Journal of Multivariate Analysis 160 (2017) 31–41

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On stochastic comparisons of maximum order statistics from the location-scale family of distributions

Nil Kamal Hazra [a], Mithu Rani Kuiti [b], Maxim Finkelstein [a,c,*], Asok K. Nanda [d]

[a] Department of Mathematical Statistics and Actuarial Science, University of the Free State, 339 Bloemfontein 9300, South Africa
[b] Department of Operations Management, Indian Institute of Management Calcutta (IIMC), Diamond Harbour Road, Kolkata-700104, India
[c] ITMO University, St. Petersburg, Russia
[d] Department of Mathematics and Statistics, Indian Institute of Science Education and Research (IISER) Kolkata, Mohanpur-741246, India

ARTICLE INFO

Article history:
Received 25 October 2016
Available online 9 June 2017

AMS 2010 subject classifications:
60E15

Keywords:
Location-scale family
Majorization orders
Parallel system
Schur-convex/concave function
Stochastic orders

ABSTRACT

We consider the location-scale family of distributions, which contains many standard lifetime distributions. We give conditions under which the largest order statistic of a set of random variables with different/the same location as well as different/the same scale parameters dominates that of another set of random variables with respect to various stochastic orders. Along with general results, we consider important special cases, namely, the Feller–Pareto, generalized Pareto, Burr, exponentiated Weibull, Power generalized Weibull, generalized gamma, Half-normal and Fréchet distributions.

1. Introduction and preliminaries

Order statistics play an important role in diverse areas of statistics, probability and other related fields; see [1,2,9] for encyclopedic information on order statistics. Different order statistics can be used in different applications; for example, the maximum is of interest in the study of floods and other meteorological phenomena while the minimum is often used in reliability and survival analysis, etc.; see [11,12,14] for more discussion on this topic.

Let $X_1, \ldots, X_n$ be a collection of random variables, and $X_{1:n} \leq \cdots \leq X_{n:n}$ be their corresponding order statistics. For each $k \in \{1, \ldots, n\}$, $X_{k:n}$ is called the kth order statistic. If the variables $X_1, \ldots, X_n$ represent the lifetimes of components of a system, say, then the reliability (survival) function of a k-out-of-n system formed by $X_1, \ldots, X_n$ is the same as that of $X_{n-k+1:n}$. Thus, to study a k-out-of-n system, it is sufficient to deal with the $(n-k+1)$th order statistic, and vice versa. The theory and applications of order statistics from homogeneous samples have been well studied in the literature; see [9] and references therein. However, due to mathematical complexity, much less attention was paid to the order statistics from heterogeneous samples, and this topic needs a thorough investigation.

Stochastic comparisons between different order statistics from homogeneous/ heterogeneous samples have been explored by different researchers in the last few decades; see [4] for a review. However, most results in the literature are obtained with respect to specific distributions, e.g., exponential, gamma, generalized exponential; see [3,6,13,17,31,33,34]

* Corresponding author at: Department of Mathematical Statistics and Actuarial Science, University of the Free State, 339 Bloemfontein 9300, South Africa.
E-mail address: FinkelM@ufs.ac.za (M. Finkelstein).

http://dx.doi.org/10.1016/j.jmva.2017.06.001
0047-259X/

and references therein. Much fewer studies have been reported with respect to general families of distributions, e.g., location-scale families or proportional hazard rates families; see [15,20,26,27] and references therein. Special attention has been paid to the location-scale family as it contains most of the popular lifetime distributions. Results for the scale family of distributions can be found in [10,15,16,19,20]. However, stochastic comparisons of order statistics when location and scale parameters are different for different sets of random variables were not considered in the literature so far. This is the major challenge to be addressed in the present work.

The foregoing refers both to the largest and the smallest order statistics. Due to analytical complexity and substantial differences in the corresponding proofs and reasoning, it is unrealistic to combine these two cases in a single paper. Therefore, in the current paper, we will focus on comparisons between maximum order statistics, whereas the study for the minimum order statistics will be reported separately.

Specifically, in this paper, we compare stochastically two maximum order statistics formed from two different sets of random variables having different/the same location and scale parameters. General results are obtained for the location-scale family of distributions. In the case of the location-scale families, we consider a number of well known lifetime distributions. For each type of stochastic comparison considered, we present important practical results showing for which baseline distributions (and the corresponding admissible regions of parameters) our comparisons hold. A number of intermediate results of independent interest are needed along the way.

Thus, our paper develops a theory of stochastic comparisons for two maximum order statistics that are formed from two different sets of random variables having different/the same location as well as different/the same scale parameters. We believe that it is meaningful from a theoretical and a methodological point of view; in particular, it unifies some of the existing literature. Before proceeding, we provide below some notation and basic definitions.

1.1. Preliminaries

Given an absolutely continuous random variable Y, we denote its probability density function (pdf) by f_Y, its cumulative distribution function (cdf) by F_Y, and its reversed hazard rate function by $\tilde{r}_Y$. The survival function of Y is written as $\bar{F}_Y = 1 - F_Y$. We denote the set of real numbers by $\mathbb{R}$, and the set of positive real numbers by $\mathbb{R}_+$.

A random variable X is said to follow the location-scale family, written as $X \sim \mathcal{LS}(\lambda, \theta)$, if its distribution function can be written, for all $x > \lambda$, by

$$F_X(x; \lambda, \theta) = F\left(\frac{x-\lambda}{\theta}\right),$$

where $\lambda \in \mathbb{R}$ and $\theta > 0$ are the location and the scale parameters, respectively, and F is the baseline distribution function; see [21]. We also denote the pdf and the reversed hazard rate function of the baseline distribution by f and $\tilde{r}$, respectively. It is well known that incorporating the location and the scale parameters in a distribution makes it more flexible and general as far as modeling and data analysis are concerned. Note that many well known distributions could be considered as baseline distributions, for example:

(a) The Feller–Pareto distribution, denoted $\mathcal{FP}(\alpha, \beta, \gamma)$, with pdf is given by

$$f(x) = \frac{1}{\gamma B(\alpha, \beta)} x^{\beta/\gamma - 1}(1 + x^{1/\gamma})^{-(\alpha+\beta)}, \quad x > 0,\ \alpha > 0,\ \beta > 0,\ \gamma > 0,$$

where B is the Beta function; see [24]. Special cases include the Pareto II ($\beta = \gamma = 1$), Pareto III ($\alpha = \beta = 1$) and Pareto IV ($\beta = 1$) distributions, the Transformed Beta ($\gamma = 1/\gamma^*$) and Burr distribution ($\beta = 1, \gamma = 1/c$).

(b) The Generalized Pareto distribution, denoted $\mathcal{GP}(\xi)$, with cdf given by $F(x) = 1 - (1 + \xi x)^{-1/\xi}$, $x > 0$, $\xi > 0$; see [8]. It includes the exponential distribution ($\xi \to 0$) as a special case.

(c) The Burr distribution, denoted $\mathcal{BU}(c, k)$, with cdf given by $F(x) = 1 - (1 + x^c)^{-k}$, $x > 0$, $c > 0$, $k > 0$; see [7]. It includes the Pareto II ($c = 1$) and log–logistic ($k = 1$) distributions as special cases.

(d) The Exponentiated Weibull distribution, denoted $\mathcal{EW}(\alpha, \beta)$, with cdf given by $F(x) = (1 - e^{-x^\alpha})^\beta$, $x > 0$, $\alpha > 0$, $\beta > 0$; see [23]. It includes the exponential ($\alpha = \beta = 1$), Weibull ($\beta = 1$) and exponentiated exponential ($\alpha = 1$) distributions as special cases.

(e) The Power-generalized Weibull distribution, denoted $\mathcal{PGW}(c, k)$, with cdf given by $F(x) = 1 - \exp\{1 - (1 + x^c)^{1/k}\}$, $x > 0$, $c > 0$, $k > 0$; see [15]. It includes the Weibull ($k = 1$) and exponential ($c = k = 1$) distributions as special cases.

(f) The Generalized gamma distribution, denoted $\mathcal{GG}(p, q)$, with pdf given by $f(x) = px^{q-1}e^{-x^p}/\Gamma(q/p)$, $x > 0$, $p > 0$, $q > 0$, where Γ is Euler's gamma function; see [15]. It includes the exponential ($p = q = 1$), Weibull ($p = q$) and gamma ($p = 1$) distributions as special cases.

(g) The Half-normal distribution with pdf given by $f(x) = \sqrt{2/\pi}\, e^{-x^2/2}$, $x > 0$.

(h) The Fréchet distribution, denoted $\mathcal{FR}(\alpha)$, with cdf given by $F(x) = \exp(-x^{-\alpha})$, $x > 0$, $\alpha > 0$.

(i) The Pareto distribution, denoted $\mathcal{PD}(\alpha)$, with cdf given by $F(x) = 1 - x^{-\alpha}$, $x \geq 1$, $\alpha > 0$.