** Postal address: Department of Mathematics Education, Chungbuk National University, 1 Chungdae-ro, Seowon-gu, Cheongju, Chungbuk, 28644, Korea. Email address: jeongsimkim@chungbuk.ac.kr

defined by

$$h_X(k) = \frac{\mathbb{P}(X = k)}{\mathbb{P}(X \geq k)}, \quad k \in \mathbb{N}.$$

Let X and Y be two (nonnegative) discrete random variables. Then X is said to be smaller than Y in the hazard rate order, denoted by $X \leq_{hr} Y$, if

$$h_X(k) \geq h_Y(k) \quad \text{for all } k \in \mathbb{N}.$$

For more details on stochastic orderings, we refer the reader to Shaked and Shanthikumar [7].

Mao and Hu [6] left the question as to whether $X_{n:n} \geq_{hr} Y_{n:n}$ holds or not as an open problem. Du *et al.* [1] proved that $X_{n:n} \geq_{hr} Y_{n:n}$ when $n = 2$. Wang [8] proved that $X_{n:n} \geq_{hr} Y_{n:n}$ for $2 \leq n \leq 9$. In this paper we prove that $X_{n:n} \geq_{hr} Y_{n:n}$ for any value of n .

2. Main result

In this section we present our main result.

Theorem 1. *Let X_1, \dots, X_n be independent geometric random variables with parameters p_1, \dots, p_n , respectively, and let Y_1, \dots, Y_n be i.i.d. geometric random variables with common parameter $\bar{p} = (\sum_{i=1}^n p_i)/n$. Then, for any value of $n \geq 1$,*

$$X_{n:n} \geq_{hr} Y_{n:n}. \tag{1}$$

Since the distribution function of $X_{n:n}$ is

$$\mathbb{P}(X_{n:n} \leq k) = \prod_{i=1}^n \mathbb{P}(X_i \leq k) = \prod_{i=1}^n (1 - q_i^k),$$

where $q_i = 1 - p_i$, the hazard rate function of $X_{n:n}$ is

$$h_{X_{n:n}}(k) = \frac{\mathbb{P}(X_{n:n} \leq k) - \mathbb{P}(X_{n:n} \leq k - 1)}{1 - \mathbb{P}(X_{n:n} \leq k - 1)} = \frac{\prod_{i=1}^n (1 - q_i^k) - \prod_{i=1}^n (1 - q_i^{k-1})}{1 - \prod_{i=1}^n (1 - q_i^{k-1})}.$$

For $k = 1, 2, \dots$, define a function $h_k : (0, 1)^n \rightarrow [0, \infty)$ by

$$h_k(x_1, \dots, x_n) = \frac{\prod_{i=1}^n (1 - x_i^k) - \prod_{i=1}^n (1 - x_i^{k-1})}{1 - \prod_{i=1}^n (1 - x_i^{k-1})}.$$

Then we have

$$h_{X_{n:n}}(k) = h_k(q_1, \dots, q_n) \quad \text{and} \quad h_{Y_{n:n}}(k) = h_k(\bar{q}, \dots, \bar{q}),$$

where $\bar{q} = (\sum_{i=1}^n q_i)/n$. Without loss of generality, we assume that $1 > p_1 \geq \dots \geq p_n > 0$ and so $0 < q_1 \leq \dots \leq q_n < 1$.

To prove Theorem 1, we need a series of lemmas. These lemmas can be found in the next section.

Proof of Theorem 1. By the definition of hazard rate order, to prove (1), we need to show that

$$h_k(q_1, \dots, q_n) \leq h_k(\bar{q}, \dots, \bar{q}), \quad k = 1, 2, \dots \tag{2}$$

We prove (2) by induction on the number of distinct values in the set $\{q_1, \dots, q_n\}$. If q_1, \dots, q_n all have the same value then (2) holds trivially. As the induction hypothesis, we assume that (2) holds for all q_1, \dots, q_n in which the number of distinct values in the set $\{q_1, \dots, q_n\}$ is less than or equal to $m - 1$ ($m \geq 2$). Suppose that the number of distinct values in the set $\{q_1, \dots, q_n\}$ is m . Let a be the number of indices i such that $q_i = q_1$, and let b be the number of indices i such that $q_i = q_n$. We define \tilde{q}_i as follows. If $m = 2$ then $\tilde{q}_i = \bar{q}$. If $m \geq 3$ then

$$\tilde{q}_i = \begin{cases} \min \left\{ q_{a+1}, q_1 + \frac{b}{a}(q_n - q_{n-b}) \right\} & \text{if } 1 \leq i \leq a, \\ q_i & \text{if } a + 1 \leq i \leq n - b, \\ \max \left\{ q_{n-b}, q_n - \frac{a}{b}(q_{a+1} - q_1) \right\} & \text{if } n - b + 1 \leq i \leq n. \end{cases}$$

Note that

$$\tilde{q}_1 = \dots = \tilde{q}_a \leq \tilde{q}_{a+1} = q_{a+1} \leq \dots \leq q_{n-b} = \tilde{q}_{n-b} \leq \tilde{q}_{n-b+1} = \dots = \tilde{q}_n,$$

and that either $\tilde{q}_a = \tilde{q}_{a+1}$ or $\tilde{q}_{n-b} = \tilde{q}_{n-b+1}$ hold. Therefore, the number of distinct values in the set $\{\tilde{q}_1, \dots, \tilde{q}_n\}$ is either $m - 2$ or $m - 1$. It can be shown that

$$a(\tilde{q}_1 - q_1) = b(q_n - \tilde{q}_n), \tag{3}$$

which implies that $(\tilde{q}_1 + \dots + \tilde{q}_n)/n = \bar{q}$. By the induction hypothesis,

$$h_k(\tilde{q}_1, \dots, \tilde{q}_n) \leq h_k(\bar{q}, \dots, \bar{q}).$$

Therefore, it suffices to show that

$$h_k(q_1, \dots, q_n) \leq h_k(\tilde{q}_1, \dots, \tilde{q}_n).$$

Let

$$\psi(x) = h_k(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax).$$

Then

$$\begin{aligned} \psi'(x) = & -b \sum_{i=1}^a h_{k,i}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax) \\ & + a \sum_{i=n-b+1}^n h_{k,i}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \\ & \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax), \end{aligned} \tag{4}$$

where $h_{k,i}(x_1, \dots, x_n) = (\partial/\partial x_i)h_k(x_1, \dots, x_n)$. Since $\tilde{q}_1 = \dots = \tilde{q}_a$ and $\tilde{q}_{n-b+1} = \dots = \tilde{q}_n$, we have, for $i = 1, \dots, a$,

$$\begin{aligned} h_{k,i}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax) \\ = h_{k,1}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax), \end{aligned}$$

and, for $i = n - b + 1, \dots, n$,

$$\begin{aligned} h_{k,i}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax) \\ = h_{k,n}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax). \end{aligned}$$

Thus, (4) becomes

$$\begin{aligned} \psi'(x) = & ab(h_{k,n}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax) \\ & - h_{k,1}(\tilde{q}_1 - bx, \dots, \tilde{q}_a - bx, \tilde{q}_{a+1}, \dots, \tilde{q}_{n-b}, \tilde{q}_{n-b+1} + ax, \dots, \tilde{q}_n + ax)). \end{aligned}$$

By Lemma 3, $\psi'(x) \leq 0$. Hence,

$$h_k(\tilde{q}_1, \dots, \tilde{q}_n) = \psi(0) \geq \psi\left(\frac{\tilde{q}_1 - q_1}{b}\right) = h_k(q_1, \dots, q_n),$$

where we have used (3) in the last equality. The proof is complete. □

Remark 1. The smallest order statistics can be used in the survival function of a series system. Let $X_{1:n} = \min\{X_1, \dots, X_n\}$ and $Y_{1:n} = \min\{Y_1, \dots, Y_n\}$. It can be easily shown that when $n \geq 1$, $X_{1:n} \leq_{hr} Y_{1:n}$ holds.

3. Lemmas

In this section we state and prove the following three lemmas. The first two lemmas are used in the proof of the last lemma. The last lemma was used in the proof of Theorem 1.

Lemma 1. For $0 < x \leq y < 1$ and $k = 1, 2, \dots$,

$$\begin{aligned} & (1 - x^{k-1})\left(\frac{\partial}{\partial x}(1 - x^k)(1 - y^k) - \frac{\partial}{\partial y}(1 - x^k)(1 - y^k)\right) \\ & \geq y(1 - x^k)\left(\frac{\partial}{\partial x}(1 - x^{k-1})(1 - y^{k-1}) - \frac{\partial}{\partial y}(1 - x^{k-1})(1 - y^{k-1})\right). \end{aligned} \tag{5}$$

Proof. A direct calculation shows that (5) holds when $k = 1, 2$. Now we suppose that $k \geq 3$. For $0 < x \leq y < 1$, let

$$\begin{aligned} f(x, y) = & k(1 - x^{k-1})(x^{k-1} - y^{k-1} + x^{k-1}y^{k-1}(x - y)) \\ & - (k - 1)(1 - x^k)y(x^{k-2} - y^{k-2} + x^{k-2}y^{k-2}(x - y)) \end{aligned}$$

and

$$g(x, y) = \frac{(\partial/\partial x)f(x, y)}{x^{k-3}}.$$

Then we have

$$\begin{aligned} g(x, y) = & -(k - 1)(k - 2)y(1 - y^{k-1}) + [(k - 1)y^{k-1} + (k - 1)k(1 - y^k)]x \\ & + ky^{k-1}x^2 - [2(k - 1)^2(1 - y) + 2(k - 1)(1 - y^k)]x^k - (2k - 1)y^{k-1}x^{k+1}. \end{aligned}$$

Note that

$$g(0^-, y) < 0. \tag{6}$$

Also, it can be easily shown that

$$g(y, y) = y[2(k - 1) - (k^2 - 1)y^{k-1} + (k - 1)^2y^k] + y^{k+1}(1 - y^{k-1}) > 0. \tag{7}$$

By taking the partial derivative of $g(x, y)$ with respect to x , we have

$$\begin{aligned} \frac{\partial}{\partial x}g(x, y) = & (k - 1)y^{k-1} + (k - 1)k(1 - y^k) + 2ky^{k-1}x \\ & - k[2(k - 1)^2(1 - y) + 2(k - 1)(1 - y^k)]x^{k-1} - (k + 1)(2k - 1)y^{k-1}x^k. \end{aligned}$$

An examination of $(\partial^3/\partial x^3)g(x, y)$ reveals that $(\partial^2/\partial x^2)g(x, y)$ is strictly decreasing in x . Since $(\partial^2/\partial x^2)g(0^-, y) = 2ky^{k-1} > 0$, we have either

- (i) $(\partial^2/\partial x^2)g(x, y) > 0$ for all $x \in (0, y)$, or
- (ii) there exists $x_1(y) \in (0, y)$ such that $(\partial^2/\partial x^2)g(x, y) > 0$ for all $x \in (0, x_1(y))$ and $(\partial^2/\partial x^2)g(x, y) < 0$ for all $x \in (x_1(y), y]$.

We will consider these two cases separately.

(i) Since $(\partial/\partial x)g(0^-, y) > 0$, we have $(\partial/\partial x)g(x, y) > 0$ for all $x \in (0, y]$. By (6) and (7), there exists $x_2(y) \in (0, y)$ such that $g(x, y) < 0$ for all $x \in (0, x_2(y))$ and $g(x, y) > 0$ for all $x \in (x_2(y), y]$. Since $f(0^-, y) = -y^{k-1} < 0$ and $f(y, y) = 0$, we have $f(x, y) \leq 0$ for all $x \in (0, y]$, which implies (5).

(ii) Since $(\partial/\partial x)g(0^-, y) > 0$, we have either

- (a) $(\partial/\partial x)g(x, y) > 0$ for all $x \in (0, y)$, or
- (b) there exists $x_3(y) \in (0, y)$ such that $(\partial/\partial x)g(x, y) > 0$ for all $x \in (0, x_3(y))$ and $(\partial/\partial x)g(x, y) < 0$ for all $x \in (x_3(y), y]$.

For both (a) and (b), by (6) and (7), there exists $x_4(y) \in (0, y)$ such that $g(x, y) < 0$ for all $x \in (0, x_4(y))$ and $g(x, y) > 0$ for all $x \in (x_4(y), y]$. Since $f(0^-, y) < 0$ and $f(y, y) = 0$, we have $f(x, y) \leq 0$ for all $x \in (0, y]$, which implies (5). □

Lemma 2. For $k = 1, 2, \dots, n \geq 2$ and $0 < q_1 \leq \dots \leq q_n < 1$,

$$q_n \left(\prod_{i=1}^{n-1} (1 - q_i^k) \right) \left(1 - \prod_{i=1}^n (1 - q_i^{k-1}) \right) \geq \left(\prod_{i=1}^{n-1} (1 - q_i^{k-1}) \right) \left(1 - \prod_{i=1}^n (1 - q_i^k) \right).$$

Proof. By induction on n , we can show that for any $k \geq 1$,

$$1 - \prod_{i=1}^n (1 - q_i^{k-1}) = \sum_{i=1}^n \left(q_i^{k-1} \prod_{j=1}^{i-1} (1 - q_j^{k-1}) \right) \tag{8}$$

with the convention that $\prod_{j=l}^{l-1} a_j = 1$. To prove the lemma, we have to show that

$$\frac{\prod_{i=1}^{n-1} (1 - q_i^{k-1}) (1 - \prod_{i=1}^n (1 - q_i^k))}{\prod_{i=1}^{n-1} (1 - q_i^k) (1 - \prod_{i=1}^n (1 - q_i^{k-1}))} \leq q_n. \tag{9}$$

By (8), we have

$$\begin{aligned} \frac{\prod_{i=1}^{n-1} (1 - q_i^{k-1}) (1 - \prod_{i=1}^n (1 - q_i^k))}{\prod_{i=1}^{n-1} (1 - q_i^k) (1 - \prod_{i=1}^n (1 - q_i^{k-1}))} &= \frac{\prod_{i=1}^{n-1} (1 - q_i^{k-1}) \sum_{i=1}^n (q_i^k \prod_{j=1}^{i-1} (1 - q_j^k))}{\prod_{i=1}^{n-1} (1 - q_i^k) \sum_{i=1}^n (q_i^{k-1} \prod_{j=1}^{i-1} (1 - q_j^{k-1}))} \\ &= \frac{\sum_{i=1}^n (q_i^k \prod_{j=1}^{i-1} (1 - q_j^k)) / \prod_{i=1}^{n-1} (1 - q_i^k)}{\sum_{i=1}^n (q_i^{k-1} \prod_{j=1}^{i-1} (1 - q_j^{k-1})) / \prod_{i=1}^n (1 - q_i^{k-1})} \\ &= \frac{\sum_{i=1}^n (q_i^k / \prod_{j=i}^{n-1} (1 - q_j^k))}{\sum_{i=1}^n (q_i^{k-1} / \prod_{j=i}^{n-1} (1 - q_j^{k-1}))}. \end{aligned}$$

It can be easily shown that

$$\sum_{i=1}^n \left(\frac{q_i^k}{\prod_{j=i}^{n-1} (1 - q_j^k)} \right) \leq q_n \sum_{i=1}^n \left(\frac{q_i^{k-1}}{\prod_{j=i}^{n-1} (1 - q_j^{k-1})} \right).$$

Therefore, (9) is proved. □

Lemma 3. For $k = 1, 2, \dots$ and $0 < q_1 \leq \dots \leq q_n < 1$,

$$\frac{\partial}{\partial q_n} h_k(q_1, \dots, q_n) - \frac{\partial}{\partial q_1} h_k(q_1, \dots, q_n) \leq 0.$$

Proof. By taking the partial derivatives of $h_k(q_1, \dots, q_n)$ with respect to q_1 and q_n , we obtain

$$\begin{aligned} & \left(\frac{\partial}{\partial q_1} h_k(q_1, \dots, q_n) - \frac{\partial}{\partial q_n} h_k(q_1, \dots, q_n) \right) \left(1 - \prod_{i=1}^n (1 - q_i^{k-1}) \right)^2 \\ &= \left(\prod_{i=2}^{n-1} (1 - q_i^k) \right) \left(\frac{\partial}{\partial q_1} (1 - q_1^k)(1 - q_n^k) - \frac{\partial}{\partial q_n} (1 - q_1^k)(1 - q_n^k) \right) \left(1 - \prod_{i=1}^n (1 - q_i^{k-1}) \right) \\ & \quad - \left(\prod_{i=2}^{n-1} (1 - q_i^{k-1}) \right) \\ & \quad \times \left(\frac{\partial}{\partial q_1} (1 - q_1^{k-1})(1 - q_n^{k-1}) - \frac{\partial}{\partial q_n} (1 - q_1^{k-1})(1 - q_n^{k-1}) \right) \left(1 - \prod_{i=1}^n (1 - q_i^k) \right). \end{aligned}$$

Then we have

$$\begin{aligned} & \frac{((\partial/\partial q_1)h_k(q_1, \dots, q_n) - (\partial/\partial q_n)h_k(q_1, \dots, q_n))(1 - \prod_{i=1}^n (1 - q_i^{k-1}))^2}{(\partial/\partial q_1)(1 - q_1^{k-1})(1 - q_n^{k-1}) - (\partial/\partial q_n)(1 - q_1^{k-1})(1 - q_n^{k-1})} \\ & \geq \frac{q_n(1 - q_1^k)(\prod_{i=2}^{n-1} (1 - q_i^k))(1 - \prod_{i=1}^n (1 - q_i^{k-1}))}{1 - q_1^{k-1}} - \prod_{i=2}^{n-1} (1 - q_i^{k-1}) \left(1 - \prod_{i=1}^n (1 - q_i^k) \right) \\ & \geq \frac{\prod_{i=1}^{n-1} (1 - q_i^{k-1})(1 - \prod_{i=1}^n (1 - q_i^k))}{1 - q_1^{k-1}} - \prod_{i=2}^{n-1} (1 - q_i^{k-1}) \left(1 - \prod_{i=1}^n (1 - q_i^k) \right) \\ & = 0, \end{aligned}$$

where we have used Lemma 1 in the first inequality and Lemma 2 in the second inequality. Since

$$\frac{\partial}{\partial q_1} (1 - q_1^{k-1})(1 - q_n^{k-1}) - \frac{\partial}{\partial q_n} (1 - q_1^{k-1})(1 - q_n^{k-1}) \geq 0,$$

we have

$$\frac{\partial}{\partial q_1} h_k(q_1, \dots, q_n) - \frac{\partial}{\partial q_n} h_k(q_1, \dots, q_n) \geq 0,$$

which completes the proof. □

Acknowledgements

We are grateful to the anonymous reviewer for the careful reading of the manuscript. B. Kim's research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (grant number NRF-2017R1A2B4012676). J. Kim's research was supported by the Basic Science Research Program through the NRF funded by the Ministry of Education (grant number 2017R1D1A1B03029542).

References

- [1] DU, B., ZHAO, P. AND BALAKRISHNAN, N. (2012). Likelihood ratio and hazard rate orderings of the maxima in two multiple-outlier geometric samples. *Prob. Eng. Inf. Sci.* **26**, 375–391, 613.
- [2] DYKSTRA, R., KOCHAR, S. AND ROJO, J. (1997). Stochastic comparisons of parallel systems of heterogeneous exponential components. *J. Statist. Planning Infer.* **65**, 203–211.
- [3] KHALEDI, B.-E. AND KOCHAR, S. (2000). Some new results on stochastic comparisons of parallel systems. *J. Appl. Prob.* **37**, 1123–1128.
- [4] KOCHAR, S. AND XU, M. (2007). Stochastic comparisons of parallel systems when components have proportional hazard rates. *Prob. Eng. Inf. Sci.* **21**, 597–609.
- [5] KOCHAR, S. AND XU, M. (2009). Comparisons of parallel systems according to the convex transform order. *J. Appl. Prob.* **46**, 342–352.
- [6] MAO, T. AND HU, T. (2010). Equivalent characterizations on orderings of order statistics and sample ranges. *Prob. Eng. Inf. Sci.* **24**, 245–262.
- [7] SHAKED, M. AND SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. Springer, New York.
- [8] WANG, J. (2015). A stochastic comparison result about hazard rate ordering of two parallel systems comprising of geometric components. *Statist. Prob. Lett.* **106**, 86–90.