



Sharp bounds on distribution functions and expectations of mixtures of ordered families of distributions

Tomasz Rychlik¹

Received: 30 October 2017 / Accepted: 19 July 2018 / Published online: 17 August 2018
© The Author(s) 2018

Abstract

We compare two mixtures of arbitrary collection of stochastically ordered distribution functions with respect to fixed mixing distributions. Under the assumption that the first mixture distribution is known, we establish optimal lower and upper bounds on the latter mixture distribution function and present single families of ordered distributions which attain the bounds uniformly at all real arguments. Furthermore, we determine sharp upper and lower bounds on the differences between the expectations of the mixtures expressed in various scale units. General results are illustrated by several examples.

Keywords Stochastic order · Mixture · Sharp bound · Nonparametric Bayes inference · Nonparametric regression

Mathematics Subject Classification Primary: 60E15 · Secondary: 62C10 · 62E15

1 Introduction

Let $\{F_\theta\}_{\theta \in \mathbb{R}}$ be an arbitrary family of stochastically ordered distribution functions, i.e., ones that satisfy

$$\theta_1 < \theta_2 \Rightarrow F_{\theta_1}(x) \geq F_{\theta_2}(x), \quad x \in \mathbb{R}. \quad (1.1)$$

We assume that the family is not known. We further introduce two distribution functions S and T which are assumed to be known. We do not impose any restrictions on them. They may have discrete and/or continuous components, and supports concentrated on possibly partially or even fully different subsets of the real line. The purpose of the paper is to compare the mixture distribution functions

✉ Tomasz Rychlik
trychlik@impan.pl

¹ Institute of Mathematics, Polish Academy of Sciences, Śniadeckich 8, 00656 Warsaw, Poland

$$G(x) = \int_{\mathbb{R}} F_{\theta}(x) S(d\theta), \quad (1.2)$$

$$H(x) = \int_{\mathbb{R}} F_{\theta}(x) T(d\theta), \quad (1.3)$$

and their expectations.

Our model is motivated by nonparametric Bayes and regression ideas. The relation between the predictor θ and random response with distribution function F_{θ} is completely unknown except for a frequently natural and practically justified premise that the greater value of predictor results in a (stochastically) greater response. We investigate consequences of various choices of prior S and T for the wide nonparametric response model restricted merely by the above order constraints. Classic Bayes procedures are focused on identifying the predictor value managing a single random experiment. Our approach is global: we analyze the final consequences of random choice of predictor, and random response to the selected predictor. This corresponds to the situation of multiple repetitions of the experiment where various values of predictors are chosen according to the prior selection rule.

Precisely, in Sect. 2 we determine sharp lower and upper bounds on distribution functions (1.3) under the constraint that condition (1.2) is satisfied for an arbitrarily fixed G . The lower and upper bounds constitute proper distribution functions iff the mixing distribution function T does not have a positive mass right and left, respectively, to the support of S . We also show that the bounds are attained uniformly for any G , i.e., there exist single families $\{\underline{F}_{\theta}\}_{\theta \in \mathbb{R}}$ and $\{\overline{F}_{\theta}\}_{\theta \in \mathbb{R}}$ such that (1.2) holds, and the lower and upper bounds on (1.3) are attained by $\{\underline{F}_{\theta}\}_{\theta \in \mathbb{R}}$ and $\{\overline{F}_{\theta}\}_{\theta \in \mathbb{R}}$, respectively, for every real argument x . In Sect. 3, we determine the greatest possible lower and upper deviations of $\mathbb{E}Y = \int_{\mathbb{R}} xH(dx)$ from $\mathbb{E}X = \int_{\mathbb{R}} xG(dx)$ (cf (1.2) and (1.3)), measured in various scale units $(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p} = (\int_{\mathbb{R}} |x - \mathbb{E}X|^p G(dx))^{1/p}$, $p \geq 1$, based on central absolute moments of (1.2). We illustrate theoretical results by a number of examples in Sect. 4. Section 5 is devoted to the proof of Theorem 1 of Sect. 2 that is the basic result of the paper providing the tools for establishing the expectation bounds of Sect. 3.

Mixture models have multiple applications in probability and statistics. It was shown in a review paper of Karlis and Xekalaki (2003) that they are exploited in data modeling, discriminant and cluster analysis, outlier and robustness studies, analysis of variance, random effects and related models, factor analysis, and latent structure models, random variable generation, and approximation of distributions. Here we merely mention several nonparametric Bayesian inference and regression analysis applications. One of the pioneering works on nonparametric Bayesian estimation methods was that of Ferguson (1973) who provided Bayes estimates of the response distribution function and several its functionals under the Dirichlet process prior. Ghosh and Mukherjee (2005) presented a sequential version of the distribution function estimate. Susarla and Ryzin (1978), Tiwari and Zalkikar (1990), Gasparini (1996), and Zhou (2004) determined nonparametric Bayes estimators of distribution function under various censoring schemes. Sethuraman and Hollander (2009) estimated the nominal lifetime distribution of an unit exposed to various repair treatments. Hansen and Lauritzen

(2002) proposed a method of Bayesian estimation of arbitrary concave distribution function when the prior is a proper mixture of Dirichlet processes. For the Bayesian density estimation, we refer to Escobar and West (1995) and Vieira et al. (2015), whereas the hazard rate function was treated by McKeague and Tighiouart (2002).

From the vast nonparametric Bayesian regression literature, we mention only the following. Choi (2008) studied convergence of posterior distributions when the response–predictor relation was modeled by mixtures of parametric densities, and a sieve prior was assumed. Chung and Dunson (2009) estimated the conditional response distribution and identified significant predictors under the assumption of probit stick-breaking process as a prior. Zhu and Dunson (2013) used nested Gaussian processes as locally adaptive priors for the regression model. Jo et al. (2016) considered quantile regression problems with the Dirichlet process mixture modeling the error distribution.

A small proportion of nonparametric Bayes research is devoted to order restricted inference. Assuming restricted dependent Dirichlet prior for a collection of partially ordered distributions, Dunson and Peddada (2008) tested equalities in the homogenous groups and estimated group-specific distributions. Yang et al. (2011) developed posterior computations for stochastically ordered latent class model based on the nested Dirichlet process prior. Nashimoto and Wright (2007, 2008) performed Bayesian multiple comparisons for ordered medians and means, respectively. The model of stochastically ordered mixtures presented here was earlier examined by Miziula (2017) who established tight lower and upper bounds on the ratios of various dispersion measures of mixtures. Some similarities to our approach can be also found in Robertson and Wright (1982) where lower and upper bounds on mixtures of stochastically ordered Chi-squared distributions useful in testing homogeneity of normal means against their ordering were established.

2 Distribution functions of mixtures

The main result of this section is following.

Theorem 1 *Let S , T , and G be fixed distribution functions on \mathbb{R} . Then for every family $\{F_\theta\}_{\theta \in \mathbb{R}}$ of stochastically ordered distribution functions (cf (1.1)) satisfying (1.2), the following inequalities hold*

$$\underline{H}(G(x)) \leq \int_{\mathbb{R}} F_\theta(x) T(d\theta) \leq \overline{H}(G(x)), \quad x \in \mathbb{R}, \quad (2.1)$$

where $\underline{H}, \overline{H} : [0, 1] \mapsto [0, 1]$ are the greatest convex minorant and the smallest concave majorant, respectively, of the set

$$\mathcal{U} = \mathcal{U}(S, T) = \{(u, v) = (S(\theta), T(\theta)) : \theta \in \mathbb{R}\}. \quad (2.2)$$

Moreover, the bounds in (2.1) are optimal and uniformly attainable which means that there exist stochastically ordered families $\{\underline{F}_\theta\}_{\theta \in \mathbb{R}}$ and $\{\overline{F}_\theta\}_{\theta \in \mathbb{R}}$ satisfying (1.2) such that for every real x yields

$$\begin{aligned} \underline{H}(G(x)) &= \int_{\mathbb{R}} \underline{F}_{\theta}(x)T(d\theta), \\ \overline{H}(G(x)) &= \int_{\mathbb{R}} \overline{F}_{\theta}(x)T(d\theta). \end{aligned} \tag{2.3}$$

Remark 1 The set of stochastically ordered families satisfying (1.2) is non-empty. A trivial example is one-element family $\{G\}_{\theta \in \mathbb{R}}$.

Remark 2 Set (2.2) is certainly a subset of the standard unit square $[0, 1]^2$. It only contains ordered planar points which means that for every $(u_1, v_1), (u_2, v_2) \in \mathcal{U}$, inequality $u_1 \leq (\geq, \text{ resp.}) u_2$ implies the same relation between v_1 and v_2 . Points $(0, 0)$ and $(1, 1)$ belong to $\mathcal{U}(S, T)$ iff both S and T have left and right bounded supports, respectively.

Besides, (2.2) may have very different forms for various S and T . Note first that since S and T are arbitrary, and we can interchange their roles. Except for given $\mathcal{U}(S, T)$ we also consider $\mathcal{U}(T, S)$ which is a symmetric transformation of $\mathcal{U}(S, T)$ about the diagonal $u = v$.

In the simplest case, when both S and T are continuous, (2.2) is represented by a single curve joining $(0, 0)$ and $(1, 1)$. It does not necessarily contain the end-points. The curve may contain horizontal line segments if S is increasing on some interval, and T is constant there. By the symmetry property mentioned above, vertical line segments are possible as well.

Another simple image set (2.2) arises when S and T are discrete. It consists of separate points. Notice that some of them may lie on some horizontal (or vertical) lines. If S is continuous and T is discrete, then $\mathcal{U}(S, T)$ consists of horizontal line segments. Sometimes the set is identical with the graph of non-decreasing right-continuous step-wise function which means that the steps have left end-points, and do not have the right ones. However, if S is constant on some left neighborhood of a jump of T , then the right end-point belongs to the line segment as well. Clearly, under the assumptions $\mathcal{U}(T, S)$ is built of vertical intervals.

After analyzing these simple cases, we are prepared for examining general S and T , possibly possessing some continuity intervals, and some discontinuities. Set (2.2) consists of at most countably many pieces. For each pair of pieces, one is located right and above the other. Only one coordinate of the left lower end-point of the former piece may be equal to the respective coordinate of the right upper end-point of the latter one. The pieces of (2.2) are either curves or separate points. The curves may contain horizontally and vertically oriented line segments as well as some their pieces can represent the graphs of strictly increasing functions. Each curve contains the point with minimal values of both coordinates, but it does not necessarily contain the right upper end-point.

Remark 3 The points of the graphs of the greatest convex minorant and the smallest concave majorant of (2.2) either coincide with the points of the set, or belong to straight line segments joining the elements of the set (more precisely, the limiting values of the right upper ends of the set pieces may also serve as the left and right ends of the line segments). By definition of $\mathcal{U}(S, T)$, the greatest convex minorant and the smallest concave majorant are non-decreasing functions. They are obviously continuous as

well. Accordingly, compositions $\underline{H} \circ G$ and $\overline{H} \circ G$ are well-defined functions on the whole real axis, non-decreasing, right-continuous, and have all values in $[0, 1]$.

Their actual image sets can be smaller, though. We explain when it occurs with use of a natural notion of ordering intervals which we further often use in the proof of Theorem 1. We say that interval Θ_1 precedes interval Θ_2 , and write $\Theta_1 < \Theta_2$ iff for every $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$ we have $\theta_1 < \theta_2$. We admit that the intervals have or do not have their left and right end-points. Accordingly, $\underline{H} \circ G$ is a distribution function iff

$$\forall \Theta_1 < \Theta_2 \quad S(\Theta_1) = 1 \Rightarrow T(\Theta_2) = 0. \tag{2.4}$$

(for simplicity of notation, we use the same symbol for the distribution function and respective probability distribution). Clearly, $\underline{H} \circ G$ is not a proper distribution function iff

$$\exists \Theta_1 < \Theta_2 \quad S(\Theta_1) = 1, T(\Theta_2) > 0. \tag{2.5}$$

Then

$$1 > \bar{u} = \bar{u}(S, T) = 1 - \max\{T(\Theta_2) : \Theta_2 > \Theta_1, S(\Theta_1) = 1\}. \tag{2.6}$$

is the supremum point of $\underline{H} \circ G(\mathbb{R})$. It means that $\underline{H} \circ G$ is not a distribution function when T has a positive probability mass right to the support of S . This may even happen when the support of T is contained in that of S , but they have a joint right support end-point, at which T has an atom, and S has not. Formally we admit the situation that $\underline{u}(S, T) = 0$ if the whole mass of T is located right to the mass of S , and then $\underline{H} \circ G$ simply vanishes. Analogously, $\overline{H} \circ G$ is a proper distribution function iff

$$\forall \Theta_1 < \Theta_2 \quad S(\Theta_2) = 1 \Rightarrow T(\Theta_1) = 0. \tag{2.7}$$

In the general case, $\inf \overline{H} \circ G(\mathbb{R}) = \underline{u} \geq 0$ where

$$\underline{u} = \max\{T(\Theta_1) : \Theta_1 < \Theta_2, S(\Theta_2) = 1\}.$$

It is clear that $\underline{u} \geq \bar{u}$.

Remark 4 Observe that the inverses

$$\underline{\underline{H}}(u) = \begin{cases} 0, & 0 \leq u \leq \bar{u}, \\ \overline{H}^{-1}(u), & \bar{u} \leq u \leq 1, \end{cases}$$

and

$$\overline{\overline{H}}(u) = \begin{cases} \underline{H}^{-1}(u), & 0 \leq u \leq \underline{u}, \\ 1, & \underline{u} \leq u \leq 1, \end{cases}$$

are the greatest convex minorant and the smallest concave majorant, respectively, of the set $\mathcal{U}(T, S)$ which are exploited in minimizing and maximizing mixture distribution function $\int_{\mathbb{R}} F_{\theta}(x)S(d\theta)$ under restriction on fixed form of $\int_{\mathbb{R}} F_{\theta}(x)T(d\theta)$.

The proofs of the lower and upper bounds, and their attainability are rather long, but similar. Therefore in the proof presented in Sect. 5, we restrict ourselves to the lower bound case. Families $\{F_{\theta}\}_{\theta \in \mathbb{R}}$ and $\{\bar{F}_{\theta}\}_{\theta \in \mathbb{R}}$ which attain the bounds in (2.1) strongly depend on the particular shape of set (2.2), and resulting forms of the greatest convex minorant and the smallest concave majorant. It is impossible to describe them in a concise form for general S and T (dependence on G is much simpler here) in the statement of Theorem 1. Precise construction of $\{F_{\theta}\}_{\theta \in \mathbb{R}}$ is presented in the proof of Theorem 1. Construction of $\{\bar{F}_{\theta}\}_{\theta \in \mathbb{R}}$ is analogous to that of $\{F_{\theta}\}_{\theta \in \mathbb{R}}$, and it not presented there.

3 Expectations of mixtures

In this section we examine variations of the expectations of mixtures under stochastic ordering of mixed distributions. Suppose that we consider a mixture of arbitrary unknown family of ordered distribution functions, and the actual mixing distribution is S . The resulting random variable X has an unknown distribution function (1.2). However, we assume that the mixing distribution is T , different from S . This generates a random variable Y with different distribution function (1.3) and expectation $\mathbb{E}Y$. Our purpose is to evaluate the maximal possible differences between the assumed and actual expectations of mixtures $\mathbb{E}Y - \mathbb{E}X$. This is measured in various scale units $\sigma_p = (\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}$, $p \geq 1$, generated by the central absolute moments of the actual mixture variable X . The bounds depend merely on the mixing distributions S and T , and on the parameter p of the measurement units σ_p . They are valid for all possible $\{F_{\theta}\}_{\theta \in \mathbb{R}}$, and resulting mixture distributions G and H . The only restriction is that X has a finite p th raw moment $\mathbb{E}|X|^p$ of chosen order $p \geq 1$, and Y has a finite expectation. Similarly, under condition that actual mixture variable X has a bounded support, we determine upper and lower bounds on $\mathbb{E}Y - \mathbb{E}X$ gauged in the scale units $\sigma_{\infty} = \text{ess sup } |X - \mathbb{E}X| = \sup_{\{0 < x < 1\}} |G^{-1}(x) - \mathbb{E}X|$.

First we exclude the possibilities for which we obtain trivial infinite bounds on $\mathbb{E}Y - \mathbb{E}X$, independently of the choice of the scale units. Assume first that S and T are such that respective $H([0, 1]) = [0, \bar{u}] \subsetneq [0, 1]$, i.e., condition (2.5) holds. Then clearly one can find the partition Θ_1, Θ_2 of \mathbb{R} such that (2.5) is satisfied with $T(\Theta_2) = 1 - \bar{u}$. Take the family of distributions

$$F_{\theta} = \begin{cases} F_U, & \theta \in \Theta_1, \\ F_P, & \theta \in \Theta_2, \end{cases}$$

where $F_U(x) = x, 0 \leq x \leq 1$, and $F_P(x) = 1 - \frac{1}{x}, x \geq 1$, are the standard uniform and Pareto distributions with shape parameter 1, respectively. Then obviously $G = F_U$, $H = (1 - \bar{u})F_U + \bar{u}F_P$, and so $\mathbb{E}X = \frac{1}{2}, \frac{1}{4} \leq \sigma_p \leq \frac{1}{2}$ for all $1 \leq p \leq +\infty$ whereas $\mathbb{E}Y = +\infty$. Similarly, contradicting (2.7), we are able to choose an ordered family

such that its mixtures with respect to S and T have bounded support, and expectation $-\infty$, respectively. Therefore assumptions (2.4) and (2.7) are necessary for getting nontrivial finite upper and lower bounds on $\mathbb{E}Y - \mathbb{E}X$.

Other extremities may occur when S and T are stochastically ordered. If $S(\theta) \geq T(\theta)$, $\theta \in \mathbb{R}$, then all the points of (2.2) lie either above or on the diagonal $\{(u, u) : 0 \leq u \leq 1\}$ of the unit square. If we additionally assume (2.4), we obtain $H(u) = u$, $0 \leq u \leq 1$. By Theorem 1, we get $H(x) \geq G(x)$, $x \in \mathbb{R}$, and $\mathbb{E}Y \leq \mathbb{E}X$ for an arbitrary choice of ordered $\{F_\theta\}_{\theta \in \mathbb{R}}$ for which the expectations exist. Similarly, assumptions $S \leq T$ and (2.7) imply $\mathbb{E}Y \geq \mathbb{E}X$ for any $\{F_\theta\}_{\theta \in \mathbb{R}}$. Determination of the nonpositive upper bounds and nonnegative lower ones for $(\mathbb{E}Y - \mathbb{E}X)/\sigma_p$, $1 \leq p \leq +\infty$, when $S \geq T$ and $S \leq T$, respectively, needs more subtle tools than these used below, and will be treated elsewhere.

Now we present strictly positive upper bounds and strictly negative lower bounds on $(\mathbb{E}Y - \mathbb{E}X)/\sigma_p$ for $1 < p < \infty$, $p = 1$ and $p = +\infty$ in Propositions 1, 2, and 3, respectively. Due to its importance and relative simplicity, case $p = 2$ is especially distinguished in Corollary 1. The bounds depend on the mixing distributions S and T , and the scale unit parameter p , and all of them are optimal. We describe conditions of their attainability by specifying mixture distribution $G = G(S, T, p)$ (see (1.2)). Construction of the families of mixing distributions $\{F_\theta\}_{\theta \in \mathbb{R}}$ and $\{\bar{F}_\theta\}_{\theta \in \mathbb{R}}$ providing stochastically minimal and maximal mixture distribution H (and, in consequence, maximal and minimal expectations $\mathbb{E}Y$, respectively) for given S, T , and $G = G(S, T, p)$ are described in the proof of Theorem 1.

Proposition 1 *Let X and Y be random variables with distribution functions which are the mixtures of some stochastically ordered families of distributions $\{F_\theta\}_{\theta \in \mathbb{R}}$ with respect to fixed distribution functions S and T , respectively. Assume that X has a finite p th moment for some fixed $1 < p < \infty$.*

- (i) *Suppose that $S \not\leq T$ satisfy (2.4), and denote by \underline{h} the right-continuous version of the derivative of the greatest convex minorant of set $\mathcal{U}(S, T)$ defined in (2.2). If $\int_0^1 \underline{h}^{p/(p-1)}(u) du$ is finite, then for all mixture distribution functions (1.2) for which $\mathbb{E}|X|^p < \infty$ we have*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}} \leq B_p(S, T) = \left(\int_0^1 |\underline{h}(u) - C_p|^{p/(p-1)} du \right)^{(p-1)/p}, \tag{3.1}$$

where $\underline{h}(0) < C_p < \underline{h}(1)$ uniquely solves the equation

$$\int_{\{\underline{h}(u) < c\}} [c - \underline{h}(u)]^{1/(p-1)} du = \int_{\{\underline{h}(u) > c\}} [\underline{h}(u) - c]^{1/(p-1)} du. \tag{3.2}$$

The equality in (3.1) is attained if (1.2) has the right-continuous quantile function satisfying

$$\frac{G^{-1}(u) - \mu}{\sigma_p} = \left| \frac{\underline{h}(u) - C_p}{B_p(S, T)} \right|^{1/(p-1)} \times \operatorname{sgn} \{ \underline{h}(u) - C_p \}, \quad 0 < u < 1, \tag{3.3}$$

where $\mu \in \mathbb{R}$ and $\sigma_p^p \in \mathbb{R}_+$ denote arbitrarily chosen values of expectation and p th absolute central moment of G , respectively.

- (ii) Let $S \not\leq T$ satisfy (2.7), and \bar{h} denote the right derivative of the smallest concave majorant of (2.2). If $\int_0^1 \bar{h}^{p/(p-1)}(u) \, du < \infty$, then for all mixture distributions (1.2) with finite p th moment yields

$$\frac{\mathbb{E}Y - \mathbb{E}X}{(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}} \geq -b_p(S, T) = - \left(\int_0^1 |\bar{h}(u) - c_p|^{p/(p-1)} \, du \right)^{(p-1)/p} \tag{3.4}$$

where $\bar{h}(0) > c_p > \bar{h}(1)$ is determined by Eq. (3.2) with \underline{h} replaced by \bar{h} . The equality in (3.4) is attained by (1.2) with the quantile function

$$\frac{G^{-1}(u) - \mu}{\sigma_p} = \left| \frac{\bar{h}(u) - c_p}{b_p(S, T)} \right|^{1/(p-1)} \times \operatorname{sgn} \{ c_p - \bar{h}(u) \}, \quad 0 < u < 1,$$

and μ and σ_p^p denoting arbitrarily chosen expectation and p th absolute central moments of G .

Proof We prove statement (i). Note first that by definition \underline{h} is well-defined nonnegative, non-decreasing and right continuous function. Assumption $\int_0^1 \underline{h}^{p/(p-1)}(u) \, du < \infty$ implies that the left hand-side of (3.2) is a continuous strictly decreasing function of c ranging from $\int_0^1 \underline{h}^{1/(p-1)}(u) \, du$ at 0 to 0 as $c \nearrow \underline{h}(1)$ even if it is infinite. Similarly, the right-hand side of (3.2) continuously increases from 0 to $\int_0^1 \underline{h}^{1/(p-1)}(u) \, du$ as c varies over the same range.

We can write

$$\begin{aligned} \mu &= \mathbb{E}X = \int_{\mathbb{R}} x G(dx) = \int_0^1 G^{-1}(u) \, du, \\ \sigma_p^p &= \mathbb{E}|X - \mu|^p = \int_0^1 |G^{-1}(u) - \mu|^p \, du \end{aligned} \tag{3.5}$$

whereas owing to Theorem 1

$$\begin{aligned} \mathbb{E}Y &= \int_{\mathbb{R}} x H(dx) \leq \int_{\mathbb{R}} x \underline{H}(G(dx)) \\ &= \int_0^1 G^{-1}(u) \underline{H}(du) = \int_0^1 G^{-1}(u) \underline{h}(u) \, du. \end{aligned}$$

Since we also have $\mu = \mu \underline{H}(1) = \int_0^1 \mu \underline{h}(u) du$, the following yields

$$\begin{aligned} \mathbb{E}Y - \mathbb{E}X &\leq \int_0^1 [G^{-1}(u) - \mu] \underline{h}(u) du = \int_0^1 [G^{-1}(u) - \mu] [\underline{h}(u) - C_p] du \\ &\leq \left(\int_0^1 |G^{-1}(u) - \mu|^p du \right)^{1/p} \left(\int_0^1 |\underline{h}(u) - C_p|^{p/(p-1)} du \right)^{(p-1)/p} \end{aligned}$$

which, due to (3.5), gives (3.1).

We obtain equality in the above Hölder inequality iff $|\underline{h}(u) - C_p|^{1/(p-1)} \operatorname{sgn} \{\underline{h}(u) - C_p\}$ and $G^{-1}(u) - \mu$ are proportional with a nonnegative proportionality factor which is fulfilled in (3.3). By (3.2), the right-hand side of (3.3) has zero Lebesgue integral over $[0, 1]$ which guarantees that the expectation of G is equal to μ . Moreover,

$$1 = \int_0^1 \left| \frac{\underline{h}(u) - C_p}{B_p(S, T)} \right|^{p/(p-1)} du = \int_0^1 \frac{|G^{-1}(u) - \mu|^p}{\sigma_p^p} du,$$

which means that the p th absolute central moment of G amounts to σ_p^p , as desired.

Similar proof of claim (ii) is left to the reader. □

Remark 5 Assumption $\int_0^1 \underline{h}^{p/(p-1)}(u) du < \infty$ is necessary for finiteness of the upper bound in Proposition 1 even if (2.4) holds. Indeed, infiniteness of the integral implies that \underline{H} is not linear on some left neighborhood $(1 - \varepsilon, 1)$ of 1, and its graph coincides with a piece of set (2.2) there. This means that

$$\forall 1 - \varepsilon < u < 1 \quad \{(u, \underline{H}(u))\} = \{(S(\theta), T(\theta)) : \theta \in \Theta^u\}$$

where $\{\Theta^u\}_{1-\varepsilon < u < 1}$ is a family of ordered intervals so that $1 - \varepsilon < u_1 < u_2 < 1$ implies $\Theta^{u_1} < \Theta^{u_2}$, and $\bigcup_{1-\varepsilon < u < 1} \Theta^u$ constitutes a single right-unbounded interval. From the proof of Theorem 1 we conclude that the family $\{F_\theta\}_{\theta \in \mathbb{R}}$ attaining uniformly the lower bound in (2.1) consists of degenerate distributions for all $\theta \in \bigcup_{1-\varepsilon < u < 1} \Theta^u$.

If we take $1 - \varepsilon < u_1 < 1$, and replace all the degenerate distribution functions $F_\theta, \theta \in \bigcup_{u_1 < u < 1} \Theta^u$, by a single distribution function

$$F_{u_1}(x) = \begin{cases} 0, & G(x) \leq S(\theta_{u_1}), \\ \frac{G(x) - S(\theta_{u_1})}{1 - S(\theta_{u_1})} & S(\theta_{u_1}) \leq G(x) \leq 1, \end{cases}$$

where θ_{u_1} is any element of Θ^{u_1} , we do not disturb ordering of the family, and the shape of \underline{H} on $[0, u_1]$, and modify the latter into the linear function $\frac{1 - \underline{H}(u_1)}{1 - u_1}(u - u_1) + \underline{H}(u_1)$ on $(u_1, 1]$. The resulting distribution function \underline{H}_1 preserves convexity, and its right derivative \underline{h}_1 is bounded. There exists a unique $C_{p,1}$ being the solution to (3.2) with \underline{h} replaced by \underline{h}_1 . For finite

$$B_{p,1}(S, T) = \left(\int_0^1 |\underline{h}_1(u) - C_{p,1}|^{p/(p-1)} du \right)^{(p-1)/p}$$

and G_1 satisfying

$$\frac{G_1^{-1}(u) - \mu}{\sigma_p} = \left| \frac{h_1(u) - C_{p,1}}{B_{p,1}(S, T)} \right|^{1/(p-1)} \times \text{sgn} \{h_1(u) - C_{p,1}\}, \quad 0 < u < 1,$$

we get

$$\frac{\mathbb{E}Y - \mathbb{E}X}{(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}} = B_{p,1}(S, T)$$

where X and Y have distribution functions G_1 and $\underline{H}_1 \circ G_1$, respectively, arising by mixing ordered families of distribution functions with respect to S and T . We repeat the procedure for consecutive elements of increasing sequence $(u_k)_{k=1}^\infty$ tending to 1. Then bounded \underline{h}_k tend increasingly to \underline{h} , and

$$\int_0^1 |\underline{h}_k(u) - c|^{p/(p-1)} du \rightarrow +\infty, \quad \text{as } k \rightarrow +\infty,$$

for arbitrary c , and for all $C_{p,k}$ in particular. This implies that the difference $\mathbb{E}Y - \mathbb{E}X$ measured in $(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}$ units does not have a finite bound.

Analogously, we verify that under conditions (2.5) and $\int_0^1 \bar{h}^{p/(p-1)}(u) du = \infty$, the respective lower bound amounts to $-\infty$.

Corollary 1 *Assume notation of Proposition 1, and specify $p = 2$.*

(i) *If (2.4), $S \not\leq T$, and $\int_0^1 \underline{h}^2(u) du < \infty$ hold, then*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\text{Var } X}} \leq B_2(S, T) = \sqrt{\int_0^1 \underline{h}^2(u) du} - 1,$$

and equality is attained iff

$$\frac{G^{-1}(u) - \mu}{\sigma_2} = \frac{\underline{h}(u) - 1}{B_2(S, T)}, \quad 0 < u < 1.$$

(ii) *Under assumptions (2.7), $S \not\leq T$, and $\int_0^1 \bar{h}^2(u) du < \infty$*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\text{Var } X}} \geq -b_2(S, T) = -\sqrt{\int_0^1 \bar{h}^2(u) du} - 1,$$

and equality is attained iff

$$\frac{G^{-1}(u) - \mu}{\sigma_2} = \frac{1 - \bar{h}(u)}{b_2(S, T)}, \quad 0 < u < 1.$$

An essential simplification in case $p = 2$ follows from the fact that under the above assumptions we get $\int_0^1 \underline{h}(u)du = \underline{H}(1) - \underline{H}(0) = 1 = \int_0^1 \bar{h}(u)du = \bar{H}(1) - \bar{H}(0)$. This implies that $C_2 = c_2 = 1$ then. Moreover, $\int_0^1 [\underline{h}(u) - 1]^2 du = \int_0^1 \underline{h}^2(u)du - 1$, and the same equality holds for \bar{h} .

Proposition 2 *Let X, Y , and \underline{h}, \bar{h} be defined as in Proposition 1, and suppose that X has a finite mean.*

(i) *If (2.4) holds, $S \not\leq T$, and \underline{h} is bounded, then*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \leq \frac{\underline{h}(1) - \underline{h}(0)}{2}, \tag{3.6}$$

where $\underline{h}(1)$ is the left derivative of \underline{H} at 1. The equality is attained in the limit by the three-point distributions G_ε , where

$$G_\varepsilon \left(\left\{ \mu \pm \frac{\sigma_1}{2\varepsilon} \right\} \right) = \varepsilon, \quad G_\varepsilon(\{\mu\}) = 1 - 2\varepsilon,$$

as $\varepsilon \rightarrow 0$, and μ and σ_1 are arbitrarily chosen common values of the expectation and mean absolute deviation from the mean of all mixture distributions G_ε .

(ii) *Under (2.7), $S \not\leq T$, and boundedness of \bar{h} , we have*

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \geq -\frac{\bar{h}(0) - \bar{h}(1)}{2}, \tag{3.7}$$

with $\bar{h}(1)$ denoting the left derivative of \bar{H} at 1. The bound is attained in the limit by the family of mixture distributions $\{G_\varepsilon\}_{0 < \varepsilon < 1/2}$ described in Proposition 2(i) as $\varepsilon \rightarrow 0$.

The difference between the attainability conditions in Proposition 2(i) and (ii) consists in different choice of mixed distribution functions $\{F_\theta\}_{\theta \in \mathbb{R}}$ and $\{\bar{F}_\theta\}_{\theta \in \mathbb{R}}$ leading to different distributions $\underline{H} \circ G_\varepsilon$ and $\bar{H} \circ G_\varepsilon$ of Y .

Proof of Proposition 2 We focus on the proof of statement (i), because that of (ii) is similar, and we omit it here. We start with the relations

$$\begin{aligned} \mathbb{E}Y - \mathbb{E}X &\leq \int_0^1 [G^{-1}(u) - \mu] \left[\underline{h}(u) - \frac{\underline{h}(0) + \underline{h}(1)}{2} \right] du \\ &\leq \sup_{0 < u < 1} \left| \underline{h}(u) - \frac{\underline{h}(0) + \underline{h}(1)}{2} \right| \int_0^1 |G^{-1}(u) - \mu| du \\ &= \frac{\underline{h}(1) - \underline{h}(0)}{2} \sigma_1. \end{aligned}$$

If the distribution function of Y is G_ε , then

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\sigma_1} = \frac{1}{2\varepsilon} \left[\int_0^\varepsilon \left(\frac{\underline{h}(1) - \underline{h}(0)}{2} - \underline{h}(u) \right) du + \int_{1-\varepsilon}^1 \left(\underline{h}(u) - \frac{\underline{h}(1) - \underline{h}(0)}{2} \right) du \right]. \tag{3.8}$$

Function \underline{h} and its left continuous versions are monotone and continuous at 0 and 1, respectively. Accordingly, they tend to $\underline{h}(0)$ and $\underline{h}(1)$ uniformly over intervals $[0, \varepsilon]$ and $[1 - \varepsilon, 1]$ as $\varepsilon \rightarrow 0$. Therefore the right-hand side of (3.8) tends to $\frac{\underline{h}(1) - \underline{h}(0)}{2}$. \square

Obviously, the families of distributions described in Proposition 2 are not the only ones which attain the bounds in the limit. One can modify them in many ways without disturbing desired asymptotic properties. If $\underline{h}(u) = \underline{h}(0)$ and $\underline{h}(u) = \underline{h}(1)$ on some non-degenerate intervals $[0, u_0]$ and $[u_1, 1]$, say, then bound (3.6) is attained non-asymptotically by a single family of ordered distributions. The necessary and sufficient conditions of attainability are then $G(\{\mu\}) = 1 - u_0 - u_1$ with

$$- \int_0^{u_0} [G^{-1}(u) - \mu] du = \int_{u_1}^1 [G^{-1}(u) - \mu] du = \frac{\sigma_1}{2}.$$

This happens when \underline{H} is linear on some vicinities of 0 and 1. Accordingly, the condition for non-limit attainability of lower bound (3.7) is linearity of \overline{H} about 0 and 1. It is worth pointing out that such behavior of \underline{H} and \overline{H} is not especially extraordinary, and occurs for a significant proportion of pairs S and T .

Observe that

$$\underline{h}(0) = \inf_{\theta \in \mathbb{R}} \frac{T(\theta)}{S(\theta)} \leq 1 \leq \underline{h}(1) = \sup_{\theta \in \mathbb{R}} \frac{1 - T(\theta)}{1 - S(\theta)} \tag{3.9}$$

under condition (2.4), and

$$\overline{h}(0) = \sup_{\theta \in \mathbb{R}} \frac{T(\theta)}{S(\theta)} \geq 1 \geq \overline{h}(1) = \inf_{\theta \in \mathbb{R}} \frac{1 - T(\theta)}{1 - S(\theta)} \tag{3.10}$$

when (2.7) holds. Miziula and Solnický (2018, Theorem 1) (see Miziula 2015, Section 2.3, for the proof completely different from the above) used these notions for presenting the inequalities

$$\min \left\{ 0, \frac{\overline{h}(1) - \overline{h}(0)}{2} \right\} \leq \frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \leq \max \left\{ 0, \frac{\underline{h}(1) - \underline{h}(0)}{2} \right\},$$

when both (2.4) and (2.7) are assumed. In fact, the min, max, and zeros can be dropped, because $\overline{h}(1) - \overline{h}(0) \leq 0 \leq \underline{h}(1) - \underline{h}(0)$ due to (3.9) and (3.10).

Proposition 3 *Take X, Y , and $\underline{h}, \overline{h}$ defined above and suppose that X has a bounded support.*

(i) If (2.4) holds, $S \not\leq T$, and $\int_0^1 \underline{h}(u)du < \infty$, we have

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\text{ess sup}|X - \mathbb{E}X|} \leq 1 - 2\underline{H}\left(\frac{1}{2}\right). \tag{3.11}$$

Furthermore, define

$$0 \leq \underline{\lambda}_- = \lambda\left(\left\{\underline{h}(u) < \underline{h}\left(\frac{1}{2}\right)\right\}\right) \leq \frac{1}{2}, \tag{3.12}$$

$$0 \leq \underline{\lambda}_+ = \lambda\left(\left\{\underline{h}(u) > \underline{h}\left(\frac{1}{2}\right)\right\}\right) \leq \frac{1}{2}, \tag{3.13}$$

for $\lambda(A)$ denoting the Lebesgue measure of measurable set $A \subset \mathbb{R}$. The equality in (3.11) is attained for (1.2) satisfying $G(\{\mu - \sigma_\infty\}) = \underline{\lambda}_-$, $G(\{\mu + \sigma_\infty\}) = \underline{\lambda}_+$, and

$$\int_{\{\underline{h}(u)=\underline{h}(\frac{1}{2})\}} [G^{-1}(u) - \mu]du = \sigma_\infty(\underline{\lambda}_- - \underline{\lambda}_+) \tag{3.14}$$

with arbitrary $\mu \in \mathbb{R}$ and $\sigma_\infty \in \mathbb{R}_+$ serving as the mean and maximal absolute deviation from the mean, respectively, of distribution function G .

(ii) Assume (2.7), $S \not\leq T$, and $\int_0^1 \bar{h}(u)du < \infty$. It follows that

$$\frac{\mathbb{E}Y - \mathbb{E}X}{\text{ess sup}|X - \mathbb{E}X|} \geq 1 - 2\bar{H}\left(\frac{1}{2}\right). \tag{3.15}$$

Let $\bar{\lambda}_-$ and $\bar{\lambda}_+$ be defined as in (3.12) and (3.13), respectively, except for \underline{h} replaced by \bar{h} . Bound (3.15) is attained if $G(\{\mu - \sigma_\infty\}) = \bar{\lambda}_+$, $G(\{\mu + \sigma_\infty\}) = \bar{\lambda}_-$, and

$$\int_{\{\bar{h}(u)=\bar{h}(\frac{1}{2})\}} [G^{-1}(u) - \mu]du = \sigma_\infty(\bar{\lambda}_+ - \bar{\lambda}_-)$$

with the above meaning of μ and σ_∞ .

Proof (i) We easily obtain (3.11), because

$$\begin{aligned} \mathbb{E}Y - \mathbb{E}X &\leq \int_0^1 [G^{-1}(u) - \mu] \left[\underline{h}(u) - \underline{h}\left(\frac{1}{2}\right) \right] du \\ &\leq \sup_{0 < u < 1} |G^{-1}(u) - \mu| \int_0^1 \left| \underline{h}(u) - \underline{h}\left(\frac{1}{2}\right) \right| du \\ &= \sigma_\infty \left[1 - 2\underline{H}\left(\frac{1}{2}\right) \right]. \end{aligned}$$

The former inequality follows from Theorem 1, and is valid and attainable for any G under appropriate choice of mixed distributions. The latter becomes the

equality iff $G^{-1}(u) - \mu = -\sigma_\infty$ on $\{\underline{h}(u) < \underline{h}(\frac{1}{2})\}$ and $G^{-1}(u) - \mu = +\sigma_\infty$ on $\{\underline{h}(u) > \underline{h}(\frac{1}{2})\}$. If $\underline{\lambda}_- = \underline{\lambda}_+ = \frac{1}{2}$, and, in consequence, $\{\underline{h}(u) = \underline{h}(\frac{1}{2})\}$ is either empty or a degenerate interval, the optimal G is the unique two-point symmetric distribution function supported on $\mu \pm \sigma_\infty$. Otherwise (3.14) is the only assumption necessary for fulfilling the first moment condition

$$0 = \int_0^1 [G^{-1}(u) - \mu]du = \sigma_\infty \underline{\lambda}_- + \int_{\{\underline{h}(u)=\underline{h}(\frac{1}{2})\}} [G^{-1}(u) - \mu]dx + \sigma_\infty \underline{\lambda}_+.$$

(ii) The proof is omitted due to its similarity to the above. □

If $\underline{\lambda}_- + \underline{\lambda}_+ < 1$, the necessary condition (3.14) is in particular satisfied by two- or three-point distribution $G(\{\mu - \sigma_\infty\}) = \underline{\lambda}_-$, $G(\{\mu + \sigma_\infty\}) = \underline{\lambda}_+$, and $G(\{\mu + \frac{\underline{\lambda}_- - \underline{\lambda}_+}{1 - \underline{\lambda}_- - \underline{\lambda}_+} \sigma_\infty\}) = 1 - \underline{\lambda}_- - \underline{\lambda}_+$. We easily check that $-1 \leq \frac{\underline{\lambda}_- - \underline{\lambda}_+}{1 - \underline{\lambda}_- - \underline{\lambda}_+} \leq 1$ which assures that actually $\text{ess sup } |X - \mathbb{E}X| = \sigma_\infty$. An analogous construction is possible in Proposition 3(ii).

Also, one can verify that dropping the assumptions on boundedness and integrability of \underline{h} (and \bar{h}) in Propositions 2 and 3 implies that the respective upper (and lower) bounds are infinite. We finally point out that strengthening the assumptions on moments of X with mixture distribution G in Propositions 1–3 allows to relax respective conditions on integrability of powers of functions \underline{h} and \bar{h} , and, in consequence, on moments of Y with distribution functions $\underline{H} \circ G$ and $\bar{H} \circ G$, respectively.

4 Examples

We first mention applications of our mixtures in the reliability theory. If a system is composed of n identical elements with arbitrary exchangeable joint lifetime distribution, then the single component lifetime is the discrete uniform mixture of n stochastically ordered component lifetime order statistics distributions, whereas the system lifetime is another convex combination of these distributions whose probability vector, called the Samaniego signature, depends on the structure of the system (see Samaniego 1985; Navarro et al. 2008). These representations were used by Navarro and Rychlik (2007) for evaluating the distribution functions and expectations of system lifetimes by means of distribution functions and moments of component lifetimes. A similar idea can be used for calculating bounds on expectations of linear combinations of order statistics based on identically distributed samples (cf Rychlik 1993).

In the examples below, we restrict ourselves to calculating the sharp upper and lower bounds expressed with scale units with parameters $p = 1, 2$ and ∞ , because for the other values of p we merely obtain numerical approximations.

Example 1 We start with a simple pair of discrete mixing distributions

$$\begin{aligned} S(\{1\}) &= 0.2, & S(\{2\}) &= 0.3, & S(\{3\}) &= 0.3, & S(\{4\}) &= 0.2, \\ T(\{1\}) &= 0.2, & T(\{2\}) &= 0.4, & T(\{3\}) &= 0.1, & T(\{4\}) &= 0.3. \end{aligned}$$

The image of the parametric transformation $\{(S(\theta), T(\theta)) : \theta \in \mathbb{R}\}$ consists of 5 separate points

$$\mathcal{U}(S, T) = \{(0, 0), (0.2, 0.2), (0.5, 0.6), (0.8, 0.7), (1, 1)\}.$$

We immediately notice that the greatest convex minorant and the smallest concave majorant of the points are piecewise linear functions

$$\begin{aligned} \underline{H}(u) &= \begin{cases} \frac{7}{8}u, & 0 < u \leq 0.8, \\ \frac{3}{2}(u - 1) + 1, & 0.8 \leq u < 1, \end{cases} \\ \overline{H}(u) &= \begin{cases} \frac{6}{5}u, & 0 < u \leq 0.5, \\ \frac{4}{5}(u - 1) + 1, & 0.5 \leq u < 1, \end{cases} \end{aligned}$$

with respective stepwise right-continuous derivatives

$$\begin{aligned} \underline{h}(u) &= \begin{cases} \frac{7}{8}, & 0 < u < 0.8, \\ \frac{3}{2}, & 0.8 \leq u < 1, \end{cases} \\ \overline{h}(u) &= \begin{cases} \frac{6}{5}, & 0 < u < 0.5, \\ \frac{4}{5}, & 0.5 \leq u < 1. \end{cases} \end{aligned}$$

Applying Corollary 1, we calculate

$$\begin{aligned} B_2(S, T) &= \sqrt{\int_0^1 \underline{h}^2(u) du} - 1 = \frac{1}{4}, \\ b_2(S, T) &= -\sqrt{\int_0^1 \overline{h}^2(u) du} - 1 = -\frac{1}{5}. \end{aligned}$$

Conclusions of Propositions 2 and 3 are following here

$$\begin{aligned} -\frac{1}{5} &\leq \frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \leq \frac{5}{16}, \\ -\frac{1}{5} &\leq \frac{\mathbb{E}Y - \mathbb{E}X}{\text{ess sup } |X - \mathbb{E}X|} \leq \frac{1}{8}. \end{aligned}$$

After changing the roles of S and T , we get

$$\begin{aligned} \underline{h}(u) &= \begin{cases} \frac{5}{6}, & 0 < u < 0.6, \\ \frac{5}{4}, & 0.6 \leq u < 1, \end{cases} \\ \overline{h}(u) &= \begin{cases} \frac{8}{7}, & 0 < u < 0.7, \\ \frac{2}{3}, & 0.7 \leq u < 1, \end{cases} \end{aligned}$$

and so

$$\begin{aligned}
 -\frac{5}{21} &\leq \frac{\mathbb{E}X - \mathbb{E}Y}{\mathbb{E}|Y - \mathbb{E}Y|} \leq \frac{5}{24}, \\
 -\frac{1}{\sqrt{21}} &\leq \frac{\mathbb{E}X - \mathbb{E}Y}{\sqrt{\text{Var } Y}} \leq \frac{1}{2\sqrt{6}}, \\
 -\frac{1}{7} &\leq \frac{\mathbb{E}X - \mathbb{E}Y}{\text{ess sup } |Y - \mathbb{E}Y|} \leq \frac{1}{6}.
 \end{aligned}$$

The bounds shrink down towards 0 as p increases. This is evident due to the fact that by the Hölder inequality the scale units $(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}$ increase in p . Besides, the classes of distributions G with finite p th moments decrease with increase of p . The only exception make here the lower bounds on $\mathbb{E}Y - \mathbb{E}X$ which are still equal to $-\frac{1}{5}$. This is caused by the fact that then the bounds are attained for two-point symmetric distributions G for which $(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}$ are equal for all $1 \leq p \leq +\infty$.

Example 2 Suppose that the actual mixing distribution function is the power one $S(\theta) = 1 - (1 - \theta)^\alpha$, $0 < \theta < 1$, for some $\alpha > 0$, whereas it is assumed that the mixing distribution function is another member of the power family $T(\theta) = 1 - (1 - \theta)^\beta$ for $0 < \beta \neq \alpha$. Set $\mathcal{U}(S, T)$ (see (2.2)) is the graph of increasing function

$$v(u) = T(S^{-1}(u)) = 1 - (1 - u)^{\beta/\alpha}, \quad 0 \leq u \leq 1,$$

transforming $[0, 1]$ onto $[0, 1]$. It has the derivatives

$$\begin{aligned}
 v'(u) &= \frac{\beta}{\alpha} (1 - u)^{\beta/\alpha - 1}, \\
 v''(u) &= -\frac{\beta}{\alpha} \left(\frac{\beta}{\alpha} - 1 \right) (1 - u)^{\beta/\alpha - 2},
 \end{aligned}$$

which implies that v itself is convex if $\alpha > \beta$, and concave when $\alpha < \beta$. Accordingly, the greatest convex minorant and the smallest concave majorant of (2.2) have the forms $\underline{H}(u) = 1 - (1 - u)^{\beta/\alpha}$ and $\overline{H}(u) = u$, and their derivatives are $\underline{h}(u) = \frac{\beta}{\alpha} (1 - u)^{\beta/\alpha - 1}$ and $\overline{h}(u) = 1$, when $\alpha > \beta$, and the functions change their roles if $\alpha < \beta$. By Theorem 1, every mixture distribution H of stochastically ordered family $\{F_\theta\}_{\theta \in \mathbb{R}}$ with respect to T satisfies the inequalities

$$G(x) \leq H(x) \leq 1 - [1 - G(x)]^{\beta/\alpha},$$

when $\alpha > \beta$ and the mixture of $\{F_\theta\}_{\theta \in \mathbb{R}}$ with respect to S is G , and the inequalities are reversed for $\alpha < \beta$.

Below we discuss the expectation bounds for the case $\alpha > \beta$ in greater detail, and merely mention the final results for $\alpha < \beta$ which are derived analogously. By Proposition 1, the upper bound on $\frac{\mathbb{E}Y - \mathbb{E}X}{(\mathbb{E}|X - \mathbb{E}X|^p)^{1/p}}$ is finite for given $1 < p < \infty$ when

$$\int_0^1 (1 - u)^{\left(\frac{\beta}{\alpha} - 1\right) \frac{p}{p-1}} du < \infty, \text{ i.e., when the exponent is greater than } -1, \text{ which is}$$

equivalent to $p > \frac{\alpha}{\beta}$. However, the bound does not have an analytic form except for the case $p = 2$. The lower one is 0 for all $1 < p < \infty$ (and for $p = 1, \infty$ as well). When $\alpha < \beta$, the upper bound vanishes, and the lower one is negative finite for all $1 \leq p \leq \infty$.

If we have $\beta < \alpha < 2\beta$, and fix $p = 2$, we get

$$\int_0^1 \underline{h}^2(u) du = \left(\frac{\beta}{\alpha}\right)^2 \int_0^1 (1-u)^{2\beta/\alpha-2} du = \frac{\beta^2}{\alpha(2\beta-\alpha)},$$

and consequently

$$0 \leq \frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\text{Var } X}} \leq \frac{\alpha - \beta}{\sqrt{\alpha(2\beta - \alpha)}}.$$

For $\alpha \geq 2\beta$, the right-hand side expression is replaced by $+\infty$. Similarly, we obtain

$$\frac{\alpha - \beta}{\sqrt{\alpha(2\beta - \alpha)}} \leq \frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\text{Var } X}} \leq 0$$

for $\alpha < \beta$.

Proposition 2 asserts that

$$0 \leq \frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \leq +\infty$$

for $\alpha > \beta$, but in the case $\alpha < \beta$ the respective lower bound is nontrivial, and we have

$$-\frac{\beta}{2\alpha} \leq \frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \leq 0.$$

For $p = \infty$, we have $\underline{H}(\frac{1}{2}) = 1 - 2^{-\beta/\alpha}$, and so

$$0 \leq \frac{\mathbb{E}Y - \mathbb{E}X}{\text{ess sup } |X - \mathbb{E}X|} \leq 2^{1-\beta/\alpha} - 1$$

for $\alpha > \beta$. We easily check that the inequalities are reversed for $\alpha < \beta$. However, we do not claim that the zero bounds are optimal here. We finally observe that all the nonzero bounds tend to 0 as $\beta \rightarrow \alpha$.

Example 3 Suppose that one mixing distribution function is exponential with expectation 2, i.e., $S(\theta) = 1 - \exp(-\theta/2)$, $\theta > 0$, and the alternative one is that of the sum of two independent standard exponentials, that is $T(\theta) = 1 - (1 + \theta) \exp(-\theta)$, $\theta > 0$. The composition

$$v(u) = T(S^{-1}(u)) = 1 - (1-u)^2[1 - 2 \ln(1-u)], \quad 0 < u < 1,$$

represents set (2.2) for the above choice S and T . It is easily verified that v ranges over the whole standard unit interval, is strictly increasing, and convex on $(0, 1 - e^{-1})$, and concave on $(1 - e^{-1}, 1)$. The greatest convex minorant \underline{H} of v is first coinciding with some convex part of v , and then linear. The breaking point is determined by the equation

$$1 - v(u) = (1 - u)^2[1 - 2 \ln(1 - u)] = (1 - u)v'(u) = -4(1 - u)^2 \ln(1 - u),$$

and amounts to $1 - e^{-1/2} \approx 0.39347$ which satisfies $v'(1 - e^{-1/2}) = 2e^{-1/2} \approx 1.21306$. Therefore the greatest convex minorant and its derivative are following

$$\underline{H}(u) = \begin{cases} 1 - (1 - u)^2[1 - 2 \ln(1 - u)], & 0 < u \leq 1 - e^{-1/2}, \\ 2e^{-1/2}(u - 1) + 1, & 1 - e^{-1/2} \leq u < 1, \end{cases}$$

and

$$\underline{h}(u) = \begin{cases} -4(1 - u) \ln(1 - u), & 0 < u \leq 1 - e^{-1/2}, \\ 2e^{-1/2}, & 1 - e^{-1/2} \leq u < 1, \end{cases}$$

respectively. The smallest concave majorant \overline{H} of v is first linear, and then identical with v . The change point satisfies

$$v(u) = 1 - (1 - u)^2[1 - 2 \ln(1 - u)] = uv'(u) = -4u(1 - u) \ln(1 - u).$$

This does not have an analytic representation, and equals to $u_0 \approx 0.87242$ so that $v'(u_0) \approx 1.05075$. The smallest concave majorant

$$\overline{H}(u) = \begin{cases} v'(u_0)u, & 0 < u \leq u_0, \\ 1 - (1 - u)^2[1 - 2 \ln(1 - u)], & u_0 \leq u < 1, \end{cases}$$

has the derivative

$$\overline{h}(u) = \begin{cases} v'(u_0), & 0 < u \leq u_0, \\ -4(1 - u) \ln(1 - u), & u_0 \leq u < 1. \end{cases}$$

Note that both functions \underline{h} and \overline{h} are bounded here, and hence all the bounds of Propositions 1–3 are finite.

For establishing the standard deviation bounds with $p = 2$, we use the following indefinite integral

$$\begin{aligned} v_2(u) &= \int_1^u [v'(t)]^2 dt = 16 \int_1^u (1 - t)^2 \ln^2(1 - t) dt \\ &= -16(1 - u)^3 \left[\frac{1}{3} \ln^2(1 - u) - \frac{2}{9} \ln(1 - u) + \frac{2}{27} \right]. \end{aligned}$$

The upper and lower bounds are

$$\begin{aligned}
 B_2(S, T) &= \sqrt{\int_0^1 \underline{h}^2(u)du} - 1 = \sqrt{v_2(1 - e^{-1/2}) - v_2(0) + 4e^{-3/2} - 1} \\
 &= \sqrt{\frac{5 - 8e^{-3/2}}{27}}
 \end{aligned}$$

and

$$b_2(S, T) = -\sqrt{\int_0^1 \bar{h}^2(u)du} - 1 = -\sqrt{u_0[v'(u_0)]^2 + v_2(1) - v_2(u_0) - 1},$$

respectively. In the case $p = 1$, we immediately check that $e^{-1/2}$ and $-\frac{1}{2}v'(u_0)$ are the upper and lower optimal evaluations. Since $\underline{H}(\frac{1}{2}) = 1 - e^{-1/2}$ and $\bar{H}(\frac{1}{2}) = \frac{1}{2}v'(u_0)$, bounds (3.11) and (3.15) of Proposition 3 take on the forms $2e^{-1/2}$ and $1 - v'(u_0)$, respectively. Below we present numerical approximations of the above bounds

$$\begin{aligned}
 -0.52538 &\leq \frac{\mathbb{E}Y - \mathbb{E}X}{\mathbb{E}|X - \mathbb{E}X|} \leq 0.60653, \\
 -0.16684 &\leq \frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\text{Var } X}} \leq 0.34507, \\
 -0.05075 &\leq \frac{\mathbb{E}Y - \mathbb{E}X}{\text{ess sup } |X - \mathbb{E}X|} \leq 0.21306.
 \end{aligned}$$

Example 4 We finally consider a pair of seemingly similar symmetric unimodal mixing distributions supported on $[0, 1]$. One is

$$S(\theta) = \begin{cases} \frac{3}{2}\theta^2, & 0 < \theta < \frac{1}{2}, \\ 1 - \frac{3}{2}(1 - \theta)^2, & \frac{1}{2} \leq \theta < 1, \end{cases}$$

which has a symmetric triangular density, and an atom $\frac{1}{4}$ at the mode $\frac{1}{2}$. The other one

$$T(\theta) = \begin{cases} 4\theta^3, & 0 < \theta < \frac{1}{2}, \\ 1 - 4(1 - \theta)^3, & \frac{1}{2} \leq \theta < 1, \end{cases}$$

has quadratic density of different forms on $(0, \frac{1}{2})$ and $(\frac{1}{2}, 1)$, and a sharp peak at $\frac{1}{2}$. Then

$$\begin{aligned}
 \mathcal{U}(S, T) &= \left\{ \left(u, \frac{8}{3}\sqrt{\frac{2}{3}}u^{3/2} \right) : 0 \leq u < \frac{3}{8} \right\} \\
 &\cup \left\{ \left(u, 1 - \frac{8}{3}\sqrt{\frac{2}{3}}(1 - u)^{3/2} \right) : \frac{5}{8} \leq u < 1 \right\}.
 \end{aligned}$$

The first pairs represent the graph of convex increasing function, and that of the latter is concave increasing. The greatest convex minorant of $\mathcal{U}(S, T)$ has three different parts: function $v(u) = \frac{8}{3}\sqrt{\frac{2}{3}}u^{3/2}$ on some right neighborhood of 0, then a straight line joining the graph of v with point $(\frac{5}{8}, \frac{1}{2})$, and ultimately another piece of line passing between $(\frac{5}{8}, \frac{1}{2})$ and $(1, 1)$. The point of change from strictly convex v into the first linear one is determined by the slope equality condition

$$v'(u) = \frac{\frac{1}{2} - v(u)}{\frac{5}{8} - u}$$

for some $0 < u < \frac{3}{8}$ which can be rewritten as

$$8\sqrt{2}(\sqrt{u})^3 - 15\sqrt{2}\sqrt{u} + 3\sqrt{3} = 0.$$

The cubic function of real argument replacing \sqrt{u} has three real roots $-\frac{\sqrt{3}}{2} - \frac{\sqrt{6}}{4} < \frac{\sqrt{3}}{2} - \frac{\sqrt{6}}{4} < \frac{\sqrt{6}}{2}$, and only the middle one is located in $(0, 1)$. Therefore our change point is $(\frac{\sqrt{3}}{2} - \frac{\sqrt{6}}{4})^2 = \frac{9}{8} - \frac{3}{4}\sqrt{2} \approx 0.064340$ so that $v'(\frac{9}{8} - \frac{3}{4}\sqrt{2}) = 2\sqrt{2} - 2 \approx 0.82843$. Accordingly,

$$\underline{H}(u) = \begin{cases} \frac{8}{3}\sqrt{\frac{2}{3}}u^{3/2}, & 0 \leq u \leq \frac{9}{8} - \frac{3}{4}\sqrt{2}, \\ (2\sqrt{2} - 2)(u - \frac{5}{8}) + \frac{1}{2}, & \frac{9}{8} - \frac{3}{4}\sqrt{2} \leq u \leq \frac{5}{8}, \\ \frac{4}{3}(u - 1) + 1, & \frac{5}{8} \leq u \leq 1, \end{cases}$$

and its derivative is

$$\underline{h}(u) = \begin{cases} 4\sqrt{\frac{2u}{3}}, & 0 \leq u \leq \frac{9}{8} - \frac{3}{4}\sqrt{2}, \\ 2\sqrt{2} - 2, & \frac{9}{8} - \frac{3}{4}\sqrt{2} \leq u < \frac{5}{8}, \\ \frac{4}{3}, & \frac{5}{8} \leq x < 1. \end{cases}$$

Therefore

$$\begin{aligned} B_2^2(S, T) &= \int_0^{\frac{9}{8} - \frac{3}{4}\sqrt{2}} \frac{32u}{3} du + (2\sqrt{2} - 2)^2 \left(\frac{3}{4}\sqrt{2} - \frac{1}{2}\right) \\ &\quad + \left(\frac{4}{3}\right)^2 \frac{3}{8} - 1 = 4\sqrt{2} - 5\frac{7}{12}, \end{aligned}$$

and $B_2(S, T) = \sqrt{4\sqrt{2} - 5\frac{7}{12}} \approx 0.27115$ is the sharp upper bound on $\frac{\mathbb{E}Y - \mathbb{E}X}{\sqrt{\text{Var } X}}$. Furthermore, $B_1(S, T) = \frac{1}{2}[\underline{h}(1) - \underline{h}(0)] = \frac{2}{3}$ and $B_\infty(S, T) = 1 - 2\underline{H}(\frac{1}{2}) = \frac{\sqrt{2}-1}{2} \approx 0.20711$ provide analogous bounds expressed in $\mathbb{E}|X - \mathbb{E}X|$ and $\text{ess sup } |X - \mathbb{E}X|$ units, respectively.

Since $\bar{H}(u) = 1 - \underline{H}(1-u)$, and $\bar{h}(u) = \underline{h}(1-u)$, we easily deduce that $b_p(S, T) = -B_p(S, T)$, $1 \leq p \leq +\infty$, which means that the sharp lower bounds are the negatives of their upper counterparts.

5 Proof of Theorem 1

We prove the first inequality in (2.1) for every fixed value $u = G(x) \in [0, 1]$. It is evident that for $u = G(x) = \int_{\mathbb{R}} F_{\theta}(x)S(d\theta) = 0$, we can take $F_{\theta}(x) = 0$, $\theta \in \mathbb{R}$, which guarantees the smallest possible value of $\int_{\mathbb{R}} F_{\theta}(x)T(d\theta) = 0$.

For fixed $0 < u \leq 1$, we solve an auxiliary minimization problem for discrete approximations of $F_{\theta}(x)$, $\theta \in \mathbb{R}$. Consider a finite partition of real axis $\mathbb{R} = \bigcup_{i=1}^n \Theta_i$, where $\Theta_1 < \dots < \Theta_n$ are disjoint ordered intervals. Each interval either has or does not have the left and right end-point. Define $s_i = S(\Theta_i)$, $t_i = T(\Theta_i)$, and denote by u_i unknown constant values of $F_{\theta}(x)$ on Θ_i , $i = 1, \dots, n$. The only restrictions on them are $1 \geq u_1 \geq \dots \geq u_n \geq 0$, and $\int_{\mathbb{R}} F_{\theta}(x)S(d\theta) = \sum_{i=1}^n s_i u_i = u$. Our auxiliary problem is to minimize $\int_{\mathbb{R}} F_{\theta}(x)T(d\theta) = \sum_{i=1}^n t_i u_i$ under the constraints on u_1, \dots, u_n . By the change of variables

$$\begin{aligned} v_n &= u_n \geq 0, \\ v_i &= u_i - u_{i+1} \geq 0, \quad i = 1, \dots, n - 1, \end{aligned}$$

satisfying also $\sum_{i=1}^n v_i = u_1 \leq 1$, we have

$$\begin{aligned} \sum_{i=1}^n u_i s_i &= \sum_{i=1}^n \left(\sum_{j=i}^n v_j \right) s_i = \sum_{i=1}^n \left(\sum_{j=1}^i s_j \right) v_i = \sum_{i=1}^n S_i v_i, \\ \sum_{i=1}^n u_i t_i &= \sum_{i=1}^n T_i v_i, \end{aligned}$$

where $S_i = \sum_{j=1}^i s_j$, $T_i = \sum_{j=1}^i t_j$, $i = 1, \dots, n$, are known and satisfy $0 \leq S_1 \leq \dots \leq S_n = 1$ and $0 \leq T_1 \leq \dots \leq T_n = 1$.

Our problem can be rewritten as to minimize $\sum_{i=1}^n T_i v_i$ over the intersection of the simplex

$$S_n = \left\{ (v_1, \dots, v_n) : v_i \geq 0, \quad i = 1, \dots, n, \quad \sum_{i=1}^n v_i \leq 1 \right\}$$

and hyperplane

$$\mathcal{H}_S = \left\{ (v_1, \dots, v_n) : \sum_{i=1}^n S_i v_i = u \right\}.$$

The simplex is a convex polyhedron with vertices $\mathbf{0}$, being the zero vector, and \mathbf{e}_i , $i = 1, \dots, n$, which are the standard unit vectors in \mathbb{R}^n . Clearly, the line segments joining the extreme points constitute the edges of \mathcal{S}_n . The intersection $\mathcal{S}_n \cap \mathcal{H}_S$ is a convex set as well, and all its extreme points belong to some edges of \mathcal{S}^n . All the edges can be represented as

$$\begin{aligned} &\alpha \mathbf{e}_i + (1 - \alpha) \mathbf{0}, \quad 1 \leq i \leq n, \\ &\alpha \mathbf{e}_i + (1 - \alpha) \mathbf{e}_j, \quad 1 \leq i < j \leq n, \end{aligned}$$

for all $0 \leq \alpha \leq 1$. In the first case, $\alpha \mathbf{e}_i \in \mathcal{H}_S$ iff $S_i \geq u$ and

$$v_k = \begin{cases} \frac{u}{S_i}, & k = i, \\ 0, & \text{otherwise.} \end{cases} \tag{5.1}$$

In the latter one, $\alpha \mathbf{e}_i + (1 - \alpha) \mathbf{e}_j \in \mathcal{H}_S$ iff $S_i \leq u \leq S_j$ and $\alpha S_i + (1 - \alpha) S_j = u$. If either $u = S_i$ or $u = S_j$, then either $\mathbf{e}_i \in \mathcal{H}_S$ or $\mathbf{e}_j \in \mathcal{H}_S$, respectively. These come under the provisions of the previous case. When $S_i < u < S_j$, the respective vector has the coordinates

$$v_k = \begin{cases} \frac{S_j - u}{S_j - S_i}, & k = i, \\ \frac{u - S_i}{S_j - S_i}, & k = j, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

The linear functional $\sum_{i=1}^n T_i v_i$ attains its minimal value on the convex compact set $\mathcal{S}_n \cap \mathcal{H}_S$ at some its extreme points. This means that it suffices to confine analysis to the finite set of points (5.1) if $u \leq S_i$, and (5.2) if $S_i < u < S_j$. Coming back to the original variables we conclude that the possible candidates for the minimum points are (u_1, \dots, u_n) with

$$u_k = \begin{cases} \frac{u}{S_i}, & k = 1, \dots, i, \\ 0, & k = i + 1, \dots, n, \end{cases} \quad \text{if } u \leq S_i, \tag{5.3}$$

and

$$u_k = \begin{cases} 1, & k = 1, \dots, i, \\ \frac{u - S_i}{S_j - S_i}, & k = i + 1, \dots, j, \\ 0, & k = j + 1, \dots, n, \end{cases} \quad \text{if } S_i < u < S_j. \tag{5.4}$$

Adding $i = 0$ with $S_0 = 0$, we can jointly represent (5.3) and (5.4) by the formula

$$u_k = \begin{cases} 1, & k = 1, \dots, i, \\ \frac{u - S_i}{S_j - S_i}, & k = i + 1, \dots, j, \\ 0, & k = j + 1, \dots, n, \end{cases} \quad \text{if } S_i < u \leq S_j, \quad 0 \leq i < j \leq n. \tag{5.5}$$

Note that for $i = 0$ and $j = n$, the first and last options in (5.5) are empty, respectively.

Accordingly, for given $u = G(x)$ and $F_\theta(x)$, $\theta \in \mathbb{R}$, constant on the intervals of finite partition $\Theta_1 < \dots < \Theta_n$ there exist $0 \leq i < j \leq n$ such that $S_i = S\left(\bigcup_{k=1}^i \Theta_k\right) < u \leq S_j = S\left(\bigcup_{k=1}^j \Theta_k\right)$ which provide the minimal value of

$$\begin{aligned} \sum_{k=1}^n t_k u_k &= \sum_{k=1}^i t_k + \frac{u - S_i}{S_j - S_i} \sum_{k=i+1}^j t_k \\ &= T\left(\bigcup_{k=1}^i \Theta_k\right) + \frac{T\left(\bigcup_{k=i+1}^j \Theta_k\right)}{S\left(\bigcup_{k=i+1}^j \Theta_k\right)} \left[u - S\left(\bigcup_{k=1}^i \Theta_k\right) \right]. \end{aligned}$$

This means that in this case it suffices to partition the parameter set \mathbb{R} onto at most three intervals $\Theta_1 < \Theta_2 < \Theta_3$ such that $S(\Theta_1) < u \leq S(\Theta_1 \cup \Theta_2)$. Observe that Θ_1 and/or Θ_3 may be empty so that $S(\Theta_1) = 0$ and/or $S(\Theta_1 \cup \Theta_2) = 1$. For given S , T , and $0 < u \leq 1$, we define

$$\underline{H}(u) = \inf \left\{ T(\Theta_1) + \frac{T(\Theta_2)}{S(\Theta_2)} [u - S(\Theta_1)] \right\}, \tag{5.6}$$

where the infimum is taken over all partitions $\Theta_1 < \Theta_2 < \Theta_3$ of the real line satisfying $S(\Theta_1) < u \leq S(\Theta_1 \cup \Theta_2)$. Observe that

$$S(\Theta_1) = \begin{cases} 0, & \text{if } \Theta_1 = \emptyset, \\ S(\theta_1-), & \text{if } \Theta_1 = (-\infty, \theta_1), \\ S(\theta_1), & \text{if } \Theta_1 = (-\infty, \theta_1], \end{cases}$$

and

$$S(\Theta_2) = \begin{cases} S(\theta_2-) - S(\Theta_1), & \text{if } \Theta_1 \cup \Theta_2 = (-\infty, \theta_2), \\ S(\theta_2) - S(\Theta_1), & \text{if } \Theta_1 \cup \Theta_2 = (-\infty, \theta_2], \end{cases}$$

for some $\theta_1 < \theta_2$, and similar relations hold for $T(\Theta_1)$ and $T(\Theta_2)$. This means that for given $0 < u \leq 1$, $\underline{H}(u)$ is the infimum over the values of all linear functions joining points $(S(\theta_1\mp), T(\theta_1\mp))$ and $(S(\theta_2\mp), T(\theta_2\mp))$ for which $S(\theta_1\mp) < u \leq S(\theta_2\mp)$ at argument u . This defines the greatest convex minorant of the set of points $\{(S(\theta-), T(\theta-)), (S(\theta), T(\theta)) : \theta \in \mathbb{R}\}$ on interval $[0, 1]$ when we let vary u over the interval. Obviously, the greatest convex minorant does not change if we drop the left-limit values.

A trivial but important observation is that the same partition $\Theta_1 < \Theta_2 < \Theta_3$ may provide the infimum in (5.6) for various u . In particular, if the infimum is attained when $S(\Theta_2) = 1$, then $\underline{H}(u) = \underline{u} u$ with \underline{u} defined in (2.6) for every $u \in [0, 1]$. Another extremity is that $\underline{H}(u)$ is approached as $S(\Theta_2) \searrow 0$, and the linear functions

in (5.6) tend to the tangent of $\mathcal{U}(S, T)$ at u . In this case, one needs to perform separate approximations for single values of u .

We prove that (5.6) satisfies the left hand-side inequality in (2.1). Let $F_\theta(x)$ for fixed x be a non-increasing function of $\theta \in \mathbb{R}$ taking values in $[0, 1]$ and satisfying $\int_{\mathbb{R}} F_\theta(x)S(d\theta) = u$. If $F_\theta(x), \theta \in \mathbb{R}$, take on finitely many values (u_1, \dots, u_n) , we proved above that there exists $\underline{F}_\theta(x), \theta \in \mathbb{R}$, with no more than three values $0, 0 < c < 1$ and 1 , satisfying the assumptions and such that

$$\int_{\mathbb{R}} \underline{F}_\theta(x)T(d\theta) \leq \int_{\mathbb{R}} F_\theta(x)T(d\theta),$$

and the infimum over these at most three-valued functions is just $\underline{H}(u)$.

If $F_\theta(x), \theta \in \mathbb{R}$, has an infinite image set, we need more subtle arguments. Since it is monotone, nonnegative and bounded above, we can write

$$u = \int_{\mathbb{R}} F_\theta(x)S(d\theta) = \inf_{A_1, \dots, A_m} \sum_{i=1}^m \sup_{\theta \in A_i} F_\theta(x)S(A_i),$$

$$\int_{\mathbb{R}} F_\theta(x)T(d\theta) = \inf_{B_1, \dots, B_m} \sum_{i=1}^m \sup_{\theta \in B_i} F_\theta(x)T(B_i),$$

where $\{A_1, \dots, A_m\}$ and $\{B_1, \dots, B_m\}$ are arbitrary finite partitions of \mathbb{R} onto disjoint intervals. This means that there exist increasing sequences of partitions $\{A_1^m, \dots, A_{i_m}^m\}, \{B_1^m, \dots, B_{j_m}^m\}, m = 1, 2, \dots$, such that

$$u_m = \sum_{i=1}^{i_m} \sup_{\theta \in A_i^m} F_\theta(x)S(A_i^m) \searrow u = \int_{\mathbb{R}} F_\theta(x)S(d\theta),$$

$$\sum_{j=1}^{j_m} \sup_{\theta \in B_j^m} F_\theta(x)T(B_j^m) \searrow \int_{\mathbb{R}} F_\theta(x)T(d\theta),$$

as $m \rightarrow \infty$. Let $\{C_1^m, \dots, C_{k_m}^m\}$ be the partition of \mathbb{R} composed of intersections $A_i^m \cap B_j^m, i = 1, \dots, i_m, j = 1, \dots, j_m$. Then

$$\sum_{i=1}^{i_m} \sup_{\theta \in A_i^m} F_\theta(x)S(A_i^m) \geq \sum_{k=1}^{k_m} \sup_{\theta \in C_k^m} F_\theta(x)S(C_k^m) = v_m$$

$$\searrow u = \int_{\mathbb{R}} F_\theta(x)S(d\theta) > 0,$$

$$\sum_{j=1}^{j_m} \sup_{\theta \in B_j^m} F_\theta(x)T(B_j^m) \geq \sum_{k=1}^{k_m} \sup_{\theta \in C_k^m} F_\theta(x)T(C_k^m) \searrow \int_{\mathbb{R}} F_\theta(x)T(d\theta).$$

Functions

$$F_\theta^m(x) = \frac{u}{v_m} \sum_{k=1}^{k_m} \sup_{\theta \in C_k^m} F_\theta(x) \mathbf{1}_{C_k^m}(\theta), \quad \theta \in \mathbb{R},$$

are non-increasing, have finite numbers of values in $[0, 1]$, and integrate to u with respect to measure S . Therefore

$$\begin{aligned} \underline{H}(u) &\leq \int_{\mathbb{R}} F_\theta^m(x) T(d\theta) = \frac{u}{v_m} \sum_{k=1}^{k_m} \sup_{\theta \in C_k^m} F_\theta(x) T(C_k^m) \\ &\rightarrow \int_{\mathbb{R}} F_\theta^m(x) T(d\theta) \geq \underline{H}(u), \end{aligned}$$

which is our claim.

We proved that for every fixed $x \in \mathbb{R}$, there exists function $F_\theta(x)$, $\theta \in \mathbb{R}$, with at most three values in $[0, 1]$ which satisfies (1.2) and provides the minimal value of (1.3) which amounts to (2.3). Letting x vary over \mathbb{R} , we obtain two-variable function $\underline{F} : \mathbb{R}^2 \mapsto [0, 1]$ which can be treated as a family of functions $\{F_\theta\}_{\theta \in \mathbb{R}}$ in variable x . Our aim is to prove that the construction defines a family of stochastically ordered distribution functions.

To this end, we first perform some auxiliary considerations. Define

$$\bar{A} = \{0 \leq u \leq 1 : \exists \theta \in \mathbb{R} (u, \underline{H}(u)) = (S(\theta \pm), T(\theta \pm))\}.$$

This notation means that either $(u, \underline{H}(u)) = (S(\theta), T(\theta))$ or $(u, \underline{H}(u)) = (S(\theta-), T(\theta-))$. This is a closed subset of $[0, 1]$ which can be represented as an at most countable sum of disjoint closed, possibly degenerate intervals $\bar{A} = \bigcup_j \bar{A}_j$. Then its interior is a sum of disjoint open intervals $A = \text{int } \bar{A} = \bigcup_j A_j$. Note that the number of open intervals can be less than the number of the closed ones. The border of A consists of at most countably many separate points

$$C = \partial \bar{A} = \bar{A} \setminus A = \bigcup_j \{c_j\}.$$

We finally introduce

$$B = [0, 1] \setminus \bar{A} = \{0 \leq u \leq 1 : \forall \theta \in \mathbb{R} (u, \underline{H}(u)) \neq (S(\theta \pm), T(\theta \pm))\}$$

which means that neither $(u, \underline{H}(u)) = (S(\theta), T(\theta))$ nor $(u, \underline{H}(u)) = (S(\theta-), T(\theta-))$. Obviously, this is a sum of no more than countably many disjoint intervals $B = \bigcup_j B_j$. The closures of B and B_j are denoted by \bar{B} , and \bar{B}_j , respectively. Summing up, each pair (S, T) generates a partition of the unit interval

$$[0, 1] = \bigcup_j A_j \cup \bigcup_j B_j \cup \bigcup_j \{c_j\}. \tag{5.7}$$

It is obvious that for every $A_j, A_k \subset A, A_j \prec A_k$, there exists $B_l \subset B$ such that $A_j \prec B_l \prec A_k$. This is not necessarily true if we replace the roles of subintervals of A and B . However, for every pair of subintervals from A and B there exists points c_l separating them.

For every $G(x) = u \in A$ there exists $\theta = \theta_u \in \mathbb{R}$ such that $(u, \underline{H}(u)) = (S(\theta), T(\theta))$. This is not uniquely determined, though. We denote the set of all θ satisfying the condition by Θ^u . We also define $\Theta^{A_j} = \bigcup_{u \in A_j} \Theta^u$. Similarly, for every $c_j \in C$ there exists real θ_{c_j} satisfying either $(c_j, \underline{H}(c_j)) = (S(\theta_{c_j}), T(\theta_{c_j}))$ or $(c_j, \underline{H}(c_j)) = (S(\theta_{c_j}^-), T(\theta_{c_j}^-))$. This is unique in the latter case, but otherwise may be many θ 's satisfying the condition. We denote the set of all θ sharing the property by Θ^{c_j} . Finally, for $B_j = \overline{B}_j \setminus \{c_k, c_l\}$ with $c_k < c_l$, define

$$\Theta^{B_j} = \{\theta \in \mathbb{R} : \forall \theta_1 \in \Theta^{c_k}, \theta_2 \in \Theta^{c_l} \theta_1 < \theta < \theta_2\}.$$

Then we have a partition of the parameter set

$$\mathbb{R} = \bigcup_j \Theta^{A_j} \cup \bigcup_j \Theta^{B_j} \cup \bigcup_j \Theta^{c_j}.$$

For any two elements $D_k \prec D_l$ of the partition (5.7) we have $\Theta^{D_k} \prec \Theta^{D_l}$. Moreover, inequality $u < v$ for some $u, v \in A_j$ implies $\Theta^u \prec \Theta^v$ as well.

From the solution of the local minimization problems for various $u = G(x)$, we conclude the following minimization conditions. If $u \in A_j = (c_k, c_l)$, say, then

$$\underline{F}_\theta(x) = \begin{cases} 1, & \text{if } \theta \preceq \Theta^u, \\ 0, & \text{if } \theta \succ \Theta^u. \end{cases} \tag{5.8}$$

The above notation, introduced for brevity, means that $\theta \prec (\succ) \Theta^u$ when $\{\theta\} \prec (\succ) \Theta^u$, and $\theta \preceq (\succeq) \Theta^u$ when either $\{\theta\} \prec (\succ) \Theta^u$ or $\theta \in \Theta^u$. For $u \in \overline{B}_j = [c_k, c_l]$, say, we have four possible cases. We take

$$\underline{F}_\theta(x) = \begin{cases} 1, & \text{if } \theta \preceq \Theta^{c_k}, \\ \frac{G(x) - S(\theta_{c_k})}{S(\theta_{c_l}) - S(\theta_{c_k})}, & \text{if } \Theta^{c_k} \prec \theta \preceq \Theta^{c_l}, \\ 0, & \text{if } \theta \succ \Theta^{c_l}, \end{cases} \tag{5.9}$$

if $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}), T(\theta_{c_k}))$ and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}), T(\theta_{c_l}))$ for θ_{c_k} and θ_{c_l} , being some representatives of Θ^{c_k} and Θ^{c_l} , respectively. Moreover,

$$\underline{F}_\theta(x) = \begin{cases} 1, & \text{if } \theta \prec \Theta^{c_k}, \\ \frac{G(x) - S(\theta_{c_k}^-)}{S(\theta_{c_l}^-) - S(\theta_{c_k}^-)}, & \text{if } \Theta^{c_k} \preceq \theta \prec \Theta^{c_l}, \\ 0, & \text{if } \theta \succeq \Theta^{c_l}, \end{cases} \tag{5.10}$$

when $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}-), T(\theta_{c_k}-))$ and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}-), T(\theta_{c_l}-))$, and

$$\underline{F}_\theta(x) = \begin{cases} 1, & \text{if } \theta \leq \Theta^{c_k}, \\ \frac{G(x)-S(\theta_{c_k})}{S(\theta_{c_l}-)-S(\theta_{c_k})}, & \text{if } \Theta^{c_k} < \theta < \Theta^{c_l}, \\ 0, & \text{if } \theta \geq \Theta^{c_l}, \end{cases} \tag{5.11}$$

if $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}), T(\theta_{c_k}))$ and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}-), T(\theta_{c_l}-))$, and finally

$$\underline{F}_\theta(x) = \begin{cases} 1, & \text{if } \theta < \Theta^{c_k}, \\ \frac{G(x)-S(\theta_{c_k}-)}{S(\theta_{c_l})-S(\theta_{c_k}-)}, & \text{if } \Theta^{c_k} \leq \theta \leq \Theta^{c_l}, \\ 0, & \text{if } \theta > \Theta^{c_l}, \end{cases} \tag{5.12}$$

if $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}-), T(\theta_{c_k}-))$ and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}), T(\theta_{c_l}))$. We easily check that in all the cases (5.8)–(5.12) imply

$$\int_{\mathbb{R}} \underline{F}_\theta(x) S(d\theta) = G(x),$$

$$\int_{\mathbb{R}} \underline{F}_\theta(x) T(d\theta) = \underline{H}(G(x)).$$

where $\underline{H}(G(x)) = T(\theta_u)$ for (5.8), and $\underline{H}(G(x)) = T(\theta_{c_k \pm}) + \frac{T(\theta_{c_l \pm}) - T(\theta_{c_k \pm})}{S(\theta_{c_l \pm}) - (\theta_{c_k \pm})} [u - S(\theta_{c_k \pm})]$ with properly chosen limits at c_k and c_l in the other cases.

Now we rewrite the above formulae fixing $\theta \in \mathbb{R}$, and letting x vary. Assume first **(A1)** $\theta \in \Theta^u \subset \Theta^{A_j}$ for some $u = G(x) \in A_j \subset A$.

Then $\underline{F}_\theta(x) = 1$ by (5.8). Take $u > v = G(y) \in A$ (we admit $v \in A_j$ as well as $v \in A_k < A_j$). Since

$$\underline{F}_\theta(y) = \begin{cases} 0, & \text{if } \theta > \Theta^v, \\ 1, & \text{otherwise,} \end{cases}$$

and $\Theta^u > \Theta^v$, we obtain $\underline{F}_\theta(y) = 0$. Assume now that $v = G(y) \in \overline{B}_j = [c_k, c_l]$ for some $\overline{B}_j < A_j$. Then $\Theta^u > \Theta^{c_l}$ and any of (5.9)–(5.12) implies that $\underline{F}_\theta(y) = 0$ for all $\theta \in \Theta^u$. Similar arguments show that $\underline{F}_\theta(y) = 1$ if $G(y) > G(x)$ and either $G(y) \in A$ or $G(y) \in \overline{B}$. This shows that for every $\theta \in \Theta^{G(x)}$ such that $G(x) \in A$, we have

$$\underline{F}_\theta(y) = \begin{cases} 0, & \text{if } G(y) < G(x), \\ 1, & \text{if } G(y) \geq G(x), \end{cases} \tag{5.13}$$

Consider now the set of parameters θ such that

(A2) $\Theta^{c_k} < \theta \leq \Theta^{c_l}$ for some $(c_k, c_l) = B_j$, with $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}), T(\theta_{c_k}))$, and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}), T(\theta_{c_l}))$.

Then for every x such that $G(x) = u \in [c_k, c_l]$ we have

$$F_\theta(x) = \frac{G(x) - S(\theta_{c_k})}{S(\theta_{c_l}) - S(\theta_{c_k})}.$$

(cf (5.9)). If $G(x) = v \in A_k < \overline{B}_j$, then $\Theta^v < \Theta^{c_k} < \theta$, and $F_\theta(x) = 0$ by (5.8). Certainly, $F_\theta(x) = 1$ if $G(x) = v \in A_k > \overline{B}_j$, because $\Theta^v > \Theta^{c_l}$.

If $G(x) \in \overline{B}_i = [c_p, c_q] < \overline{B}_j$, we obtain $\Theta^{c_q} < \Theta^{c_k} < \theta$, and each of (5.9)–(5.12) implies $F_\theta(x) = 0$. If $G(x) \in \overline{B}_i \setminus \{c_k\} = [c_p, c_k)$, relation $\Theta^{c_k} < \theta$ provides the same conclusion by (5.9) and (5.12). When $G(x) \in \overline{B}_i = [c_p, c_q]$ with $c_p \geq c_l$, we have two possibilities. When $c_p > c_l$, we apply the relation $\Theta^{c_l} < \Theta^{c_p}$, possibly together with $\theta < \Theta^{c_l}$, to obtain $F_\theta(x) = 1$. Suppose now that $G(x) \in (c_l, c_q]$. Applying formulae (5.9) and (5.11), we conclude again $F_\theta(x) = 1$ from condition $\theta \leq \Theta^{c_l}$. Accordingly, for θ 's satisfying assumption (A2) we have

$$F_\theta(x) = \begin{cases} 0, & \text{if } G(x) \leq S(\theta_{c_k}), \\ \frac{G(x) - S(\theta_{c_k})}{S(\theta_{c_l}) - S(\theta_{c_k})}, & \text{if } S(\theta_{c_k}) \leq G(x) \leq S(\theta_{c_l}), \\ 1, & \text{if } G(x) \geq S(\theta_{c_l}), \end{cases} \tag{5.14}$$

which is a proper distribution function on \mathbb{R} .

Using much the same arguments, we deduce that for parameters θ satisfying the conditions

- (A3) $\Theta^{c_k} \leq \theta < \Theta^{c_l}$ for some $(c_k, c_l) = B_j$, with $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}-), T(\theta_{c_k}-))$, and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}-), T(\theta_{c_l}-))$,
- (A4) $\Theta^{c_k} < \theta < \Theta^{c_l}$ for some $(c_k, c_l) = B_j$, with $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}), T(\theta_{c_k}))$, and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}-), T(\theta_{c_l}-))$,
- (A5) $\Theta^{c_k} \leq \theta \leq \Theta^{c_l}$ for some $(c_k, c_l) = B_j$, with $(c_k, \underline{H}(c_k)) = (S(\theta_{c_k}-), T(\theta_{c_k}-))$, and $(c_l, \underline{H}(c_l)) = (S(\theta_{c_l}), T(\theta_{c_l}))$,

the minimal distribution functions F_θ have the following forms

$$F_\theta(x) = \begin{cases} 0, & \text{if } G(x) \leq S(\theta_{c_k}-), \\ \frac{G(x) - S(\theta_{c_k}-)}{S(\theta_{c_l}-) - S(\theta_{c_k}-)}, & \text{if } S(\theta_{c_k}-) \leq G(x) \leq S(\theta_{c_l}-), \\ 1, & \text{if } G(x) \geq S(\theta_{c_l}-), \end{cases} \tag{5.15}$$

$$F_\theta(x) = \begin{cases} 0, & \text{if } G(x) \leq S(\theta_{c_k}), \\ \frac{G(x) - S(\theta_{c_k})}{S(\theta_{c_l}-) - S(\theta_{c_k})}, & \text{if } S(\theta_{c_k}) \leq G(x) \leq S(\theta_{c_l}-), \\ 1, & \text{if } G(x) \geq S(\theta_{c_l}-), \end{cases} \tag{5.16}$$

$$F_\theta(x) = \begin{cases} 0, & \text{if } G(x) \leq S(\theta_{c_k}-), \\ \frac{G(x) - S(\theta_{c_k}-)}{S(\theta_{c_l}) - S(\theta_{c_k}-)}, & \text{if } S(\theta_{c_k}-) \leq G(x) \leq S(\theta_{c_l}), \\ 1, & \text{if } G(x) \geq S(\theta_{c_l}), \end{cases} \tag{5.17}$$

respectively. Assumptions (A1)–(A5) guarantee that the family of distribution functions described in (5.13)–(5.17) is defined for all $\theta \in \mathbb{R}$. For verifying that this is

stochastically ordered it suffices to refer to formulae (5.8)–(5.12). They show that for $F_\theta(x)$ is decreasing in θ every fixed $x \in \mathbb{R}$.

The proof of the right-hand side inequality in (2.1) as well as its optimality is analogous to the above, and therefore it is omitted. \square

Acknowledgements The research was supported by National Science Centre, Poland, Grant 2015/19/B/ST1/03100. The author thanks the anonymous referees and associate editor for valuable comments that allowed him to improve presentation of the results.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Choi T (2008) Convergence of posterior distribution in the mixture of regressions. *J Nonparametr Stat* 20:337–351
- Chung Y, Dunson DB (2009) Nonparametric Bayes conditional distribution modeling with variable selection. *J Am Stat Assoc* 104:1646–1660
- Dunson DB, Peddada SD (2008) Bayesian nonparametric inference on stochastic ordering. *Biometrika* 95:859–874
- Escobar MD, West M (1995) Bayesian density estimation and inference using mixtures. *J Am Stat Assoc* 90:577–588
- Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *Ann Stat* 1:209–230
- Gasparini M (1996) Nonparametric Bayes estimation of a distribution function with truncated data. *J Stat Plan Inference* 55:361–369
- Ghosh M, Mukherjee B (2005) Nonparametric sequential Bayes estimation of the distribution function. *Seq Anal* 24:389–409
- Hansen MB, Lauritzen SL (2002) Nonparametric Bayes inference for concave distribution functions. *Stat Neerl* 56:110–127
- Jo S, Roh T, Choi T (2016) Bayesian spectral analysis models for quantile regression with Dirichlet process mixtures. *J Nonparametr Stat* 28:177–206
- Karlis D, Xekalaki E (2003) Mixtures everywhere. In: Panaretos J (ed) *Stochastic musings: perspectives from the pioneers of the late 20th century*. Laurence Erlbaum, Mahwah, pp 78–95
- McKeague IW, Tighiouart M (2002) Nonparametric Bayes estimators for hazard functions based on right censored data. *Tamkang J Math* 33:173–189
- Miziula P (2015) Moment comparisons for mixtures of ordered distributions. Ph. D. dissertation. Institute of Mathematics of the Polish Academy of Sciences, Warsaw
- Miziula P (2017) Comparison of dispersions of mixtures of ordered distributions. *Statistics* 51:862–877
- Miziula P, Solnický R (2018) Sharp bounds on change in expected values and variances for single risk analysis in the flood catastrophe model. *Scand Actuar J* 2018:64–75
- Nashimoto K, Wright FT (2007) Nonparametric multiple-comparison methods for simply ordered medians. *Comput Stat Data Anal* 51:5068–5076
- Nashimoto K, Wright FT (2008) Bayesian multiple comparisons of simply ordered means using priors with a point mass. *Comput Stat Data Anal* 52:5143–5153
- Navarro J, Balakrishnan N, Samaniego FJ, Bhattacharya D (2008) On the application and extension of system signatures to problems in engineering reliability. *Naval Res Logist* 55:313–327
- Navarro J, Rychlik T (2007) Reliability and expectation bounds for coherent systems with exchangeable components. *J Multivar Anal* 98:102–113
- Robertson T, Wright FT (1982) Bounds on mixtures of distributions arising in order restricted inference. *Ann Stat* 10:302–306
- Rychlik T (1993) Sharp bounds on L -estimates and their expectations for dependent samples. *Commun Stat Theory Methods* 22:1053–1068

- Samaniego FJ (1985) On closure of the IFR class under formation of coherent systems. *IEEE Trans Reliab* 34:69–72
- Sethuraman J, Hollander M (2009) Nonparametric Bayes estimation in repair models. *J Stat Plan Inference* 139:1722–1733
- Susarla V, Van Ryzin J (1978) Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. *Ann Stat* 6:755–768
- Tiwari RC, Zalkikar JN (1990) Nonparametric Bayes estimation of the survival function from a record of failures and follow-ups. *Commun Stat Theory Methods* 19:4331–4354
- Vieira CC, Loschi RH, Duarte D (2015) Nonparametric mixtures based on skew-normal distributions: an application to density estimation. *Commun Stat Theory Methods* 44:1552–1570
- Yang H, O'Brien S, Dunson DB (2011) Nonparametric Bayes stochastically ordered latent class models. *J Am Stat Assoc* 106:807–817
- Zhou M (2004) Nonparametric Bayes estimator of survival functions for doubly/interval censored data. *Stat Sin* 14:533–546
- Zhu B, Dunson DB (2013) Locally adaptive Bayes nonparametric regression via nested Gaussian processes. *J Am Stat Assoc* 108:1445–1456