Operations Research Letters 45 (2017) 475–480

Contents lists available at ScienceDirect

Operations Research Letters

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

Stochastic comparisons on two finite mixture models

Ebrahim Amini-Seresht [a,*], Yiying Zhang [b]

[a] Department of Statistics, Bu-Ali Sina University, Hamedan, Iran
[b] Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China

ARTICLE INFO

Article history:
Received 23 February 2016
Received in revised form 23 July 2017
Accepted 24 July 2017
Available online 3 August 2017

Keywords:
Stochastic orders
Finite mixture model
Aging notions
Lead time model

ABSTRACT

This paper carries out stochastic comparisons for two classical finite mixture models with different baseline random variables and different mixing proportions in the sense of the hazard rate, reversed hazard rate, likelihood ratio, mean residual lifetime and mean inactivity time orders. An application in lead time model is presented to reveal the significance of the theoretical findings as well. The results generalize and extend some known in the literature such as Hernandez (2007), Navarro et al. (2009) and Navarro (2016).

1. Introduction

Let $\mathbf{X} = (X_1, \ldots, X_k)$ be a random vector comprised of k non-negative random variables. The distribution function of the classical finite (positive) mixture of $X_1, \ldots, X_k$ is defined as

$$F_{\mathbf{X},\boldsymbol{\alpha}}(x) = \sum_{i=1}^{k} \alpha_i F_i(x), \tag{1}$$

where F_i is the distribution function of X_i and α_i is the mixing proportion (weight) such that $\sum_{i=1}^{k}\alpha_i = 1$ and $\alpha_i \geq 0$, for $i \in \{1, 2, \ldots, k\}$. The survival and probability density functions of the mixture of $X_1, \ldots, X_k$ can be expressed as

$$\overline{F}_{\mathbf{X},\boldsymbol{\alpha}}(x) = \sum_{i=1}^{k} \alpha_i \overline{F}_i(x) \text{ and } f_{\mathbf{X},\boldsymbol{\alpha}}(x) = \sum_{i=1}^{k} \alpha_i f_i(x),$$

respectively, where $\overline{F}_i = 1 - F_i$ and f_i are the corresponding survival and probability density functions of X_i, for $i = 1, \ldots, k$.

Finite mixture model plays an important role in the fields of reliability theory, actuarial science, medicine and economics since there exist practical situations where the sample of data are drawn from a finite number of different populations. Therefore, the finite mixture model is an effective tool for modelling the distribution of random sample from heterogeneous populations. Here, we present two practical examples where the finite mixture models can be applied.

- Usually, there is more than one reason (see, for example, [7]) causing the failures of a sample of items in the context of industrial engineering. The failure distribution for each reason can be adequately approximated by a simple density function such as the negative exponential. Then, the overall distribution can be modelled as a finite mixture of negative exponential random variables.
- A system is coherent if each component is relevant and its structure function is increasing in each component (see [3]). In the context of reliability theory, the distribution and survival functions of a coherent system having k independent and identically distributed components can be expressed as a linear combination of the distributions and survival functions of the ordered lifetimes of these k components, respectively.

A finite mixture model allows us to identify and estimate the parameters of interest for each component (subpopulation). Interested readers may find some other applications of finite mixture models in [12,25,26], and the references therein.

Stochastic ordering, as a powerful tool to compare the magnitude and variability of different random variables or vectors, has been widely used in reliability theory for comparing two coherent systems with different components (cf. [16] and [4]), and in actuarial science for comparing the riskiness of different insurance portfolios (cf. [8] and [2]). Hereafter, we will use "$\geq_{\rm lr}$", "$\geq_{\rm hr}$", "$\geq_{\rm rh}$", "$\geq_{\rm st}$", "$\geq_{\rm mrl}$", "$\geq_{\rm mit}$" and "$\geq_{\rm disp}$" to represent the likelihood ratio order, hazard rate order, reversed hazard rate order, usual stochastic order, mean residual lifetime order, mean inactivity time order and dispersive order, respectively. For explicit definitions and their applications, one may refer to [19] and [24].

* Corresponding author.
E-mail addresses: e.amini64@yahoo.com (E. Amini-Seresht), zhyy@hku.hk (Y. Zhang).

http://dx.doi.org/10.1016/j.orl.2017.07.009

Let $F_{\mathbf{X},\boldsymbol{\alpha}}$ [$G_{\mathbf{Y},\boldsymbol{\beta}}$] be distribution function of k-component finite mixture model with baseline random variables $X_1, \ldots, X_k$ [$Y_1, \ldots, Y_k$] and mixing proportions $\alpha_1, \ldots, \alpha_k$ [$\beta_1, \ldots, \beta_k$], where X_i [Y_i] has distribution function F_i [G_i], for $i = 1, \ldots, n$. [10] considered the mixture hazard rate ordering for the ordered mixing distributions under different environments. [13] and [23] obtained some comparison results for generalized mixtures (the mixing proportions may be negative) with only two baseline random variables. [22] obtained some comparison results for generalized mixtures having different mixing proportions and extended the preceding result to k-component finite mixture models when $k > 2$. Specifically, it was proved that if $F_1 \ge_{\mathrm{hr[rh,lr]}} \cdots \ge_{\mathrm{hr[rh,lr]}} F_k$, then

$$\alpha_i\beta_j \le \alpha_j\beta_i \text{ whenever } 1 \le i \le j \le k \Longrightarrow F_{\mathbf{X},\boldsymbol{\alpha}} \le_{\mathrm{hr[rh,lr]}} F_{\mathbf{X},\boldsymbol{\beta}}. \quad (2)$$

Recently, [6] defined various local stochastic orderings to compare the hazard rate functions of two mixture models with dynamic mixing measures.

To the best of our knowledge, there are few results on stochastic comparisons of finite mixture models with different baseline random variables and different mixing proportions. Motivated by this, the present paper is of the aim at studying the ordering results between two finite mixture models $F_{\mathbf{X},\boldsymbol{\alpha}}$ and $G_{\mathbf{Y},\boldsymbol{\beta}}$ in the sense of the hazard rate, reversed hazard rate, likelihood ratio order, mean residual lifetime and mean inactivity time orders. The rest of the paper is organized as follows: The main results are presented in Section 2. Section 3 provides a practical application of our theoretical findings in the lead time model. Section 4 concludes the paper.

Throughout the paper, we shall use "*increasing*" and "*decreasing*" to denote "*non-decreasing*" and "*non-increasing*", respectively. All expectations, inverse functions and ratio functions are assumed to exist whenever they appear.

2. Main results

In this section, some ordering results are established for two k-component finite mixture models both having different baseline random variables and different sets of mixing proportions. In the first place, sufficient conditions are provided for the reversed hazard rate and hazard rate orders between two finite mixture models.

Theorem 2.1. *Let $F_{\mathbf{X},\boldsymbol{\alpha}}$ and $G_{\mathbf{Y},\boldsymbol{\alpha}}$ be two k-component finite mixture models with common mixing proportions $\alpha_1, \ldots, \alpha_k$. Suppose that*

(i) $G_1 \ge_{\mathrm{rh[hr]}} \cdots \ge_{\mathrm{rh[hr]}} G_k$ *or* $F_1 \ge_{\mathrm{rh[hr]}} \cdots \ge_{\mathrm{rh[hr]}} F_k$;

(ii) $\frac{G_j(x)}{F_j(x)}$ $\left[\frac{\overline{G}_j(x)}{\overline{F}_j(x)}\right]$ *is decreasing in* $j \in \{1, \ldots, k\}$;

(iii) $F_j \le_{\mathrm{rh[hr]}} G_j$ *for all* $j \in \{1, \ldots, k\}$.

Then, it follows that $F_{\mathbf{X},\boldsymbol{\alpha}} \le_{\mathrm{rh[hr]}} G_{\mathbf{Y},\boldsymbol{\alpha}}$.

Proof. We only give the proof for the reversed hazard rate order since the proof for the hazard rate order is a trivial case of Theorem 9 in [6]. Without loss of generality, it is assumed that $G_1 \ge_{\mathrm{rh}} \cdots \ge_{\mathrm{rh}} G_k$. The reversed hazard rate function of $F_{\mathbf{X},\boldsymbol{\alpha}}$ is given by

$$\tilde{r}_{\mathbf{X},\boldsymbol{\alpha}}(x) = \frac{\sum_{j=1}^k \alpha_j f_j(x)}{\sum_{j=1}^k \alpha_j F_j(x)} = \sum_{j=1}^k \tilde{r}_{X_j}(x) p_j(x),$$

where $\tilde{r}_{X_j}(x)$ is the reversed hazard rate function of X_j and $p_j(x) = \frac{\alpha_j F_j(x)}{\sum_{i=1}^k \alpha_i F_i(x)}, j = 1, \ldots, k$. The expression of the reversed hazard rate function of $G_{\mathbf{Y},\boldsymbol{\alpha}}$ can be written as

$$\tilde{r}_{\mathbf{Y},\boldsymbol{\alpha}}(x) = \frac{\sum_{j=1}^k \alpha_j g_j(x)}{\sum_{j=1}^k \alpha_j G_j(x)} = \sum_{j=1}^k \tilde{r}_{Y_j}(x) q_j(x),$$

where $q_j(x) = \frac{\alpha_j G_j(x)}{\sum_{i=1}^k \alpha_i G_i(x)}$. To reach the desired result, it is enough to prove that $\psi(x) = \tilde{r}_{\mathbf{Y},\alpha}(x) - \tilde{r}_{\mathbf{X},\boldsymbol{\alpha}}(x)$ is non-negative for all $x \in \mathcal{R}_+ = [0, +\infty)$. Note that

$$\psi(x) = \sum_{j=1}^k \tilde{r}_{Y_j}(x) q_j(x) - \sum_{j=1}^k \tilde{r}_{X_j}(x) p_j(x)$$
$$\ge \sum_{j=1}^k \tilde{r}_{X_j}(x) q_j(x) - \sum_{j=1}^k \tilde{r}_{X_j}(x) p_j(x) =: \xi(x),$$

where the inequality follows from condition (iii). Thus, it suffices to show that $\xi(x)$ is non-negative for all $x \in \mathcal{R}_+$. Consider two non-negative discrete random variables $\mathbb{W}$ and $\mathbb{V}$ on a sample space $\{1, \ldots, k\}$ with probability mass functions $q_j(x)$ and $p_j(x)$, $j = 1, \ldots, k$, respectively. Using these observations, one can see that

$$\xi(x) = \mathsf{E}[\phi(\mathbb{W})] - \mathsf{E}[\phi(\mathbb{V})], \quad (3)$$

where $\phi(j) = \tilde{r}_{X_j}(\cdot), j = 1, \ldots, k$. According to the definition of the usual stochastic order (see [24]), we need to prove that (3) is non-negative by showing that $\phi(j)$ is decreasing in j and $\mathbb{W} \le_{\mathrm{st}} \mathbb{V}$. On the one hand, based on condition (i), it follows that $\tilde{r}_{X_1}(x) \le \cdots \le \tilde{r}_{X_k}(x)$ for all $x \in \mathcal{R}_+$. Thus, we know that $\phi(j) = \tilde{r}_{X_j}(x)$ is decreasing in $j \in \{1, 2, \ldots, k\}$. On the other hand, one can see that for all $x \in \mathcal{R}_+$,

$$\frac{q_j(x)}{p_j(x)} \propto \frac{G_j(x)}{F_j(x)}, \text{ whenever } j \in \{1, \ldots, k\}.$$

Hence, the condition (ii) implies that q_j/p_j is decreasing in $j \in \{1, 2, \ldots, k\}$ and this concludes that $\mathbb{W} \le_{\mathrm{lr}} \mathbb{V}$, which in turn implies $\mathbb{W} \le_{\mathrm{st}} \mathbb{V}$. Thus, the desired result follows immediately. For the case of $F_1 \ge_{\mathrm{rh}} \cdots \ge_{\mathrm{rh}} F_k$, it can be proved in similar arguments. □

The following example illustrates the assumptions (i), (ii) and (iii) of Theorem 2.1.

Example 2.2. For some constant $\lambda > 0$, suppose that X_j and Y_j, $j = 1, \ldots, k$, have respective survival functions

$$\overline{F}_j(x) = e^{-\frac{2}{j\lambda}\sqrt{x} - \frac{x}{j\lambda}}, \text{ whenever } x \in \mathcal{R}_+, \text{ and}$$

$$\overline{G}_j(x) = (2/x)^{j\lambda}, \text{ whenever } x \ge 2.$$

It is easy to check that $\frac{\overline{G}_j(x)}{\overline{F}_j(x)}$ is decreasing in $j \in \{1, \ldots, k\}$ and increasing in $x \ge 2$. Thus, conditions (ii) and (iii) of Theorem 2.1 are satisfied. It can be examined that $\overline{G}_j(x)$ is decreasing in $j \in \{1, \ldots, k\}$ with respect to the hazard rate order, which implies condition (i) of Theorem 2.1.

Let X be a non-negative random variable with density function f. Then, the Glaser's function η_X of X is defined by $\eta_X(x) = -\frac{f'(x)}{f(x)}$ for all $x \in \mathcal{R}_+$ whenever $f(x) > 0$ (see [11]). The following lemma is introduced to establish the likelihood ratio order between two finite mixture models.

Lemma 2.3 (*[21]*). *Let X and Y be non-negative random variables with distribution functions F and G, density functions f and g, and Glaser's functions η_X and η_Y, respectively. Then*

$$F \le_{\mathrm{lr}} G \Longleftrightarrow \eta_X(x) \ge \eta_Y(x) \text{ for all } x \in \mathcal{R}_+.$$

Theorem 2.4. *Let $F_{\mathbf{X},\boldsymbol{\alpha}}$ and $G_{\mathbf{Y},\boldsymbol{\alpha}}$ be two k-component finite mixture models with common mixing proportions $\alpha_1, \ldots, \alpha_k$. Suppose that*

(i) $G_1 \ge_{\mathrm{lr}} \cdots \ge_{\mathrm{lr}} G_k$ *or* $F_1 \ge_{\mathrm{lr}} \cdots \ge_{\mathrm{lr}} F_k$;

(ii) $\frac{g_j(x)}{f_j(x)}$ *is decreasing in* $j \in \{1, \ldots, k\}$;

(iii) $F_j \le_{\mathrm{lr}} G_j$ *for all* $j \in \{1, \ldots, k\}$.

Then, it follows that $F_{\mathbf{X},\boldsymbol{\alpha}} \le_{\mathrm{lr}} G_{\mathbf{Y},\boldsymbol{\alpha}}$.