Sample range - some stochastic comparison results

Baha-Eldin Khaledi
Department of Mathematics
University of Arizona
Tucson, AZ 85721, USA
e-mail : bkhaledi@math.arizona.edu

Subhash Kochar
Indian Statistical Institute
7, SJS Sansanwal Marg
New Delhi-110016, India
e-mail : kochar@.isid.ac.in

AMS classification: 60E15, 62N05, 62D05

Key words : Hazard rate ordering, stochastic ordering, proportional hazard rates, exponential distribution.

Abstract

Let $X_1, \ldots, X_n$ be independent random variables following the proportional hazards model so that X_i has survival function $\overline{F}^{\lambda_i}(x)$, $\quad i = 1, \ldots, n$. Let $Y_1, \ldots, Y_n$ be a random sample of size n from a distribution with survival function $\overline{F}^{\tilde{\lambda}}(x)$, where $\tilde{\lambda} = (\prod_{i=1}^n \lambda_i)^{1/n}$, is the geometric mean of the λ_i's. It is shown that if the baseline survival function $\overline{F}$ is *new worse than used* (NWU) then the sample range of X_i's is stochastically greater than the sample range of Y_i's. This result gives a simple upper bound on the distribution function of the sample range of X_i's in terms of $\tilde{\lambda}$ and $\overline{F}$.

The resulting bound when applied to exponentials is sharper than the one given by Kochar and Rojo (1996), which is in terms of the arithmetic mean of the λ_i's.

1 Introduction

Order statistics and statistics based on them, play an important role in reliability theory and statistics. The time to failure of a k-out-of-n system of n components corresponds to the $(n-k+1)$th order statistic. In particular, the lifetime of a parallel system is the same as the largest order statistic. Sample range is a simple and popular statistic for comparing variabilities in distributions and it is important to study its stochastic properties. These statistics have been studied extensively in the literature in case the components are independent and identically distributed. But in real life, often, the systems are made up of components with non-identically distributed lifetimes. Since the distribution theory becomes quite complicated then, relatively fewer results are available in the general case.

The exponential distribution plays a very important role in statistics. Because of its non-aging property, it has many nice properties and it often gives very convenient bounds on survival probabilities and other characteristics of interest for systems with non-exponential components. Pledger and Proschan (1971) studied the problem of stochastically comparing the order statistics and spacings of non-identically distributed independent exponential random variables with those corresponding to independent and identically distributed exponential random variables. This topic has been followed up by many researchers including Proschan and Sethuraman (1976), Boland, El-Neweihi and Proschan (1994), Kochar and Korwar (1996), Kochar and Rojo (1996), Dykstra, Kochar and Rojo (1997), Kochar and Ma (1999), Bon and Paltanea (1999) and Khaledi and Kochar (2000), among others. For a recent review of this area, see Kochar (1998).

In this note we study the stochastic properties of the sample range when the parent observations come from different distributions. First we review some well-known notions of stochastic orders. These can be found at one place in Shaked and Shanthikumar (1994).

Let X and Y be two random variables with distribution functions F and G; and survival functions $\overline{F}$ and $\overline{G}$, respectively. Let F^{-1} and G^{-1} be the right continuous inverses of F and G, defined by $F^{-1}(u) = sup\{x : F(x) \leq u\}$ and $G^{-1}(u) = sup\{x : G(x) \leq u\}, \;\; u \in [0,1]$. We shall assume that densities exist and shall denote by f and g the densities of X and Y, respectively. Throughout this paper the term *increasing* is used for *monotone nondecreasing* and *decreasing* for *monotone nonincreasing.*

Definition 1.1 *Y is said to be stochastically smaller than X (denoted by $Y \leq_{st} X$) if*

$$\overline{G}(x) \leq \overline{F}(x) \quad \text{for all } x. \tag{1.1}$$

It is well known that (1.1) *is equivalent to*

$$E[\phi(Y)] \leq E[\phi(X)] \quad \text{for all increasing functions } \phi : \mathcal{R} \to \mathcal{R}, \tag{1.2}$$

for which the expectations exist.

Definition 1.2 *Y is said to be smaller than X in the sense of hazard rate ordering (denoted by $Y \leq_{hr} X$) if*

$$\frac{\overline{F}(x)}{\overline{G}(x)} \quad \text{is increasing in } x. \tag{1.3}$$

In the continuous case this is equivalent to

$$r_F(x) \leq r_G(x) \quad \text{for all } x, \tag{1.4}$$

where $r_F = f/\overline{F}$ and $r_G = g/\overline{G}$ are the hazard (or failure) rates of F and G, respectively.

One of the basic criteria for comparing variability in two probability distributions is that of dispersive ordering.

Definition 1.3 *Y is less dispersed than X ($Y \overset{disp}{\preceq} X$) if*

$$G^{-1}(v) - G^{-1}(u) \leq F^{-1}(v) - F^{-1}(u), \tag{1.5}$$

$\forall\ 0 < u \leq v < 1.$

This means that the difference between any two quantiles of G is smaller than the difference between the corresponding quantiles of F. A consequence of $Y \overset{disp}{\preceq} X$ is that $|Y_1 - Y_2| \leq_{st} |X_1 - X_2|$ and which in turn implies $var(Y) \leq var(X)$ as well as $E[|Y_1 - Y_2|] \leq E[|X_1 - X_2|]$, 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).

Bagai and Kochar (1986) established the following connections between hazard rate ordering and dispersive ordering under some restrictions on the shapes of the distributions.

Theorem **1.1** *(a) If $Y \leq_{hr} X$ and either F or G is DFR (decreasing failure rate), then $Y \overset{disp}{\preceq} X$,*

(b) if $Y \overset{disp}{\preceq} X$ and either F or G is IFR (increasing failure rate), then $Y \leq_{hr} X$.

First we study the stochastic properties of the range of a random sample from a continuous distribution. Let $X_1, \ldots, X_n$ be a random sample from F and let $Y_1, \ldots, Y_n$ be an independent random sample from another distribution G. It follows from Lemma 3(c) of Bartoszewic (1986) that $Y \overset{disp}{\preceq} X \Rightarrow Y_{n:n} - Y_{1:n} \leq_{st} X_{n:n} - X_{1:n}$. This observation along with Theorem 1.1(a) leads to the following theorem.

Theorem **1.2** *Let $Y \leq_{hr} X$ and let either F or G be DFR. Then*

$$Y_{n:n} - Y_{1:n} \leq_{st} X_{n:n} - X_{1:n}. \tag{1.6}$$

Next we consider the case when the parent observations are independent exponentials but with unequal parameters. Let $X_1, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i, $i = 1, \ldots, n$. Let $Y_1 \ldots, Y_n$ be a random sample of size n from an exponential distribution with hazard rate $\overline{\lambda} = \sum_{i=1}^n \lambda_i / n$. Kochar and Rojo (1996) proved that in this case $Y_{n:n} - Y_{1:n} \leq_{st} X_{n:n} - X_{1:n}$. In Section 2 we improve upon this bound by replacing $\overline{\lambda}$ with $\tilde{\lambda} = (\prod_{i=1}^n \lambda_i)^{1/n}$, the geometric mean of the λ_i's. In Section 3, we extend this result to proportional hazards model.

2 The case of heterogenous independent exponentials

Let $\{x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}\}$ denote the increasing arrangement of the components of the vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. The vector $\mathbf{y}$ is said to *majorize* the vector $\mathbf{x}$ (written as $\mathbf{x} \overset{m}{\preceq} \mathbf{y}$) if $\sum_{i=1}^{j} y_{(i)} \leq \sum_{i=1}^{j} x_{(i)}$, for $j = 1, \ldots, n-1$ and $\sum_{i=1}^{n} y_{(i)} = \sum_{i=1}^{n} x_{(i)}$. The following result is proved in Kochar and Rojo (1996).

THEOREM **2.1** *Let $X_1, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i, for $i = 1, \ldots, n$. Let $X_1^*, \ldots, X_n^*$ be another set of exponential random variables with $\lambda_i^\star$ as the hazard rate of X_i^*, $i = 1, \ldots, n$. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ and $\boldsymbol{\lambda}^\star = (\lambda_i^\star, \ldots, \lambda_n^\star)$, then $\boldsymbol{\lambda}^\star \overset{m}{\preceq} \boldsymbol{\lambda}$ implies*

$$X_{n:n}^* - X_{1:n}^* \leq_{st} X_{n:n} - X_{1:n} \tag{2.1}$$

This theorem immediately leads to the following corollary.

COROLLARY 2.1 *Let $X_1, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i, $i = 1, \ldots, n$. Let $Y_1 \ldots, Y_n$ be a random sample of size n from an exponential distribution with hazard rate $\overline{\lambda} = \sum_{i=1}^n \lambda_i / n$. Then*

$$Y_{n:n} - Y_{1:n} \leq_{st} X_{n:n} - X_{1:n}. \tag{2.2}$$

In this paper we improve upon this bound by replacing $\overline{\lambda}$ with $\tilde{\lambda} = (\prod_{i=1}^{n} \lambda_i)^{1/n}$, the geometric mean of the λ's. We shall need the following lemmas to prove it.

LEMMA 2.1 *(Theorem 2.1 of Khaledi and Kochar, 2000) Let $X_1, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i, $\quad i = 1, \ldots, n$. Let $Z_1, \ldots, Z_n$ be a random sample of size n from an exponential distribution with common hazard rate $\tilde{\lambda} = (\prod_{i=1}^{n} \lambda_i)^{1/n}$. Then*

(a) $\quad Z_{n:n} \leq_{disp} X_{n:n}$,

(b) $\quad Z_{n:n} \leq_{hr} X_{n:n}$ *(and hence $Z_{n:n} \leq_{st} X_{n:n}$).*

The proof of the following inequality can be found in Theorem 2.1 of Dykstra, Kochar and Rojo (1997).

LEMMA 2.2 *For $y_i > 0$, $i = 1, \ldots, n$;*

$$\sum_{i=1}^{n} \frac{y_i}{1 - e^{-y_i}} \leq \left(\sum_{i=1}^{n} y_i \right) \prod_{i=1}^{n} (1 - e^{-y_i})^{-\frac{1}{n}}. \tag{2.3}$$

Now we prove the main result of this paper.

THEOREM **2.2** *Let $X_1, \ldots, X_n$ be independent exponential random variables with X_i having hazard rate λ_i, for $i = 1, \ldots, n$. Let $Z_1, \ldots, Z_n$ be a random sample of size n from an exponential distribution with common hazard rate $\tilde{\lambda}$, the geometric mean of the λ_i's. Then,*

$$Z_{n:n} - Z_{1:n} \leq_{st} X_{n:n} - X_{1:n}.$$

PROOF : The distribution function of the sample range $X_{n:n} - X_{1:n}$ (see David, 1981, p. 26) is

$$F_{R_n^X}(x) = \frac{1}{\sum_{i=1}^{n} \lambda_i} \sum_{i=1}^{n} \frac{\lambda_i}{1 - e^{-\lambda_i x}} \prod_{i=1}^{n} (1 - e^{-\lambda_i x}), \quad x > 0, \tag{2.4}$$

and that of $Z_{n:n} - Z_{1:n}$ is

$$G_{R_n^Z}(x) = \left(1 - e^{-\tilde{\lambda}x)}\right)^{n-1}, \quad x > 0. \tag{2.5}$$

Using (2.4) and (2.5), we have to show that for $x > 0$,

$$\sum_{i=1}^{n} \frac{\lambda_i}{1 - e^{-\lambda_i x}} \prod_{i=1}^{n}(1 - e^{-\lambda_i x}) \le \sum_{i=1}^{n} \lambda_i \left(1 - e^{-\tilde{\lambda}x)}\right)^{n-1}. \tag{2.6}$$

Multiplying both sides of (2.6) by x (> 0), it is sufficient to prove that for $x > 0$,

$$\sum_{i=1}^{n} \frac{\lambda_i x}{1 - e^{-\lambda_i x}} \prod_{i=1}^{n}(1 - e^{-\lambda_i x}) \le \left(\sum_{i=1}^{n} \lambda_i x\right) \left(1 - e^{-\tilde{\lambda}x)}\right)^{n-1}. \tag{2.7}$$

From Lemma 2.2, for $x > 0$,

$$\sum_{i=1}^{n} \frac{\lambda_i x}{1 - e^{-\lambda_i x}} \le \left(\sum_{i=1}^{n} \lambda_i x\right) \prod_{i=1}^{n}(1 - e^{-\lambda_i x})^{-\frac{1}{n}}.$$

and hence

$$\sum_{i=1}^{n} \frac{\lambda_i x}{1 - e^{-\lambda_i x}} \prod_{i=1}^{n}(1 - e^{-\lambda_i x}) \le \left(\sum_{i=1}^{n} \lambda_i x\right) \prod_{i=1}^{n} \left(1 - e^{-\lambda_i x}\right)^{\frac{n-1}{n}}, \tag{2.8}$$

for $x > 0$. From Lemma 2.1 (b), under the given conditions, $Z_{n:n} \le_{st} X_{n:n}$. That is, for $x > 0$, $\prod_{i=1}^{n}(1 - e^{-\lambda_i x})^{1/n} \le 1 - e^{-\tilde{\lambda}x}$. Using this result, we find that the expression on the R.H.S. of (2.8) is less than or equal to that on the R.H.S. of (2.7) and from which the required result follows.

■

As a consequence of this result we get the following upper bound on the distribution function of $X_{n:n} - X_{1:n}$ in terms $\tilde{\lambda}$.

COROLLARY 2.2 *Under the conditions of Theorem 2.2, for $x > 0$,*

$$P[X_{n:n} - X_{1:n} \le x] \le \left[1 - e^{-\tilde{\lambda}x}\right]^{n-1}. \tag{2.9}$$

This bound is better than the one obtained in Kochar and Rojo (1996) (which was in terms of the arithmetic mean $\overline{\lambda}$) since the expression on the R.H.S. of (2.9) is increasing in $\tilde{\lambda}$ and $\tilde{\lambda} \le \overline{\lambda}$.

3 Extension to the PHR model

In this section we extend Theorem 2.2 to the proportional hazard rates (PHR) model. Let $\overline{F}$ denote the survival function of a nonnegative random variable X with hazard rate h. According to the PHR model, the random variables $X_1, \ldots, X_n$ are independent with X_i having survival function $\overline{F}^{\lambda_i}(.)$, so that its hazard rate is $\lambda_i h(.)$, $i = 1, \ldots, n$. In the next theorem we assume that F is *new worse than used* (NWU), that is,

$$\overline{F}(x+y) \geq \overline{F}(x)\overline{F}(y), \quad \text{for } x, y \geq 0,$$

or equivalently,

$$H(x+y) \leq H(x) + H(y), \quad \text{for } x, y \geq 0,$$

where $H(x) = -\log \overline{F}(x)$ denotes the cumulative hazard of F.

THEOREM **3.1** *Let $X_1, \ldots, X_n$ be independent random variables with X_i having survival function $\overline{F}^{\lambda_i}(x)$, $i = 1, \ldots, n$. Let $Y_1, \ldots, Y_n$ be a random sample of size n from a distribution with survival function $\overline{F}^{\tilde{\lambda}}(x)$, where $\tilde{\lambda} = (\prod_{i=1}^n \lambda_i)^{1/n}$. If F is NWU, then $X_{n:n} - X_{1:n} \geq_{st} Y_{n:n} - Y_{1:n}$.*

PROOF :

The distribution function of the sample range $X_{n:n} - X_{1:n}$ (see David, 1981, p. 26) is

$$\begin{aligned}
F_{R_n^X}(x) &= \sum_{i=1}^n \int_0^{+\infty} \lambda_i h(t) e^{-\lambda_i H(t)} \prod_{j \neq i}^n \left(e^{-\lambda_j H(t)} - e^{-\lambda_j H(t+x)} \right) dt \\
&\leq \sum_{i=1}^n \int_0^{+\infty} \lambda_i h(t) e^{-\lambda_i H(t)} \prod_{j \neq i}^n \left(e^{-\lambda_j H(t)} - e^{-\lambda_j H(t)} e^{-\lambda_j H(x)} \right) dt \\
&\quad \text{(since } F \text{ is NWU)} \\
&= \sum_{i=1}^n \lambda_i \prod_{j \neq i} (1 - e^{-\lambda_j H(x)}) \int_0^{+\infty} h(t) \prod_{j=1}^n e^{-\lambda_j H(t)} dt
\end{aligned}$$

$$= \sum_{i=1}^{n} \lambda_i \prod_{j \neq i} (1 - e^{-\lambda_j H(x)}) \int_0^{+\infty} h(t) e^{-H(t) \sum_{j=1}^{n} \lambda_j}$$

$$= \frac{1}{\sum_{i=1}^{n} \lambda_i} \sum_{i=1}^{n} \frac{\lambda_i}{1 - e^{-\lambda_i H(x)}} \prod_{i=1}^{n} (1 - e^{-\lambda_i H(x)}), \quad x > 0,$$

Now, replacing x with $H(x)$ in the proof of Theorem 2.2, it is easy to see that

$$F_{R_n^X}(x) \leq F_{R_n^Y}(x).$$

■

References

Bagai, I. and Kochar, S. C. (1986). On tail ordering and comparison of failure rates. *Commun. Statist. Theory and Methods* **15**, 1377-1388.

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

Boland, P.J., El-Neweihi, E. and Proschan, F. (1994). Applications of the hazard rate ordering in reliability and order statistics. *J. Appl. Prob.* **31**, 180-192.

Bon, J. L. and Paltanea, E. (1999). Ordering properties of convolutions of exponential random variables. *Lifetime Data Anal.* **5**, 185-192.

David, H. A. (1981). *Order Statistics* (2nd ed.). Wiley, New York.

Dykstra, R., Kochar, S. C. and Rojo, J. (1997). Stochastic comparisons of parallel systems of heterogeneous exponential components. *J. Statist. Plann. Inference*, **65**, 203-211.

Khaledi, B. and Kochar, S. (2000). Some new results on stochastic comparisons of parallel systems. *J. Appl. Probab.*, **37**, 1123- 1128.

Kochar, S. C. (1998). Stochastic comparisons of spacings and order statistics. *Frontiers in Reliability.* World Scientific : Singapore. 201-216. eds., Basu, A. P., Basu, S. K. and Mukhopadhyay, S.

Kochar, S. C. and Korwar, R. (1996). Stochastic orders for spacings of heterogeneous exponential random variables. *J. Mult. Anal.* **57**, 69-83.

Kochar, S. C. and Ma, C. (1999). Dispersive ordering of convolutions of exponential random variables. *Statist. Prob. Letters.* **43**, 321-324.

Kochar, S. and Rojo, J. (1996). Some new results on stochastic comparisons of spacings from heterogeneous exponential distributions. *J. Mult. Anal.* **59**, 272-281.

Pledger, P. and Proschan, F. (1971). Comparisons of order statistics and of spacings from heterogeneous distributions. *Optimizing Methods in Statistics.* Academic Press : New York., 89-113. ed. Rustagi, J. S.

Proschan, F. and Sethuraman, J. (1976). Stochastic comparisons of order statistics from heterogeneous populations, with applications in reliability. *J. Mult. Analysis*, **6**, 608-616.

Shaked, M. and Shanthikumar, J.G. (1994). *Stochastic Orders and their Applications.* Academic Press, San Diego, CA.